# Looking at the BiG picture: Incorporating bipartite graphs in drug response prediction

David Earl Hostallero[1,2], Yihui Li[1], and Amin Emad[1,2,3,*]

[1] Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada

[2] Mila, Quebec AI Institute, Montréal, QC, Canada

[3] The Rosalind and Morris Goodman Cancer Institute, Montréal, QC, Canada

* Corresponding Author:

Amin Emad

755 McConnell Engineering Building

3480 University Street, Montreal, Quebec, Canada, H3A 0E9

Email: amin.emad@mcgill.ca

**ABSTRACT**

Motivation: The increasing number of publicly available databases containing drugs' chemical structures, their response in cell lines, and molecular profiles of the cell lines has garnered attention to the problem of drug response prediction. However, many existing methods do not fully leverage the information that is shared among cell lines and drugs with similar structure. As such, drug similarities in terms of cell line responses and chemical structures could prove to be useful in forming drug representations to improve drug response prediction accuracy.

Results: We present two deep learning approaches, BiG-DRP and BiG-DRP+, for drug response prediction. Our models take advantage of the drugs' chemical structure and the underlying relationships of drugs and cell lines through a bipartite graph and a heterogenous graph convolutional network that incorporate sensitive and resistant cell line information in forming drug representations. Evaluation of our methods and other state-of-the-art models in different scenarios shows that incorporating this bipartite graph significantly improves the prediction performance. Additionally, genes that contribute significantly to the performance of our models also point to important biological processes and signaling pathways. Analysis of predicted drug response of patients' tumors using our model revealed important associations between mutations and drug sensitivity, illustrating the utility of our model in pharmacogenomics studies.

Availability and Implementation: An implementation of the algorithms in Python is provided in github.com/ddhostallero/BiG-DRP.

Contact: amin.emad@mcgill.ca

Supplementary Information: Online-only supplementary data is available at the journal's website.

## INTRODUCTION

Utilization of machine learning and statistical analyses in precision medicine has gained attention in the past decade. Prediction of drug response based on samples' molecular profiles is a major problem in this domain and various approaches have been proposed for this purpose [1-5]. Gene expression profile of samples is widely used for this purpose due to their higher predictive ability compared to other molecular profiles [1]. The curation of large public databases of gene expression profiling of hundreds of cancer cell lines (CCLs) and their response to hundreds of different drugs (e.g., GDSC [6]) has accelerated the development of novel methodologies in this domain.

Due to the similarity in molecular and chemical structure of different drugs and their mechanisms of action, machine learning (ML) methods that can take advantage of these similarities are of great interest. Instead of training a different ML model for each drug, one can formulate the drug response prediction as a paired prediction problem, such that a model takes in a (drug, CCL) pair as input and trains a single model for all drugs and CCLs [7-9]. This increases the number of samples, and enables information sharing across many drugs and drug families. Chemical structure data (e.g., PubChem [10], ChEMBL [11]) is particularly useful for representing the drugs, and models have been developed to take advantage of these [12-14].

53 Some approaches [15-17] have formulated this as a matrix factorization problem, forming a
54 matrix of drugs and CCLs. One advantage of this is that these methods directly work with the
55 “entities” (i.e., drugs and CCLs) and responses, and do not need to map feature representations
56 of the entities to their responses, although available features can be utilized for regularization
57 [15, 17]. However, this formulation is inherently transductive, since samples and drugs are
58 expected to be present in the matrix. As a result, these models cannot be directly used to predict
59 the response of a new CCL to a drug unless the CCL has drug response information in the training
60 set for some other drugs prior to training. Another group of methods utilize collaborative filtering
61 [18, 19] and predictions are calculated using an entity's neighborhood, which are defined by the
62 similarities calculated from gene expressions, molecular fingerprints, and drug responses. Since
63 these approaches require the calculation of drug response similarities, an inherent assumption
64 is to have at least a few known responses for each unique CCL and drug in the test set, which is
65 a more relaxed assumption compared to that of matrix factorization methods.
66
67 Taking inspiration from the concept of “entity” from the matrix factorization approaches and to
68 overcome their shortcoming due to their transductive nature, we propose to utilize the
69 underlying matrix by transforming these entities into drug and CCL nodes and form a bipartite
70 graph. We hypothesized that incorporating cell line information that are highly sensitive or
71 resistant to a drug could improve the drug representation for drug response prediction. In our
72 approach called Bipartite Graph-represented Drug Response Predictor (BiG-DRP and BiG-DRP+),
73 we formed this graph by filtering the most sensitive and resistant CCLs for each drug, and linking
74 them through an edge. Although drugs are not directly connected to each other through an edge,

2-hop message passing incorporates information on drug similarities. The model accepts drugs' descriptors and CCLs' gene expression profiles as input, and utilizes them as node attributes for the bipartite graph and as sample features. The output is a continuous drug response value pertaining to the predicted normalized log IC50.

To evaluate the performance of BiG-DRP and BiG-DRP+, we used 5-fold cross validation and compared these results across different baselines and other drug response prediction approaches, namely NRL2DRP [20], PathDNN [7], and tCNN [9]. We tested on two data-splitting methods, 5-fold leave-pairs-out and 5-fold leave-cell lines-out, which represent two possible scenarios of data availability. In both scenarios, we have shown significant improvement compared to other approaches. In addition, using a computational pipeline that we developed for identifying the most contributing features, we identified genes that pointed to biological processes and signaling pathways involved in drugs' mechanisms of action.

## METHODS

### Bipartite Graph-based Drug Response Prediction

We developed a novel deep learning-based drug response prediction model that takes advantage of a bipartite graph between drugs and cell lines, which we called Bipartite Graph-represented Drug Response Predictor (BiG-DRP). We also proposed an extension of BiG-DRP, called BiG-DRP+, which accounts for constantly changing drug representations in the former approach. An overview of the (shared) architecture of these models are provided in Figure 1.

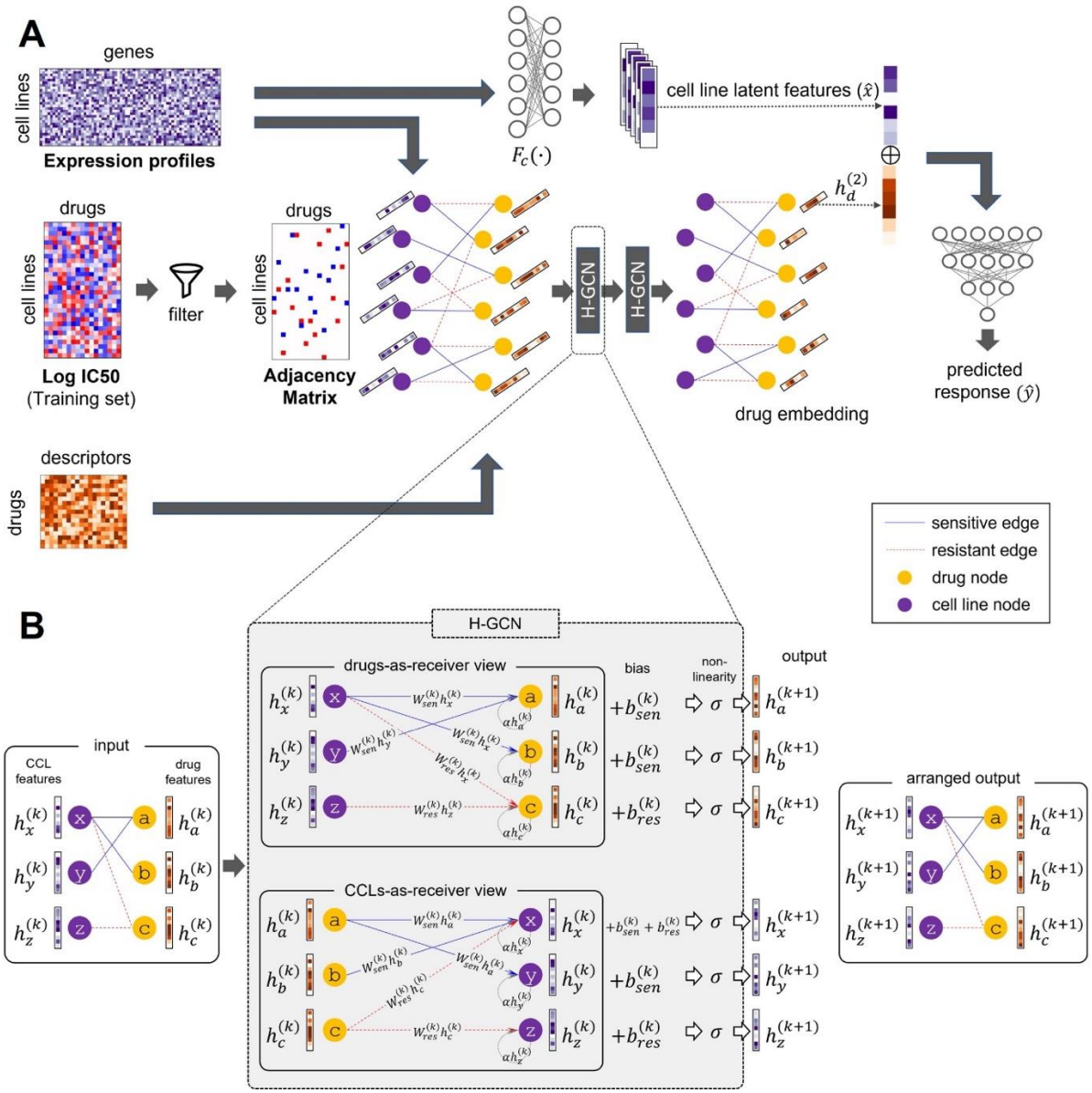



**Figure 1:** The computational pipeline and architecture of BiG-DRP and BiG-DRP+. A) Latent drug embeddings are generated using a heterogenous graph convolutional network based on a bipartite graph of drug-CCLs and drug descriptors. In parallel, CCL embeddings are generated using an encoder neural network based on their gene expression profile. These embeddings are then used by a predictor neural network to predict the drug response values. B) An overview of a single H-GCN layer is shown (our models use two stacked H-GCN layers). The H-GCN propagates information to neighbouring nodes by taking the graph as an input. Node attributes are multiplied to the weight matrices ($\boldsymbol{W_{sen}}$ and $\boldsymbol{W_{res}}$) to produce the messages, which will be sent

to their neighbours, depending on the type of edge between two nodes. Each of the nodes will then aggregate their received messages, along with the biases and self-information. The output is then the same graph, whose nodes' attributes include information from neighbouring nodes and their connectivity.

The BiG-DRP pipeline first obtains latent embeddings of CCLs and drugs and uses them in the drug response prediction task. To obtain drug embeddings, first a heterogeneous bipartite graph composed of CCL nodes and drug nodes is formed. The nodes of the bipartite graph are connected via two types of edges: sensitive edges or resistant edges. These edges are based on the log IC50 values of each CCL-drug pair. A sensitive edge implies that the CCL is likely to be sensitive to the drug, while a resistant edge implies that it is likely to be resistant to the drug. In addition, each CCL node is assigned attributes corresponding to its gene expression (GEx) profile and each drug node is assigned attributes corresponding to its drug descriptors. Then, a heterogenous graph convolutional network (H-GCN) generates embeddings of each drug (denoted as ${h_d}^{(2)}$ in Figure 1) using this bipartite graph. For each drug of interest, the H-GCN obtains an embedding that not only captures the molecular characteristic of the drug itself, but also captures the characteristics of other drugs that induce a similar sensitive/resistant pattern in CCLs. Inclusion of the GEx profiles of CCLs as node attributes in the bipartite graph allows the model to define the "similar pattern" mentioned above in a broader sense: instead of requiring a similar pattern in the exact same CCLs, the model can identify such patterns in CCLs that have a similar GEx profile.

To obtain embeddings of the CCLs based on their GEx profiles (denoted as $\hat{x}$ in Figure 1), the model uses a neural network that is separate from the H-GCN. While it is possible to use the

132 bipartite graph and the H-GCN to obtain CCL embeddings, such a choice would limit the
133 applicability of the pipeline to only CCLs that are already present in the training set. The reason
134 is that a CCL that is not present in the training set will be in the form of a single disconnected
135 node in the bipartite graph and no embedding can be found for it using the H-GCN. However, in
136 many practical applications (e.g., prediction of clinical drug response of patients based on models
137 trained on preclinical CCLs [4, 5]), a model must be able to predict drug response of samples that
138 are not seen by the model during the training for any drug. To avoid this limitation, the CCL
139 embeddings are obtained independent of the H-GCN network and the bipartite graph. The drug
140 and CCL embeddings are then concatenated, representing each (drug, CCL) pair. Then, a series of
141 neural network layers (collectively called the *predictor*) are used to predict the drug response of
142 each such pair using the concatenated embeddings.
143
144 The BiG-DRP+ is an extension of BiG-DRP with the exact same architecture, which aims to stabilize
145 the trained model. After the "last" training epoch of BiG-DRP (i.e., starting from BiG-DRP's trained
146 weights), we train the model for one more epoch but with a smaller learning rate and "frozen"
147 drug embeddings. The lower learning rate prevents the predictor from overfitting while the
148 freezing of the embeddings allows the predictor to learn the finite set of drugs instead of
149 constantly changing representations of the exact same drugs.
150
151 **Construction of the Heterogenous Bipartite Graph**
152 We denote the heterogenous bipartite graph as $G(V_C, V_D, E_r, E_s)$, where $V_C$ is the set of CCL
153 nodes used to build the graph (a subset of all the CCLs in the study) and $V_D$ is the set of drug

154 nodes. $E_r$ is the set of edges that connect drugs to their "most resistant" CCLs, while $E_s$ is the set
155 of edges that connect drugs to their "most sensitive" CCLs. For a fixed value of $k$, a drug is
156 connected via a resistant edge to CCLs whose log IC50 is among the top $k$ percent and is
157 connected via a sensitive edge to CCLs whose log IC50 is among the bottom $k$ percent of the CCLs.
158 The set $V_C$ is then the union of all such CCLs whose drug response are among the top $k$ or bottom
159 $k$ percent of all cell lines for at least one drug. It is worth noting that the edges in this graph are
160 unweighted and the log IC50 values are only used to determine whether a resistant (or sensitive)
161 edge exists or not. We used $k = 1$ in our analysis, but the performance of BiG-DRP and BiG-DRP+
162 were not sensitive to the choice of $k$, as discussed in Results.
163

164 **Drug embedding using heterogenous graph convolutions**

165 We used a 2-layer heterogenous graph convolutional network (H-GCN) to find a network-based
166 embedding of the drugs. An H-GCN is a variation of graph convolutional network [21], which
167 allows multiple edge types. A forward pass of an H-GCN can be summarized using the following
168 equation:

169
$$h_v^{(l+1)} = \sigma\left(\sum_{r\in\mathcal{R}}\left(b_r^{(l)} + \frac{1}{\sqrt{|\mathcal{N}(v,r)|}}\sum_{u\in\mathcal{N}(v,r)} h_u^{(l)} W_r^{(l)}\right) + \alpha h_v^{(l)}\right)$$

170 where $h_v^{(l)}$ is node $v$'s embedding at the $l$th layer, $\sigma$ is a non-linearity function, $\mathcal{N}(v,r)$ is node
171 $v$'s set of neighbours connected using the edge type $r$. $W_r^{(l)}$ and $b_r^{(l)}$ are the weights and biases
172 at the $l$th H-GCN layer for edge type $r$, respectively. Intuitively, this allows a separation of GCN
173 parameters for each edge type, and thus creates context during the message passing. The

174 normalization factor $\sqrt{|\mathcal{N}(v,r)|}$ prevents the embedding values from exploding due to a large
175 number of neighbours.
176
177 Although we constructed a bipartite graph, artificially adding self-loops to the graph is a common
178 practice in GCNs to retain some information from the previous layer, avoiding the complete
179 dependence of the node's embedding to its neighbours. However, in the case of H-GCN, self-
180 loops increase the complexity of the model by adding another set of parameters. To avoid this,
181 we injected a residual term ($\alpha h_v^{(l)}$) to the forward pass to simulate self-loops. Here, $\alpha$ is a
182 hyperparameter (we fixed the value to $\alpha = 0.5$ ) pertaining to the amount of information to be
183 retained for the next layer.
184
185 The bipartite graph and the H-GCN allow us to find a drug embedding that captures relevant
186 information about the CCLs that are generally resistant/sensitive to it (its 1-hop neighbours), as
187 well as information on other drugs to which these CCLs have a similar or inverse pattern of
188 response (its 2-hop neighbours). These embeddings enable sharing of information across drugs
189 that are connected to similar set of cell lines via similar edge types.
190
191 **Data Acquisition and Preprocessing**
192 We obtained the drug response data in the form of log IC50 values from the Genomics of Drug
193 Sensitivity in Cancer (GDSC) database [6]. We only selected drugs with known log IC50 values as
194 well as binarized responses that allow us to calculate the key performance metrics used for
195 evaluation of different methods. We also filtered out duplicate drugs that came from different

196 batches, which are tagged with different drug IDs, named with synonyms, or labeled as
197 "rescreens". In cases of such duplicates, we only kept the one for which the drug response in a
198 larger number of cell lines was measured. We collected the Simplified Molecular Input Line Entry
199 System (SMILES) encoding [22] of these drugs and used the RDKit software [12] to generate drug
200 descriptors (e.g. molecular weight, number of aromatic rings) from these encodings. Descriptors
201 with missing values were excluded from the analysis. At the end of these data cleaning steps, we
202 were left with 237 unique drugs, each with feature vectors of length 198 (representing their drug
203 descriptors).
204
205 We performed z-score normalization on drug descriptors, one feature at a time across all drugs.
206 We also z-score normalized the log IC50 values of each drug (one drug at a time) across cell lines.
207 This is necessary since the log IC50 values of different drugs have significantly different means
208 and standard deviations, which renders the calculated metrics incomparable across drugs and
209 inflates the overall correlation coefficient. For example, a relatively small mean squared error for
210 a certain drug, or a high overall spearman correlation do not necessarily indicate good
211 performance without such a normalization. This drug-wise normalization allows us to compare
212 results across different drugs, and prevents overestimation of the models' performance.
213
214 For the 237 drugs above, we obtained the RNA-seq GEx profile of 1001 CCLs from the Cell Model
215 passports [23]. We performed $\log_2$(FPKM +1) transformation on the FPKM values. We excluded
216 genes that showed a small variability across the cell lines (genes with standard deviation <0.1) as
217 well as genes with missing values in some cell lines. After these preprocessing steps, we ended

218 up with 944 unique CCLs and their GEx values of 13,823 genes. This amounted to a total of
219 181,380 labeled CCL-drug pairs.
220
221 **Training Procedure**
222 As discussed earlier, to enable the model to generalize to completely new CCLs (those that are
223 not seen by the model for any drug during training), we used a separate neural network, parallel
224 to the H-GCN. As input, this network accepts the CCLs' gene expression vector $x$ and produces a
225 latent representation $\hat{x} = F_c(x)$. We then concatenate $\hat{x}$ with the drug $d$'s embedding, $h_d^{(2)}$,
226 and use it as input for our predictor, a 3-layer neural network that outputs the predicted drug
227 response values ($\hat{y}$).
228
229 The model was trained end-to-end using the mean squared error $\mathcal{L} = (y - \hat{y})^2$ and Adam as the
230 optimizer [24]. We also fixed the learning rate to 0.0001 and batch size to 128 (see Results for
231 the effect of different choices of hyperparameters on the performance). We used Leaky ReLU for
232 all non-linearity functions (i.e. $\sigma(x) = \max(0, x) + 0.01 \times \min(0, x)$). The number of training
233 steps were decided by randomly selecting samples from the training data and using them as a
234 validation set for early stopping. The model was then re-trained with the entire training set and
235 the previously identified optimal number of training steps. For BiG-DRP+, the extra epoch's
236 learning rate was set to 0.00001.
237
238 In our approach, elements of a batch are (drug, CCL) pairs, although all drug embeddings can be
239 generated simultaneously for each forward pass. Embeddings generated using graph

240 convolutional networks rely on the node connectivity. This generally means that a small
241 perturbation of a node's embedding may affect the embeddings of its neighbours in the next GCN
242 (or H-GCN) layer. Unlike regular dense neural networks, it is possible that a dramatic change
243 would occur in the embeddings, even with a relatively small learning rate. In such cases, the
244 predictor may not easily map the "new" embedding to the "known" ones, especially if the drug
245 was not part of the batch during the previous training step. The predictor could see this as having
246 an infinite number of drugs, increasing the level of complexity to the learning process. To address
247 this "moving embedding problem", we developed BiG-DRP+, which slightly modifies the training
248 of BiG-DRP.
249
250 The idea of BiG-DRP+ is to stop the training of the H-GCN component after several epochs but
251 continue the training of the predictor using the "frozen" drug embeddings. In our BiG-DRP+
252 model, we froze the drug embeddings obtained by BiG-DRP (after the number of epochs
253 determined by early stopping), but continued the training of other components of the
254 architecture for one extra epoch (we used a lower learning rate for this epoch). This stabilizes the
255 training of the predictor and enables it to identify CCLs that were treated by the same drug (since
256 the half of the input to the predictor pertaining to the drug features are now fixed). The lower
257 learning rate is a preventative measure to avoid overfitting.
258
259 **Evaluation and Cross-validation**
260 To evaluate the performance of our model we used 5-fold cross validation (CV), in which one fold
261 was kept aside as the test set for evaluation and was not used during training of the model nor

262 for the selection of hyperparameters. This process was repeated five times (each time with a
263 different fold as the test set) to ensure that the specific choice of the test set does not bias the
264 results. We adopted two types of data splitting techniques to form the folds, namely leave-pairs-
265 out (LPO) and leave-cell lines-out (LCO).
266
267 In the LPO-CV, the disjoint folds were randomly selected from the set of all (CCL, drug) pairs,
268 while in the LCO-CV the folds contained randomly selected sets of mutually exclusive CCLs. Prior
269 to training, GEx values were z-score normalized per gene. We used only the training folds' unique
270 CCLs to calculate the means and standard deviations to prevent data leakage between training
271 and test sets. Imposing the uniqueness criterion above ensures that the models are not biased
272 towards CCLs that exists in a larger number of (drug, CCL) pairs. To ensure a fair comparison,
273 identical folds were used for all methods. For each drug, predictions of the five folds on their
274 respective test sets were collected and were used to evaluate different methods.
275
276 To assess generalizability of our models to independent datasets, we obtained GEx profile (in the
277 form of FPKM) of cancer tumours and their RECIST clinical drug response from The Cancer
278 Genome Atlas (TCGA) [25]. Similar to previous studies [26], we considered "stable disease" or
279 "progressive disease" as resistant and "complete response" or "partial response" as sensitive.
280 We selected cisplatin (n = 398), paclitaxel (n = 233), gemcitabine (n = 226), and doxorubicin (n =
281 208), since they were present in our training dataset, had a large number of samples with known
282 clinical drug response, and had more than 50 samples in each category of resistant and sensitive.
283 Similar to the preprocessing steps used for GDSC dataset, the expression of each gene in the

284 testing set (in the form of $\log_2$(FPKM +1)) was z-score normalized across the samples. We used
285 PyCombat [27] to reduce the statistical discrepancies between the GDSC and TCGA samples.

286

287 **Baseline Methods**

288 We compared our method against several baseline algorithms including both deep learning-
289 based and traditional machine learning methods, detailed below. First, we used a multilayer
290 perceptron (MLP) with a similar architecture and hyperparameters as BiG-DRP. Similar to BiG-
291 DRP, the MLP also utilized the GEx and drug features. However, instead of an H-GCN, we replaced
292 it with a dense neural network, which takes the drug features as input. We also used a support
293 vector regressor (SVR) with a linear kernel as well as a SVR with a radial basis function (RBF) as
294 traditional ML baselines. The concatenation of the GEx and drug features were used as the input
295 to SVR models. Due to the large size of the data, we used Nystroem's transformation [28] to
296 approximate the SVR's kernels to improve its efficiency. Hyperparameters, namely the number
297 of Nystroem components, regularization factor, and gamma for RBF were selected using nested
298 cross validation. In addition to the SVR models above, we used recursive feature elimination (RFE)
299 [29] to identify the most relevant features to be used with the linear and non-linear SVR models.

300

301 NRL2DRP [20] is a graph representation learning-based method that uses a graph composed of
302 genes, drugs, and CCL nodes, connected by edges according to their sensitivity, mutation, and
303 protein-protein interactions. However, NRL2DRP uses a topology-based graph embedding called
304 LINE [30], which is typically used for transductive learning. We slightly modified NRL2DRP to
305 predict continuous values instead of binary values (so that it can be applied to our data). PathDNN

306 [7] is another deep learning method that proposes to add some level of explainability to the drug
307 response prediction problem by constraining the neural network connectivity using a pathway
308 mask. This method uses drug targets and gene expressions, both of which should be in any of the
309 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways [31]. We obtained the drug targets
310 and pathway information from the PathDNN's repository. The drug targets were represented by
311 their normalized STITCH [32] confidence score, which indicates a non-zero value for genes in the
312 drug's targets. However, we removed three compounds because they did not have known targets
313 in the KEGG pathways. Another deep learning approach is tCNN [9], which utilizes 1-dimensional
314 CNNs. The canonical SMILES string of the compound is encoded into a sequence of one-hot
315 vectors, each of which represents a single character. Since the SMILES strings vary in length, the
316 resulting binary encoding is padded by zeros to the right to match the length of the longest
317 encoding, resulting in a matrix of size $m \times n$, where $m$ is the number of unique characters and
318 $n$ is the length of the longest encoding. Mutations and copy number alterations, which GDSC
319 dubs as "genetic features," were used as the features of the CCLs.
320
321 In order to ensure a fair comparison, in our cross-validations we fixed the folds and used identical
322 folds for each method. In addition, when an algorithm required extra information that was not
323 used in Big-DRP, we provided those datasets as inputs to the baseline models, following the
324 descriptions provided in each method's manuscript. This was done to ensure we give each
325 baseline model a fair chance.
326

327 **Identification of genes that are most predictive of drug response**

328 To identify genes that are predictive of drug response, we used a neural network explainer called
329 CXPlain [33] and a similar approach which we previously developed to aggregate contribution
330 across CCLs and identify top contributors [5]. CXplain uses Granger's causality [34] as the basis of
331 the feature attribution. Intuitively, for each of the features, it tries to predict the increase in the
332 sample's loss if that specific feature is zeroed-out. We trained separate explainers for each of the
333 drugs, since this eliminates the unnecessary complexity of learning attributions for multiple drugs,
334 as well as the additional feature dimensions (i.e., drug features). We pooled the scores by
335 calculating the mean attribution across all the CCLs for each of these drugs. The top genes were
336 identified by a threshold calculated using kneedle [35], with sensitivity S=2.

337

338 **Pathway characterization analysis**

339 We used KnowEnG's gene set characterization pipeline [36] to perform pathway enrichment
340 analysis (using Reactome pathways [37]). The p-values of Fisher's exact test were corrected for
341 multiple tests (i.e., multiple pathways) using Benjamini-Hochberg false discovery rate (FDR).

342

343 **Analysis of TCGA tumor mutations and their relationship with predicted drug responses**

344 From TCGA database, we selected primary tumor samples that had both GEx profiles and
345 mutation data (n = 9067). We utilized BiG-DRP+ to predict response of 237 drugs for each of the
346 tumor samples using their GEx as input (see the Evaluation and Cross-validation section). Using
347 the Mutation Annotation Format (MAF) file, a binary matrix indicating the existence of a mutation
348 for a sample was formed. Similar to previous studies [38], we focused on four types of mutations:

349 nonsense, missense, frameshift insertions and frameshift deletions. Only mutations that exist in
350 at least 10% of the samples were included in the analysis.
351
352 **RESULTS**
353 **Performance of BiG-DRP and BiG-DRP+ based on leave-pair-out cross validation**
354 First, we evaluated BiG-DRP, BiG-DRP+, and other baseline algorithms using a five-fold LPO-CV,
355 in which the folds were randomly selected among the set of all possible (CCL, drug) pairs. Table
356 1 shows a summary of the performance results using area under the receiver operating
357 characteristic curve (AUROC), root mean squared error (RMSE), Pearson's correlation coefficient
358 (PCC) and Spearman's correlation coefficient (SCC). To calculate these metrics across all drugs,
359 we first calculated them separately for each drug (Supplementary Table S1) and then obtained
360 mean and standard deviation across the drugs. BiG-DRP+ outperforms all other methods
361 according to all metrics, and BiG-DRP outperforms all baselines but has a slightly worse
362 performance compared to BiG-DRP+. BiG-DRP+ has a ~5% higher AUROC and ~11% higher SCC
363 and PCC compared to that of MLP, which utilizes a similar architecture to BiG-DRP+ (except for
364 the usage of the bipartite graph and the H-GCN). This highlights the importance of this novel
365 aspect of the algorithm.
366
367 As mentioned earlier, in our models the H-GCN is used to obtain drug representations and a
368 separate encoder is used to obtain cell line representations. We were interested to determine
369 how the performance of the models change if we substitute the role of these two components:
370 use the encoder to obtain drug embeddings and use the H-GCN to obtain cell line embeddings

371 (called inverted BiG-DRP, henceforth). Our analysis showed that inverted Big-DRP outperforms
372 all baselines, except for BiG-DRP and BiG-DRP+ (Table 1). However, it is important to note that
373 inverted BiG-DRP has two shortcomings compared to BiG-DRP and BiG-DRP+. First, it cannot be
374 used to predict the response of a new CCL (i.e., it cannot be used in the LCO framework), since a
375 new CCL would not be part of the bipartite graph and as a result a representation for it cannot
376 be obtained. Second, the bipartite graph used in inverted BiG-DRP connects each CCL to most
377 sensitive and most resistant drugs and as a result is less reliable than the bipartite graph of BiG-
378 DRP (that connects each drug to CCLs that are most sensitive or resistant to it). The reason is that
379 log IC50 of different drugs for the same CCL are not directly comparable and making a bipartite
380 graph based on this criterion may introduce errors in the network.

381

382 **Table 1:** The performance of BiG-DRP, BiG-DRP+ and baseline methods using 5-fold LPO-CV
383 evaluation. Best performance values are in bold-face and underlined. The mean and standard
384 deviations are calculated across the drugs. *Since PathDNN requires availability of at least one
385 drug target in any of the signaling pathways, we could only apply it to 234 drugs.
386

| | **Drug Attributes** | **Other inputs** | **Num. Drugs** | **AUROC mean (± std.)** | **RMSE mean (± std.)** | **SCC mean (± std.)** | **PCC mean (± std.)** |
|---|---|---|---|---|---|---|---|
| **BiG-DRP+** | Descriptors | GEx | 237 | **<u>0.878</u>** (±0.068) | **<u>0.843</u>** (±0.241) | **<u>0.748</u>** (±0.100) | **<u>0.758</u>** (±0.102) |
| **BiG-DRP** | Descriptors | GEx | 237 | 0.875 (±0.068) | 0.855 (±0.244) | 0.742 (±0.099) | 0.752 (±0.102) |
| **Inverted BiG-DRP** | Descriptors | GEx | 237 | 0.862 (±0.075) | 0.888 (±0.253) | 0.721 (±0.110) | 0.730 (±0.110) |
| **MLP** | Descriptors | GEx | 237 | 0.835 (±0.083) | 0.954 (±0.273) | 0.675 (±0.120) | 0.681 (±0.119) |
| **NRL2DRP** | None | Drug-CCL-Gene network | 237 | 0.804 (±0.085) | 1.153 (±0.345) | 0.516 (±0.119) | 0.514 (±0.123) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **tCNN** | SMILES One-hot encoding | Genetic Features | 237 | 0.787 (±0.082) | 1.086 (±0.336) | 0.587 (±0.119) | 0.591 (±0.117) |
| **PathDNN** | Drug Targets | GEx, pathway information | 234* | 0.766 (±0.083) | 1.165 (±0.355) | 0.516 (±0.115) | 0.529 (±0.117) |
| **SVR-RBF (w/RFE)** | Descriptors | GEx | 237 | 0.738 (±0.101) | 1.182 (±0.384) | 0.503 (±0.125) | 0.500 (±0.130) |
| **SVR-Linear (w/RFE)** | Descriptors | GEx | 237 | 0.738 (±0.101) | 1.181 (±0.393) | 0.498 (±0.130) | 0.497 (±0.135) |
| **SVR-RBF** | Descriptors | GEx | 237 | 0.737 (±0.100) | 1.182 (±0.383) | 0.502 (±0.123) | 0.499 (±0.129) |
| **SVR-Linear** | Descriptors | GEx | 237 | 0.736 (±0.101) | 1.184 (±0.393) | 0.494 (±0.129) | 0.493 (±0.134) |

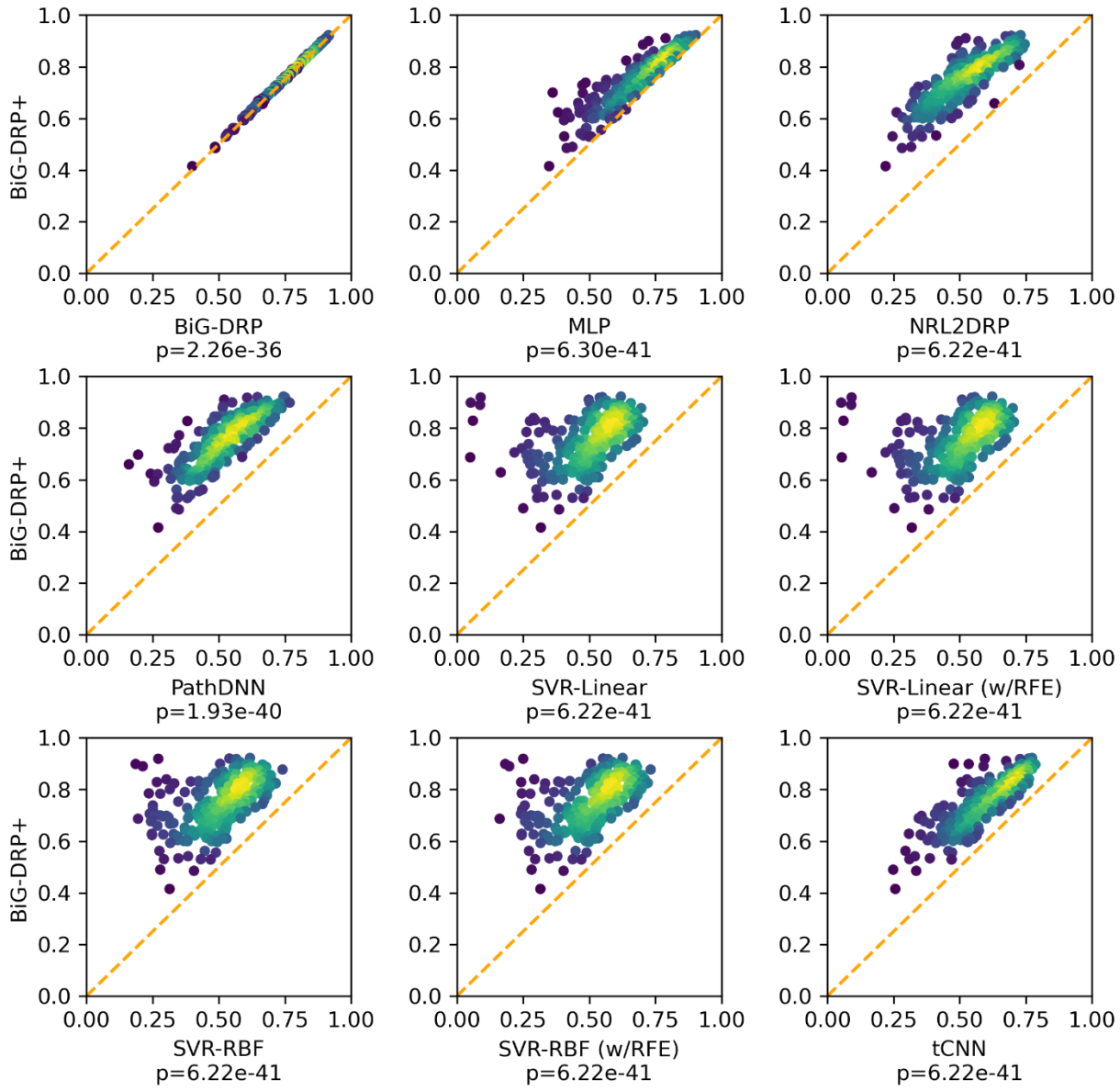



**Figure 2:** The drug-wise performance of BiG-DRP+ compared to baseline methods using five-fold LPO-CV evaluation. Each circle represents a drug, and the color of the circles reflect the density of other circles in their vicinity (yellow shows that there are many circles concentrated in that area). The coordinates reflect SCC for BiG-DRP+ (y-axis) and baseline methods (x-axis). The p-values are obtained using a one-sided Wilcoxon signed rank test, comparing the SCC of BiG-DRP+ and the baselines across drugs.

397 Figure 2 compares the performance of BiG-DRP+ against other methods for individual drugs
398 (measured based on SCC). Each circle in the scatter plots reflects a drug, and the color of the
399 circles reflect the density of other circles in their vicinity. Comparing BiG-DRP+ and BiG-DRP
400 shows that the drug-specific SCC values are generally close to each other (concentrated around
401 the diagonal line); however, the one-sided Wilcoxon signed rank test (p=2.26E-36) suggests that
402 the performance for the majority of the drugs have improved in BiG-DRP+, albeit a small amount.
403 Comparing to other baselines, the figure shows that the majority (and in many cases all) of the
404 circles are above the diagonal line, suggesting a substantial improvement of their response
405 prediction by BiG-DRP+. One-sided Wilcoxon signed rank tests also confirmed this observation,
406 resulting in statistically significant p-values (Figure 2 and Supplementary Table S2).
407
408 **Performance of BiG-DRP and BiG-DRP+ based on leave-cell line-out cross validation**
409 Next, we evaluated the performance of different models using a five-fold LCO-CV. This is a stricter
410 evaluation, since unlike LPO-CV, a CCL in the test set is never seen by the models during training,
411 since folds are randomly selected based on the CCLs and not based on (CCL, drug) pairs. Table 2
412 shows the summary of the results using our performance metrics. Note that due to the
413 transductive nature of NRL2DRP's embedding method (LINE [30]), this method could not be
414 applied to the LCO-CV evaluation and hence is not included in this table.
415
416 Based on these evaluations, BiG-DRP+ has the best performance using all metrics, while BiG-DRP
417 has the second-best performance. The BiG-DRP+ clearly outperforms MLP, further highlighting
418 the importance of the bipartite graph and H-GCN in the drug response prediction task. Similar to

419 LPO-CV evaluation, a drug-wise analysis using SCC for each drug showed a significantly superior
420 performance of BiG-DRP+ compared to all baseline methods (one-sided Wilcoxon Signed-Rank
421 test, Supplementary Table S2). Supplementary Table S1 provides the drug-specific performance
422 metrics for all drugs.

423

424 **Table 2:** The performance of BiG-DRP+, BiG-DRP and baseline methods using five-fold LCO-CV
425 evaluation. Best performance values are underlined. The mean and standard deviations are
426 calculated across the drugs. *Since PathDNN requires availability of at least one drug target in
427 any of the signaling pathways, we could only apply it to 234 drugs.
428

| | **Drug Attributes** | **Other input features** | **Num. Drugs** | **AUROC mean (± std.)** | **RMSE mean (± std.)** | **SCC mean (± std.)** | **PCC mean (± std.)** |
|---|---|---|---|---|---|---|---|
| **BiG-DRP+** | Descriptors | GEx | 237 | **0.746** (±0.077) | **1.204** (±0.367) | **0.431** (±0.094) | **0.450** (±0.105) |
| **BiG-DRP** | Descriptors | GEx | 237 | 0.743 (±0.077) | 1.210 (±0.368) | 0.426 (±0.095) | 0.443 (±0.106) |
| **MLP** | Descriptors | GEx | 237 | 0.730 (±0.086) | 1.219 (±0.374) | 0.413 (±0.100) | 0.430 (±0.111) |
| **SVR-RBF (w/RFE)** | Descriptors | GEx | 237 | 0.682 (±0.107) | 1.276 (±0.404) | 0.354 (±0.116) | 0.360 (±0.127) |
| **SVR-RBF** | Descriptors | GEx | 237 | 0.680 (±0.110) | 1.278 (±0.403) | 0.348 (±0.120) | 0.354 (±0.135) |
| **SVR-Linear** | Descriptors | GEx | 237 | 0.666 (±0.102) | 1.292 (±0.420) | 0.324 (±0.119) | 0.331 (±0.126) |
| **SVR-Linear (w/RFE)** | Descriptors | GEx | 237 | 0.666 (±0.102) | 1.293 (±0.421) | 0.322 (±0.118) | 0.330 (±0.124) |
| **PathDNN** | Drug Targets | GEx, pathway information | 234* | 0.612 (±0.074) | 2.201 (±0.698) | 0.193 (±0.061) | 0.170 (±0.078) |
| **tCNN** | SMILES one-hot encoding | Genetic Features | 237 | 0.586 (±0.060) | 1.369 (±0.427) | 0.147 (±0.068) | 0.147 (±0.072) |

429

430 To assess the generalizability of BiG-DRP+ to independent datasets, we used it to predict the drug
431 response of patient tumours from the TCGA dataset treated with cisplatin, gemcitabine,
432 doxorubicin, and paclitaxel. Given the predicted log IC50 values, we used a one-sided statistical
433 test to determine if our models can distinguish between the patients that are resistant from those
434 that are sensitive to these two drugs (using all TCGA samples with known clinical drug response).
435 Our statistical analysis (Mann Whitney U test, since data corresponding to one of the drugs did
436 not pass test of normality) showed significant p-values for three drugs (p = 2.19E-7 for cisplatin,
437 p = 8.80E-3 for doxorubicin, and p = 3.40E-2 for gemcitabine). Next, we removed any tumour
438 sample that had received a different drug beforehand or during the period that our drug of
439 interest was administered. Even though this significantly reduced the number of samples, the
440 results (Welch's t-test, since data corresponding to all drugs passed test of normality) were
441 significant for cisplatin (p = 1.82E-2) and doxorubicin (p = 4.29E-2). Supplementary Table S3
442 provides detailed information regarding the samples and the results of different statistical tests.

443

444 **Detailed Evaluation of BiG-DRP+**

445 Since one major component of the BiG-DRP and BiG-DRP+ pipeline is the bipartite graph of the
446 CCLs and drugs, we sought to evaluate the effect of different thresholds for forming this graph.
447 As explained in Methods, a drug is connected to a CCL with a sensitive (resistant) edge if the log
448 IC50 of the CCL is among the bottom (top) k% of all the CCLs. In our analysis, we fixed this value
449 to be k = 1. To assess the robustness of the results to this parameter, we formed different
450 bipartite graphs with different choices of k = 0.5, 1, 2, 5, 10 and repeated the LPO-CV and LCO-

CV. Supplementary Table S4 provides the SCC and AUROC of BiG-DRP and BiG-DRP+ for these evaluations for different values of k. These results suggest that the performance of our proposed methods remain stable for these different choices of k, with a slight deterioration as k increases (less than 1% in all evaluations when comparing k=10 to k=1). This deterioration is expected, since an increase in k increases potentially erroneous edges in the bipartite graph.

Next, we asked whether the choice of drug features as attributes in the bipartite graph has a significant effect on the performance of BiG-DRP+. To address this question, we used Morgan fingerprints [13] of the drugs, alone or in addition to the drug descriptors, as the attributes of the drug nodes in the bipartite graph. The results (Table 3) revealed that there is not a substantial difference between any of these choices, but simultaneously using both types of drug features slightly improves the results.

**Table 3:** The performance of BiG-DRP+ with different drug attributes. The rows show the results of BiG-DRP+ when drug descriptors (vectors of length 198), Morgan fingerprints (vectors of length 512), or the combination of both (vectors of length 710) are used as node attributes.

| **Method** | **Drug Attribute** | **LPO-CV** | | **LCO-CV** | |
|---|---|---|---|---|---|
| | | **AUROC mean (± std.)** | **SCC mean (± std.)** | **AUROC mean (± std.)** | **SCC mean (± std.)** |
| BiG-DRP+ | Descriptors | 0.878 (±0.068) | 0.748 (±0.100) | 0.746 (±0.077) | 0.431 (±0.094) |
| | Morgan | 0.878 (±0.068) | 0.748 (±0.100) | 0.743 (±0.080) | 0.426 (±0.098) |
| | Both | 0.879 (±0.068) | 0.748 (±0.099) | 0.746 (±0.077) | 0.433 (±0.095) |

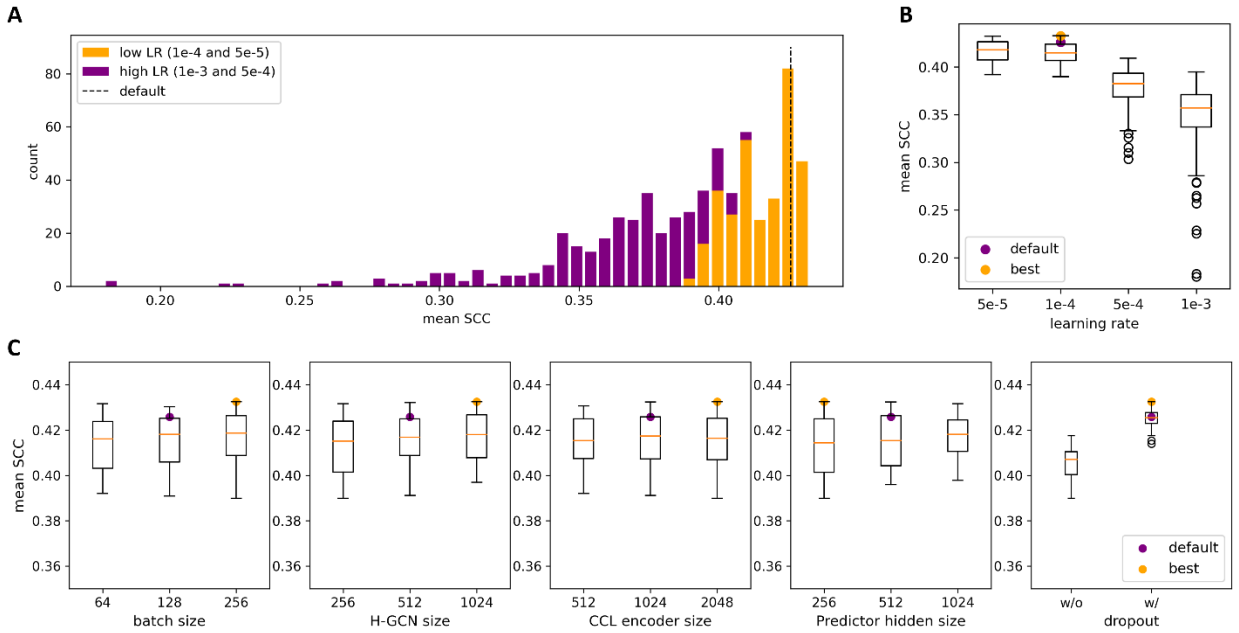



**Figure 3:** Performance comparison of BiG-DRP for different combinations of hyperparameters. A) The distribution of mean SCCs of the models in 5-fold LCO-CV. The colors correspond to the fraction of the bin that utilized either a high or low learning rate. B) The boxplots of the mean SCCs grouped by the learning rate. C) The boxplots of the mean SCCs for combinations that used 1E-4 and 5E-5 as learning rates, grouped by specific hyperparameters. The boxes range from the first to the third quartile, while the horizonal line corresponds to the median. The purple datapoint represents the default hyperparameter combinations and the orange datapoint pertains to the combination of hyperparameters that performed best.

Finally, we asked how different choices of hyperparameters influence the performance of BiG-DRP+. For this purpose, we ran our model with 648 different combinations of learning rate (5E-5, **1E-4**, 5E-4, 1E-3), batch size (64, **128**, 256), CCL encoder size (512, **1024**, 2048), H-GCN size (256, **512**, 1024), predictor hidden layer size (256, **512**, 1024), and dropout (**with** or without). (The bold-face options represent the default values used for our models). The stacked histogram in Figure 3A shows the mean SCC value of these combinations in a 5-fold LCO-CV framework. Interestingly, there are 82 combinations that perform on par with the default parameters and 47

486 combinations that perform better. This suggests that if computational complexity is not of a
487 concern, one may improve the performance of BiG-DRP+ by tuning the hyperparameters.

488

489 Further analysis revealed that the learning rate is the most influential hyperparameter (Figure 3B)
490 and a relatively large learning rate deteriorates the performance; however, learning rates of 5E-
491 5 or 1E-4 (the default) work well. More importantly, if the learning rate is selected appropriately
492 (the two choices mentioned above), the effect of other hyperparameters is relatively small and
493 the majority of choices result in good performance (orange fraction of the histogram in Figure
494 3A). Figure 3C better illustrates this by depicting the mean SCC for different choices of
495 hyperparameters when only learning rates of 5E-5 and 1E-4 are included. The only other
496 hyperparameter that seems to play an important role is dropout, where its inclusion (slightly)
497 improves the performance.

498

499 **Characterization of the bipartite graph**

500 Next, we sought to better characterize the bipartite graph and the drugs that have most
501 benefited from using this graph in the drug response prediction task. For this purpose, we first
502 formed a single bipartite graph by aggregating the bipartite graphs corresponding to each of the
503 five folds in our LCO-CV evaluation (i.e., by finding the union of edges). Then, we used a nested
504 stochastic block model (NSBM) [39] to infer the modular substructure of the graph, while taking
505 into account the edge type (i.e., resistant and sensitive) connecting each two nodes. This
506 approach automatically identifies the number of clusters by maximizing the likelihood of the

507 graph being generated from the partitioning. The final partitioning is based on running the
508 stochastic algorithm many times (in our case 1000 times) and selecting the number of clusters
509 and the partitioning that is most frequently supported by these runs. The number of clusters
510 varied between 17 to 20 (Figure 4A), with 18 selected by the algorithm as the final number of
511 clusters (5 drug clusters and 13 CCL clusters). Comparing the clusters identified by each run of
512 the algorithm with the final clusters using Rand Index (RI) [40] revealed a high degree of
513 concordance (Figure 4B, mean RI = 0.89 ± 0.01).

514

515 Figure 4C illustrates the bipartite graph and clusters identified using this method (see
516 Supplementary Table S5 for the cluster assignment of drugs and CCLs). In particular, five drug
517 clusters were identified. Comparing the performance of BiG-DRP+ compared to MLP (SCC-LCO),
518 revealed that all these clusters significantly benefit from the use of the bipartite graph (one-sided
519 Wilcoxon signed rank test, Figure 4B). In particular, Cluster 3 had the highest median
520 improvement in SCC (8.4%) and had a significant improvement p-value (p = 5.25 E-5). The
521 majority of the drugs in this cluster (13 out of 20) are protein kinase inhibitors, with 8 of them
522 targeting members of serine/threonine protein kinase family and 5 of them targeting members
523 of tyrosine kinase family. These observations suggest that information sharing across the
524 bipartite graph used in our methods benefit certain groups of drugs more than others and this
525 may be dependent on the similarity between drugs' mechanisms of action.

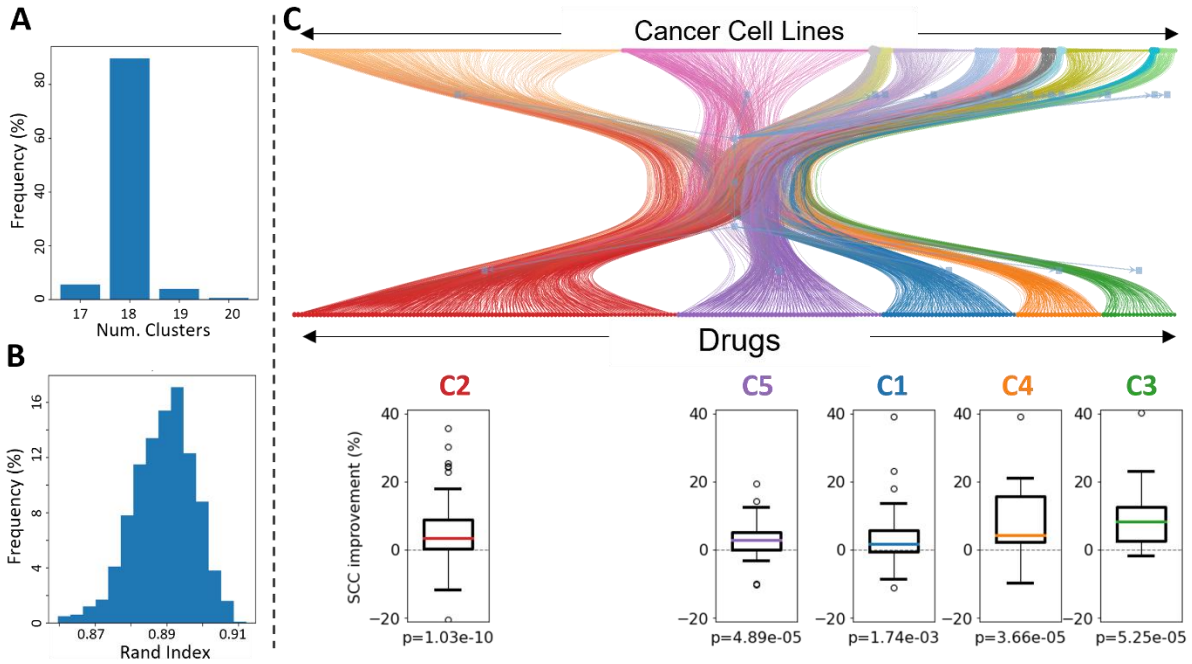



**Figure 4:** The CCL-drug bipartite graph and its clusters. A) The histogram shows the number of clusters obtained by NSBM in each run (total of 1000 runs). B) The histogram shows the Rand Index between each clustering (in each run) with the final cluster assignment. C) The graph represents the bipartite graph and the boxplots show the distribution of SCC improvements obtained for each drug using BiG-DRP+ compared to MLP in the LCO evaluation. The p-values are obtained using a one-sided Wilcoxon signed rank test.

Next, we sought to characterize the CCL clusters. Supplementary Table S6 shows the enrichment of CCL clusters in tissue type, cancer type, and driver mutations (hypergeometric test, corrected for multiple tests using Benjamini-Hochberg FDR). The analysis revealed that while only two clusters (out of 13) were enriched in cancer type (FDR < 0.05), namely cluster 1 in B-Lymphoblastic Leukemia and cluster 4 in Chronic Myelogenous Leukemia, the majority of clusters (9 out of 13) were enriched in at least one driver gene mutation. For example, cluster 1 was enriched in CCLs with mutations in RBM38 and GNA13, while cluster 2 was enriched in CCLs with mutations in POLQ and BRCA1. These observations suggest that the patterns captured by the

542 bipartite graph goes beyond tissue or cancer types and is able to capture patterns at the
543 molecular level.

544

545 **Identification of biomarkers of drug sensitivity**

546 To identify genes whose expression substantially contribute to the predictive model, we used a
547 pipeline similar to the one we proposed in a previous study [5]. This approach provides an
548 aggregate contribution score for each gene in the model and uses these scores to systematically
549 identify the set of top contributing genes in each model. We focused on 15 drugs for which BiG-
550 DRP+ provided the highest SCC values in the LCO-CV evaluation. Supplementary Table S7 provides
551 the ranked list of genes that were implicated for each of the 15 drugs. We clustered the drugs
552 based on the contribution scores of all implicated genes (Figure 5). Interestingly, four drugs
553 formed a clear cluster, separate from the others: trametinib, refametinib, selumetinib, and
554 pd0325901. Further investigation revealed that these drugs all are MEK inhibitors (i.e., inhibit the
555 mitogen-activated protein kinase kinase enzymes) and involve some similar mechanisms of
556 action [10].

557

558 Next, we focused on genes implicated for trametinib, a MEK-inhibitor for which BiG-DRP+ had
559 the best performance (SCC in LCO-CV). For this drug, *ETV5* had the highest prediction contribution.
560 ETV5 and ETV4 (the fourth highest contributor) are among the ETS family of oncogenic
561 transcription factors. The expression of this family has been shown to be upregulated in solid
562 tumours and they have been shown to be involved in tumour's progression, metastasis and
563 chemoresistance [41]. Previous studies have shown ETV5 to be regulated by ALK, a receptor

564 tyrosine kinase, in a MEK/ERK-dependent manner in neuroblastoma cell lines [42]. In addition,
565 treatment of various cancer cell lines with trametinib has been shown to downregulate ETV5 [42-
566 44]. Moreover, the overexpression of ETV4 and ETV5 have been shown to reduce the sensitivity
567 different cancer cell lines to this drug [44].
568
569 To obtain a better functional characteristic of the genes implicated for trametinib, we also
570 performed pathway enrichment analysis on genes implicated for this drug (see Supplementary
571 Table S8 for results of pathway enrichment analysis of all 15 drugs). Several important pathways
572 related to MAPK signaling, EGFR signaling, and IL2 signaling were identified (Fisher's exact test,
573 FDR<0.05). Taken together, these results suggest that genes that contribute to the predictive
574 ability of BiG-DRP+ for trametinib point to important genes and signaling pathways involved in
575 its mechanism of action.
576

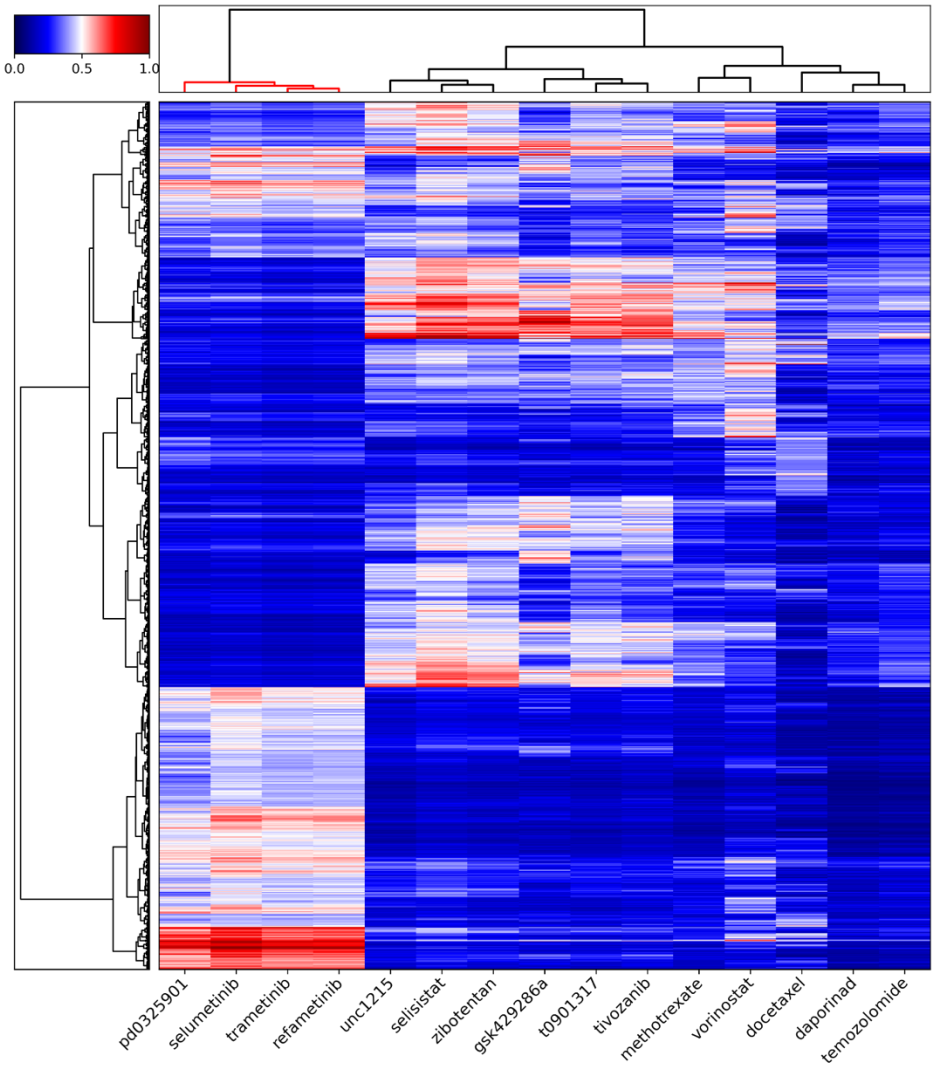



**Figure 5:** The clustering of 15 drugs based on contribution scores of their genes. The contribution scores of the union of genes implicated for these drugs is used to cluster drugs using hierarchical clustering. The heatmap shows the contribution scores.

**Mutation landscape of TCGA tumor samples and their association with drug response**

Next, we sought to evaluate the mutation landscape of tumors in TCGA dataset and their associations with drug response predicted using Big-DRP+. For this purpose, we predicted the normalized logIC50 of 9067 TCGA tumors (that had both mutation and GEx data) corresponding to 32 cancer types to 237 drugs in our training dataset (Supplementary Table S9, Methods). We

587 identified 10 genes that were mutated in more than 10% of the samples (Supplementary Table
588 S9) and performed two-sided Mann–Whitney U tests to assess the association between
589 mutations in these genes and drug response (the FDR values reported in this study correspond
590 to this test). In this section we focus on the insights obtained from *PIK3CA* mutation due to its
591 important role in determining the drug response in various cancers and its potential as a
592 therapeutic target [45] (results of statistical tests for all genes are provided in Supplementary
593 Table S9).

594

595 *PIK3CA* is an oncogene whose mutation leads to hyperactivation of PI3K/AKT/mTOR pathway,
596 associated with cancer progression and poor outcome in many cancer types [46-49]. Various
597 targeted therapies have been developed to target and inhibit this pathway in patients with
598 deregulation and hyperactivity of PI3K/AKT/mTOR pathway (due to *PIK3CA* mutation or other
599 mechanisms such as loss or inactivation of *PTEN*) [50]. Additionally, various studies have shown
600 that mutation in this gene is associated with better response to PI3K inhibitors both *in vitro* and
601 *in vivo* [50, 51]. Consistent with these, our pan-cancer analyses showed that tumors that harbor
602 this mutation are significantly more sensitive to drugs targeting PI3K/AKT/mTOR pathway (Figure
603 6A, one-sided Wilcoxon signed rank test P = 1.14E-5) such as the pan-AKT kinase inhibitor
604 GSK690693 (FDR = 2.13E-59) and the pan-class I PI3K inhibitor ZSTK474 (FDR = 1.15E-29)
605 (Supplementary Table S9).

606

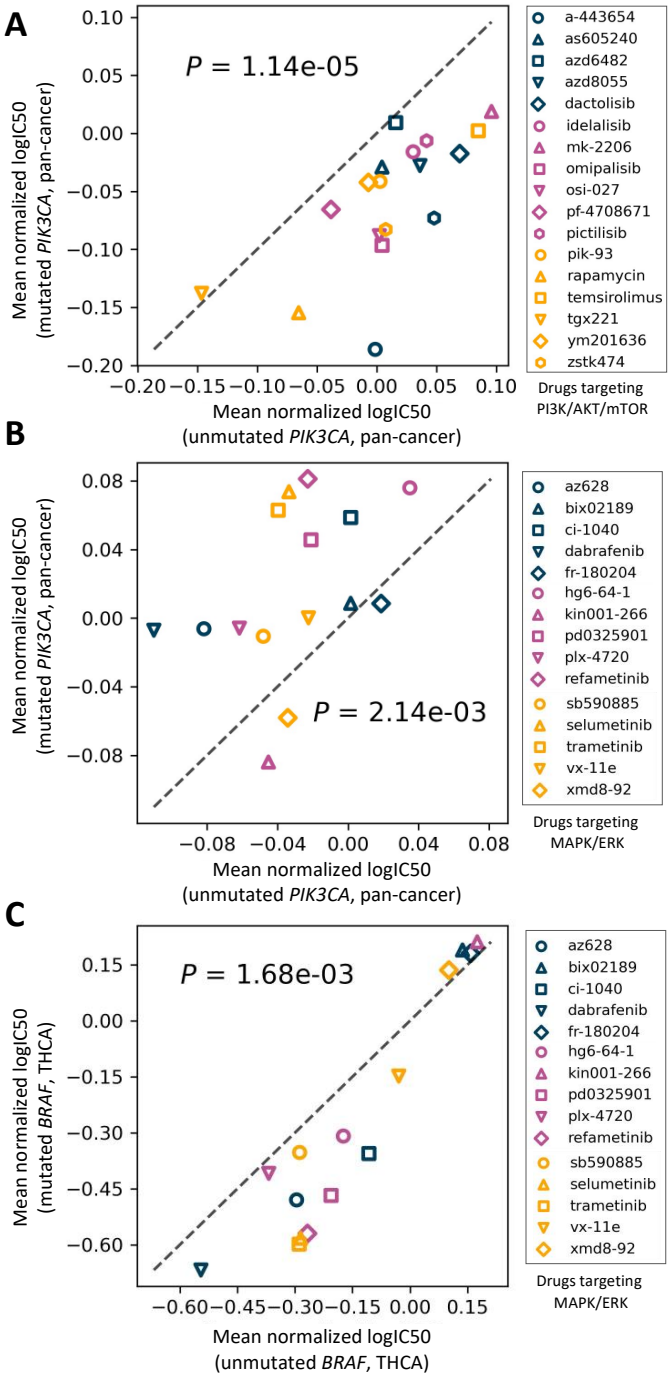
607 On the other hand, mutation in this gene was associated with increase in resistance to drugs
608 targeting the MAPK/ERK signaling pathway (Supplementary Table S9). In particular, the mean
609 predicted normalized logIC50 of *PIK3CA*-mutated tumors were significantly larger for drugs
610 targeting this pathway compared to tumors that did not harbor this mutation (one-sided
611 Wilcoxon signed ranked test P = 2.14E-3, Figure 6B). Various studies (both *in vivo* and *in vitro)*
612 have shown a regulatory link between MAPK/ERK and PI3K/AKT/mTOR pathways and inhibition
613 of MAPK/ERK signaling has been linked to an increase in the activity of PI3K/AKT/mTOR pathway
614 ([52] and references therein). Previous studies have shown that hyperactivity in PI3K/AKT/mTOR
615 pathway as a result of *PIK3CA* mutation increases drug resistance to dabrafenib and trametinib
616 (drugs targeting MAPK/ERK pathway), supporting our observations (dabrafenib FDR = 2.93E-18,
617 trametinib FDR = 7.64E-9, Supplementary Table S9). Mutation in *PIK3CA* has been shown to
618 confer resistance to PD0325901 [53], a MEK-inhibitor that decreases MAPK/ERK pathway activity,
619 and genetic ablation of the mutant allele of this gene has been shown to increase sensitivity to
620 this drug in MEK-inhibitor resistant cells [53]. Our analysis also showed that tumors harboring
621 *PIK3CA* mutation are more resistant to this drug (FDR = 1.33E-5).

622

623 Among the four drugs that target IGF1R, three showed a significantly higher predicted logIC50
624 value in *PIK3CA*-mutated tumors. Previous studies have shown a link between this protein and
625 *PIK3CA*-driven ovarian cancer [54] and breast cancer tumors harboring mutation in this gene [55],
626 suggesting the dual inhibition of PI3K and IGF1R as a new therapeutic approach. Another
627 noteworthy example identified by our analyses is cetuximab (FDR = 5.98E-3), which is an

628 epidermal growth factor receptor inhibitor. Previous studies have shown an association between

629 activity of PI3K/AKT/mTOR pathway and resistance to this drug [56].




630

631 **Figure 6:** The association between mutations and drug response in TCGA. The scatter plots show
632 the mean predicted normalized logIC50 for mutated and unmutated tumors. P-values are

633 calculated using a one-sided Wilcoxon signed rank test. A) The association between *PIK3CA*
634 mutation and response to drugs targeting the PI3K/AKT/mTOR pathway in our pan-cancer study.
635 B) The association between *PIK3CA* mutation and response to drugs targeting the MAPK/ERK
636 pathway in our pan-cancer study. C) The association between *BRAF* mutation and response to
637 drugs targeting the MAPK/ERK pathway in THCA.

638

639 To assess the effect of mutations on drug resistance in a cancer type-specific manner, we focused
640 on thyroid carcinoma (THCA), which is the most common endocrine malignancy, as an illustrating
641 example [57]. In this cancer type, only *BRAF* was mutated in more than 10% of the samples
642 (mutated in 57.7% of tumors). The mutation in this gene, the most frequent of which in thyroid
643 cancer is V600E mutation [57, 58], activates the MAPK/ERK pathway resulting in sustained cell
644 proliferation adverse phenotypes [57]. Various studies have proposed this pathway as a
645 therapeutic target, and have shown that cancer cells (including those corresponding to thyroid
646 cancers) harboring this mutation are much more sensitive to BRAF-inhibitors (e.g., AZ628 [59])
647 and various MEK-inhibitors [60]. Our analyses also showed that THCA tumors harboring *BRAF*
648 mutation are significantly more sensitive to drugs targeting MAPK/ERK pathway (Figure 6C, one-
649 sided Wilcoxon signed rank test P = 1.68E-3), including BRAF-inhibitors AZ628 (FDR = 2.25E-21)
650 and HG6-64-1 (FDR = 5.62E-12), and MEK-inhibitors such as trametinib (FDR = 2.83E-26),
651 refametinib (FDR = 1.69E-25), and selumetinib (FDR = 1.75E-25).

652

653 Taken together, these results suggest the utility of our proposed model in providing insights in
654 pharmacogenomics studies.

655

656 **DISCUSSION AND CONCLUSION**

657 In this study, we proposed two novel graph representation deep learning methods to incorporate
658 information regarding the sensitivity and resistance of cell lines, their gene expression profiles
659 and chemical drug attributes to obtain better drug representations. Using cross-validation and
660 different data splitting methods we showed significant improvement compared to traditional and
661 state-of-the-art methods. Using a computational pipeline to make neural networks explainable,
662 we identified a set of genes that substantially contribute to the predictive model. These genes
663 implicated important signaling pathways and pointed to shared and unique mechanisms of action
664 in the drugs. In addition, we performed a study on the association between the mutation status
665 of cancer tumors from TCGA and the predicted drug response. These analyses revealed various
666 insights, many of which were confirmed by independent studies, which further illustrates the
667 utility of our pipeline in pharmacogenomics studies.

668

669 Moreover, detailed evaluation of our methods showed a high degree of robustness towards
670 changes in the threshold used to form the bipartite graph. This further supports the importance
671 of different techniques we used to ensure stability of our proposed architecture: the
672 normalization factor and the injected self-loop in our H-GCN's forward pass. More specifically,
673 due to the injected self-loop, the nodes retain a portion of their own information, which forces
674 the embeddings to have some level of separation. The normalization factor also helps by
675 preventing the received messages from becoming too large and overpowering the self-loop. It is
676 important to note that this robustness may not be applicable to some corner cases. For example,
677 when a drug's connected CCLs are not connected to any other drug (i.e., it forms a disconnected

678 star subgraph), this drug's embedding would not benefit from the existence of the second H-GCN
679 layer. As another example, the second H-GCN layer will be obsolete if all the drugs happen to
680 form disconnected stars, and thus no information sharing will take place across drugs. Another
681 example is when we add a new drug that results in a disconnected node. A disconnected node
682 will not be able to incorporate CCL information into the drug embedding, which defeats the
683 purpose of the H-GCN.
684
685 Unlike many previous models (e.g., NRL2DRP [30]) that require both cell lines and drugs to be
686 present in the training set, BiG-DRP is designed to enable prediction of unseen cell lines (those
687 that are not present in the training set). However, the drug embedding part of the model (the H-
688 GCN) requires the drugs to be part of the bipartite graph. This constraint implies that the drugs
689 present in the test set must be also present in the training set. As a result, this model generally
690 is not applicable to predict the response of CCLs to unseen new drugs. Although this could be
691 naively remedied by assuming known edges involving the unseen drug in the bipartite graph, this
692 kind of solution is impractical and would be difficult to enact without reducing the test set.
693 However, in most practical applications (e.g., prediction of drug response of cancer patients [4]
694 and [5]), it is more crucial for the model to generalize to unseen samples (CCLs or patients). The
695 reason is that before a drug enters clinical trial or enters clinical usage, many *in vitro* studies on
696 CCLs are first performed. Consequently, one can expect to have access to molecular description
697 and drug response of a drug for which the drug responses of a new set of samples (CCLs or
698 patients) are to be predicted.
699

700 In this study, instead of directly using the logIC50 of drugs, we normalized the logIC50 of each
701 drug (separately) across the CCLs. This was done first to ensure that our prediction performance
702 results are not artificially inflated and second to make the drug response ranges of different drugs
703 comparable to allow the model to learn useful representations across drugs. However, this
704 normalization means that the predicted values should not be used to compare the potency of
705 different drugs, but rather should be used to compare the sensitivity of different CCLs to a specific
706 drug. This is why when we reported our prediction performance results, we calculated them one
707 drug at a time (across CCLs). If one wants to recover logIC50 values, these predictions can be
708 easily modified to reverse the normalization and allow comparison of different drugs for the
709 same CCL.
710
711 One of the main motivations of this study was to improve the representations of drugs for the
712 task of drug response prediction. While direct drug targets or SMILES chemical information of
713 drugs are common approaches for representing drugs, we believe these representations can be
714 improved by capturing the effects these drugs have on CCLs, either by measuring the changes in
715 the GEx profiles of CCLs after administration of the drug (e.g., LINCS dataset [61]) or using the
716 bipartite graph formulation proposed in this study. Improved drug representations are
717 particularly important in more challenging tasks such as prediction of response to drug
718 combinations, in which the sheer number of possible drug combinations (even for drug-pairs)
719 means that experimental measurements can only capture a very small portion of all possibilities.
720 As a result of this small sample size problem, more informative and robust drug representations

become crucial in developing generalizable machine learning models for drug combinations, a direction that we will pursue in the future by generalizing the models introduced in this study.

**Funding:**

This work was supported by the Government of Canada's New Frontiers in Research Fund (NFRF) [NFRFE-2019-01290] (AE), by Natural Sciences and Engineering Research Council of Canada (NSERC) grant RGPIN-2019-04460 (AE), and by McGill Initiative in Computational Medicine (MiCM) (AE). This work was also funded by Génome Québec, the Ministère de l'Économie et de l'Innovation du Québec, IVADO, the Canada First Research Excellence Fund and Oncopole, which receives funding from Merck Canada Inc. and the Fonds de Recherche du Québec – Santé (AE). This research was enabled in part by support provided by Calcul Québec (www.calculquebec.ca) and Compute Canada (www.computecanada.ca).

**Authors' contributions:**

AE and DEH conceived the study, designed the project and the algorithms. DEH implemented the pipeline. DEH and YL ran the baseline methods and performed the statistical analyses of the results. All authors read and approved the final manuscript.

## SUPPLEMENTARY TABLES

**Supplementary Table S1:** The performance of BiG-DRP+, BiG-DRP and baseline methods (columns) using five-fold CV for each drug (rows). Values presented are the mean across five folds.

Each tab corresponds to a performance metric (Spearman's correlation coefficient and area under the receiver operating characteristic) for each data-splitting scenario: leave-cell lines-out (LCO) and leave-pairs-out (LPO).

**Supplementary Table S2:** The results of the one-sided Wilcoxon signed rank test, represented by the p-values. The test compares the per-drug Spearman's correlation coefficient (SCC) of the BiG-DRP+ and the other methods, where the alternative is that BiG-DRP+'s SCC is significantly larger.

**Supplementary Table S3:** The description of samples and results of statistical tests for prediction of clinical drug response of TCGA cancer patients.

**Supplementary Table S4:** The performance of BiG-DRP and BiG-DRP+ using for $k \in \{0.5, 1, 2, 5, 10\}$. The performance metrics were calculated independently per drug and presented as the mean and standard deviation.

**Supplementary Table S5:** The cluster assignments of the aggregated bipartite graph using the nested stochastic block model. The first tab shows that cluster indices for the drugs and the second tab shows the cluster indices for the cell lines. Although the graph partitioning was performed on the entire graph, drug and cell line clusters were mutually exclusive; as a result, similar indices in different tabs do not pertain to the same cluster.

**Supplementary Table S6:** Characterization of CCL clusters in the bipartite graph with respect to tissue type, cancer type and driver mutations.

**Supplementary Table S7:** The top genes and their normalized contribution scores for top-performing drugs in the leave-cell lines-out scenario. Each tab corresponds to the top genes for a specific drug.

**Supplementary Table S8:** The results of the pathway enrichment analysis of the top genes (Supplementary Table S7) on the Reactome pathways using Fisher's exact test. Each tab corresponds to a specific drug. The corrected p-values are indicated in the *pvalue_cor* column.

**Supplementary Table S9:** The association between predicted drug response and mutation status of TCGA samples for our pan-cancer and THCA study.

## REFERENCES


1. Costello JC, Heiser LM, Georgii E, Gonen M, Menden MP, Wang NJ, Bansal M, Ammad-ud-din M, Hintsanen P, Khan SA, et al: **A community effort to assess and improve drug sensitivity prediction algorithms.** *Nat Biotechnol* 2014, **32:**1202-1212.
2. Sharifi-Noghabi H, Zolotareva O, Collins CC, Ester M: **MOLI: multi-omics late integration with deep neural networks for drug response prediction.** *Bioinformatics* 2019, **35:**i501-i509.

783 3. Geeleher P, Zhang Z, Wang F, Gruener RF, Nath A, Morrison G, Bhutra S, Grossman RL, Huang RS:
784 **Discovering novel pharmacogenomic biomarkers by imputing drug response in cancer patients**
785 **from large genomics studies.** *Genome research* 2017, **27:**1743-1751.
786 4. Huang EW, Bhope A, Lim J, Sinha S, Emad A: **Tissue-guided LASSO for prediction of clinical drug**
787 **response using preclinical samples.** *PLoS computational biology* 2020, **16:**e1007607.
788 5. Hostallero DE, Wei L, Wang L, Cairns J, Emad A: **A Deep Learning Framework for Prediction of**
789 **Clinical Drug Response of Cancer Patients and Identification of Drug Sensitivity Biomarkers**
790 **using Preclinical Samples.** *bioRxiv* 2021, **2021.07.06.451273**.
791 6. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, Bindal N, Beare D, Smith JA,
792 Thompson IR, others: **Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic**
793 **biomarker discovery in cancer cells.** *Nucleic acids research* 2012, **41:**D955-D961.
794 7. Deng L, Cai Y, Zhang W, Yang W, Gao B, Liu H: **Pathway-Guided Deep Neural Network toward**
795 **Interpretable and Predictive Modeling of Drug Sensitivity.** *J Chem Inf Model* 2020, **60:**4497-4505.
796 8. Nguyen T-T, Nguyen GTT, Nguyen T, Le D-H: **Graph convolutional networks for drug response**
797 **prediction.** *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2021.
798 9. Liu P, Li H, Li S, Leung KS: **Improving prediction of phenotypic drug response on cancer cell lines**
799 **using deep convolutional network.** *BMC Bioinformatics* 2019, **20:**408.
800 10. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, et al:
801 **PubChem in 2021: new data content and improved web interfaces.** *Nucleic Acids Res* 2021,
802 **49:**D1388-D1395.
803 11. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, Mutowo P, Atkinson F, Bellis
804 LJ, Cibrian-Uhalte E, et al: **The ChEMBL database in 2017.** *Nucleic Acids Res* 2017, **45:**D945-D954.
805 12. Landrum G: **RDKit: Open-source cheminformatics.** 2010.
806 13. Rogers D, Hahn M: **Extended-connectivity fingerprints.** *Journal of chemical information and*
807 *modeling* 2010, **50:**742-754.
808 14. Duvenaud D, Maclaurin D, Aguilera-Iparraguirre J, Gómez-Bombarelli R, Hirzel T, Aspuru-Guzik A,
809 Adams RP: **Convolutional networks on graphs for learning molecular fingerprints.** In *Advances*
810 *in Neural Information Processing Systems (NeurIPS)*. pp. 2224–2232. Montreal, Canada: MIT Press;
811 2015:2224–2232.
812 15. Wang L, Li X, Zhang L, Gao Q: **Improved anticancer drug response prediction in cell lines using**
813 **matrix factorization with similarity regularization.** *BMC Cancer* 2017, **17:**513.
814 16. Suphavilai C, Bertrand D, Nagarajan N: **Predicting cancer drug response using a recommender**
815 **system.** *Bioinformatics* 2018, **34:**3907-3914.
816 17. Guan NN, Zhao Y, Wang CC, Li JQ, Chen X, Piao X: **Anticancer Drug Response Prediction in Cell**
817 **Lines Using Weighted Graph Regularized Matrix Factorization.** *Mol Ther Nucleic Acids* 2019,
818 **17:**164-174.
819 18. Zhang L, Chen X, Guan NN, Liu H, Li JQ: **A Hybrid Interpolation Weighted Collaborative Filtering**
820 **Method for Anti-cancer Drug Response Prediction.** *Front Pharmacol* 2018, **9:**1017.
821 19. Liu H, Zhao Y, Zhang L, Chen X: **Anti-cancer Drug Response Prediction Using Neighbor-Based**
822 **Collaborative Filtering with Global Effect Removal.** *Mol Ther Nucleic Acids* 2018, **13:**303-311.
823 20. Yang J, Li A, Li Y, Guo X, Wang M: **A novel approach for drug response prediction in cancer cell**
824 **lines via network representation learning.** *Bioinformatics* 2019, **35:**1527-1535.
825 21. Kipf TN, Welling M: **Semi-supervised classification with graph convolutional networks.** In
826 *International Conference on Learning Representations (ICLR)*. 2017
827 22. Weininger D: **SMILES, a chemical language and information system. 1. Introduction to**
828 **methodology and encoding rules.** *Journal of Chemical Information and Modeling* 1988, **28:**31-36.

23. van der Meer D, Barthorpe S, Yang W, Lightfoot H, Hall C, Gilbert J, Francies HE, Garnett MJ: **Cell Model Passports-a hub for clinical, genetic and functional datasets of preclinical cancer models.** *Nucleic Acids Res* 2019, **47:**D923-D929.
24. Kingma DP, Ba J: **Adam: A method for stochastic optimization.** In *International Conference on Learning Representations (ICLR)*. 2015
25. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM: **The Cancer Genome Atlas Pan-Cancer analysis project.** *Nat Genet* 2013, **45:**1113-1120.
26. Geeleher P, Cox NJ, Huang RS: **Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines.** *Genome Biol* 2014, **15:**R47.
27. Behdenna A, Haziza J, Azencott C-A, Nordor A: **pyComBat, a Python tool for batch effects correction in high-throughput molecular data using empirical Bayes methods.** *bioRxiv* 2021.
28. Williams C, Seeger M: **Using the Nyström method to speed up kernel machines.** In *Advances in Neural Information Processing Systems (NeurIPS)*. 2001: 682-688.
29. Guyon I, Weston J, Barnhill S, Vapnik V: **Gene Selection for Cancer Classification using Support Vector Machines.** *Machine Learning* 2002, **46:**389-422.
30. Tang J, Qu M, Wang M, Zhang M, Yan J, Mei Q: **Line: Large-scale information network embedding.** In *International Conference on World Wide Web (WWW)*. 2015: 1067-1077.
31. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2000, **28:**27-30.
32. Szklarczyk D, Santos A, von Mering C, Jensen LJ, Bork P, Kuhn M: **STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data.** *Nucleic Acids Res* 2016, **44:**D380-384.
33. Schwab P, Karlen W: **CXPlain: Causal Explanations for Model Interpretation under Uncertainty.** In *Advances in Neural Information Processing Systems (NeurIPS)*. 2019
34. Granger CW: **Investigating causal relations by econometric models and cross-spectral methods.** *Econometrica: journal of the Econometric Society* 1969**:**424-438.
35. Satopaa V, Albrecht J, Irwin D, Raghavan B: **Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior.** In *International Conference on Distributed Computing Systems Workshops*. 2011: 166-171.
36. Blatti C, 3rd, Emad A, Berry MJ, Gatzke L, Epstein M, Lanier D, Rizal P, Ge J, Liao X, Sobh O, et al: **Knowledge-guided analysis of "omics" data using the KnowEnG cloud platform.** *PLoS Biol* 2020, **18:**e3000583.
37. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, Sidiropoulos K, Cook J, Gillespie M, Haw R, et al: **The reactome pathway knowledgebase.** *Nucleic Acids Res* 2020, **48:**D498-D503.
38. Chiu Y-C, Chen H-IH, Zhang T, Zhang S, Gorthi A, Wang L-J, Huang Y, Chen Y: **Predicting drug response of tumors from integrated genomic profiles by deep neural networks.** *BMC Medical Genomics* 2019, **12:**18.
39. Peixoto TP: **Hierarchical block structures and high-resolution model selection in large networks.** *Physical Review X* 2014, **4:**011047.
40. Rand WM: **Objective Criteria for the Evaluation of Clustering Methods.** *Journal of the American Statistical Association* 1971, **66**.
41. Sizemore GM, Pitarresi JR, Balakrishnan S, Ostrowski MC: **The ETS family of oncogenic transcription factors in solid tumours.** *Nat Rev Cancer* 2017, **17:**337-351.
42. Lopez-Delisle L, Pierre-Eugene C, Louis-Brennetot C, Surdez D, Raynal V, Baulande S, Boeva V, Grossetete-Lalami S, Combaret V, Peuchmaur M, et al: **Activated ALK signals through the ERK-ETV5-RET pathway to drive neuroblastoma oncogenesis.** *Oncogene* 2018, **37:**1417-1429.

876 43. Ranzani M, Alifrangis C, Perna D, Dutton-Regester K, Pritchard A, Wong K, Rashid M, Robles-
877 Espinoza CD, Hayward NK, McDermott U, et al: **BRAF/NRAS wild-type melanoma, NF1 status and**
878 **sensitivity to trametinib.** *Pigment Cell Melanoma Res* 2015, **28:**117-119.
879 44. Wang B, Krall EB, Aguirre AJ, Kim M, Widlund HR, Doshi MB, Sicinska E, Sulahian R, Goodale A,
880 Cowley GS, et al: **ATXN1L, CIC, and ETS Transcription Factors Modulate Sensitivity to MAPK**
881 **Pathway Inhibition.** *Cell Rep* 2017, **18:**1543-1557.
882 45. Gustin JP, Cosgrove DP, Park BH: **The PIK3CA gene as a mutated target for cancer therapy.** *Curr*
883 *Cancer Drug Targets* 2008, **8:**733-740.
884 46. Alzahrani AS: **PI3K/Akt/mTOR inhibitors in cancer: At the bench and bedside.** *Semin Cancer Biol*
885 2019, **59:**125-132.
886 47. West KA, Castillo SS, Dennis PA: **Activation of the PI3K/Akt pathway and chemotherapeutic**
887 **resistance.** *Drug Resist Updat* 2002, **5:**234-248.
888 48. Liu R, Chen Y, Liu G, Li C, Song Y, Cao Z, Li W, Hu J, Lu C, Liu Y: **PI3K/AKT pathway as a key link**
889 **modulates the multidrug resistance of cancers.** *Cell Death & Disease* 2020, **11:**797.
890 49. Dong C, Wu J, Chen Y, Nie J, Chen C: **Activation of PI3K/AKT/mTOR Pathway Causes Drug**
891 **Resistance in Breast Cancer.** *Front Pharmacol* 2021, **12:**628690.
892 50. Yang J, Nie J, Ma X, Wei Y, Peng Y, Wei X: **Targeting PI3K in cancer: mechanisms and advances in**
893 **clinical trials.** *Molecular Cancer* 2019, **18:**26.
894 51. Wang M, Li J, Huang J, Luo M: **The Predictive Role of PIK3CA Mutation Status on PI3K Inhibitors**
895 **in HR+ Breast Cancer Therapy: A Systematic Review and Meta-Analysis.** *Biomed Res Int* 2020,
896 **2020:**1598037.
897 52. Jokinen E, Koivunen JP: **MEK and PI3K inhibition in solid tumors: rationale and evidence to date.**
898 *Ther Adv Med Oncol* 2015, **7:**170-180.
899 53. Halilovic E, She QB, Ye Q, Pagliarini R, Sellers WR, Solit DB, Rosen N: **PIK3CA mutation uncouples**
900 **tumor growth and cyclin D1 regulation from MEK/ERK and mutant KRAS signaling.** *Cancer Res*
901 2010, **70:**6804-6814.
902 54. Zorea J, Prasad M, Cohen L, Li N, Schefzik R, Ghosh S, Rotblat B, Brors B, Elkabets M: **IGF1R**
903 **upregulation confers resistance to isoform-specific inhibitors of PI3K in PIK3CA-driven ovarian**
904 **cancer.** *Cell Death Dis* 2018, **9:**944.
905 55. Leroy C, Ramos P, Cornille K, Bonenfant D, Fritsch C, Voshol H, Bentires-Alj M: **Activation of**
906 **IGF1R/p110β/AKT/mTOR confers resistance to α-specific PI3K inhibition.** *Breast Cancer*
907 *Research* 2016, **18:**41.
908 56. Beadnell TC, Nassar KW, Rose MM, Clark EG, Danysh BP, Hofmann MC, Pozdeyev N, Schweppe RE:
909 **Src-mediated regulation of the PI3K pathway in advanced papillary and anaplastic thyroid**
910 **cancer.** *Oncogenesis* 2018, **7:**23.
911 57. Crispo F, Notarangelo T, Pietrafesa M, Lettini G, Storto G, Sgambato A, Maddalena F, Landriscina
912 M: **BRAF Inhibitors in Thyroid Cancer: Clinical Impact, Mechanisms of Resistance and Future**
913 **Perspectives.** *Cancers (Basel)* 2019, **11**.
914 58. Xing M: **BRAF mutation in thyroid cancer.** *Endocr Relat Cancer* 2005, **12:**245-262.
915 59. McDermott U, Sharma SV, Dowell L, Greninger P, Montagut C, Lamb J, Archibald H, Raudales R,
916 Tam A, Lee D, et al: **Identification of genotype-correlated sensitivity to selective kinase inhibitors**
917 **by using high-throughput tumor cell line profiling.** *Proceedings of the National Academy of*
918 *Sciences* 2007, **104:**19936-19941.
919 60. Solit DB, Garraway LA, Pratilas CA, Sawai A, Getz G, Basso A, Ye Q, Lobo JM, She Y, Osman I, et al:
920 **BRAF mutation predicts sensitivity to MEK inhibition.** *Nature* 2006, **439:**358-362.
921 61. Koleti A, Terryn R, Stathias V, Chung C, Cooper DJ, Turner JP, Vidovic D, Forlin M, Kelley TT, D'Urso
922 A, et al: **Data Portal for the Library of Integrated Network-based Cellular Signatures (LINCS)**

923 **program: integrated access to diverse large-scale cellular perturbation response data.** *Nucleic*
924 *Acids Res* 2018, **46:**D558-D566.

925