NAR Genomics and Bioinformatics, 2024, **6**, lqae043
https://doi.org/10.1093/nargab/lqae043
Advance access publication date: 27 April 2024
Standard Article

Cancer drug sensitivity estimation using modular deep Graph Neural Networks

Pedro A. Campana [1,*], Paul Prasse [1], Matthias Lienhard [2], Kristina Thedinga[2], Ralf Herwig [2] and Tobias Scheffer [1]

[1]University of Potsdam, Department of Computer Science, Potsdam, Germany
[2]Max Planck Institute for Molecular Genetics, Department Computational Molecular Biology, Berlin, Germany
[*]To whom correspondence should be addressed. Tel: +49 331 977 3128; Email: pedro.alonso.campana@uni-potsdam.de

Abstract

Computational drug sensitivity models have the potential to improve therapeutic outcomes by identifying targeted drugs components that are tailored to the transcriptomic profile of a given primary tumor. The SMILES representation of molecules that is used by state-of-the-art drug-sensitivity models is not conducive for neural networks to generalize to new drugs, in part because the distance between atoms does not generally correspond to the distance between their representation in the SMILES strings. Graph-attention networks, on the other hand, are high-capacity models that require large training-data volumes which are not available for drug-sensitivity estimation. We develop a modular drug-sensitivity graph-attentional neural network. The modular architecture allows us to separately pre-train the graph encoder and graph-attentional pooling layer on related tasks for which more data are available. We observe that this model outperforms reference models for the use cases of precision oncology and drug discovery; in particular, it is better able to predict the specific interaction between drug and cell line that is not explained by the general cytotoxicity of the drug and the overall survivability of the cell line. The complete source code is available at https://zenodo.org/doi/10.5281/zenodo.8020945. All experiments are based on the publicly available GDSC data.

Introduction

The treatment of cancer relies on standard-of-care therapies that fail to address the diverse nature of the disease. Cancer is caused by combinations of genomic alterations of cells; in combination with the biochemical mechanisms underlying drugs these alterations cause high variations in sensitivity to drug compounds across cancer cells, and ultimately lead to diverse therapeutic outcomes for seemingly similar clinical presentations (1). Advances in sequencing technology and the availability of large-scale drug-sensitivity screening databases—such as the Genomics of Drug Sensitivity in Cancer Database (GDSC) (2,3) and the Cancer Cell Line Encyclopedia (CCLE) (4)—are driving the development of *precision oncology* (5,6). Precision oncology aims at replacing broadly applicable chemotherapies that are toxic for healthy cells as well (7) by targeted drugs that are tailored to the transcriptomic profile of a given primary tumor.

Increasingly, machine-learning approaches are used to facilitate drug-sensitivity estimation (8). Specifically, deep-learning models show promise at predicting the sensitivity of cell lines to drugs (9). Cell lines can be represented by their transcriptomic features as input to neural networks, and the neural network can learn to map this input to a representation that captures properties which determine sensitivity to specific drug mechanisms.

Representing chemical molecules in a way that enables the neural network to generalize well across drugs, on the other hand, is more challenging. State-of-the-art models such as PaccMann (9) use the Simplified Molecular-Input Line-Entry System (SMILES) (101). However, this representation has less desirable properties that impede generalization. Firstly, the code is not unique; a molecule generally has multiple SMILES strings. This necessitates data augmentation and imposes a challenge on the neural network that has to learn a meaningful internal representation. Secondly, the distance between elements in the string does not always correspond to physical distance between the entities, which makes it more challenging for the network to understand physical interactions between elements that are far apart in the string.

Recently, diverse deep-learning architectures for graph data that are referred to as *graph neural networks (GNNs)* have been created. They can predict properties of the nodes, of the edges or properties of the entire graph (11). Specifically, GNNs estabish new state-of-the-art results for drug-discovery-related benchmarks (12,13). GNNs have also been applied for cancer-related tasks, using graph representations for either drugs (14–17), cell lines (18,19), or both (20,21). In this work, we developed and evaluated graph neural network architectures that allow neural drug sensitivity models to process the graph structure of candidate drug molecules, in addition to a transcriptomic tumor profile.

In order to comprehensively evaluate the predictive performance of the models, we consider two distinct use cases that reflect highly relevant challenges in cancer research: *precision oncology* and *drug discovery*. In *precision oncology*, the primary goal is to identify the optimal treatment strategy for an individual patient. In this scenario, the transcriptomic profile of the patient's primary tumor is available, enabling a personalized approach to treatment selection. However, the challenge

Received: November 23, 2023. **Revised:** March 1, 2024. **Editorial Decision:** April 15, 2024. **Accepted:** April 17, 2024

lies in the absence of drug screening data for the patient's specific case. Therefore, the task in precision oncology can be framed as predicting the sensitivity of a new, previously unseen cell line to a known panel of drugs. By contrast, the *drug discovery* use case focuses on identifying promising candidate drug compounds for further development and clinical studies. Here, the sensitivity of a known panel of cell lines to a new, previously unseen drug molecule is predicted. By prioritizing potential candidates for further investigation, the predictions help reducing the reliance on resource-intensive and time-consuming experimental validations.

Most prior work has evaluated drug sensitivity models in terms of their Pearson correlation between predicted and observed sensitivity (9,18). We argue that this criterion is misaligned with the actual goal of either use case. The sensitivity of a pair of a cell line and a drug can be decomposed into a mean value for the cell line that reflects its general ability to resist treatment, a mean value for the drug which reflects its general toxicity, and a specific interaction residual. In the precision-oncology use case, the model should not be rewarded for delivering accurate predictions of general drug toxicities, because in this setting the toxicity of available drugs is well established. Similarly, in the drug development use case, precise predictions of cell-line survivability should not be the basis for evaluating the model's performance. We argue that the correlation between predicted and observed interaction residuals is a better quantification of the model's merit; it quantifies the extent to which the drug can impede specific cellular mechanisms rather than the drug's general cytotoxicity or the cell line's survivability.

In total, this manuscript makes the following contributions. (i) We propose a new performance metric for drug sensitivity models that is closely tied to the models' merit in the use cases of precision oncology and drug discovery. (ii) We develop a modular deep neural network for drug-sensitivity estimation. The modular architecture allows the drug module to be pre-trained on separate tasks with abundant training data. This enables the model to process drug molecules in a rich graph representation instead of as SMILES codes, and to encode molecules using high-capacity graph-attention layers. (iii) We find that the developed network architecture is substantially better than known reference methods. An ablation study sheds light on the contributions of the modular architecture, pre-training, and the graph encoder. (iv) We observe that the new model assigns importance to genes that are functionally more focused than the state of the art.

Materials and methods

Problem setting and performance metrics

We studied two variants of problem settings that represent the use cases of *precision oncology* and *drug discovery*, respectively. The goal of *precision medicine* is to predict the effect of a range of available drugs for a given, previously unseen, tumor case. Therefore, performance metrics for this use case were measured for cell lines that did not occur in the training data. Drug sensitivity is measured in terms of the inhibitory concentration IC_{50}. For precision medicine, we used the mean value, across cell lines, of the *Pearson correlationR* between predicted and measured inhibitory concentration for all drugs as the first reference performance measure. We refer to this quantity as *R (precision oncology)*. In addition, the *mean squared error (MSE)* of the predicted IC_{50} served as reference measure.

For *drug discovery*, performance metrics were calculated for drug compounds that were not present in the training data. Here, we measured the mean value, across drugs, of the *Pearson correlationR* between predicted and measured inhibitory concentration for all cell lines as the first reference performance measure. We refer to this measure as *R (drug discovery)*. As above, the *mean squared error (MSE)* was used as second reference measure.

In both settings, the sensitivity y_{ij} of cell line j to drug i can be decomposed into a mean value α_i of drug i that reflects its toxicity, a mean value β_j of cell line j that reflects is susceptibility, and a residual term γ_{ij} that reflects the specific interaction of drug and cell line:

$$y_{ij} = \alpha_i + \beta_j + \gamma_{ij}. \tag{1}$$

Even though the true values of α_i and β_j are not known, they can be estimated robustly in practice by fitting a linear model with parameters α_i and β_j to a given matrix of predicted or observed sensitivity values y_{ij}; the interaction residuals γ_{ij} follow immediately from $\gamma_{ij} = y_{ij} - \alpha_i - \beta_j$. In later sections, we will refer to estimations of these values based on the existing data as $\widehat{y_{ij}}$, $\widehat{\alpha_i}$, $\widehat{\beta_j}$ and $\widehat{\gamma_{ij}}$, and in more abstract references to this decomposition they will be denoted by $\widetilde{y_{ij}}$, $\widetilde{\alpha_i}$, $\widetilde{\beta_j}$ and $\widetilde{\gamma_{ij}}$. Note that the linear decomposition of Equation 1 is always possible and generally has multiple solutions: setting all α_i and β_j to zero leads to the trivial but exact solution $y_{ij} = \gamma_{ij}$. Applying ℓ_2-regularization to the α_i and β_j and using the square of the residuals $\gamma_{ij} = y_{ij} - \alpha_i - \beta_j$ as loss function favors solutions that place large weights on the effects of α_i and β_j.

Since any drug sensitivity y_{ij} is an aggregate of drug mean, cell-line mean, and interaction term, the Pearson correlation penalizes any deviation between prediction and ground truth in any of these constituents equally. We argue that this misaligns the performance metric from the actual, use-case-driven goal. For precision oncology, the model's ability to predict the overall cytotoxicity α_i is less relevant because it is generally known, and the ability to predict the tumor survivability β_j is less relevant because it is not the subject of the therapeutic decision. For drug discovery, both drug toxicity and tumor survivability are issues that are relevant, but disparate from identifying tumors that a drug candidate might target effectively. In both cases, the interaction residuals quantify the extent to which the drug is an effective match for the tumor at hand.

For *precision medicine*, we therefore measured the mean, across cell lines, of the *Pearson correlationR* between predicted and measured interaction residuals, $\hat{\gamma}_{ij}$ and γ_{ij} respectively, for all drugs as an additional performance measure. We refer to this quantity as *R interaction (precision oncology)*. For *drug discovery*, we measured the mean value, across drugs, of the *Pearson correlationR* between predicted and measured inhibitory concentration for all cell lines as a additional performance measure that we refer to as *R interaction (drug discovery)*.

Note that the *Pearson correlation* is translation and scale invariant. A predictive model that scales the inhibitory concentrations incorrectly while sorting compounds for precision oncology—or cell lines, for drug discovery—perfectly according to their true inhibitory concentrations can achieve a perfect Pearson correlation.

We judged the statistical significance of differences in the performances of models with two-sided paired Student's *t*-tests with *P*-value threshold of between 0.05 and 10^{-4}, and the resulting *P*-values were corrected for repeated testing using the Holm-Šídák method.

Model architecture

This section develops the architecture of *CANDELA*, a *cancer drug sensitivity estimation modular graph neural network*, visualized in Figure 1. The network accepts a graph that encodes a drug component *i*, gene expression data of a cell line *j*, and produces a prediction $\hat{y}_{ij}$ of the inhibitory concentration IC_{50}. The network architecture follows the intuition of a decomposition of the inhibitory effect into a mean value α_i that reflects the cytotoxicity of drug *i*, a mean value β_j that reflects the survivability of tumor cell line *j*, and an interaction residual γ_{ij} that quantifies the extent to which molecule *i* specifically impedes cellular mechanisms employed by tumor cell *j*. The following sections describe the individual modules; all hyper-parameter values are the result of a tuning procedure described in the Sections below.

Drug encoder and drug module

Graph representations of molecules have been explored for different biochemical problems (22), and a wide range of descriptors encodes chemical properties of atoms, bonds and their neighborhood in the molecule as node and edge attributes, respectiely (23–25). In our experiments, node features are the degree of each node, the atom type, the number of neighboring heavy atoms, the charge of that atom, the hybridization type, binary variables that indicate whether the atom is contained in a ring and whether it is contained in an aromatic ring, the mass of the atom, its scaled van der Waals radius, and its scaled covalent radius. Edge annotations are the type of edge (single, double, triple or aromatic), and binary variables indicating whether the edge is conjugated or not, and whether the edge is part of a ring. All discrete annotations are one-hot encoded.

In the *drug encoder*, three consecutive stacks of a graph attention layers (26) with eight attention heads each, interleaved by a LeakyReLU (leaky rectified linear) activation functions, generate an embedding of the nodes with a dimensionality of 128 per node. In the *drug module*, self-attention pooling (27) with a hidden dimension of 1024 is used to aggregate the node embeddings into an embedding of the drug molecule with 128 dimensions. Finally, a dense layer of 2027 units for drug discovery and 1152 units for precision medicine, followed by a sigmoidal activation function and another dense layer that generates a drug score is applied to the drug embedding. The generated scalar value corresponds to $\widetilde{\alpha}_i$ in equation 1.

Expression encoder and expression module

Drawing inspiration from PaccMann (9), cell lines are represented in terms of the expression levels of genes selected using network propagation. The list of 2,089 genes was extracted from the original publication. They correspond to the genes with the highest random-walk probability to the genes targeted by each drug. Details are elaborated by Manica *et al.* (9). In the *expression encoder*, two fully-connected layers with ReLU (rectified linear) activations generate an embedding of the expressions for the 2,089 genes, with an initial dimension of 1024 and a bottleneck dimension of 128.

In the *expression module*, a multi-layer perceptron (MLP) transforms the encoded gene expression into an scalar score, corresponding to $\widetilde{\beta}_i$ in equation 1. The MLP has an embedding size of 660 for the precision-oncology setting and 1310 for drug discovery.

Interaction module and score generation

This subnetwork combines the expression and drug embeddings into a single score, which captures the specific pairwise interaction between the drug and the cell-line. We adapt the pooling strategy of Manica *et al.* (9) and apply cross-attentional pooling (27) of the embeddings of each node with the cell-line embedding into a joint embedding of drug *i* and cell line *j*. Then, the graph-level features are transformed by a ReLU activation function, followed by a linear layer with hidden size 1,212 (precision oncology) or 3744 (drug discovery), a sigmoid activation function, and a linear layer that maps the hidden features to a scalar value which corresponds to $\widetilde{\gamma}_{ij}$ in Equation (1).

Finally, the three scalar scores resulting from the gene-expression, drug, and interaction modules are simply summed into a scalar prediction $\hat{y}_{ij}$ of the final output of $\log(IC_{50})$ predictions. In consequence, this architecture offers additional possibilities for interpretability, regularization, and separate pre-training of drug and expression modules.

Pre-training the graph encoder

The *Genomics of Drug Sensitivity in Cancer project (GDSC)* data (2,3)—the largest publicly available drug-sensitivity database—includes fewer that 1000 cell lines and 282 drugs. *CANDELA*, on the other hand, has >40 million parameters that allow it to model complex relationships between the molecular structure of drugs, transcriptomes, and inhibitory effects. In order to bridge this disproportion between training data and model complexity, the drug encoder can be pre-trained on tasks for which training data are more abundant. Pre-training encoders on different but related tasks has proven to be a powerful tool for computer vision and natural language processing, and it can be a useful strategy for graph-learning tasks (28).

The drug module that generates the node embeddings as well as the self-attentional pooling layer that aggregates these node embeddings into a drug embedding were pre-trained. Specifically, we study the merits of the following two pre-training tasks.

Pre-training the drug encoder to predict metabolite properties

Molecular properties, such as their sizes or octanol–water partition coefficients are known to play a determinant role in their interaction with proteins, and hence in their biological properties (29). For this reason, and given the large amount of labeled data, we designed an initial pre-training task where the model learns to predict such properties.

First, 123 559 drug- and metabolite-like compounds of the human interactome were extracted from the STITCH database (30). Compounds that also occur in GDSC were discarded to prevent leakage of test data into the training process. Then, using the PubChem API (31), seven numerical features that were available for all compounds were retrieved: the molecular weight, octanol/water partition coefficient, polar surface area, complexity, hydrogen bond donor count,

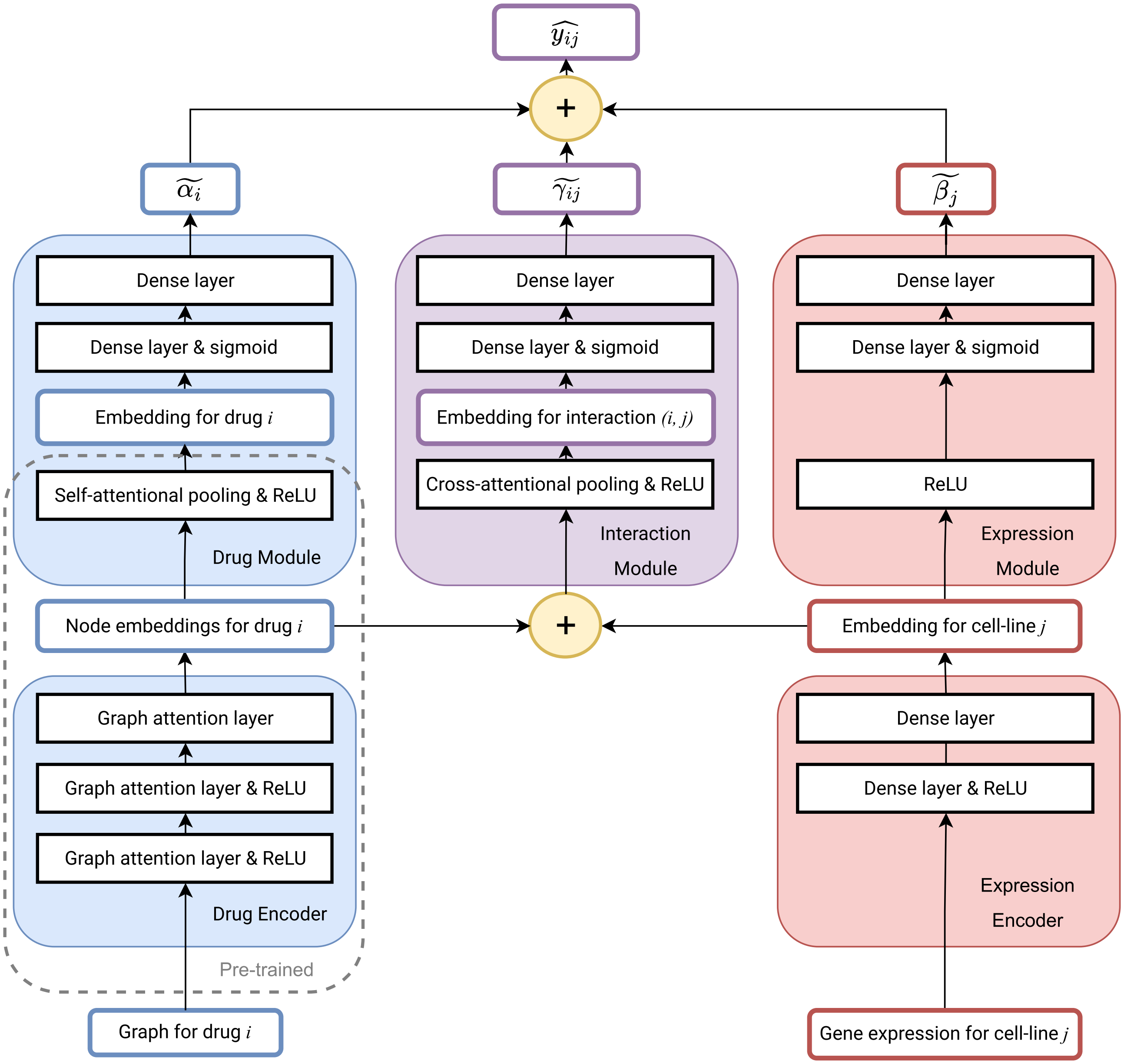

Figure 1. Outline of the proposed modular *CANDELA* architecture.

hydrogen bond acceptor count, and rotatable bond count. This data served as ancillary data for this pre-training step.

These features were scaled to the range of –1 and 1, and were used as multi-output regression target for pre-training the drug encoder. To this end, an attention pooling layer, a fully-connected layer with 512 hidden units with a dropout rate of 0.4, and an output layer with seven units were stacked on top of the embedding layer of the drug encoder. The network was trained to minimize the mean squared error for 1000 epochs using Adam with a learning rate of 3×10^{-4} and a batch size of 256. The pre-trained graph-attentional encoder layers are subsequently used for initializing the weights of the drug encoder in the next pre-training task.

During this pre-training task, the hyper-parameters of the architecture were selected using Bayesian Optimization Hyperband (BOHB) (32), which combines a Bayesian tree-structure Parzen estimator for the selection of promising configurations and Hyperband for early stopping and resource allocation into a single, robust framework. The pre-training data sets were split into a training portion, a validation portion for early stopping, and an evaluation portion.

Hyper-parameters included parameters of the Adam optimizer (the learning rate, weight decay and the beta terms), the drug encoder architecture (number of graph-attention layers, number of attention heads and dimension of the node embeddings), the self-attentional pooling layer (number of attention heads and dimension of the drug embedding), and the ancillary part of the network (number of units in the hidden layer and dropout probability in the regression or classification head). The hyper-parameters selected were those minimizing the loss on the evaluation portion of the ancillary data.

Pre-training the drug encoder to predict toxicity

The toxicity of drugs plays a vital role in their success as cancer therapies (33), and furthermore, many anticancer drugs are specifically cytotoxic compounds (34). For these reasons, and given the amount of labeled data in public databases, we studied a second pre-training strategy related to toxicity.

The CATMoS database (35) that contains 11 992 molecules classified as non-toxic, moderately toxic and highly toxic compounds constituted the ancillary data for this second pre-training step. A network consisting of the pre-trained drug encoder obtained in the previous step, an additional attention pooling layer, and a fully-connected classification head with 512 hidden units and a dropout probability of 0.3 was built. It was trained to minimize the cross-entropy loss between the predicted and observed classes for 30 epochs, using the Adam optimizer with a learning rate of 2.5×10^{-4} and a batch size of 512. The weights learned by the drug encoder and the attention-pooling module are then used as initialization weights for the task of drug sensitivity prediction.

During this pre-training task, only hyper-parameters of the Adam optimizer and architecture parameters of the ancillary part of the network were selected using BOHB. The pre-training data sets were split into a training portion, a validation portion for early stopping, and an evaluation portion of the ancillary data.

Training the CANDELA model

After pre-training the drug encoder and drug self-attentional pooling following two different pre-training strategies, the CANDELA model was integrated and trained on GDSC. For the first pre-trained model only the molecular property prediction task was considered, whereas the second model was pre-trained on both tasks sequentially. The remaining hyper-parameters were obtained using BOHB and a three-way split cross-validation schema consisting of 25-fold, where iteratively 23-fold were used for training the model, one was used for selecting hyper-parameters and the last one was used for evaluating the performance attained after hyper-parameter selection. The hyper-parameters tuned were the configuration of the Adam optimizer (learning rate, weight decay and beta terms), the number of hidden units and dropout probabilities in the corresponding fully-connected networks, the number of hidden units found in the cell-line encoder, and the architecture parameters of model components that were not pre-trained (number of graph-attention layers, number of attention heads, number of hidden units). In all cases, the selected hyper-parameters were those minimizing the mean squared error averaged over the tuning portions of the 25 validation folds.

Fitting the modular scores

Determining the Pearson correlation of the interaction residuals (R interaction) requires the true coefficients α_i, β_j and γ_{ij} and the predicted coefficients $\hat{\alpha}_i$, $\hat{\beta}_j$ and $\hat{\gamma}_{ij}$ to be known for all drugs and cell lines. These coefficients were fitted to the observed and predicted values of IC_{50}, respectively, using a ridge regression with regularization factor $\alpha = 10^{-5}$.

Reference models

This section introduces reference methods that represent the state of the art and serve as baselines in the experimental comparison.

DrugCGN

represents cell lines as graphs that are obtained from the protein-protein interaction (PPI) network (18). A separate model is learned for each drug using the cell-line graphs as input and a GNN as an encoder. In consequence, this approach can only generalize to unseen cell lines. Since no information is shared across models, drugs for which the screening data contained a large number of missing values were discarded and the model was trained to minimize the MSE of IC_{50} values using only a subset of GDSC.

GraphDRP

represents cell lines by a vector of binary features that indicates which of their genes are mutated (14). Drugs are represented using graphs. Note that this representation can generalize to unseen drugs and cell lines as long as mutation data are available. An MLP encodes gene expression levels, a GNN performs message passing on the drug graphs, and node features are max-pooled over all nodes, which results in a graph-level representation. Finally, the features are combined via concatenation, and the predictions are obtained thanks to a series of fully-connected layers. The learning objective is the MSE of the predicted IC_{50}s.

PaccMann

represents cell lines by their expression profiles in the form of continuous values, and drugs by the SMILES string for each drug (9). This representation can generalize to both unseen drugs and unseen cell lines. Although they present different versions of the model, the gene expression data is encoded using a multi-layer perceptron (MLP) and the drugs are encoded using CNNs. Both representations are fused using attention over the cell-lines to pool different convoluted sequences obtained from each drug, and the learning objective is to minimize the mean squared error (MSE) between the observed and predicted IC_{50}.

SubCDR

represents cell lines by cancer-driving genes taken from the COSMIC database (36), masks genes that are not relevant for a particular tumor type, and applies a one-dimensional CNN to infer an embedding. Drugs are fragmented using the BRICS algorithm (37), molecular fragments represented by extended-connectivity fingerprints (38) and are embedded by a gated recurrent layer. Additionally, this method extracts side information by applying Single Value Decomposition to the drug-responses matrix. Finally, the features are concatenated and an MLP generates the final $\log(IC_{50})$ predictions (39). Note that SVD generates representations that do not generalize for unseen drugs or cell lines; for this reason, we discard the drug embeddings during the drug discovery experiments, and the cell-line embeddings during the precision oncology experiments.

PaccMann

In PaccMann (9), cell lines are represented by their expression profiles in the form of continuous values, and drugs are represented by the SMILES string for each drug. This representation can generalize to both unseen drugs and unseen cell lines. Although they present different versions of the model, the gene expression data is encoded using a multi-layer perceptron (MLP) and the drugs are encoded using CNNs. Both representations are fused using attention over the cell-lines to pool different convoluted sequences obtained from each drug, and the learning objective is to minimize the mean squared error (MSE) between the observed and predicted IC_{50}.

3D Infomax

pre-trains a graph neural network to encode information about the 3D structure of molecules on databases of molecular properties. To this end, it is trained to maximize the mutual information between 2D and 3D representations of molecules (40). The resulting molecular embedding can be used as input representation for a range of downstream tasks. The molecular architecture of CANDELA was designed specifically to make use of pre-trained drug embeddings. As an alternative to pre-training the drug encoder on matabolite properties and toxicity, we have studied the use of the 3D Infomax encoding for drugs.

Data

The *Genomics of Drug Sensitivity in Cancer project (GDSC)* database (2,3) contains screening data for tumoral cell lines under different anticancer treatments. It contains two different experiments: GSDC1 and GDSC2 (41). GDSC1 has a larger number of experiments (up to 345 different compounds were screened), whereas the GDSC2 started later and tried to improve and standardize the quality of the screening methodology. Since it still contains the lower number of 192 screened compounds, GDSC1 is used in our experiments. Cell lines cover the spectrum of common and rare types of adult and childhood cancers of epithelial, mesenchymal and haematopoietic origin. Cell lines were characterized using data from six different sources (whole exome sequencing, gene expression, copy number alterations, DNA methylation, gene fusions and microsatellite stability). GDSC1 contains 310 904 IC_{50} values for 987 cell lines and 367 compounds. The IC_{50} is measured using fluorescence-based cell viability assays following 72 h of drug treatment.

Results

In this section, the performance of CANDELA will be compared to that of state-of-the-art models under the precision medicine and drug discovery settings, and also against ablated versions. The global performance of the models and their ability to recover the different latent features found in the data will be analyzed. To further understand how CANDELA works, the features driving the predictions will be studied.

Comparison of methods

For the use-case of precision oncology, Figure 2A shows that both versions of CANDELA outperform all reference methods with respect to MSE, Pearson correlation R between predicted and measured IC_{50} values, and Pearson correlation between predicted and measured interaction residuals of IC_{50} values at Holm–Šídák-corrected significance levels of 0.01 and lower. For drug discovery, Figure 2B again shows that both versions of CANDELA outperform all reference models with respect to all three performance measures at significance levels of 0.05 and below.

In both settings, CANDELA shows a significantly higher Pearson correlation between predicted and measured interaction residuals of IC_{50} values. This means that CANDELA's performance cannot just be attributed to a better ability to predict drug cytotoxicity or cell-line survivability; it specifically excels at predicting whether a drug has an inhibitory interaction with a particular cell line. When comparing in detail the distributions of observations vs. predictions, and their squared residuals in Figure 3, it can be seen that under both settings CANDELA has a higher density of predictions with low residual errors, and that furthermore the predictions with low value—which correspond to strong inhibitions—are more accurately predicted.

Extended omics-representations

CANDELA represents the cell-lines using exclusively the expression level of genes. While mutation and copy-number variations are often known for established cell lines, restricting CANDELA to expression-level input makes the method more broadly applicable in precision-oncology and drug-discovery scenarios. Nevertheless, we further investigated if the addition of other sources of omics data leads to better performance. We added the copy number variations (CNVs) and binary features representing mutation events as additional input features. These features are available for 902 of the cell lines in GDSC; we performed 10-fold cross validation on these cell lines along cell lines and drugs for precision oncology and drug discovery, respectively.

Table 1 shows that performance differences between CANDELA with and without copy-number-variation and mutation information are minimal and statistically insignificant.

Ablation study

We compare the performance of the CANDELA model to various ablated versions of CANDELA. We distinguish between CANDELA pre-trained on both pre-training tasks (marked by 'two tasks' in column 'pre-training' of Table 2) and CANDELA pre-trained only on metabolite properties (marked as 'one task'). We compare these models to an ablated version without any pre-training (marked as '×' in column 'pre-training'). Additionally, we also trained CANDELA with pre-trained node embeddings obtained using the 3D Infomax method (40).

In an ablated version *without score decomposition*, the drug module and expression module were removed; the interaction module is trained to predict the inhibitory concentration by itself. The modular drug encoder and expression encoder remain and can be pre-trained. The model *without score decomposition and pre-training* combines both ablations. In a model *without score decomposition, graph attention, and pre-training* we replace each graph-attention layer by a graph-convolutional layer with equally many units; as a result, this model cannot process the edge features.

CANDELA fuses the node embeddings of drug encoder and expression encoder by cross-attentional pooling (marked as 'cross-attention') in column 'fusion' of Table 2. We also study an ablated version in which the embeddings are concatenated, and a subsequent fully-connected layer can learn any nonlinear integration function (marked as 'concatenation').

Table 2 shows the results and highlights configurations that perform significantly better and worse than the CANDELA model, based on a pairwise t-test with $\alpha = 0.05$. The P-values were corrected for multiple testing using the Holm–Šídák method. For both precision oncology and drug discovery, Table 2 shows that pre-training on metabolite properties is uniformly better than both pre-training on two tasks, not pre-training and the usage of pre-trained embeddings obtained

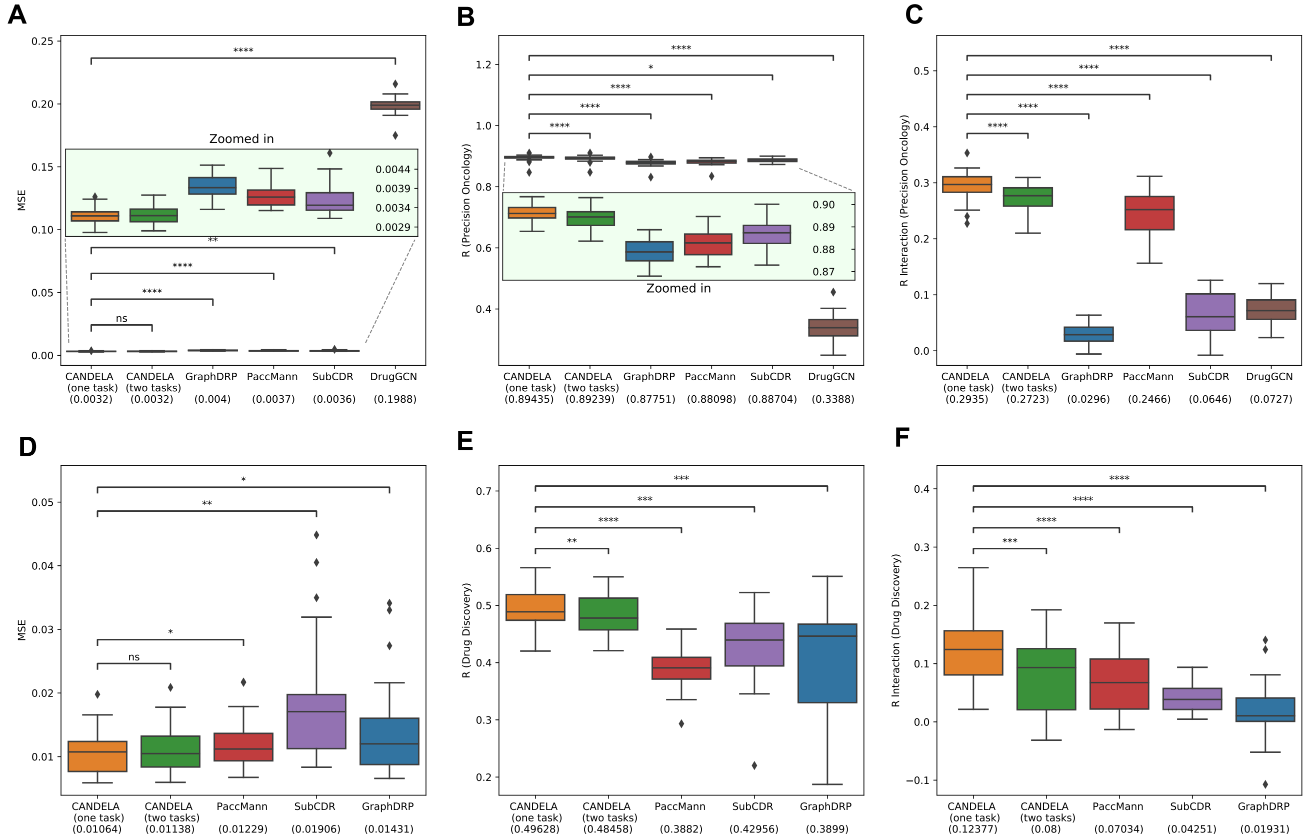

Figure 2. Performance measures for CANDELA and reference models for precision oncology (**A–C**) and drug discovery (**D–F**). 'CANDELA (one task)' refers to the model pre-trained on only metabolite properties. 'CANDELA (two tasks)' was pre-trained on metabolites properties and toxicity sequentially. Boxes display the median value and interquartile range; whiskers extend up to the most extreme data point within 1.5 IQR, other points are considered outliers. The upper bars depict the result of the corrected *P*-values obtained from *t*-tests, and the number of asterisks corresponds to the Holm–Šídák-corrected significance levels of *: $0.01 < P \leq 0.05$, **: $10^{-3} < P \leq 0.01$, ***: $10^{-4} < P \leq 10^{-03}$, ****: $P \leq 10^{-04}$.

from 3D infomax; for most configurations, the deterioration is statistically significant. For drug design, only the MSE of two pre-training tasks is slightly (but insignificantly) lower than for one pre-training tasks.

This outcome led us to the hypothesis that the number of molecules in the toxicity data (11 992) is too small, and the molecules are too different from the molecules in GDSC, for this pre-training task to benefit the model. The STITCH database of metabolite properties is larger by an order of magnitude (123 559 molecules). We quantified the similarity similarity between the molecules found in GDSC with the molecules found in STITCH and CATMos. For each query molecule from the GDSC, we selected 10 random samples of 512 target molecules from each of these datasets, and we calculated the Tanimoto coefficient between the query molecule and each target molecule. We then compared the mean over all batches for the maximum observed Tanimoto coefficient in each batch. We observed that the average maximum Tanimoton coefficient between GDSC and STITCH is 0.304 while the average maximum Tanimoto coefficient between GDSC and CATMoS is 0.255. The similarity between GDSC and STITCH is 18.7% larger than the similarity between GDSC and CATMoS, which is consistent with our hypothesis that CATMoS is too different to be beneficial for pre-training.

For precision oncology, the ablated model without score decomposition performs marginally better than CANDELA. For drug design, CANDELA significantly outperforms the ablated model without score decomposition. The ablated model without score decomposition has fewer model parameters, and while higher-capacity networks can model more complex relationships, they require more training data to avoid over-fitting. Our interpretation is that the score decomposition is beneficial for drug design, while a lower-capacity model turns out to offer a better trade-off for precision oncology.

For both precision oncology and drug design, ablated versions without graph attention perform significantly worse. Also for both problems, fusion by concatenation plus a fully connected layer performs worse than fusion by cross-attention, albeit the deterioration is not significant.

For both precision oncology and drug discovery, using the pre-trained 3D Infomax embedding deteriorates CANDELA's performance compared to pre-training the embedding on STITCH. While 3D Infomax encourages the drug embedding to preserve information about the spatial structure of the molecule, we conclude that predicting metabolites is a more closely related, meaningful pre-training task.

Cytotoxic versus targeted drugs

GDSC contains cytotoxic, targeted, and other drugs. Given the radical differences in the mechanism of action between cytotoxic and targeted drugs, one could expect differences in performance for the different models between these classes of drugs compounds. For this reason, we further study the ability

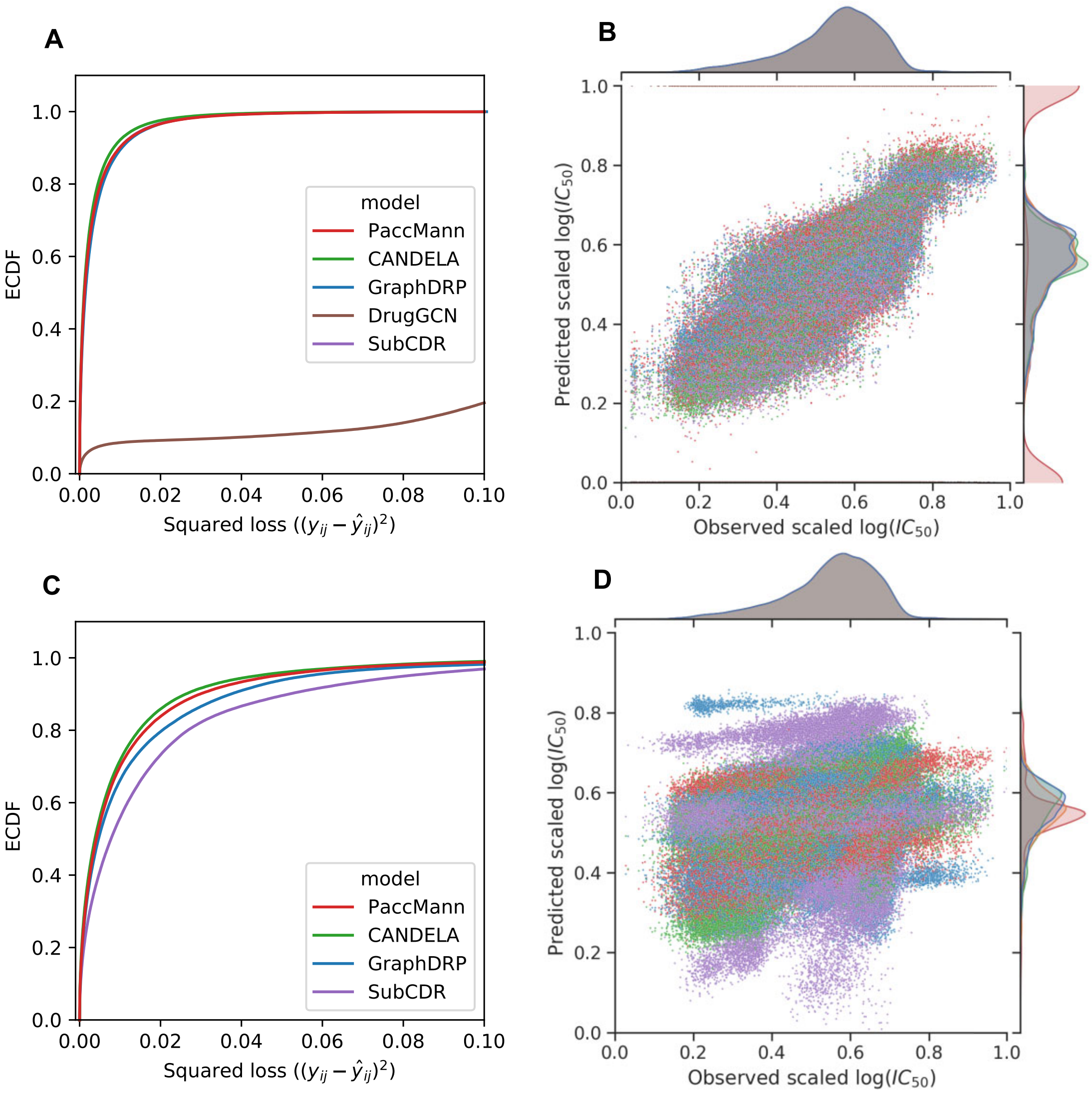

Figure 3. (**A**) Empirical cumulative distribution function of the squared loss $(y_{ij} - \hat{y}_{ij})^2$ for precision oncology; (**B**) Comparison of observed and predicted values of log(IC_{50}) for precision oncology; (**C**) empirical cumulative distribution function of the squared loss $(y_{ij} - \hat{y}_{ij})^2$ for drug discovery; and (**D**) comparison of observed and predicted values of *log*(IC_{50}) for drug discovery.

Table 1. CANDELA with and without copy-number variation and mutation features: performance measures ± standard error of the mean

		Metric		
Setting	CNVs & mutations	MSE	*R*	*R* interaction
Precision oncology	×	0.003 ± 0.000	0.901 ± 0.002	0.327 ± 0.009
	✓	0.003 ± 0.000	0.899 ± 0.001	0.327 ± 0.005
Drug discovery	×	0.012 ± 0.001	0.468 ± 0.006	0.110 ± 0.009
	✓	0.012 ± 0.001	0.475 ± 0.007	0.125 ± 0.010

of the different models to perform predictions for new drugs of these two classes.

First, we compared the performance achieved by CANDELA and reference methods in the drug discovery setting between cytotoxic and targeted compounds. Figure 4 shows that for targeted compounds, the differences in performance are very similar to the overall performance in terms of *R* (drug discovery), but the absolute values of *R* (drug discovery) are higher. This is unsurprising, due to targeted drugs constituting the larger fraction of GDSC1, and also having killing patterns that are, in terms of biological expectations, more predictable. Interestingly, for cytotoxic compounds, the performance of all models is higher than for targeted drugs, but this effect is more extreme for CANDELA and SubCDR.

We further analyzed the ability of the different ablated versions of CANDELA to predict the response of cytotoxic and targeted drugs in the drug-discovery setting. We found that the average correlations for the targeted drugs using the modular

Table 2. Ablation study: performance measures ± standard error of the mean. '✓' indicates included, '×' indicates excluded model components; the first two rows are the complete CANDELA model with the two different pre-training stragies. Models marked '†' are significantly worse than CANDELA using only one pre-training task. Models marked '*' are significantly better (pairwise *t*-test, $\alpha = 0.05$, corrected for multiple testing using Holm–Šídák). Bold values indicate the best configuration

	Model components				Metric		
	Pre-training	Score decomposition	Graph attention	Fusion	MSE	*R*	*R* interaction
Precision oncology	Two tasks	✓	✓	Cross-attention	0.0032 ± 0.0000	0.8929 ± 0.0023†	0.2743 ± 0.0057†
	One task	✓	✓	Cross-attention	0.0032 ± 0.0000	0.8941 ± 0.0022	0.2926 ± 0.0059
	3D Infomax	✓	✓	Cross-attention	0.0045 ± 0.0002†	0.8632 ± 0.0062†	0.1210 ± 0.0061†
	×	✓	✓	Cross-attention	0.0205 ± 0.0002†	0.6765 ± 0.0097†	0.0801 ± 0.0034†
	Two tasks	×	✓	Cross-attention	0.0034 ± 0.0000†	0.8865 ± 0.0024†	0.2600 ± 0.0061†
	One task	×	✓	Cross-attention	**0.0031 ± 0.0000***	**0.8957 ± 0.0023***	**0.3557 ± 0.0055***
	×	×	✓	Cross-attention	0.0035 ± 0.0001†	0.8832 ± 0.0025†	0.1228 ± 0.0157†
	×	×	×	Cross-attention	0.0034 ± 0.0000†	0.8830 ± 0.0024†	0.2070 ± 0.0074†
	Two tasks	✓	✓	Concatenation	0.0034 ± 0.0000†	0.8819 ± 0.0023†	0.0263 ± 0.0038†
	One task	✓	✓	Concatenation	0.0034 ± 0.0000†	0.8821 ± 0.0024†	0.0364 ± 0.0056†
Drug discovery	Two tasks	✓	✓	Cross-attention	0.0114 ± 0.0007	0.4846 ± 0.0071	0.0800 ± 0.01300†
	One task	✓	✓	Cross-attention	**0.0106 ± 0.0007**	**0.4963 ± 0.0068**	**0.1238 ± 0.01300**
	3D Infomax	✓	✓	Cross-attention	0.0112 ± 0.0008	0.4339 ± 0.0070†	0.0269 ± 0.0061†
	×	✓	✓	Cross-attention	0.0136 ± 0.0009	0.4829 ± 0.0069	0.0281 ± 0.0099†
	Two tasks	×	✓	Cross-attention	0.0123 ± 0.0009	0.4860 ± 0.0070	0.0335 ± 0.0109†
	One task	×	✓	Cross-attention	0.0123 ± 0.0010	0.4128 ± 0.0069†	0.0199 ± 0.0078†
	×	×	✓	Cross-attention	0.0126 ± 0.0010	0.4354 ± 0.0075†	– 0.0123 ± 0.0105†
	×	×	×	Cross-attention	0.0127 ± 0.0010	0.4038 ± 0.0095†	0.0203 ± 0.0060†
	Two tasks	✓	✓	Concatenation	0.0116 ± 0.0007	0.4675 ± 0.0068†	– 0.0122 ± 0.0098†
	One task	✓	✓	Concatenation	0.0116 ± 0.0008	0.4954 ± 0.0068	0.0091 ± 0.0074†

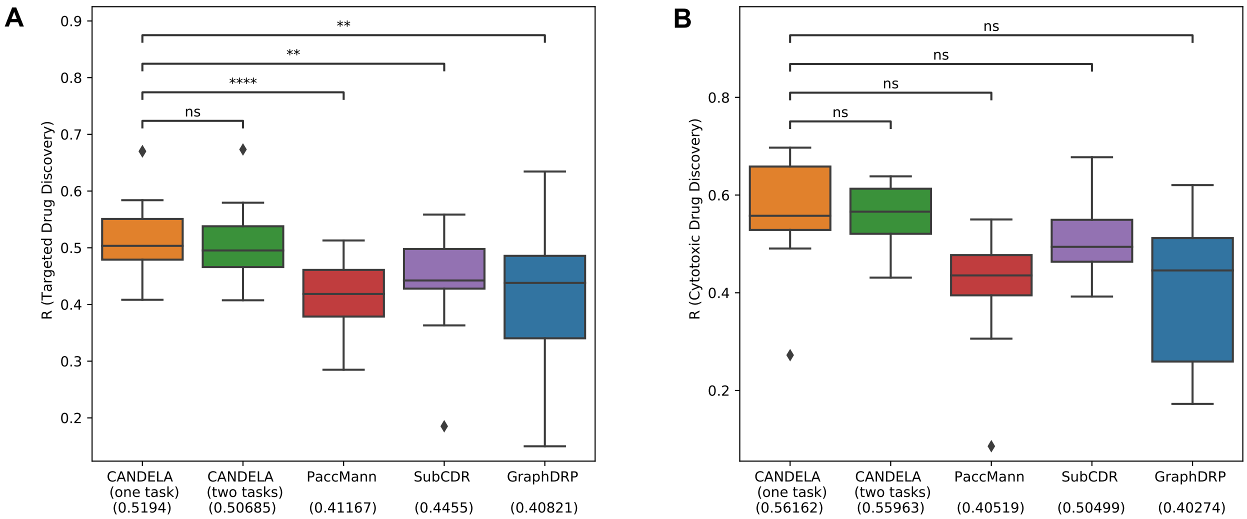

Figure 4. Pearson correlation coefficient between observed and predicted values for (**A**) targeted drugs and (**B**) cytotoxic drugs for CANDELA and reference methods. The upper bars depict the result of the corrected *P*-values obtained from *t*-tests, and the number of asterisks corresponds to the Holm–Šídák-corrected significance levels of ns: not significant, $*0.01 < P \leq 0.05$, $**10^{-3} < P \leq 0.01$, $***10^{-4} < P \leq 10^{-03}$, $****P \leq 10^{-04}$.

architecture with one pre-training task ($R = 0.519 \pm 0.02$) or two pre-training tasks ($R = 0.507 \pm 0.02$) is in a similar range when compared to other baselines without score decomposition ($R = 0.511 \pm 0.012$). In contrast, when the predictions for cytotoxic compounds were analyzed, we observed that CANDELA with one pre-training task ($R = 0.561 \pm 0.025$) and two pre-training tasks ($R = 0.560 \pm 0.013$) has better performance than the best-performing baseline without score decomposition ($R = 0.517 \pm 0.019$).

Biomarker importance

In order to challenge the biological plausibility of the features that appear important for judging either resistance or sensitivity of cell lines to drugs, we extracted the mean overall absolute feature importances for each of the 2,089 genes across all drug-cell line combinations using integrated gradients (42). Thus, a feature gained importance if it contributes to explain either resistance or sensitivity with respect to drug-cell line combinations. Next we argue that plausibility of these important features can be assessed with respect to the enrichment of the gene sets in biological pathways similar to Prasse *et al.* (43).

We compared CANDELA to PaccMann by extracting the genes with the highest overall feature importances from both models and performing over-representation analysis with respect to a comprehensive collection of 5578 human pathways integrated from various resources in the Consensus-

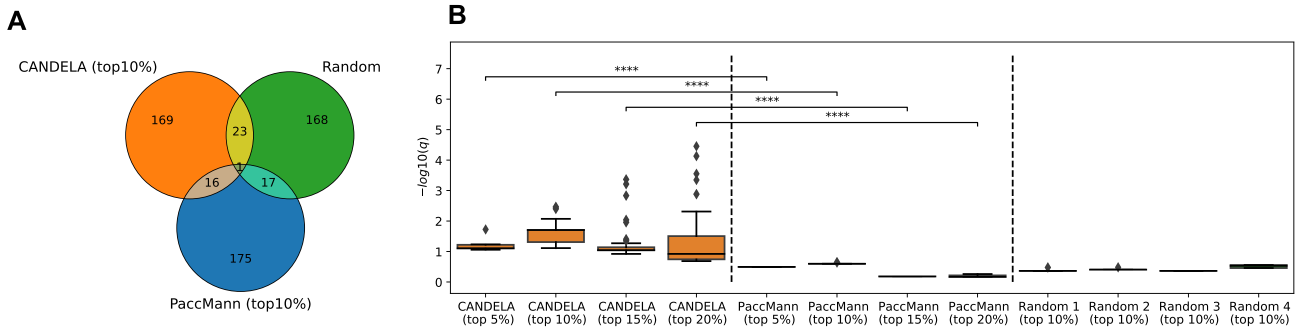

Figure 5. (**A**) Overlap between the 10% most relevant expression features of CANDELA, PaccMann, and a randomly drawn set of gene-expression features. (**B**) Strength of enrichment of biological pathways with different feature sets. X-axis: box plots for gene sets reflecting top 5% of the important features from CANDELA (left panel) and PaccMann (middle panel) respectively (104 genes), top 10% (209 genes), top 15% (313 genes) and top 20% (418 genes). Bars show max and min values, boxes show 25–75% range of $-\log_{10}$(q-values). The right panel corresponds to four randomly chosen gene sets (209 genes each). Y-axis: $-\log_{10}$ of the enrichment q-value. Significance of differences between CANDELA and PaccMann q-values is judged by an unpaired Wilcoxon test ($****P \leq 1.0e{-}04$).

PathDB (44). Figure 5(A) shows that the two methods assign overall importances to different genes. For example, only 17 out of 209 (8%) of the top 10% important genes overlap among the two methods—a fraction that resembles an overlap to randomly chosen genes as shown in Figure 5(A). Thus, both models use different features to predict drug sensitivity presumably due to redundant variables in the data. Overrepresentation analysis of different gene sets with high importances (top 5%, 104 genes, top 10%, 209 genes, top 15%, 313 genes, and top 20%, 418 genes), yields a significantly higher enrichment ($P < 10^{-4}$) of the genes selected by CANDELA compared to the gene sets selected by PaccMann or randomly chosen gene sets among the 2089 drug target genes under consideration (see Figure 5(B)). Thus, CANDELA, when compared to the PaccMann model, appears to assign importance to features that are more focused on specific cellular pathways and functions, which suggests a larger biological relevance and plausability for biomarker selection.

We can observe that CANDELA attributes high importances to genes that reflect the biological mechanism of the targeted drugs. We exemplify this with two drugs, fedratinib and refamitinib. Fedratinib is a highly specific kinase inhibitor of JAK2 and FLT3 tyrosine kinase, and we found an overall good correlation of 0.68 between predicted and ground truth log(IC_{50}) values. Fedratinib has been recently approved for the treatment of myeloproliferative neoplasm-associated myelofibrosis, a disease of the blood cells that cover different leukemias such as CML (45). From the prediction results we observe a high agreement between predictability and approved prescription with 45 different blood cell lines having been assigned a high sensitivity to that drug. Furthermore, both major targets of fedratinib, FLT3 (fms related receptor tyrosine kinase 3) and JAK2, have been assigned high importances with 2.27- and 1.35-fold above the median importance value across all 2090 drug–target genes. Additionally, many members of the JAK/STAT signalling pathway were found highly attributed by CANDELA. JAK2 signalling typically appears through cytokines and growth factors that bind to their corresponding receptors, leading to receptor dimerization and recruitment of related JAKs (46). Elevated importances attributed to key hormone and cytokine receptors reflect this signalling, for example growth hormone 2 (GH2, 1.63-fold feature importance) and growth hormone receptor (GRH, 1.75), leptin (LEP, 1.46) and prolactin receptor (PPLR, 2.01). Also, for resistance mechanisms cross-talk between the JAK/STAT pathway and other pathways is important. Such cross-talk has been reported for example with the PI3K–AKT signalling pathway (47). Interestingly, many components of that pathway have also been assigned high importances by CANDELA, most prominently FLT3 (2.27), TEK (TEK receptor tyrosine kinase, 2.18) and ERBB2 (erb-b2 receptor tyrosine kinase 2, 1.94).

Refamitinib, a targeted therapy against MEK1/2 and inhibitor of the ERK-MAPK signaling pathway, is another example with an overall correlation of 0.71 between predicted and ground truth log(IC_{50}) values. This pathway is relevant for several cancer types, such as melanoma because ERK-MAPK signalling is a crucial regulator of melanocyte proliferation and differentiation (48). Our results list melanoma as the predominant cell line type representing sensitive cell lines in agreement to the experimental $\log(IC_{50})$ measurements. CANDELA gene importances for members of the signalling pathway are enriched, genes with high importances are for example MAP4K1 (2.06), HGF (hepatocyte growth factor, 2.03), CSF1R (colony stimulating factor 1 receptor, 2.02) or FGFR2 (fibroblast growth factor receptor 2, 1.98). Major cross-talk from the highly attributed genes (>1.5× median importance value, in total 554 genes) can be observed with the PI3K-AKT signalling pathway (30 genes). Both pathways are often activated in the presence of driver mutations and lead to uncontrolled proliferation in malignant melanoma and combinatorial therapies targeting MEK and PI3K are currently tested and have shown to be more effective against metastatic melanoma compared to monotherapies (49).

Finally, the most salient features were compared for three drugs (5-Fluorouracil, Imatinib, and Masitinib). Imatinib and Masitinib are highly similar compounds targeting several serine/threonine kinases, whereas 5-Fluorouracil is an antimetabolite similar to the nucleobase Thymine. Analyzing the intersections between the top k features for each of the compounds (where k corresponds to 1%, 5%, 10% and 20% of the input genes), it is clearly seen in Figure 6 how both targeted drugs also share a higher number of important features compared to 5-fluorouracil. Furthermore, even with as many as 20% of the genes, the number of genes uniquely found important for 5-fluorouracil remains high, showing the specificity of the genes used for generating the predictions.

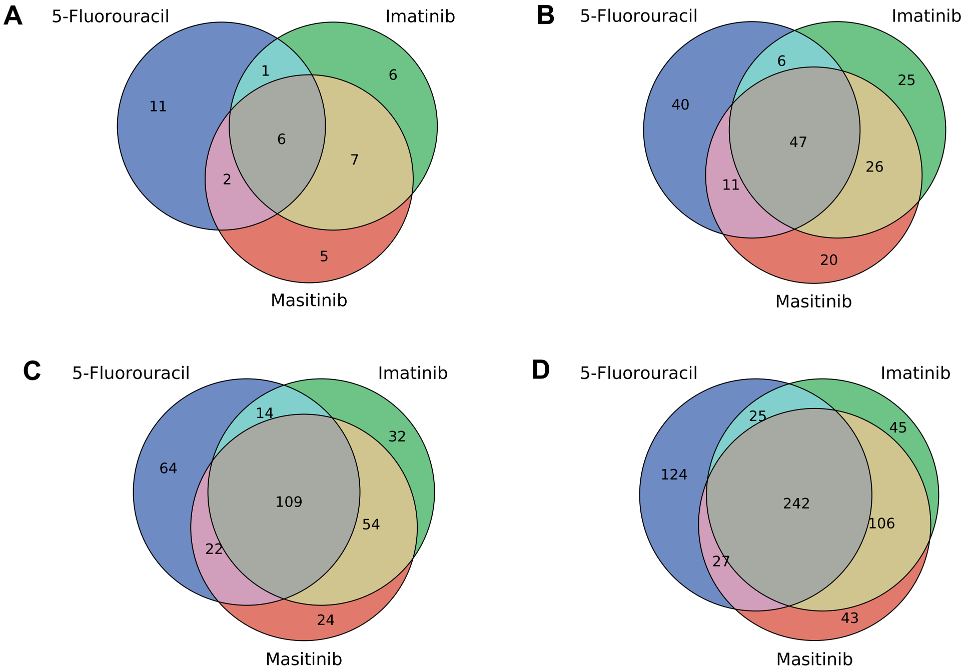

Figure 6. Overlap between top *k* features for 5-Fluorouracil, Imatinib and Masitinib (**A**) $k = 20$ (1% of all features); (**B**) $k = 104$ (5%); (**C**) $k = 209$ (10%); and, (**D**) $k = 418$ (20%).

Discussion

We have developed CANDELA, a novel cancer drug sensitivity estimation modular graph neural network. The score decomposition allows for the drug encoder and self-attentional graph pooling layer to be pre-trained on tasks for which labeled training data are more abundant than for drug-sensitivity estimation: predicting metabolite properties and predicting compound toxicity. CANDELA processes a rich graph structure and uses graph-attention layers to encode drug molecules. In the context of other biochemical applications, previous work has explored a range of graph-level features that can be extracted from chemical molecules and variations of their graphical representation, as well as attention mechanisms that are suitable for their respective applications. For instance, (23) uses super-nodes in order to represent graph-level features and a graph-edit attention mechanism that is specifically tailored to model chemical reactions. In our approach, given our multi-instance prediction setting, we leveraged the creation of graph-level feature vectors through a cross-attention mechanism that fusions cell-line and node-level attributes.

We found that CANDELA significantly outperforms PaccMann and other reference models, both for precision oncology (estimation for unseen cell lines) and drug discovery (estimation for unseen drugs). The inhibitory concentration IC_{50} of drug *i* for cell line *j* can always be decomposed into a drug toxicity α_i, a cell-line survivability β_j, and an interaction residual γ_{ij}. We argue that a model's overall ability to predict IC_{50} values can be dominated by the model's ability to predict drug toxicity and cell-line survivability. While these are relevant problems in their own right, they are not in alignment with the ultimate goal of either precision oncology or drug discovery. The toxicity of approved drugs is generally known, and the survivability of tumor cells is not subject to therapeutic decisions. By studying the models' ability to predict the interaction residuals, we can show that CANDELA excels at identifying drugs that specifically target given cell lines. In particular, the ability of the model to predict the specific pairwise interactions measured by the correlation between the observed and predicted latent interaction terms was increased by 10.8% for precision oncology and 49.3% for drug discovery.

Our ablation study shows that while pre-training on the larger database of metabolite properties is beneficial, additionally pre-training on the smaller toxicity database is detrimental. Without pre-training, CANDELA performs exceptionally poorly, because the number of its parameters are disproportionally high compared to the volume of the GDSC data that are used for training. The score decomposition has a beneficial effect only for drug discovery. Models without graph attention for the fusion of the node embeddings and cell-line embeddings show a deteriorated performance.

An analysis of the importance of the gene features relative to three drugs (Imatinib, Masitinib, 5-fluorouracil) showed that independently of the number of features selected the level of overlap of important features between Imatinib and Masitinib was consistently larger when compared to 5-Fluorouracil. This agrees with our expectations, because Masitinib and Imatinib are almost identical drugs. Furthermore, the changes in importance with respect to the global level showed that the predictions for targeted drugs such as Imatinib and Masitinib displayed extremely increased importances for several genes, whereas 5-Fluorouracil (an antimetabolite) did not.

Data availability

The original Genomics of Drug Sensitivity in Cancer Database (GDSC) (2,3) is available at https://ftp.sanger.ac.uk/project/cancerrxgene/releases/release-8.2/.

The CATMoS data is available at https://ehp.niehs.nih.gov/action/downloadSupplement?doi=10.1289%2FEHP8495&file=ehp8495.s002.codeanddata.acco.zip. The source code and the processed version of both data sets used during our experiments are available at https://zenodo.org/doi/10.5281/zenodo.8020945. The implementation for the different baselines is available in their respective repositories (https://github.com/BML-cbnu/DrugGCN, https://github.com/hauldhut/GraphDRP, https://github.com/PaccMann/paccmann_predictor_tf).

Funding

German Federal Ministry of Research and Education [01lS18044].

Conflict of interest statement

None declared.

References

1. Bucur,A., van Leeuwen,J., Christodoulou,N., Sigdel,K., Argyri,K., Koumakis,L., Graf,N. and Stamatakos,G. (2016) Workflow-driven clinical decision support for personalized oncology. *BMC Med. Inform. Decis.*, **16**, 151–162.
2. Yang,W., Soares,J., Greninger,P., Edelman,E.J., Lightfoot,H., Forbes,S., Bindal,N., Beare,D., Smith,J.A., Thompson,I.R., *et al.* (2012) Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.*, **41**, D955–D961.
3. Iorio,F., Knijnenburg,T.A., Vis,D.J., Bignell,G.R., Menden,M.P., Schubert,M., Aben,N., Gonçalves,E., Barthorpe,S., Lightfoot,H., *et al.* (2016) A landscape of pharmacogenomic interactions in cancer. *Cell*, **166**, 740–754.
4. Barretina,J., Caponigro,G., Stransky,N., Venkatesan,K., Margolin,A.A., Kim,S., Wilson,C.J., Lehár,J., Kryukov,G.V., Sonkin,D., *et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, **483**, 603–607.
5. Ahmed,F.E. (2004) Effect of diet, life style, and other environmental/chemopreventive factors on colorectal cancer development, and assessment of the risks. *J. Environ. Sci. Health, Part C*, **22**, 91–148.
6. Teer,J.K. (2014) An improved understanding of cancer genomics through massively parallel sequencing. *Transl. Cancer Res.*, **3**, 243–259.
7. Amjad,M.T., Chidharla,A. and Kasi,A. (2021) Cancer Chemotherapy. StatPearls Publishing, Treasure Island (FL).
8. Azuaje,F. (2019) Artificial intelligence for precision oncology: beyond patient stratification. *NPJ Prec. Oncol.*, **3**, 1–5.
9. Manica,M., Oskooei,A., Born,J., Subramanian,V., Sáez-Rodríguez,J. and Martínez,M.R. (2019) Toward explainable anticancer compound sensitivity prediction via multimodal attention-based convolutional encoders. *Mol. Pharm.*, **16**, 4797–4806.
10. Weininger,D. (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.*, **28**, 31–36.
11. Zhou,J., Cui,G., Zhang,Z., Yang,C., Liu,Z. and Sun,M. (2018) Graph Neural Networks: a review of methods and applications. *AI Open*, **1**, 57–81.
12. Li,P., Li,Y., Hsieh,C.-Y., Zhang,S., Liu,X., Liu,H., Song,S. and Yao,X. (2020) TrimNet: learning molecular representation from triplet messages for biomedicine. *Brief. Bioinform.*, **22**, bbaa266.
13. Li,Y., Hsieh,C.-Y., Lu,R., Gong,X., Wang,X., Li,P., Liu,S., Tian,Y., Jiang,D., Yan,J., *et al.* (2022) An adaptive graph learning method for automated molecular interactions and properties predictions. *Nat. Mach. Intel.*, **4**, 645–651.
14. Nguyen,T., Nguyen,G. T.T., Nguyen,T. and Le,D.-H. (2022) Graph convolutional networks for drug response prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **19**, 146–154.
15. Liu,Q., Hu,Z., Jiang,R. and Zhou,M. (2020) DeepCDR: a hybrid graph convolutional network for predicting cancer drug response. *Bioinformatics*, **36**, i911–i918.
16. Chu,T., Nguyen,T.T., Hai,B.D., Nguyen,Q.H. and Nguyen,T. (2022) Graph transformer for drug response prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **20**, 1065–1072.
17. Zuo,Z., Wang,P., Chen,X., Tian,L., Ge,H. and Qian,D. (2021) SWnet: a deep learning model for drug response prediction from cancer genomic signatures and compound chemical structures. *BMC Bioinformatics*, **22**, 434.
18. Kim,S., Bae,S., Piao,Y. and Jo,K. (2021) Graph convolutional network for drug response prediction using gene expression data. *Mathematics*, **9**, 772.
19. Wang,Y., Wang,Y.G., Hu,C., Li,M., Fan,Y., Otter,N., Sam,I., Gou,H., Hu,Y., Kwok,T., *et al.* (2022) Cell graph neural networks enable the precise prediction of patient survival in gastric cancer. *npj Prec. Oncol.*, **6**, 45.
20. Ma,T., Liu,Q., Li,H., Zhou,M., Jiang,R. and Zhang,X. (2022) DualGCN: a dual graph convolutional network model to predict cancer drug response. *BMC Bioinformatics*, **23**, 129.
21. Shin,J., Piao,Y., Bang,D., Kim,S. and Jo,K. (2022) DRPreter: interpretable anticancer drug response prediction using knowledge-guided Graph Neural Networks and transformer. *Int. J. Mol. Sci.*, **23**, 13919.
22. David,L., Thakkar,A., Mercado,R. and Engkvist,O. (2020) Molecular representations in AI-driven drug discovery: a review and practical guide. *J. Cheminformatics*, **12**, 56.
23. Sacha,M., Blaz,M., Byrski,P., Dabrowski-Tumanski,P., Chrominski,M., Loska,R., Wlodarczyk-Pruszynski,P. and Jastrzebski,S. (2021) Molecule edit graph attention network: modeling chemical reactions as sequences of graph edits. *J. Chem. Inf. Model.*, **61**, 3273–3284.
24. Xiong,Z., Wang,D., Liu,X., Zhong,F., Wan,X., Li,X., Li,Z., Luo,X., Chen,K., Jiang,H., *et al.* (2019) Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *J. Med. Chem.*, **63**, 8749–8760.
25. Jiang,D., Wu,Z., Hsieh,C.-Y., Chen,G., Liao,B., Wang,Z., Shen,C., Cao,D., Wu,J. and Hou,T. (2021) Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models. *J. Cheminformatics*, **13**, 12.
26. Brody,S., Alon,U. and Yahav,E. (2021) How Attentive are Graph Attention Networks? In: *International Conference on Learning Representations*.
27. Li,Y., Gu,C., Dullien,T., Vinyals,O. and Kohli,P. (2019) Graph matching networks for learning the similarity of graph structured objects. In: *Proceedings of the International Conference on Machine Learning*. PMLR, pp. 3835–3845.
28. Hu,W., Liu,B., Gomes,J., Zitnik,M., Liang,P., Pande,V. and Leskovec,J. (2020) Strategies for Pre-training Graph Neural Networks. In: *International Conference on Learning Representations*.
29. Lipinski,C.A. (2004) Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today: Technol.*, **1**, 337–341.
30. Kuhn,M., von Mering,C., Campillos,M., Jensen,L.J. and Bork,P. (2007) STITCH: interaction networks of chemicals and proteins. *Nucleic Acids Res.*, **36**, D684–D688.
31. Kim,S., Thiessen,P.A., Bolton,E.E. and Bryant,S.H. (2015) PUG-SOAP and PUG-REST: web services for programmatic access to chemical information in PubChem. *Nucleic Acids Res.*, **43**, W605–W611.
32. Falkner,S., Klein,A. and Hutter,F. (2018) BOHB: robust and efficient hyperparameter optimization at scale. In: *International Conference on Machine Learning*. PMLR, pp. 1437–1446.

33. Plenderleith,I.H. (1990) Treating the treatment: toxicity of cancer chemotherapy. *Can. Fam. Phys.*, **36**, 1827–1830.
34. Farghadani,R., Haerian,B.S., Ebrahim,N.A. and Muniandy,S. (2016) 35Year research history of cytotoxicity and Cancer: a quantitative and qualitative analysis. *Asian Pac. J. Cancer Prev.*, **17**, 3139–3145.
35. Mansouri,K., Karmaus,A.L., Fitzpatrick,J., Patlewicz,G., Pradeep,P., Alberga,D., Alepee,N., Allen,T.E., Allen,D., Alves,V.M., *et al.* (2021) CATMoS: Collaborative Acute Toxicity Modeling Suite. *Environ. Health Persp.*, **129**, 47013.
36. Sondka,Z., Bamford,S., Cole,C.G., Ward,S.A., Dunham,I. and Forbes,S.A. (2018) The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer*, **18**, 696–705.
37. Degen,J., Wegscheid-Gerlach,C., Zaliani,A. and Rarey,M. (2008) On the Art of Compiling and Using'Drug-Like'Chemical Fragment Spaces. *ChemMedChem: Chem. Enab. Drug Discov.*, **3**, 1503–1507.
38. Rogers,D. and Hahn,M. (2010) Extended-connectivity fingerprints. *J. Chem. Inf. Model.*, **50**, 742–754.
39. Liu,X. and Zhang,W. (2023) A subcomponent-guided deep learning method for interpretable cancer drug response prediction. *PLoS Comput. Biol.*, **19**, e1011382.
40. Stärk,H., Beaini,D., Corso,G., Tossou,P., Dallago,C., Günnemann,S. and Liò,P. (2021) 3D Infomax improves GNNs for molecular property prediction. In: *39th International Conference on Machine Learning*. Vol. **162**, pp. 20479–20502.
41. Yang,W., Soares,J., Greninger,P., Edelman,E.J., Lightfoot,H., Forbes,S., Bindal,N., Beare,D., Smith,J.A., Thompson,I.R., *et al.* (2012) Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.*, **41**, D955–D961.
42. Sundararajan,M., Taly,A. and Yan,Q. (2017) Axiomatic attribution for deep networks. In: *Proceedings of the 34th International Conference on Machine Learning*. Vol.**70**, pp. 3319–3328.
43. Prasse,P., Iversen,P., Lienhard,M., Thedinga,K., Herwig,R. and Scheffer,T. (2022) Pre-Training on in vitro and fine-tuning on patient-derived data improves deep neural networks for anti-cancer drug-sensitivity prediction. *Cancers*, **14**, 3950.
44. Kamburov,A. and Herwig,R. (2022) ConsensusPathDB 2022: molecular interactions update as a resource for network biology. *Nucleic Acids Res.*, **50**, D587–D595.
45. Talpaz,M. and Kiladjian,J. (2021) Fedratinib, a newly approved treatment for patients with myeloproliferative neoplasm-associated myelofibrosis. *Leukemia*, **35**, 1–17.
46. Hu,X., li,J., Fu,M., Zhao,X. and Wang,W. (2021) The JAK/STAT signaling pathway: from bench to clinic. *Signal Trans. Targ. Ther.*, **6**, 402.
47. Fruman,D., Chiu,H., Hopkins,B., Bagrodia,S., Cantley,L. and Abraham,R. (2017) The PI3K pathway in human disease. *Cell*, **170**, 605–635.
48. Wellbrock,C. and Arozarena,I. (2016) The complexity of the ERK/MAP-kinase pathway and the treatment of Melanoma Skin Cancer. *Front. Cell Dev. Biol.*, **4**, 33.
49. Aasen,S., Parajuli,H., Hoang,T., Feng,Z., Stokke,K., Wang,J., Roy,K., Bjerkvig,R., Knappskog,S. and Thorsen,F. (2019) Effective treatment of metastatic melanoma by combining MAPK and PI3K signaling pathway inhibitors. *Int. J. Mol. Sci.*, **20**, 4235.

Received: November 23, 2023. **Revised:** March 1, 2024. **Editorial Decision:** April 15, 2024. **Accepted:** April 17, 2024