GRAPH ATTENTION NETWORK BASED REPRESENTATION LEARNING FOR CANCER DRUG RESPONSE PREDICTION AND INTERPRETATION

A PREPRINT

Dionizije Fa
Laboratory for machine
learning and knowledge representation
Rudjer Boskovic Institute
dfa@irb.hr

Tomislav Šmuc
Laboratory for machine
learning and knowledge representation
Rudjer Boskovic Institute
tsmuc@irb.hr

June 13, 2021

ABSTRACT

In recent years it has been shown that protein-protein interactions are targetable by drugs which expands the potential pool of drug targets. In this work we integrate protein-protein interactions and genomic features for modeling cancer drug response, which allows us to discover cell line specific interactions that are most predictive of drug response in cancer cell lines. To that end we construct a multimodal neural network for prediction of drug response based on molecular graph structure and cancer cell line protein-protein interactions. Our model gives insight into drug response related protein-protein interactions, all the while improving on the state of the art on the common benchmark dataset.

Keywords Cancer drug response prediction · Pharmacogenomics · Deep learning · Protein-protein interactions

1 Introduction

Designing new cancer therapeutics tailored to individual's genetic makeup, or finding treatments with existing therapeutics which would provide most benefit to individuals is the goal of precision medicine. A key step in applying these treatments is understanding how drug response is mediated by different genetic factors, and how those factors interact with drug features. Many drugs have an effect on protein interaction pathways and can influence protein hubs, which makes protein-protein interactions promising as drug targets, despite numerous challenges (Goncearenco et al., 2017). Cancer pharmacogenomic databases contain a wide variety of drug response and omics data for exploration of drug response related factors, but difficulties still exist in integrating said omics profiles (Güvenç Paltun et al., 2019). For modeling interactions between different complex data structures, e.g. text, video, graph-structured data and combining these in multimodal settings, deep learning has emerged as a suitable class of machine learning models. Due to its power to learn task specific representation it has become the go-to method for molecular properties prediction and in discovery of new therapeutics (Stokes et al., 2020). Graph neural networks have shown to be one of the best approaches for learning molecular representations (Klicpera et al., 2020), however as transformers are achieving state of the art performance on numerous tasks (Lu et al., 2021), they are being successfully used in molecular properties prediction as well. (Łukasz Maziarka et al., 2020).

In our approach we have combined molecular transformers for learning drug representations and graph attention neural networks for learning cell line representations, i.e. protein-protein interactions. Graph attention networks allow us to examine interpretable relevant interactions between nodes of the protein-protein interaction graphs, which could

Code is available at https://github.com/dionizijefa/CCL-PPI-graph-attention

possibly represent biologically plausible drug response related interactions. We show that this approach improves cancer drug response prediction in pharmacogenomic databases, and allows for interpretation of the interactions. This approach solves the omics integration challenge, as additional gene-wise features such as miRNA expression, methylation, etc., can simply be added as node features of the graph. Gene dependencies and interaction strengths can also trivially be encoded as edge features. However, combining these features creates a data bottleneck which reduces the number of fully characterized cell lines available in pharmacogenomic databases.

2 Related work

(Baptista et al., 2020) give an overview of deep learning for cancer drug response prediction. We compared our approach to some of the published multimodal models, which can learn separate representations of drugs and cell line features. tCNNs (Liu et al., 2019) model uses convolutional neural networks for learning on SMILES strings and genomic features. GraphDRP (Nguyen et al., 2020) retained convolutions on genomic features as in tCNNS, but greatly improved the predictive power with learning molecular representations with graph neural networks. Both GraphDRP and tCNNS evaluate their models on different experimental settings as described in Section 5.2, however they transformed IC50 values to a small scale, making comparison difficult. DeepCDR (Liu et al., 2020) integrates omics-specific subnetworks where each subnetwork learns a representation of the cell line in latent space, which are then concatenated into a single vector and the representations are learned using 1D convolutions. PaccMann (Oskooei et al., 2019) is an interpretable multimodal attention based neural network, where genomic features are encoded using an attention-based mechanism on genes that were selected by network propagation of the protein-protein interaction network, and the molecular representations are learned by a SMILES-based encoder.

3 Data and preprocessing methods

3.1 Data

Pharmacogenomic datasets used in this study were NCI-60 (Reinhold et al., 2012), GDSC (Yang et al., 2012), CTRP (Rees et al., 2016; Seashore-Ludlow et al., 2015; Basu et al., 2013). For CTRP and GDSC data we used recalculated IC50 values, provided by PharmacoGx R-package, version 2.1.10 (Smirnov et al., 2016). Cancer cell lines omics data were extracted from DepMap, 20q3 release (doi:10.6084/m9.figshare.12931238.v1.) (Ghandi et al., 2019). List of landmark genes was obtained from the L1000 assay (Subramanian et al., 2017). "Homo sapiens" protein-protein interaction network was downloaded from the STRING database, version 11.0b (Szklarczyk et al., 2016). For regression settings we predict pIC50 values and for the classification settings we threshold the drug responses at 1 μM, so that the drug-cell line pairs with lower values are labeled sensitive, while the others are resistant.

3.2 Preprocessing protein-protein cell lines interaction networks

To construct cancer cell lines representation for drug response prediction we used the Homo sapiens protein-protein interaction network featurized by expression (TPM), copy number variation (CNV) and mutation data from DepMap. Firstly, from the whole PPI network provided by STRING we kept only experimentally confirmed interactions regardless of their combined interaction score. Secondly, we dropped all the proteins in the PPI network which don't have the associated protein coding genes in the landmark genes of the L1000 assay. Next, we merged such PPI network of landmark genes on DepMap data, where TPM, CNV and mutation data(variant classification) were available. Variant classification was one-hot encoded. To summarize, this procedure produced an undirected graph shared across all cell lines with 813 nodes and 4325 edges. Cell lines are differentiated only by the node features, not the connectivity of the protein-protein interactions. The nodes in the network are featurized by one-hot encoded mutation data, TPM and CNV values, while the edges have no features. Cell lines which were not characterized by DepMap were not used in any of the experiments.

3.3 Preprocessing molecular graphs

Drug representations were constructed by extracting features from SMILES strings using RDKit (Landrum et al., 2006). All SMILES strings which didn't produce valid *mol* RDKit objects, as well as all drugs that were individual atoms, or not explicitly connected, e.g. complexes, were not used in the experiments. Drug representations were also constructed as an undirected graph without hydrogen atoms. The numeric node features of such graphs are atomic number, number of hydrogens, while the one-hot encoded features are donor/acceptor/aromatic status, hybridization of the atom. In summary, each molecule is an undirected graph with 11 node features with corresponding distance and adjacency matrices.

4 Models

4.1 Multimodal attention neural network

Model architecture is shown in Figure 1. For learning molecular representations we used Molecule Attention Transformers (Łukasz Maziarka et al., 2020). The molecular structures are input as adjacency, distance and node feature matrices. The original transformer architecture ins Molecule Attention Transformers us modified so that the self-attention layer is an element-wise sum of self-attention, distance and adjacency matrices weighted by tunable scalar hyperparameters. For learning cell line representations from the PPI network, the input PPI graph is fed into 3 consecutive multi-headed graph attention layers (**?**) with skip connections, fisnished of by a global pooling layer. Outputs of molecular and cell line representation-learning layers are then concatenated and input into two final linear layers to produce the predicted IC50 values or class probabilities.

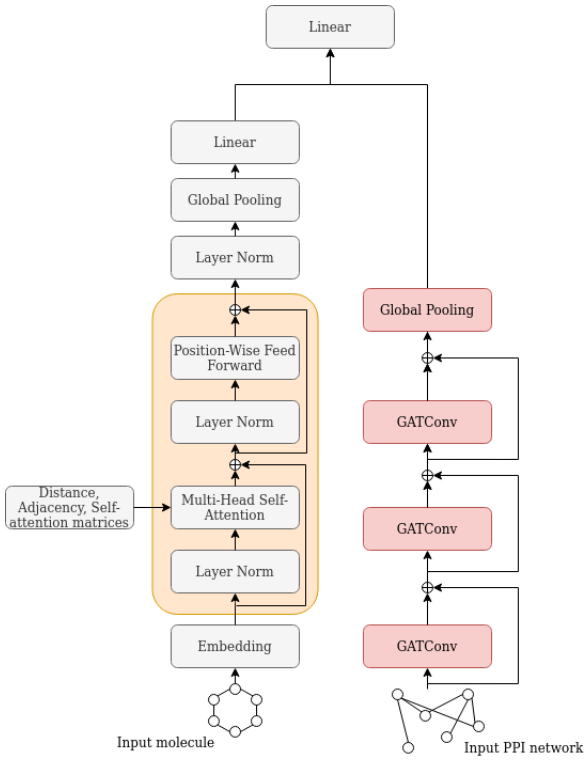

Figure 1: Model architecture

Training was run for at most 300 with a learning rate of 0.0001. We reduced the learning rate by a factor of 0.5 if the performance hadn't increased for 15 epochs. We used early stopping for regularization, either by stopping when validation loss hadn't decreased for 25 epochs in regression, or if validation average precision hadn't increased for 25 epochs in classification tasks. Models were trained on an RTX2080Ti GPU with a batch size of 32. We manually tuned the learning rate by observing the training curve, while no other hyperparameters were tuned.

5 Experiments

5.1 Comparison with state of the art models

We compared our model to the state of the art models, which were trained on the GDSC dataset. The data was obtained from (Nguyen et al., 2020), molecules with non-bonded elements were dropped from the set, cell lines which are not available in DepMap were also dropped. The data split was: train (0.9), validation (0.1) and test (0.1). Drug-cell line pairs for which there were multiple measurements were dropped from the training set, by only keeping the first occurrence of the drug-cell line pair. Results on the test set drawn randomly from all pairs of cell lines and drugs are

given in Table 1. Our model outperforms, to our knowledge, the state of the art model DeepCDR (Liu et al., 2020) on root mean squared error and on Pearson correlation coefficient on the randomly drawn test set

Table 1: Results on a randomly drawn test set

Model	**PPI multimodal**	DeepCDR	GraphDRP	PaccMann*	tCNNs
RMSE	**1.009±0.005**	1.058±0.006	1.091	0.083*	1.782±0.006
Pearson correlation	**0.931±0.001**	0.923±0.006	0.929	0.928	0.885±0.008

Models use different subset of cell lines based on the featurization used
**Results are normalized to [0, 1] interval; Our model RMSE = 0.048*

5.2 Model evaluation on different experimental settings

We further evaluated the classification models for three datasets: NCI60, GDSC and CTRP, Table 3, as well as the benchmark GDSC regression model, Table 2, on three different experimental settings. Predicting response of unknown cell lines (blind cells), predicting response of unknown drugs (blind drugs), and predicting response where both drugs and cell lines are unknown (double blind). For example, blind drugs test set was constructed by randomly sampling 15% of the drugs in the training test, which were used only for testing and not during training.

Table 2: Regression results on 3 experimental settings, GDSC dataset

Experimental setting	Blind cells	Blind drugs	Double blind
RMSE	1.548	2.998	3.098
$\mathbf{R}^2$	0.703	-0.323	-0.372

From Table 2 and Table 3 we see that the easiest task is the blind cells experimental setting, both in classification and regression settings. While in the regression settings the predicted values for blind cells show correlation to true values, in other settings the models performed worse than a horizontal line, see Figure 7. From these results it is also obvious that a random test set, which is often evaluated on the benchmark GDSC dataset, does not carry meaningful information on the usefulness of these models in practical settings, i.e. precision medicine and in discovery of new therapeutics. However in classification settings, the models show much better performance. It should be noted that although classification metrics on NCI-60 dataset show similar values to those of GDSC and CTRP, NCI-60 cell lines have on average only 6.4% of sensitive drugs, while GDSC and CTRP have, 30.2% and 17.3% respectively, all the while the chemical space of NCI-60 drugs is much more diverse. There doesn't seem to be a large difference in predicting unknown drugs and both unknown drugs and unknown cell lines at the same time, possibly due to the fact that these models can generalize well to unseen cell lines.

Table 3: Results on 3 experimental settings, averaged over all cell lines in the individual test sets

Subtask	Unknown cell lines			Unknown drugs			Double blind		
Dataset	GDSC	CTRP	NCI60	GDSC	CTRP	NCI60	GDSC	CTRP	NCI60
Average precision	0.853	0.863	0.838	0.577	0.558	0.426	0.544	0.549	0.399
AUROC	0.921	0.961	0.975	0.657	0.784	0.822	0.614	0.775	0.815

5.3 Depth of protein-protein interaction network

We evaluated how the depth of the protein-protein interaction network affects the performance of the classification models on the test set. Deeper message passing graph neural networks allow messages to be passed from higher-order neighbors, where the k-th layer propagates the messages from k-hop neighbors of the origin node. From Table 4 we see that increasing the number of graph convolution layers (GATConv) doesn't meaningfully improve the performance on the randomly drawn test set, however the interpretation and overlap of the protein-protein interactions differs greatly with the depth of the protein-protein interaction network, see Section 6.

Table 4: Results on the randomly drawn test set with different number of GATConv layers

Dataset	GDSC			CTRP			NCI60		
Num GATConv layers	1	2	3	1	2	3	1	2	3
Average precision	0.877	0.887	0.894	0.875	0.901	0.894	0.861	0.864	0.862
AUROC	0.941	0.945	0.948	0.967	0.974	0.948	0.972	0.973	0.973

6 Analyzing learned protein-protein interactions

Attention coefficients learned by the graph attention layers assign importance to the edges of the PPI network (**?**). Since the graph structure is fixed, importance is directly related to the genomic features used. We extracted attention coefficients for 1226 cell lines in DepMap which can be constructed by our featurization procedure, regardless whether they were present in the training set. To visualize these attention coefficients we performed UMAP dimensionality reduction (n_neighbors=10, min_dist=0.1, learning_rate=0.1) on the attention coefficients of the cell lines, Figure 2, colored by DepMap primary disease classification, e.g. breast cancer. UMAP plots for different number of GATConv layers are available in Supplementary materials 10. Since cell line representations are learned by minimizing the cost function over the available drugs, we would expect that the learned coefficients would be a function of the sampling of the chemical and cell line space. However, across all datasets, even with different chemical and cell line spaces we observe that the clustering of the interactions is preserved across datasets, e.g. blood cancers (left cluster on Figure 2) are seperated in an isolated cluster across all datasets, since these cancers share a large number of interactions across their primary disease. Cell lines that share a larger number of interactions across their primary diseases are generally clustered closer together, see Table 10.

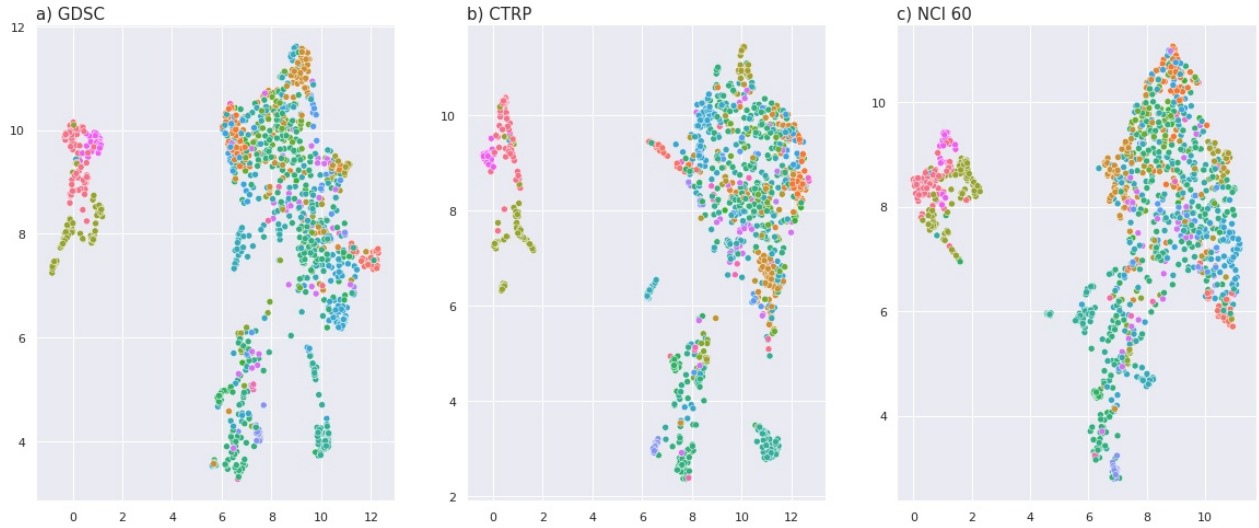

Figure 2: UMAP of the attention coefficients of the three datasets colored by disease tissue for models with PPI depth 3

We hypothesize, that as in gene expression where differentially expressed genes are correlated to drug response and not the most expressed ones, then PPIs most predictive of specific response of cell lines to drugs, would be the ones which differ the most from some baseline value. To investigate these differential PPIs, for each dataset we calculated the mean attention coefficient of the cell lines in the dataset. We then calculated the largest absolute difference from the mean for each cell line. Next, we took the 10 attention coefficients, i.e. protein-protein interactions, with the largest absolute difference from the mean.

$$\triangle a_i = a_i - \frac{1}{N}\sum_{i}^{N} a_i$$

where a_i is the vector of the normalized attention scores, $a_i = \mathbb{R}^{4325}$, and compared their overlap between the three databases and different number of graph attention layers.

Table 5: Mean interaction overlap for top 10 interactions per cell line between combinations of datasets and number of GATConv layers

Dataset	GDSC-CTRP	NCI-CTRP	NCI-GDSC	NCI-GDSC-CTRP
3 GATConv layers	0.384	0.161	0.287	0.029
2 GATConv layers	0.335	0.294	0.740	0.049
1 GATConv layer	2.485	0.071	0.200	0.037

From Table 5 we see that in general the overlap is poor across the datasets. Most of the cell lines have no overlapping interactions across datasets, see Figure 6, except for the case with models trained on GDSC and CTRP datasets with one GATConv layer, which has a much more uniform distrubition of overlapping interactions across cell lines.

Further we tried to evaluate how chemical and cell lines spaces influence the interpretation of the protein-protein interactions, with the caveat that these procedures reduced the number of available data points. We considered the case with the biggest overlap, that is between GDSC and CTRP datasets when the models have a single GATConv layer. This case coincides with models with the lowest complexity, which should, in principle, be less able to (over)fit the training data. We train 4 models for exploration of chemical space, and 4 models for exploration of the cell line space, see Table 7 for the combinations. In the models that estimated if similar chemical spaces produced higher overlap between interactions, cell lines were left as given by the databases, while the drugs were chosen by taking 125 drugs with the highest average Tanimoto similarity of each drug in a dataset (e.g. GDSC) to all of the other drugs in the other dataset (e.g. CTRP). The average Tanimoto similarity of each drug is given by

$$t_i = \frac{1}{N} \sum_{j}^{N} T_{i,j}$$

where t_i is the average Tanimoto similarity of each drug *i* in a dataset, to all the drugs *j* in the other dataset and $T_{i,j}$ is the standard Tanimoto fingerprint similarity between two drugs (default RDKit parameters). This procedure resulted in two models with shared chemical spaces, each model for one of the two datasets. When we estimated the effect of dissimilarity of chemical spaces we dropped the overlapping drugs between datasets, and took 125 drugs with the lowest average Tanimoto similarity to all the drugs in the other dataset. Additional filtering is done so that molecular scaffolds weren't shared between the datasets. To estimate the influence of the similarity of cell line space we trained 2 models, by keeping all the drugs as given by databases, and training only the overlapping cell lines, while for estimating dissimilarity we trained only on the non-overlapping cell lines.

The results of exploration of chemical and cell line spaces on per-cell overlap are given in Table 6 and Figure 3. We didn't observe a drop off in model performance on the reduced training sets under a random data split, see Table 7. Surprisingly we see that the models trained on more dissimilar drugs show larger average per-cell line overlap, 4.185, than those trained on overlapping and more similar drugs, 2.940. In the former case the mean Tanimoto similarity between the two datasets, $\bar{t}_i$, was 0.29, and in the latter 0.39. It is possible that due to variability in IC50 values, and therefore label thresholds, the models could learn different labels for similar drug-cell line pairs, and consequently throw off the interpretation of the interactions. However it would be difficult to confirm this without evaluating factors such as attention coefficients' robustness to model training, different metrics of drug similarities, different cell line spaces, etc. A counter argument would be that average Tanimoto similarity of GDSC drugs to NCI60 drugs is 0.28, similar to the setting of dissimilar drugs between GDSC and CTRP, but there are little to none overlapping interactions. However in this comparison the NCI60 dataset contains only 43 cell lines. From evaluating the cell line space we found that in the setting where we trained on the overlapping cell lines, the average per-cell line overlap was 2.542, while for the non-overlapping cell lines setting the interaction overlap was 2.210. From this it would seem that cell line space has less of an influence on the learned representations, however it could be the case that non-overlapping cell lines are still similar enough in the two databases. If we were to go further and examine the different combinations of cell line and chemical spaces, the already small number of data points would be further reduced.

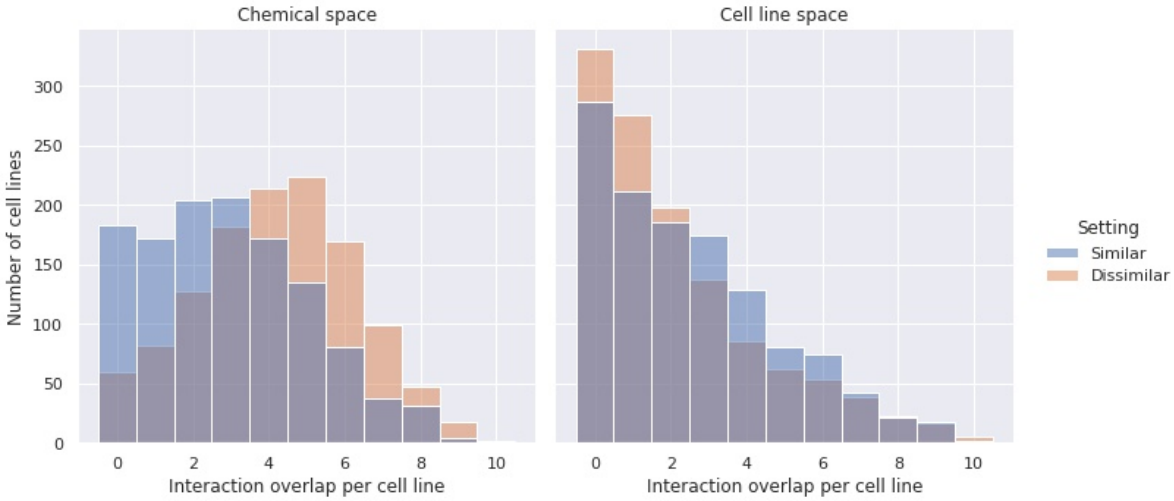

Figure 3: Influence of chemical and cell line spaces on the distribution of per-cell line overlap between GDSC and CTRP datasets for top 10 interactions with 1 GATConv layer.

7 Cancer characteristic interactions

To further investigate drug response related protein-protein interactions, we group cell lines by primary disease and subtype, then calculate how often the top 10 differential interactions appear in each of the groups, cell lines by primary disease, by subtype, and by primary disease and subtype. We consider models with 1 GATConv layer, trained on GDSC and CTRP datasets, which show the biggest overlap see Table 5. Only overlapping interactions between the datasets were considered. List of such interactions is given in Table 10. Fibroblast is the primary disease with the highest number of shared interactions across cell lines. 92.86% of fibroblast cancer cell lines have EED-TLE1 as one of top 10 most important interactions. We also observed that many types of fibroblast cancers have IL1B as an interaction hub. In colorectal adenocarcinoma, we found GLI2-USP7 interaction, which has been shown to be a promising target in cancer (Bufalieri et al., 2020). USP7 is a deubiquitylating enzyme that activates the hedgehog signaling pathway, which is associated with multiple cancers, including colorectal cancer. There have been limited data on the influence of CHECK2 on blood cancers (Stolarova et al., 2020), however we found that CHECK2 interactions were important interactions in many cancers, especially Lymphomas, Myelomas and Leukemias, where many of these cancer cell lines have a CHEK2-CHEK1 interaction. Inhibition of CHEK1 shows a potential as a therapy in blood cancers (Bryant et al., 2014). We have also found that ERBB3 is an important interaction hub in laryngeal squamous cell carcinoma. It has been shown that ERBB3 signaling can influence drug response in head and neck squamous cell carcinoma and that it could potentially be a therapeutic target. (Erjala et al., 2006).

8 Discussion

We introduced a state of the art interpretable multimodal deep learning model for prediction of drug response of cancer cell lines. To our knowledge this is the first such approach, which allows the investigation of biologically plausible drug response related protein-protein interactions, that may be helpful as a prescreening method in exploration of potential protein-protein interactions as drug targets. This approach also allows for integration of different interaction and gene-wise omics data. For some of the discovered interactions, we have found support in the existing literature. Future research could further explore the robustness of interpretation of such interactions, whether gene or protein-protein interaction networks, by using more robust models, smarter data integration and the like. As more data become available it would be interesting to train these models on denser and bigger networks of gene and protein-protein interactions.

9 Acknowledgments

9.1 Funding

This work has been fully supported by the "Research Cooperability" Program of the Croatian Science Foundation funded by the European Union from the European Social Fund under the Operational Programme Efficient Human Resources 2014-2020.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of Croatian Science Foundation, Ministry of Science and Education and European Commission.

References

Delora Baptista, Pedro G Ferreira, and Miguel Rocha. Deep learning for drug response prediction in cancer. *Briefings in Bioinformatics*, 22(1):360–379, 01 2020. ISSN 1477-4054. doi: 10.1093/bib/bbz171. URL https://doi.org/10.1093/bib/bbz171.

Amrita Basu, Nicole E Bodycombe, Jaime H Cheah, Edmund V Price, Ke Liu, Giannina I Schaefer, Richard Y Ebright, Michelle L Stewart, Daisuke Ito, Stephanie Wang, et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell*, 154(5):1151–1161, 2013.

Christopher Bryant, Kirsten Scriven, and Andrew J Massey. Inhibition of the checkpoint kinase chk1 induces dna damage and cell death in human leukemia and lymphoma cells. *Molecular cancer*, 13(1):1–12, 2014.

Francesca Bufalieri, Ludovica Lospinoso Severini, Miriam Caimano, Paola Infante, and Lucia Di Marcotullio. Dubs activating the hedgehog signaling pathway: A promising therapeutic target in cancer. *Cancers*, 12(6):1518, 2020.

Kaisa Erjala, Maria Sundvall, Teemu T Junttila, Na Zhang, Mika Savisalo, Pekka Mali, Jarmo Kulmala, Jaakko Pulkkinen, Reidar Grenman, and Klaus Elenius. Signaling via erbb2 and erbb3 associates with resistance and epidermal growth factor receptor (egfr) amplification with sensitivity to egfr inhibitor gefitinib in head and neck squamous cell carcinoma cells. *Clinical Cancer Research*, 12(13):4103–4111, 2006.

WA Falcon et al. Pytorch lightning. *GitHub. Note: https://github. com/williamFalcon/pytorch-lightning Cited by*, 3, 2019.

Matthias Fey and Jan E. Lenssen. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

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

Alexander Goncearenco, Minghui Li, Franco L Simonetti, Benjamin A Shoemaker, and Anna R Panchenko. Exploring protein-protein interactions as drug targets for anti-cancer therapy with in silico workflows. pages 221–236, 2017.

Betül Güvenç Paltun, Hiroshi Mamitsuka, and Samuel Kaski. Improving drug response prediction by integrating multiple data sources: matrix factorization, kernel and network-based approaches. *Briefings in Bioinformatics*, 22(1): 346–359, 12 2019. ISSN 1477-4054. doi: 10.1093/bib/bbz153. URL https://doi.org/10.1093/bib/bbz153.

Johannes Klicpera, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. *arXiv preprint arXiv:2003.03123*, 2020.

Greg Landrum et al. Rdkit: Open-source cheminformatics. 2006.

Pengfei Liu, Hongjian Li, Shuai Li, and Kwong-Sak Leung. Improving prediction of phenotypic drug response on cancer cell lines using deep convolutional network. *BMC bioinformatics*, 20(1):1–14, 2019.

Qiao Liu, Zhiqiang Hu, Rui Jiang, and Mu Zhou. Deepcdr: a hybrid graph convolutional network for predicting cancer drug response. *Bioinformatics*, 36(Supplement_2):i911–i918, 2020.

Kevin Lu, Aditya Grover, Pieter Abbeel, and Igor Mordatch. Pretrained transformers as universal computation engines, 2021.

Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

Tuan Nguyen, Thin Nguyen, and Duc-Hau Le. Graph convolutional networks for drug response prediction. *bioRxiv*, 2020. doi: 10.1101/2020.04.07.030908. URL https://www.biorxiv.org/content/early/2020/04/09/2020.04.07.030908.

Ali Oskooei, Jannis Born, Matteo Manica, Vigneshwari Subramanian, Julio Sáez-Rodríguez, and María Rodríguez Martínez. Paccmann: Prediction of anticancer compound sensitivity with multi-modal attention-based neural networks, 2019.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

Matthew G Rees, Brinton Seashore-Ludlow, Jaime H Cheah, Drew J Adams, Edmund V Price, Shubhroz Gill, Sarah Javaid, Matthew E Coletti, Victor L Jones, Nicole E Bodycombe, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nature chemical biology*, 12(2):109–116, 2016.

William C Reinhold, Margot Sunshine, Hongfang Liu, Sudhir Varma, Kurt W Kohn, Joel Morris, James Doroshow, and Yves Pommier. Cellminer: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the nci-60 cell line set. *Cancer research*, 72(14):3499–3511, 2012.

Brinton Seashore-Ludlow, Matthew G Rees, Jaime H Cheah, Murat Cokol, Edmund V Price, Matthew E Coletti, Victor Jones, Nicole E Bodycombe, Christian K Soule, Joshua Gould, et al. Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer discovery*, 5(11):1210–1223, 2015.

Petr Smirnov, Zhaleh Safikhani, Nehme El-Hachem, Dong Wang, Adrian She, Catharina Olsen, Mark Freeman, Heather Selby, Deena MA Gendoo, Patrick Grossmann, et al. Pharmacogx: an r package for analysis of large pharmacogenomic datasets. *Bioinformatics*, 32(8):1244–1246, 2016.

Jonathan M Stokes, Kevin Yang, Kyle Swanson, Wengong Jin, Andres Cubillos-Ruiz, Nina M Donghia, Craig R MacNair, Shawn French, Lindsey A Carfrae, Zohar Bloom-Ackermann, et al. A deep learning approach to antibiotic discovery. *Cell*, 180(4):688–702, 2020.

Lenka Stolarova, Petra Kleiblova, Marketa Janatova, Jana Soukupova, Petra Zemankova, Libor Macurek, and Zdenek Kleibl. Chek2 germline variants in cancer predisposition: Stalemate rather than checkmate. *Cells*, 9(12):2675, 2020.

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

Damian Szklarczyk, John H Morris, Helen Cook, Michael Kuhn, Stefan Wyder, Milan Simonovic, Alberto Santos, Nadezhda T Doncheva, Alexander Roth, Peer Bork, et al. The string database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic acids research*, page gkw937, 2016.

Wanjuan Yang, Jorge Soares, Patricia Greninger, Elena J. Edelman, Howard Lightfoot, Simon Forbes, Nidhi Bindal, Dave Beare, James A. Smith, I. Richard Thompson, Sridhar Ramaswamy, P. Andrew Futreal, Daniel A. Haber, Michael R. Stratton, Cyril Benes, Ultan McDermott, and Mathew J. Garnett. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Research*, 41(D1):D955–D961, 11 2012. ISSN 0305-1048. doi: 10.1093/nar/gks1111. URL https://doi.org/10.1093/nar/gks1111.

Łukasz Maziarka, Tomasz Danel, Sławomir Mucha, Krzysztof Rataj, Jacek Tabor, and Stanisław Jastrzębski. Molecule attention transformer. 2020.

10 Supplementary materials

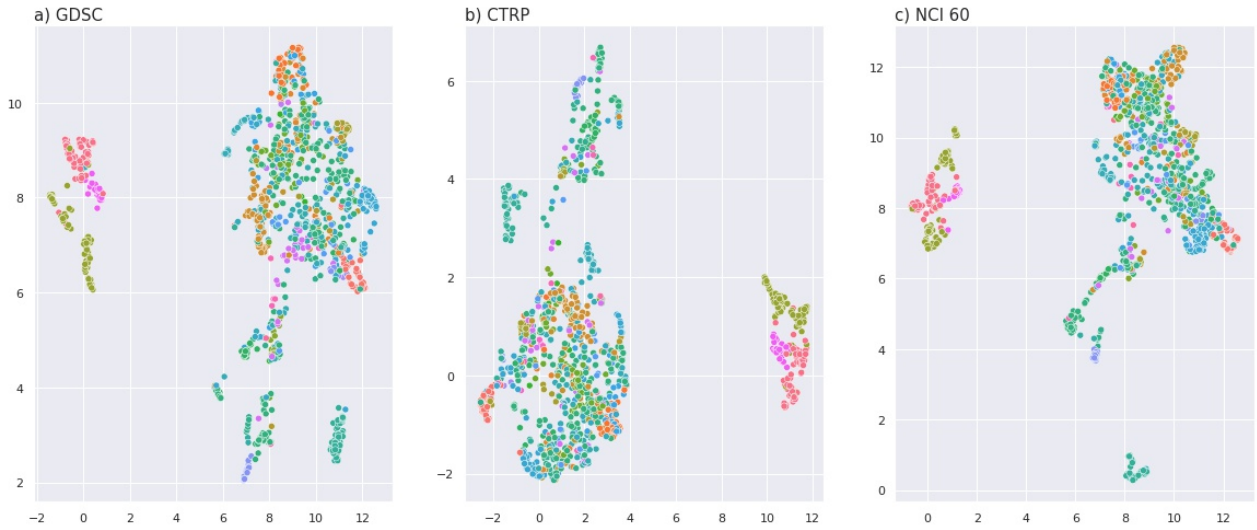

Figure 4: UMAP plot of the attention coefficients of the three datasets colored by disease tissue for models with PPI depth 2

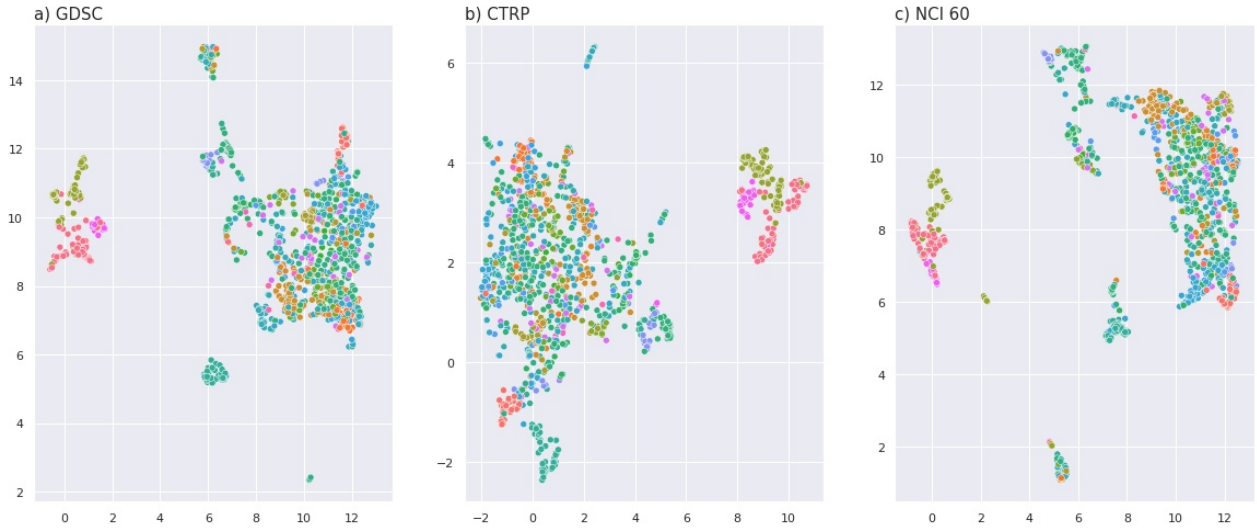

Figure 5: UMAP plot of the attention coefficients of the three datasets colored by disease tissue for models with PPI depth 1

Table 6: Influence of chemical and cell line spaces on the mean overlap per cell line between GDSC and CTRP databases

Setting	Drug similarity	Drug dissimilarity	Cell similarity	Cell dissimilarity
Overlap	2.949	4.185	2.542	2.210

Table 7: Influence of chemical and cell line spaces on the model evaluation metrics

Setting	Drug similarity		Drug dissimilarity		Cell similarity		Cell dissimilarity	
Dataset	GDSC	CTRP	GDSC	CTRP	GDSC	CTRP	GDSC	CTRP
Average precision	0.871	0.827	0.892	0.853	0.892	0.876	0.875	0.852
AUROC	0.937	0.937	0.947	0.964	0.944	0.968	0.931	0.958

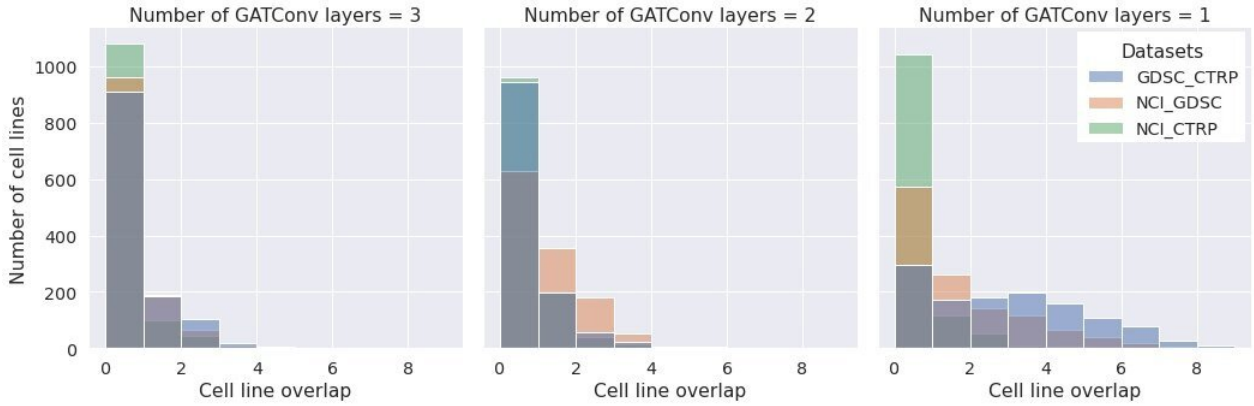

Figure 6: Distribution of interaction overlap per cell line with different number of GATConv layers

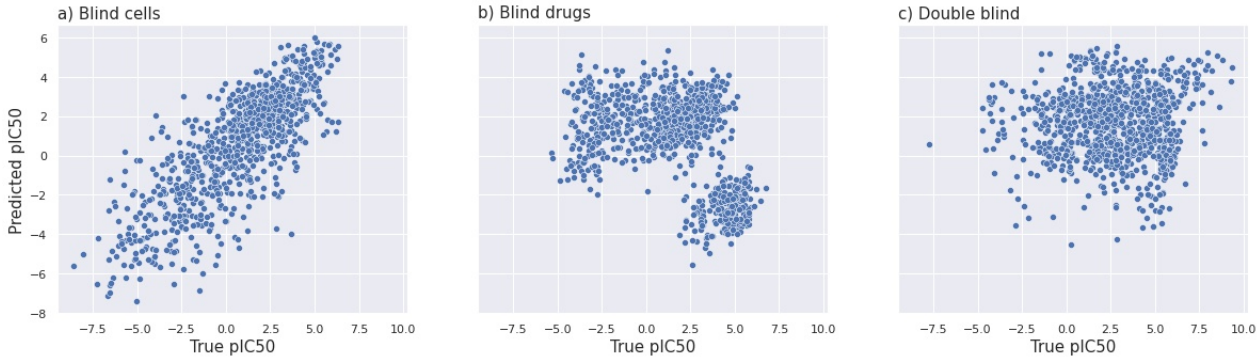

Figure 7: Scatterplot of predicted and true pIC50 values for GDSC benchmark on 3 experimental settings. Plotted values contain the first 1000 elements of the test sets

Table 8: Overlapping interactions between GDSC and CTRP datasets per primary disease for models with 1GATConv layer. Columns show overlap across primary disease in each dataset.

Primary disease	**PPI**	**% overlap in GDSC**	**% overlap in CTRP**
Ovarian Cancer	IGF1R-RASA1	8.7	5.8
Leukemia	UGDH-HSPB1	25.78	24.22
	CHEK2-RAD9A	21.09	15.62
	SHC1-IL4R	19.53	24.22
	RHOA-S100A4	18.75	13.28
	CDK4-CDKN1A	13.28	14.06

Continued on next page

Table 8 – continued from previous page

Primary disease	PPI	% overlap in GDSC	% overlap in CTRP
Colon/Colorectal Cancer	GLI2-USP7	14.47	17.11
	RAE1-MTHFD2	11.84	7.89
	GLI2-SMAD3	10.53	9.21
Skin Cancer	CHEK2-RAD9A	15.84	11.88
	CHEK2-GNAS	14.85	12.87
	NFKBIA-AKT1	14.85	10.89
	CHEK2-CHEK1	13.86	10.89
	UGDH-HSPB1	12.87	16.83
Bladder Cancer	INSIG1-HMGCR	20.51	17.95
	UBQLN2-EDEM1	17.95	17.95
	MBTPS1-HDAC2	12.82	15.38
Kidney Cancer	UBQLN2-EDEM1	25.0	26.92
	INSIG1-HMGCR	21.15	28.85
	MBTPS1-HDAC2	15.38	25.0
	HDAC2-NR3C1	7.69	15.38
	IFNAR1-CDK1	8.64	14.81
Pancreatic Cancer	HSPA8-CD40	28.07	22.81
	CCDC85B-ZFP36	17.54	8.77
Myeloma	SHC1-IL4R	47.06	52.94
	UGDH-HSPB1	47.06	44.12
	CHEK2-RAD9A	32.35	20.59
	CHEK2-GNAS	23.53	23.53
	CHEK2-CHEK1	23.53	26.47
	CASP7-NUCB2	20.59	23.53
Brain Cancer	EED-TLE1	25.49	24.51
	RPS6-NVL	24.51	5.88
	ACAT2-SPR	14.71	17.65
	IL1B-FYN	13.73	22.55
	IL1B-EGR1	9.8	18.63
Sarcoma	EED-TLE1	11.11	11.11
	FOS-EPHB2	11.11	8.33
	ACAT2-SPR	8.33	8.33
Lymphoma	UGDH-HSPB1	55.88	51.96
	CHEK2-RAD9A	41.18	24.51
	CDK4-CDKN1A	28.43	22.55
	CHEK2-CHEK1	23.53	34.31
	CHEK2-GNAS	23.53	26.47
	PTK2B-SYK	21.57	35.29
	SHC1-IL4R	16.67	20.59
	JUN-NOLC1	15.69	19.61
Bone Cancer	GOLT1B-HTRA1	15.28	22.22
	EED-TLE1	11.11	8.33
	ACOT9-EGFR	9.72	8.33
Fibroblast	EED-TLE1	92.86	83.33
	IL1B-EGR1	76.19	85.71
	IL1B-FYN	76.19	80.95
	EED-HDAC2	64.29	80.95
	ACAT2-SPR	42.86	40.48
	IL1B-MYC	26.19	28.57
Gastric Cancer	HSPA8-CD40	10.42	8.33
	FOS-ABL1	8.33	8.33
Engineered	EED-TLE1	44.44	44.44
	IL1B-FYN	33.33	22.22
Thyroid Cancer	GLRX-CASP3	21.05	21.05
	GLRX-PTPN1	21.05	21.05
	ACAT2-SPR	15.79	10.53

Continued on next page

Table 8 – continued from previous page

Primary disease	PPI	% overlap in GDSC	% overlap in CTRP
	HSPA8-CD40	10.53	10.53
Neuroblastoma	AURKA-PSRC1	15.22	19.57
	AURKA-SKP1	13.04	15.22
	PCNA-RAD9A	13.04	15.22
	ACAT2-SPR	10.87	8.7
Prostate Cancer	IGF1R-RASA1	11.11	11.11
Rhabdoid	HSPA8-CD40	65.0	50.0
	GLI2-SOX2	30.0	20.0
	HSPA1A-FOS	20.0	20.0
	HSPA1A-NFKBIB	20.0	15.0
Gallbladder Cancer	CAST-RPP38	25.0	25.0
	CAST-HTRA1	25.0	25.0
	RAE1-SMC3	25.0	25.0
	RAE1-NUP88	25.0	25.0
	CAST-PSIP1	25.0	25.0
Endometrial/Uterine Cancer	IL1B-FYN	15.15	9.09
	MRPL19-MRPS16	9.09	9.09
Head and Neck Cancer	CAST-HTRA1	14.06	18.75
	CAST-RPP38	10.94	14.06
	RAE1-SMC3	9.38	14.06
Bile Duct Cancer	EED-TLE1	11.11	11.11
Esophageal Cancer	HSPA8-CD40	7.89	7.89
	CAST-RPP38	7.89	7.89
Liver Cancer	RAE1-MTHFD2	14.81	22.22
Cervical Cancer	CAST-RPP38	23.53	17.65
	CAST-HTRA1	23.53	23.53
	RAE1-SMC3	11.76	11.76
Adrenal Cancer	FHL2-EGFR	100.0	100.0
	FHL2-MYC	100.0	100.0
	HSPA1A-FOS	100.0	100.0
	HSPA1A-NFKBIB	100.0	100.0
	DAG1-CLIC4	100.0	100.0
	HSPA1A-SHC1	100.0	100.0
Teratoma	PAF1-DNMT3A	100.0	100.0
	GABPB1-CASP2	100.0	100.0
	CDK6-AKT1	100.0	100.0
	HSPA1A-AURKB	100.0	100.0
	HSPA1A-SHC1	100.0	100.0
	CSK-SHC1	100.0	100.0
	CDK6-CCNA2	100.0	100.0