Graph attention network based representation learning for cancer drug response prediction and interpretation

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 (1).

In our approach, shown in Figure 1 we have combined molecular transformers (2) for learning drug representations and graph attention neural networks (3) for learning cell line representations, i.e. exploiting protein-protein interaction (PPI) network as a basic graph representation. Graph attention networks allow us to examine interpretable relevant interactions between nodes of the protein-protein interaction graphs. We show that this approach improves cancer drug response prediction in pharmacogenomic databases, and allows for interpretation of the interactions. This approach also 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.

We benchmarked our approach in different training and testing settings, and compared our models to the state of the art models, which were trained on the GDSC dataset. Our models outperform the state of the art deep learning models like DeepCDR (4), PaccMann (5), GraphDRP (6), tCNNS (7).

We exploited attention coefficients of the learned GAT network as means for validating and interpreting the models. We extracted normalized attention coefficients from models trained to classify cell responses as sensitive or resistant, across three different datasets, NCI-60 (8), GDSC (9) and CTRP (10). We performed UMAP (11) dimensionality reduction on the attention coefficients and found that the clustering of the interactions, by primary disease classification is largely preserved across datasets; in particular this is the case for blood cancers which are represented as a well separated cluster in all datasets. Lastly, we looked at most common overlapping interaction between datasets and cancer types to compare the attention coefficients w.r.t. findings from relevant research literature. As an example, we observed, from a number of significant interactions, that ERBB3 was an important interaction hub in laryngeal squamous cell carcinoma, which is in accordance with experimental findings that ERBB3 signaling can influence drug response in head and neck squamous cell carcinoma and that it could potentially be a therapeutic target(12).

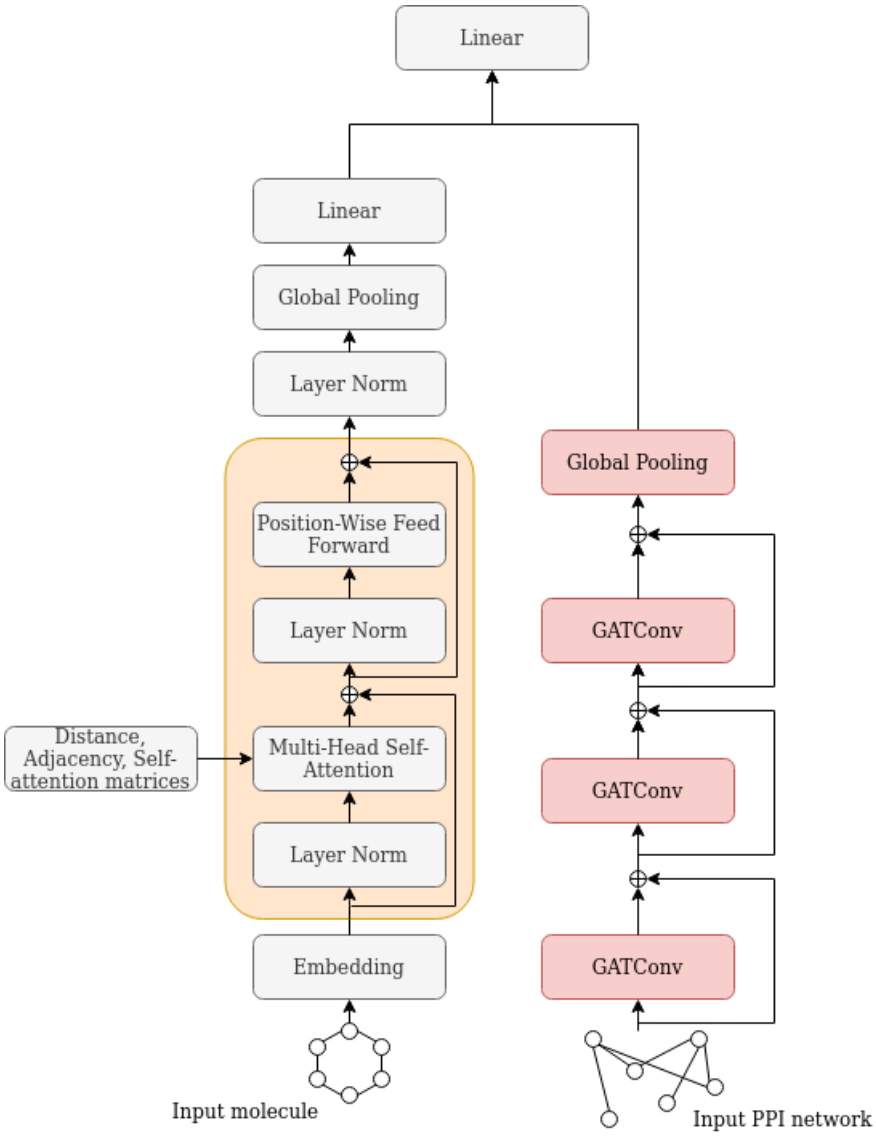

Figure 1: Model architecture

References

[1] 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.

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

[3] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks, 2018.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

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

[12] 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.