INTEGRATION OF GRAPH NEURAL NETWORK AND NEURAL-ODES FOR TUMOR DYNAMICS PREDICTION

Omid Bazgir [*1], **Zichen Wang** [*†1], **Ji Won Park** [2], **Marc Hafner** [3,4], **James Lu** [1]
[1] Modeling & Simulation/ Clinical Pharmacology, Genentech
[2] Prescient Design/ Computational Sciences, Genentech
[3] Oncology Bioinformatics/ Computational Sciences, Genentech
[4] Discovery Oncology/ Research, Genentech
{bazgir.omid, wang.zichen, park.ji_won, hafner.marc, lu.james}@gene.com

ABSTRACT

In the development of anti-cancer drugs, a major scientific challenge is disentangling the complex interplay between high-dimensional genomics data derived from patient tumor samples, the organ of origin of the tumor, the drug targets associated with the specified treatments, and the ensuing treatment response. Furthermore, to realize the aspirations of precision medicine in identifying and adjusting treatments for patients depending on the therapeutic response, there is a need for building tumor dynamics models that can integrate the longitudinal tumor size measurements with multimodal, high-throughput data. In this work, we take a step towards enhancing personalized tumor dynamics predictions by proposing a heterogeneous graph encoder that utilizes a bipartite Graph Convolutional Neural networks (GCNs) combined with Neural Ordinary Differential Equations (Neural-ODEs). We apply the methodology to a large collection of patient-derived xenograft (PDX) data, spanning a wide variety of treatments (as well as their combinations) and tumor organs of origin. We first show that the methodology is able to discover a tumor dynamic model that significantly improves upon an empirical model in current use. Additionally, we show that the graph encoder is able to effectively incorporate multimodal data to enhance tumor predictions. Our findings indicate that the methodology holds significant promise and offers potential applications in pre-clinical settings.

1 INTRODUCTION

In the development of novel anti-cancer therapies, PDX models have become an important platform for addressing key questions, such as evaluating the treatment response to therapeutic agents and the combinations thereof, identifying the relevant biomarkers of response and elucidating the mechanisms of resistance development (Byrne et al., 2017). Furthermore, as PDX models are obtained by surgically removing patients' tumor and implanting them in mice, co-clinical avatar trials Byrne et al. (2017) can be performed, whereby treatment of the patients occur simultaneously to the treatment of the corresponding (pre-clinical) PDX models generated from the same patients. Such avatar studies facilitates real-time clinical decision making and help deliver some of the promises of precision medicine (Naik et al., 2023).

Given the myriad applications of PDX models, an important computational task is the prediction of their dynamic response to treatment from the baseline -omics data and/or early tumor size data. This entails incorporating the high-dimensional omics data measured at baseline (e.g., RNA-seq pre-treatment), with the low-dimensional but serially assessed tumor size measurements under treatment. While empirical Zwep et al. (2021) and spline-based Forrest et al. (2020) tumor dynamics models have been proposed, there has been little progress in melding such dynamic models with high dimensional omics data. Zwep et al. (2021), uses lasso regression to predict tumor dynamic parameters from copy-number variations (CNVs) of genes from a large PDX data set consisting of various treatments

[*]Equal contribution
[†]Contributed to this work during internship at Genentech.

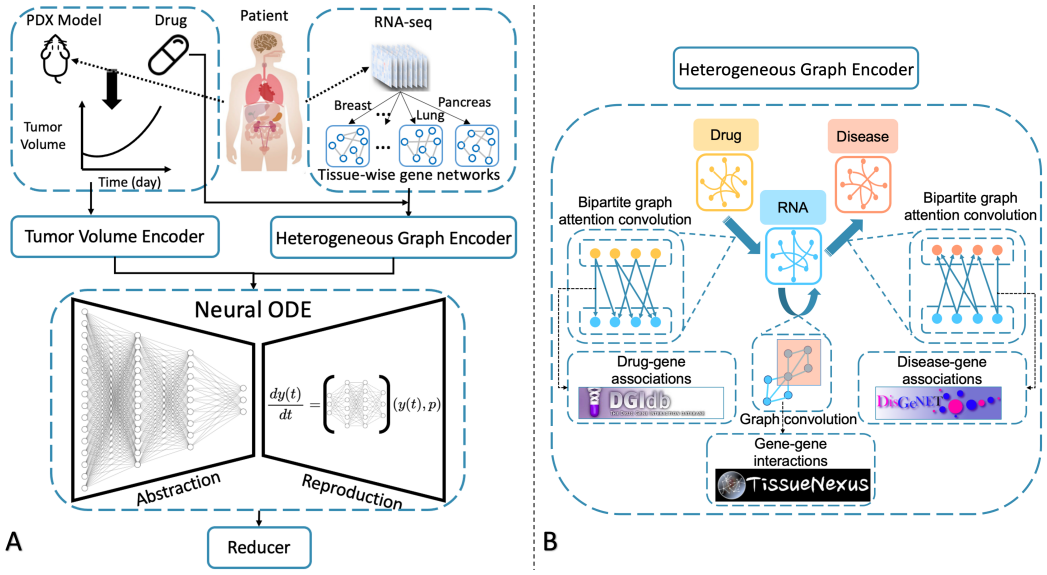

Figure 1: **Model architecture overview. A)** Integration of GCNs and Neural-ODEs for tumor dynamics prediction. **B)** Heterogeneous Graph Encoder: Two bipartite graph attention convolution NNs extract disease-gene and drug-gene associations and a graph convolution NN extracts gene-gene interactions. The output of all three are integrated into the embedding representing the baseline (pre-treatment) state.

(Gao et al., 2015). In Ma et al. (2021), a few-shot learning (multi-layer perceptron) and heterogeneous GCNs Peng et al. (2022), have been proposed to learn drug responses from in-vitro (cell-line) data and predict in-vivo (PDX) outcomes within two response categories. While the predictions made by the models in Ma et al. (2021) and Peng et al. (2022) show promising correlations with the observed responses, they do not necessarily capture the tumor dynamics, which is crucial for clinical decision making. While their proposed methods show promise, improving the predictivity of the model via the inclusion of omics data remains an important topic for further research.

Over the past few years, Neural-Ordinary Differential Equations (NODEs) Chen et al. (2018) has emerged as a promising deep learning (DL) methodology for making predictions from irregularly sampled temporal data. Recently, a methodology based on NODE has been developed for tumor dynamics modeling in the clinical trial setting and the generated embeddings from patients' tumor data have been demonstrated to be effective for predicting their Overall Survival (OS) (Laurie & Lu, 2023). While the Tumor Dynamic NODE (TDNODE) has set the mathematical foundations for modeling longitudinal tumor data Laurie & Lu (2023), a methodology for the incorporation of high-dimensional, multimodal data into such temporal models has yet to be developed.

In this work, we propose a novel way to combine the previously developed tumor volume encoder with a heterogeneous graph encoder (see Fig. 1). The latter takes as inputs multimodal data consisting of drug, disease, and RNA-seq by incorporating graphs that encapsulate drug-gene associations, disease-gene associations, and gene-gene interactions. We applied our methodology to the PDX data of Gao et al. (2015) (more details provided in Appendix A.1). In our proposed method, we show that by leveraging the multimodal data including RNA-seq in conjunction with early tumor response data, we can significantly improve future tumor response predictions. Finally, we summarize the findings and discuss the potential future applications of the proposed approach in enabling precision medicine in oncology.

2 METHOD

Within our multi-modal framework, we construct a multi-relational network using three large datasets covering interactions between drugs, genes, and diseases. We use the RNA-seq data as node features

within tissue-specific knowledge graphs and further integrate it with drug targets of treatments as a heterogeneous graph (Zhang et al., 2019).

This integration allows us to learn an embedding that captures the complex relationship among genes, tumors, and treatments. The three following graphs were utilized to represent both biological and pharmacological knowledge:

- **gene-gene graph:** tissue-specific functional gene networks obtained from TissueNexus Lin et al. (2022), which provides gene-gene associations based on gene expression across 49 human tissues.
- **drug-gene graph:** obtained from DGIdb Cotto et al. (2018), which consolidates information on drug target and interacting genes from 30 disparate sources through expert curation and text mining.
- **disease-gene graph:** obtained from DisGeNET Piñero et al. (2016), one of the largest collections of genes and variants associated with human diseases.

For each PDX instance in our dataset, the pre-treatment embeddings using the above multi-modal graphs are used in conjunction with the early, observed tumor measurements to predict the future tumor dynamic profiles. The following subsections cover these two respective aspects.

2.1 Heterogeneous Graph Encoder

We formulate the PDX representation learning task as one of graph embedding, by fusing information from a heterogeneous network that incorporates drug, disease, and gene relationships. We apply multi-layer GCN Welling & Kipf (2016) in the RNA domain to model the gene interactions (*gene-gene graph*). Additionally, we use bipartite graph attention convolutions Wang et al. (2020); Nassar (2018) for message passing from drugs to target genes (*drug-gene graph*), as well as gene embeddings to the disease domain (*disease-gene graph*). The mathematical formulation of our proposed framework is described as follows.

Bipartite Graphs Attention Convolution. Conventional GCNs assume that all nodes belong to the same category. However, in our scenario there are heterogeneous attributes across various node types, such as genes, diseases, and drug targets. This limitation becomes evident when node attributes span across different domains. We adopt a bipartite graph as a natural representation for modeling inter-domain interactions among distinct node types. In other words, our focus is on drug-disease interactions, rather than intra-disease or intra-drug interactions. We adapt GCNs to operate within a bipartite graph, where node feature aggregation exclusively occurs over inter-domain edges. Specifically, let us denote graphs as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ represents the set of vertices, given by $\{v_i\}_{i=1}^{N}$, and $\mathcal{E}$ is the set of edges. We consider a bipartite graph $\mathcal{BG}(\mathcal{U}, \mathcal{V}, \mathcal{E})$ defined as a graph $\mathcal{G}(\mathcal{U} \cup \mathcal{V}, \mathcal{E})$, where $\mathcal{U}$ and $\mathcal{V}$ represent two sets of vertices (nodes) corresponding to the two respective domains. Let u_i and v_j denote the i-th and j-th node in $\mathcal{U}$ and $\mathcal{V}$, respectively, where $i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, N$. All edges within the bipartite graph exclusively connect nodes from $\mathcal{U}$ and $\mathcal{V}$ (i.e., $\mathcal{E} = \{(u, v) | u \in \mathcal{U}, v \in \mathcal{V}\}$). The features of the two sets of nodes are denoted by X_u and X_v, where $X_u \in \mathbb{R}^{M \times P}$ is a feature matrix with $\vec{x}_{u_i} \in \mathbb{R}^P$ representing the feature vector of node u_i, and $X_v \in \mathbb{R}^{N \times Q}$ is defined similarly. For the message passing $\mathrm{MP}_{v \to u}$ from domain $\mathcal{V}$ to $\mathcal{U}$, we define a general bipartite graph convolution (bg) as:

$$bg_{\mathcal{E}}(u_i) = \rho\left(\mathrm{agg}\left(\{W_{u_i,v_j}\vec{x}_{v_j} | v_j \in \mathcal{N}_{u_i}^{\mathcal{E}}\}\right)\right), \tag{1}$$

where $\mathcal{N}_{u_i}^{\mathcal{E}}$ represents the neighborhood of node u_i connected by $\mathcal{E}$ in $\mathcal{BG}(\mathcal{U}, \mathcal{V}, \mathcal{E})$ ($\mathcal{N}_{u_i}^{\mathcal{E}} \subset \mathcal{V}$), $W_{u_i,v_j} \in \mathbb{R}^{M \times N}$ is a feature weighting kernel transforming N-dimensional features to M-dimensional features, agg is a permutation-invariant aggregation operation, and the ρ operator can be a non-linear activation function. In our work, we used element-wise mean-pooling and ReLU Nair & Hinton (2010) for agg and ρ respectively.

Our bipartite graph convolution layers utilize the graph attention network Veličković et al. (2017) as the backbone on the node features, resulting in the bipartite graph attention convolution layer (bga). Since the attention mechanism considers features of two sets of nodes, we specifically define a learnable matrix $W^u \in \mathbb{R}^{P \times S}$ (resp. $W^v \in \mathbb{R}^{Q \times S}$) for X_u (resp. X_v), where P, Q, and S are nodes

of the message passing graphs in the respective domains. The bga can be formulated as:

$$bga_{\mathcal{E}}(u_i) = \text{ReLU}\left(\sum_{v_j \in \mathcal{N}^{\mathcal{E}}_{u_i}} \alpha_{u_i,v_j} W^v \vec{x}_{v_j}\right), \tag{2}$$

The attention mechanism is a single-layer feedforward neural network, parameterized by a weight vector $\vec{a}$ and applying the LeakyReLU non-linearity function. The attention weight coefficients can be expressed as:

$$\alpha_{u_i,v_j} = \frac{\exp\left(\rho(\vec{a}^T[W^u\vec{x}_{u_i} \| W^v\vec{x}_{v_j}])\right)}{\sum_{v_k \in \mathcal{N}^{\mathcal{E}}_{u_i}} \exp\left(\rho(\vec{a}^T[W^u\vec{x}_{u_i} \| W^v\vec{x}_{v_k}])\right)}, \tag{3}$$

where T and $||$ represent the matrix transposition and concatenation operations respectively.

Information Fusion We represent the heterogeneous network as an undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with the following three sets of nodes: drugs ($\mathcal{V}^A$), diseases ($\mathcal{V}^B$), and genes ($\mathcal{V}^C$). The input features of these three sets of nodes are denoted as $X_{\mathcal{V}^A}$, $X_{\mathcal{V}^B}$, and $X_{\mathcal{V}^C}$, respectively. The edges consist of two inter-domain sets representing drug-gene associations ($\mathcal{E}^{AC}$) and disease-gene associations ($\mathcal{E}^{BC}$), as well as an intra-domain set representing gene network connections ($\mathcal{E}^{CC}$).

First, we apply multi-layer GCNs to the RNA domain to model the gene interactions. As the *gene-gene graphs* are extremely large for a graph-level prediction tasks, we first initialize the GCN model parameters by a pre-trained variational graph auto-encoder (VGAE) (Hu et al., 2019). The VGAE is fine-tuned on the RNA-seq data from all the PDX models in this study, where each PDX model was represented as a graph with genes as the nodes of the graph. The tumor-specific graphs were obtained from TissueNexus (Lin et al., 2022). We used the VGAE loss function introduced in (Kipf & Welling, 2016).

In the second step, we apply a single bipartite graph attention convolution layer to propagate the message from drugs to target genes. Conceptually, this step can be viewed as projecting information from the macro level (e.g., from the domain of drugs) to the micro level (e.g., the domain of genes). To formulate the message-passing step, we represent the hidden embeddings of node v_i^c as $h^{(k)}_{v_i^c}$, where k is the step index and when $k = 0$, $h^{(0)}_{v_i^c} = x_{v_i^c}$. The term $h^{(k)}_{v_i^c}$ is computed as follows:

$$\text{MP}^{(k)}_{\mathcal{V}^A \to \mathcal{V}^C} : h^{(k)}_{v_i^c} = bga_{\mathcal{E}^{AC}}(v_i^c) + h^{(k-1)}_{v_i^c}. \tag{4}$$

Similarly, in the last step, we utilize the non-linear graph information captured by gene nodes to update the hidden embeddings of the disease nodes. In particular, we apply another bipartite graph attention convolution layer to project gene embeddings to the disease domain. Therefore, the third step can be viewed as an attentional pooling Lee et al. (2019) of the disease gene subgraph. The updated feature representations of disease nodes are computed as follows:

$$\text{MP}^{(k)}_{\mathcal{V}^C \to \mathcal{V}^B} : h^{(k)}_{v_i^b} = bga_{\mathcal{E}^{CB}}(v_i^b) + h^{(k-1)}_{v_i^b}. \tag{5}$$

We concatenate the updated drug and disease embeddings into a unified representation for PDX experiments (tumor-treatment combinations) as β_1, used as input features for downstream tasks.

2.2 Tumor Volume Prediction

We formulate a tumor volume dynamics model using a NODE which utilizes both the embedding generated from the heterogeneous graph encoder and a tumor volume encoder which takes early tumor volume data. The model aims to utilize the early tumor volume data together with baseline embedding of the baseline state (encapsulating drug, disease and RNA-seq data) in order to make personalized predictions.

Tumor Volume Encoder In a similar manner to Laurie & Lu (2023), we implement a tumor volume encoder to inform the NODE in order to make predictions tailored to a specific PDX model. This recurrent neural network (RNN) based encoder maps a short window of early observed tumor volumes (of an arbitrary length) into an embedding that we denote as β_2.

Neural-ODE. In the clinical context, an approach to model tumor dynamics using NODE has been developed, demonstrating the ability of such a formalism to capture the right dynamical model from longitudinal patient tumor size data across treatment arms (Laurie & Lu, 2023). In this work, we generalize the methodology to integrate longitudinal measurements with the multimodal data in the form of an end-to-end framework. The proposed framework is demonstrated in the setting of modeling the PDX data dynamics. We consider a dynamical system of the following form:

$$\frac{dy(t)}{dt} = f_\theta(y(t), \beta), t \in [0, T], \tag{6}$$

where 0 and T denote the start time of PDX experimentation and the end of the prediction time respectively, f_θ is a neural network parameterized by a set of weights θ to be learned across all PDX data, and $\beta = [\beta_1 || \beta_2]$ as the PDX embedding obtained by concatenating the outputs of the heterogeneous graph encoder and tumor volume encoder. Thus, a dynamical law represented by f_θ is learned across all PDX data, with the concatenated embedding β serving to provide the initial condition for the NODE specific to the PDX instance of interest.

After simulating Equation 6 to obtain the time evolution of state $y(t)$, we then reconstruct the tumor volume data using a two-layer multi-layer perceptron (MLP). The NODE approach offers the benefits of accommodating variable-length sequences and varying observation intervals. It also allows us to incorporate positional encoding for capturing the temporal context. To enhance model simplicity and generalization, we keep the dimension of the ODE system (y dimension) no larger than the tumor growth inhibition (TGI) model proposed by (Zwep et al., 2021). The TGI model captured the longitudinal tumor volume measurements, per PDX, with two empirical ODEs governed by three estimated parameters: growth rate, treatment efficacy, and time-dependent resistance development.

3 Results

To assess the effectiveness of the heterogeneous graph encoder in describing changes in tumor volume, we trained the encoder using mRECIST Therasse et al. (2000); Gao et al. (2015) response labels. Specifically, we consolidated the response categories (CR, PR, and SD) into a single response category, and PD as the second category (Table A.1 in the Appendix). To predict these response categories, we used a binary classifier using a three-layer MLP that takes the heterogeneous graph encoder embedding as the input. The dataset was split by PDX models, and standard 5-fold cross-validation was performed. We evaluated the prediction performance using three metrics: balanced accuracy, area under the receiver operating characteristic curve (AUROC), and F1-score. Our model's performance was compared against several competing approaches, including a non-graph-based deep learning method and traditional ML methods. To investigate the impact of the pretraining strategy, we also implemented a variant of our model with a randomly initialized gene graph. More details are provided in subsection A.4.1.

3.1 Reproducing Tumor Volume Data

We evaluated the capability of our NODE in capturing the longitudinal tumor volume data, in comparison to the state-of-the-art TGI model for PDX proposed by (Zwep et al., 2021). For this experiment, we utilized all available tumor volume data for each PDX experiment. Our approach involved encoding the longitudinal tumor volume data into a latent space using tumor volume encoder. Subsequently, we used this latent space as a part of the initial condition to solve an ODE system using the NODE model to reconstruct the dynamic tumor volume data. The population-level results are shown in Panel (A) of Figure 2. Notably, our NODE outperformed the TGI model with R2 of 0.96 compared to 0.71 and Spearman correlation of 0.96 compared to 0.86.

3.2 Tumor Volume Dynamics Prediction

We evaluated the ability of our model (presented in Figure 1) to predict future tumor volume dynamics based on a limited observed longitudinal tumor data. We selected the observation windows of 7, 14, 21, and 28 days, to simulate real-world scenarios where early observations are used to forecast the (future unseen) tumor volume dynamics.

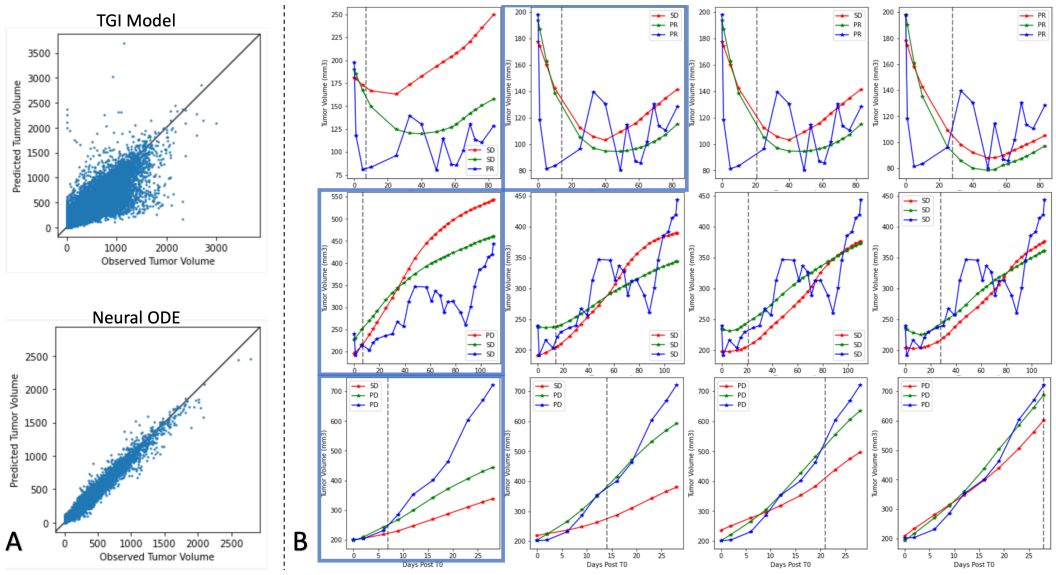

Figure 2: **Tumor dynamics prediction: (A)** A comparison of tumor volume fitting between tumor growth inhibition (TGI) model proposed by Zwep et al. (2021) and our proposed NODE model. **(B)** In each of the 3 rows: tumor volume prediction using the proposed model for an individual PDX instance is shown. The blue curves represent the measurements (ground-truth), the green curves represent our proposed model predictions, and the red curves represent the TGI model predictions. Each of the column corresponds to a single choice of observation window (of increasing size from left to right): the highlighted blocks in dark blue blue indicate the observation windows beyond which our proposed model correctly captures the classification of the mRECIST response category as the ground-truth.

We assessed the predictive performance of our model in two ways. Firstly, we employed R2 to quantify the accuracy of our model in predicting unseen tumor volumes. The results in Table A.3 indicate the following: (1) the embedding learned from the heterogeneous graph encoder enhances the predictive performance of our proposed model; and (2) as the observation window size increases, our proposed model captures the unseen tumor dynamic more accurately. Additionally, as it is demonstrated in Panel (B) of Figure 2, the model effectively captures the tumor dynamic trend.

Due to noise inherent in the tumor volume measurements and the clinical significance of mRECIST response category prediction, we also assessed our proposed model's predictive performance as a classifier. The mRECIST categories are derived from the predicted tumor volume time series by applying the response criteria. This evaluation measures the model's performance in correctly classifying the treatment responses based on the predicted tumor volume dynamics. Figure A.1 summarizes the classification results, revealing an observable trend in which incorporating the heterogeneous graph encoder embedding improves the prediction of response categories across all observation windows.

4 Conclusion

In summary, we proposed a novel approach for tumor dynamics prediction that integrates data from RNA-seq, treatment, disease as well as the longitudinal tumor volume measurements into a NODE system. In a pre-clinical PDX setting, we demonstrated that the use of NODE vastly improves the ability of the model to capture PDX tumor data than the previously proposed TGI model, when used jointly with a graph encoder that enriches the longitudinal data. As future work, elucidating how the model predictions arise from the multimodal data using explainability techniques and/or attention weights would advance our scientific understanding of the complex interplay between gene expression profiles, tumor location, and drug targets. This methodology holds significant promise and warrants further validations, including using cancer organoids Sachs et al. (2018) and in the clinical setting.

REFERENCES

Annette T Byrne, Denis G Alférez, Frédéric Amant, Daniela Annibali, Joaquín Arribas, Andrew V Biankin, Alejandra Bruna, Eva Budinská, Carlos Caldas, David K Chang, et al. Interrogating open issues in cancer precision medicine with patient-derived xenografts. *Nature Reviews Cancer*, 17 (4):254–268, 2017.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

Kelsy C Cotto, Alex H Wagner, Yang-Yang Feng, Susanna Kiwala, Adam C Coffman, Gregory Spies, Alex Wollam, Nicholas C Spies, Obi L Griffith, and Malachi Griffith. Dgidb 3.0: a redesign and expansion of the drug–gene interaction database. *Nucleic acids research*, 46(D1):D1068–D1073, 2018.

William F Forrest, Bruno Alicke, Oleg Mayba, Magdalena Osinska, Michal Jakubczak, Pawel Piatkowski, Lech Choniawko, Alice Starr, and Stephen E Gould. Generalized additive mixed modeling of longitudinal tumor growth reduces bias and improves decision making in translational oncology. *Cancer Research*, 80(22):5089–5097, 2020.

Hui Gao, Joshua M Korn, Stéphane Ferretti, John E Monahan, Youzhen Wang, Mallika Singh, Chao Zhang, Christian Schnell, Guizhi Yang, Yun Zhang, et al. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nature medicine*, 21(11): 1318–1325, 2015.

Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. Strategies for pre-training graph neural networks. *arXiv preprint arXiv:1905.12265*, 2019.

Thomas N Kipf and Max Welling. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*, 2016.

Mark Laurie and James Lu. Explainable deep learning for tumor dynamic modeling and overall survival prediction using Neural-ODE. *npj Systems Biology and Applications*, 9(1):58, 2023.

Junhyun Lee, Inyeop Lee, and Jaewoo Kang. Self-attention graph pooling. In *International conference on machine learning*, pp. 3734–3743. PMLR, 2019.

Cui-Xiang Lin, Hong-Dong Li, Chao Deng, Yuanfang Guan, and Jianxin Wang. Tissuenexus: a database of human tissue functional gene networks built with a large compendium of curated rna-seq data. *Nucleic acids research*, 50(D1):D710–D718, 2022.

Jianzhu Ma, Samson H Fong, Yunan Luo, Christopher J Bakkenist, John Paul Shen, Soufiane Mourragui, Lodewyk FA Wessels, Marc Hafner, Roded Sharan, Jian Peng, et al. Few-shot learning creates predictive models of drug response that translate from high-throughput screens to individual patients. *Nature Cancer*, 2(2):233–244, 2021.

Kunal Naik, Rahul K Goyal, Luca Foschini, Choi Wai Chak, Christian Thielscher, Hao Zhu, James Lu, Joseph Lehár, Michael A Pacanoswki, Nadia Terranova, et al. Current status and future directions: the application of artificial intelligence/machine learning (AI/ML) for precision medicine. *Clinical Pharmacology & Therapeutics*, 2023.

Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814, 2010.

Marcel Nassar. Hierarchical bipartite graph convolution networks. *arXiv preprint arXiv:1812.03813*, 2018.

Wei Peng, Hancheng Liu, Wei Dai, Ning Yu, and Jianxin Wang. Predicting cancer drug response using parallel heterogeneous graph convolutional networks with neighborhood interactions. *Bioinformatics*, 38(19):4546–4553, 2022.

Janet Piñero, Àlex Bravo, Núria Queralt-Rosinach, Alba Gutiérrez-Sacristán, Jordi Deu-Pons, Emilio Centeno, Javier García-García, Ferran Sanz, and Laura I Furlong. Disgenet: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic acids research*, pp. gkw943, 2016.

Norman Sachs, Joep De Ligt, Oded Kopper, Ewa Gogola, Gergana Bounova, Fleur Weeber, Anjali Vanita Balgobind, Karin Wind, Ana Gracanin, Harry Begthel, et al. A living biobank of breast cancer organoids captures disease heterogeneity. *Cell*, 172(1):373–386, 2018.

Patrick Therasse, Susan G Arbuck, Elizabeth A Eisenhauer, Jantien Wanders, Richard S Kaplan, Larry Rubinstein, Jaap Verweij, Martine Van Glabbeke, Allan T van Oosterom, Michaele C Christian, et al. New guidelines to evaluate the response to treatment in solid tumors. *Journal of the National Cancer Institute*, 92(3):205–216, 2000.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Zichen Wang, Mu Zhou, and Corey Arnold. Toward heterogeneous information fusion: bipartite graph convolutional networks for in silico drug repurposing. *Bioinformatics*, 36(Supplement_1): i525–i533, 2020.

Max Welling and Thomas N Kipf. Semi-supervised classification with graph convolutional networks. In *J. International Conference on Learning Representations (ICLR 2017)*, 2016.

Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V Chawla. Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 793–803, 2019.

Laura B Zwep, Kevin LW Duisters, Martijn Jansen, Tingjie Guo, Jacqueline J Meulman, Parth J Upadhyay, and JG Coen van Hasselt. Identification of high-dimensional omics-derived predictors for tumor growth dynamics using machine learning and pharmacometric modeling. *CPT: pharmacometrics & systems pharmacology*, 10(4):350–361, 2021.

Appendix

A.1 PATIENT-DERIVED XENOGRAPH (PDX) DATASET

The primary dataset utilized in this study was obtained from a large-scale pre-clinical study conducted in PDX mice models, as detailed in Gao et al. (2015). This dataset encompassed more than 1000 PDX models, each characterized by their baseline mRNA expression levels prior to treatment. In total, the dataset covered 62 distinct treatments across six different diseases, with tumor volume measurements taken every 2-3 days. For our analysis, based on the availability of RNA-seq data we included data from 191 unique tumors and 59 different treatments, resulting in a comprehensive dataset of 3470 PDX experiments (consisting of various tumor and treatment combinations) spanning 5 tumor types.

A.2 MODIFIED RECIST CATEGORIES

The time-dependent tumor response is determined by comparing tumor volume change at time t in relation to its baseline (i.e., initial) value:

$$\Delta V(t) = 100\% \times \frac{V(t) - V_{initial}}{V_{initial}}. \tag{A.1}$$

The best response is defined as the minimum value of $\Delta V(t)$ for $t \geq 10$ days.

Table A.1: **mRECIST categories:** The modified Response Evaluation Criteria in Solid Tumors (mRECIST) Therasse et al. (2000), is a tumor progression indication being calculated in less than 64 days that "captures a combination of speed, strength and durability of response into a single value" Gao et al. (2015). The best response is computed as the percentage of changes in the tumor volume using the Equation A.1. The distribution of the response categories for the dataset used in this study is provided in the third column. For further details please refer to Gao et al. (2015)

mRECIST category	Description	Distribution (%)
Complete Response (CR)	best response $\leq$ -95%	2.74%
Partial Response (PR)	-95% $<$ best response $\leq$ -50%	7.12%
Stable Disease (SD)	-50% $<$ best response $\leq$ 35%	30.29%
Progressive Disease (PD)	otherwise	59.85%

A.3 COMPARING GCN WITH NON-GCN MODELS.

As the RNA-seq data is in a tabular format, therefore we use common models such as multilayer perceptron (MLP) and random forests (RF) for training and prediction on regression or classification tasks. Hence, we used MLP and RF for comparison with a complex model such as GCNs, that requires representation of the RNA-seq data as graph. Also we compared evaluated the GCN performance with and without pretraining. The reuslts are summarized in Table A.2.

Table A.2: The summary of model performance on treatment response prediction using RNA-seq data. The mean and standard deviation over 5-fold cross-validation is reported.

Method	Balanced Accuracy	AUROC	F1
Pretrained GCN	**%68.1 $\pm$ 2.2**	**%74.1 $\pm$ 2.3**	**%62.3 $\pm$ 1.8**
GCN	%65.2 $\pm$ 1.8	%71.3 $\pm$ 1.9	%58.2 $\pm$ 1.8
MLP	%62.1 $\pm$ 2.3	%68.8 $\pm$ 2.6	%52.5 $\pm$ 2.1
RF	%61.8 $\pm$ 1.3	%67.6 $\pm$ 1.4	%53.2 $\pm$ 1.3

Table A.3: Predictive performance of our proposed model using different observation windows quantified with R2. The mean and standard deviation over 5-fold cross-validation is reported.

Observation window	w/o graph encoder	w graph encoder
7 days	%23.3 ± 5.2	**%30.2 ± 4.9**
14 days	%45.6 ± 4.8	**%47.9 ± 4.7**
21 days	%58.6 ± 4.1	**%60.8 ± 4.2**
28 days	%65.2 ± 3.9	**%65.9 ± 3.8**

A.4 RESULTS

A.4.1 GCNS OUTPERFORMS BASELINE METHODS IN MODELING RNA-SEQ FOR TREATMENT RESPONSE PREDICTION.

As shown in Table A.2, GCN-based methods demonstrated superior performance in the responder classification task across all evaluation metrics. These models outperformed non-graph-based approaches by more than 9% in F1 score and 5% in AUROC. Additionally, we observed that our model achieved a noticeable improvement of approximately 4% in F1 score and 3% in AUROC when utilizing gene graph pre-training. This result highlights the effectiveness of graph reconstruction pre-training in enhancing gene latent representations. More details are provided in Appendix A.3.

A.4.2 MRECIST CATEGORIES PREDICTION WITH AND WITHOUT HETEROGENEOUS GRAPH ENCODER

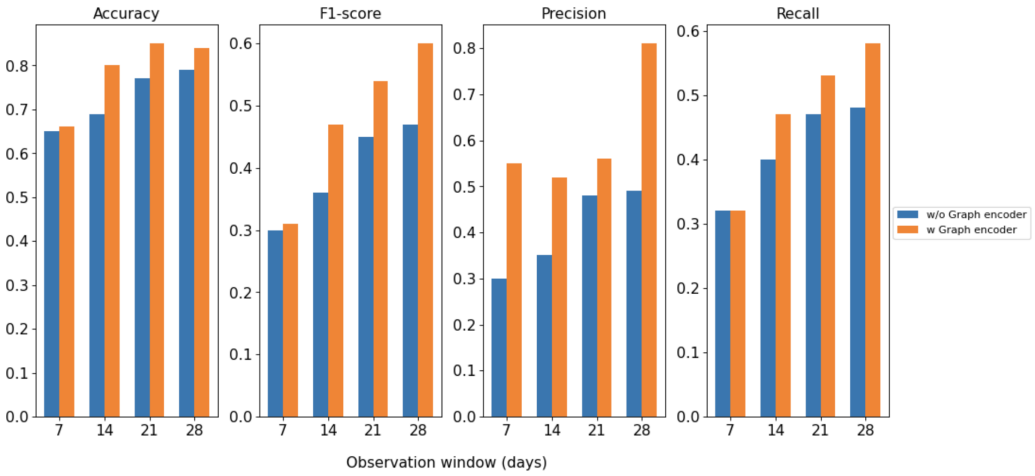

Figure A.1: Predictive performance of our proposed model as a classifier for mRECIST categories, with and without the heterogeneous graph encoder and considering different lengths of observation windows.

A.5 TRAINING PROCEDURES

A.5.1 PRETRAINED GCN

We used VGAE with two GCN layers to pretrain the *gene-gene graph* using RNA-seq data and knowledge graph. We used 3 different losses for the pretraining procedure; node reconstruction, edge reconstruction, and KL-divergence. The node reconstruction loss is the mean squared error (MSE) of genes (nodes of the input graph) and reconstructed genes (decoder output). The edge reconstruction loss is the cross entropy of positive and negative edges in the graph. The KL-divergence is used as a regularization term in the loss function to assure continuity and completeness in the latent space.

A.5.2 TREATMENT RESPONSE PREDICTION USING RNA-SEQ DATASET

In this experiment we defined the problem as a classification task and used cross entropy loss function for the pretrained GCN, GCN, and MLP. For the RF model we used *Gini* criteria as a measure of impurity.

A.5.3 TREATMENT RESPONSE PREDICTION USING HETEROGENEOUS GRAPH

The treatment response prediction experiment using the heterogeneous graph neural networks was conducted as a regression task to predict the entire tumor volume for each PDX model. Then the predicted tumor volumes were converted into response categories using the Table A.2. This end-to-end model takes all 3 modalities including RNA-seq, drug, and treatment data to the heterogeneous graph encoder, and the observed tumor volume data was fed into the tumor volume encoder. Then the predicted tumor volume from the Neural-ODE is used to be compared with the ground truth tumor volume data of the entire tumor volume using MSE loss function.