TRANSFERRING PRECLINICAL DRUG RESPONSE TO PATIENT VIA TUMOR HETEROGENEITY-AWARE ALIGNMENT AND PERTURBATION MODELING

Inyoung Sung[1] **Dongmin Bang**[1,2] **Sun Kim**[1,2,] **Sangseon Lee**[3]
[1]Seoul National University [2] AIGENDRUG Co., Ltd. [3]Inha University
{inyoung.sung,eugenomics,sunkim.bioinfo}@snu.ac.kr,
{ss.lee}@inha.ac.kr

ABSTRACT

Accurate prediction of personalized drug response is critical for precision oncology, yet limited clinical data forces reliance on preclinical datasets. However, fundamental biological differences between preclinical cell lines and patient tumors hinder direct knowledge transfer. In this work, we introduce THERAPI, a novel tumor heterogeneity-aware Domain Adaptation (DA) framework that represents patient tumors as weighted combinations of multiple cell lines with tissue-specific context. Along with our comprehensive gene expression modeling by integrating drug-induced perturbation-based and rank-based representations, THERAPI outperforms both DA-free and DA-based models and generalizes robustly to an external dataset, highlighting its potential for applications in precision medicine.

1 INTRODUCTION

Predicting personalized drug response is essential for precision medicine, as it enables tailored treatment strategies based on individual patient profiles (Feng et al., 2021). However, the limited availability of large-scale clinical datasets like The Cancer Genome Atlas (TCGA, Weinstein et al. (2013)) forces many studies to rely on preclinical data such as the Genomics of Drug Sensitivity in Cancer (GDSC, Iorio et al. (2016)) to predict cancer cell line responses (Liu et al., 2020; Bang et al., 2024). Yet, fundamental differences between cancer cell lines and patient tumors, arising from variations in the tumor microenvironment and heterogeneity, hinder direct knowledge transfer, often resulting in models that fail to generalize to real-world patient samples (Lee et al., 2018; Shen et al., 2023). These discrepancies lead to one of the core challenges in machine learning, known as domain shift, which limits the direct applicability of models trained on preclinical data to diverse patient populations. Thus, the main challenge in this task is translating knowledge from GDSC to TCGA—specifically, how to transfer transcriptomic information from GDSC cell lines to the transcriptomic profiles of patient tumors in TCGA. To address this challenge, domain adaptation (DA)-based techniques have been explored to align distributions between preclinical models and patient tumors (He et al., 2022; Kim et al., 2024).

Despite advances in DA-based techniques, two major limitations remain during transfer. First, current methods overlook tumor heterogeneity. Unlike cell lines, patient tumors exhibit distinct heterogeneity and microenvironment differences (Mourragui et al., 2019). However, most DA-based approaches align cell line data and patient data without accounting for this heterogeneity in cancer composition along with tissue-specific context. Second, existing methods fail to model and consider gene interactions, an essential element of drug response mechanism, as they operate at a low-dimensional representation level (He et al., 2022; Kim et al., 2024). Previous studies on transcriptomic modeling have shown that examining interrelationships between gene expression levels (Theodoris et al., 2023) or gene interactions from drug-induced gene expression perturbation (Bang et al., 2024) is critical for drug response prediction tasks. However, existing DA-based methods overlook this aspect and rely on simple MLP encoders for drug response prediction in a low-dimensional representation space.

To address these challenges, we propose THERAPI: Tumor Heterogeneity-aware Embedding for Response Adaptation and Patient Inference, bridging the gap between cell line transcriptomic data and patient tumor profiles. Unlike existing methods that treat patient tumors as analogous to individual cell lines, THERAPI models tumor heterogeneity by representing each patient tumor as a linear combination of multiple preclinical cancer models, further enriched with tissue-type annotations. Furthermore, to enhance drug response prediction, THERAPI integrates both drug-induced gene-gene interaction perturbations and rank-based gene expression representations, allowing for a more comprehensive modeling of gene expression without the constraints of low-dimensional latent representations. By integrating tumor heterogeneity and gene expression modeling, THERAPI outperforms both DA-free and DA-based models in translating preclinical drug response data from GDSC to predict TCGA patient drug response. In addition, performance evaluation on an external dataset demonstrated the generalizability of our framework, highlighting its potential for applications in precision medicine.

2 THERAPI

2.1 THERAPI ARCHITECTURE

Our objective is to predict patient drug response by modeling the biological characteristics of tumors and gene-gene interactions. As shown in Figure 1, our model THERAPI consists of two steps: 1) Alignment step brings the preclinical and clinical transcriptome space into a unified space through attention-based aggregation (Section 2.2) and 2) Drug response prediction step integrates perturbation-and rank-based modules for modeling gene-level knowledge (Section 2.3).

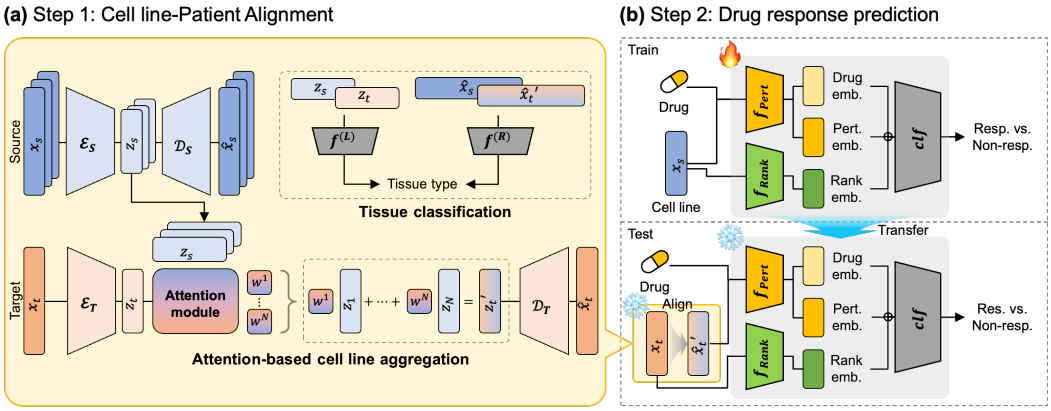

Figure 1: Overview of THERAPI framework. (a) The alignment step of preclinical (GDSC) and clinical (TCGA) transcriptomes into a unified embedding space. Through attention-based aggregation of cell lines, the alignment module represents patient samples as weighted sum of cell lines. (b) The drug response prediction step consists of a drug-induced perturbation module and a rank-based expression representation module. After the alignment module is trained in alignment step, the patient sample is encoded as weighted sum of cell lines, which is passed down to each module to yield a final drug responder label.

2.2 ALIGNMENT STEP

We represent the target domain ($\mathbf{X}_T$) consisting of patient tumors by modeling tumor heterogeneity using the source domain ($\mathbf{X}_S$) consisting of preclinical cancer models (i.e., cell lines). During the alignment, THERAPI utilizes unlabeled data from both the source and target domains to train an autoencoder that minimizes data reconstruction error. As shown in Figure 1a, unlike autoencoder-based model (He et al., 2022), THERAPI uses two unique features: Attention-based cell line aggregation and Tissue label classifiers. The aggregation module is trained to align both preclinical and clinical transcriptome profiles through accurate reconstruction. Tissue classifiers further inject the tissue information via tissue label classification and center loss (Wen et al., 2016). For detailed description of the alignment step, refer to Appendix A.1

Attention-based cell line aggregation The core idea of the aggregation module is to model tumor heterogeneity by leveraging cell line data. We use domain-specific encoders, $\mathcal{E}_S$ and $\mathcal{E}_T$, to transform cell line and patient tumor expression profiles into latent representations z_s and z_t. An attention mechanism then computes weights

$$w^i_{(t,s)} = \frac{\exp\left(\langle Qz_t,\, Kz^i_s\rangle\right)}{\sum_{j=1}^{N_s} \exp\left(\langle Qz_t,\, Kz^j_s\rangle\right)},$$

using learnable projection matrices Q and K and dot product $\langle\cdot,\cdot\rangle$, to aggregate source representations as $z'_t = \sum_{i=1}^{N_s} w^i_{(t,s)} z^i_s$. This aggregated representation is decoded with $\mathcal{D}_T$ to reconstruct the tumor sample x_t, while each cell line is reconstructed via $\mathcal{D}_S$. The reconstruction loss is defined as

$$\mathcal{L}_{\text{rec}} = \mathbb{E}_{x_s\sim X_S}\left[\|\hat{x}_s - x_s\|^2\right] + \mathbb{E}_{x_t\sim X_T}\left[\|\hat{x}_t - x_t\|^2\right]$$

where $\hat{x} = \mathcal{D}(z)$, ensuring that both domains retain essential gene expression features.

Tissue label classifiers To take tissue labels into the learning process during domain adaptation, we apply MLP-based classifiers in both the latent and reconstructed spaces. In the source domain, classifiers $f_S^{(L)}$ and $f_S^{(R)}$ operate on z_s and its reconstruction $\hat{x}_s$, respectively; similarly, for the target domain, tissue classification is applied to the attention-weighted aggregated representation z'_t and $\sum w^i_{(t,s)} x^i_s$, respectively. Thus, the overall tissue classifier loss is given by $\mathcal{L}_{\text{class}} = \mathcal{L}^{(L)} + \mathcal{L}^{(R)}$, where each term is computed via cross-entropy loss. Additionally, we used the center loss $\mathcal{L}_{\text{center}}$ to construct a compact space according to the tissue labels in the latent embedding space.

The overall loss in the alignment step $\mathcal{L}_{\text{align}}$ is formulated as:

$$\mathcal{L}_{\text{align}} = \alpha\,\mathcal{L}_{\text{rec}} + \beta\,\mathcal{L}_{\text{class}} + \gamma\,\mathcal{L}_{\text{center}},$$

with hyperparameters α, β, and γ balancing each term.

2.3 Drug response prediction step

In the drug response prediction step, THERAPI is first trained using drug-perturbed and rank-based representations derived from pre-trained models in the source domain (Figure 1b-Train). Then, this model is applied to the target domain to predict drug response in patient samples (Figure 1b-Test).

During the training of drug response predictor on the source domain, the input representation is constructed using three pre-trained models: Drug perturbation representation using CSG^{2}A (Bang et al., 2024), rank-based gene-interaction using Geneformer (Theodoris et al., 2023), and molecular encoder using MAT (Maziarka et al., 2020). The detailed explanation of each model is described in Appendix A.2. The final drug response predictor is trained by concatenating these three representations and optimizing a supervised loss function, binary cross-entropy loss, based on the binary drug response label.

After training in the source domain, the drug response predictor is applied to the target domain for patient drug response prediction. The main difference in the inference step is that the patient input x_t is first transformed into $x'_t = \sum_{i=1}^{N_s} w^i_{(t,s)} x^i_s$ using the attention-based cell lineage aggregation module during the alignment step, after which drug response is predicted. We note that the drug response-labeled patient data are not used in the alignment step to prevent data leakage.

3 Results

3.1 Experimental Setting

For patient drug response prediction, we used GDSC as the source domain and TCGA as the target domain. The GDSC dataset consists 673 cancer cell lines and 174 unique drugs, forming a total of 112,533 cell line-drug pairs. And the TCGA dataset includes 8,400 patient samples, among which only 358 patients received treatment with 21 unique drugs, resulting in 383 patient-drug pairs. To evaluate drug response prediction, we used AUROC, AUPRC, accuracy, precision, recall, and F1-score, reporting the average performance over 10-fold cross-validation. Details of the dataset, evaluation metrics, and experimental setup are provided in Appendix B.

Table 1: Performance comparison with baseline models on GDSC-TCGA dataset. Alignment column indicates whether models align the preclinical and clinical transcriptome data, and the tissue column indicates whether the model considers the tissues labels during alignment. Mean and standard deviation of 10-fold CV are provided. Best performance and its comparable results (paired t-test p-value < 0.05) are marked in bold, and second-best are underlined.

Model		AUROC	AUPRC	Accuracy	Precision	Recall	F1
DA-free (Cell-level)	DeepCDR	0.669 (0.074)	0.608 (0.058)	0.575 (0.056)	0.542 (0.044)	0.844 (0.113)	0.656 (0.046)
	CSG^{2}A	0.668 (0.053)	0.643 (0.057)	0.561 (0.060)	0.532 (0.042)	**0.857 (0.048)**	0.655 (0.032)
DA-based	CODE-AE	0.668 (0.089)	0.623 (0.067)	0.628 (0.082)	0.643 (0.092)	0.505 (0.185)	0.551 (0.144)
	PANCDR	0.714 (0.029)	0.687 (0.025)	0.638 (0.049)	0.609 (0.052)	0.740 (0.087)	0.663 (0.032)
	THERAPI	**0.775 (0.034)**	**0.710 (0.024)**	**0.716 (0.039)**	**0.713 (0.051)**	0.704 (0.105)	**0.703 (0.051)**

Table 2: Ablation study results on three modules of THERAPI, on GDSC-TCGA dataset on 10-fold CV setting. Mean and standard deviation of 10-fold CV are provided. Best performance and its comparable results (paired

Model	AUROC	AUPRC	Accuracy	Precision	Recall	F1
THERAPI	**0.775 (0.034)**	**0.710 (0.024)**	**0.716 (0.039)**	**0.713 (0.051)**	0.704 (0.105)	**0.703 (0.051)**
w/o Alignment	0.621 (0.073)	0.579 (0.061)	0.580 (0.067)	0.579 (0.085)	0.504 (0.140)	0.529 (0.101)
w/o Rank rep.	0.615 (0.162)	0.606 (0.111)	0.596 (0.133)	0.603 (0.171)	0.624 (0.227)	0.590 (0.148
w/o Perturbed rep.	0.480 (0.037)	0.574 (0.026)	0.561 (0.023)	0.587 (0.011)	**0.845 (0.074)**	0.692 (0.029)

3.2 Benchmark Dataset Performances

We evaluated THERAPI by translating preclinical drug response data from GDSC to the patient dataset TCGA. Our benchmark included two DA-free models (DeepCDR Liu et al. (2020) and CSG^{2}A Bang et al. (2024)) and two DA-based models (CODE-AE He et al. (2022) and PANCDR Kim et al. (2024)). As a result, THERAPI achieved the best performance across most metrics (Table 1). As expected, DA-based models outperform DA-free ones, underscoring the necessity of DA techniques for effective transfer learning across datasets with distinct distributions. Among the DA-based models, our approach demonstrates superior performance, which we attribute to its biologically informed construction of the shared embedding space. As shown in Appendix D.1, THERAPI not only aligns cell lines and patient samples but also preserves biological context by forming clusters consistent with tissue origin—an ability not observed in other DA-based models.

3.3 Ablation Studies

To investigate the contribution of each module to THERAPI's performance, we conducted an ablation study by removing the alignment module, the rank-based representation, and the perturbation-based representation. As shown in Table 2, the full THERAPI model achieved the highest performance. The attention-based alignment module significantly improved performance in terms of Recall and F1, demonstrating that attention-based aggregation effectively mitigates domain differences. Both the perturbation-based and rank-based modules contributed to performance gains. Notably, the gene-level perturbation representation yielded the largest improvement in terms of AUROC, suggesting that capturing drug-induced gene perturbations is critical for transferring drug response prediction. Overall, these results validate the design of THERAPI and highlight the complementary benefits of its constituent modules.

3.4 External Dataset Performances

To assess the generalizability of THERAPI, we evaluated its performance on the external I-SPY 2 dataset (Wolf et al., 2022). I-SPY 2 includes 988 transcriptome profiles from breast cancer patients, 178 of whom received paclitaxel. We aligned 810 untreated patients with the GDSC dataset and applied the GDSC-trained drug response model to I-SPY 2. As a result, THERAPI outperformed four baseline models in predicting drug response (Figure 2). Additionally, the alignment step successfully mapped the two distinct domains into a shared space (Appendix D.2).

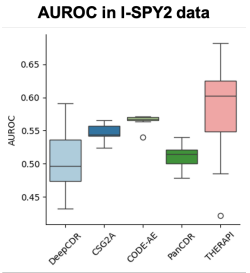

Figure 2: Results on I-SPY2 dataset. Boxplot of AUROC of 10-fold CV.

4 CONCLUSION

In this study, we introduced THERAPI, a tumor heterogeneity-aware domain adaptation framework for translating preclinical cell line data to patient tumors in personalized drug response prediction. Unlike existing methods, THERAPI models tumor heterogeneity by representing each patient tumor as a weighted combination of multiple preclinical cell lines, incorporating tissue-type annotations for greater biological relevance. Additionally, its pre-trained gene expression modeling framework integrates drug-induced perturbations and rank-based representations to enhance prediction accuracy. THERAPI outperformed both DA-free and DA-based models across multiple evaluation metrics and showed generalizability on an external dataset. These results highlight its potential for real-world clinical applications, advancing precision oncology through biologically informed domain adaptation.

REFERENCES

Dongmin Bang, Bonil Koo, and Sun Kim. Transfer learning of condition-specific perturbation in gene interactions improves drug response prediction. *Bioinformatics*, 40(Supplement_1):i130–i139, 2024.

Haotian Cui, Chloe Wang, Hassaan Maan, Kuan Pang, Fengning Luo, Nan Duan, and Bo Wang. scgpt: toward building a foundation model for single-cell multi-omics using generative ai. *Nature Methods*, 21(8):1470–1480, 2024.

Fangyoumin Feng, Bihan Shen, Xiaoqin Mou, Yixue Li, and Hong Li. Large-scale pharmacogenomic studies and drug response prediction for personalized cancer medicine. *Journal of Genetics and Genomics*, 48(7):540–551, 2021.

Di He, Qiao Liu, You Wu, and Lei Xie. A context-aware deconfounding autoencoder for robust prediction of personalized clinical drug response from cell-line compound screening. *Nature Machine Intelligence*, 4(10):879–892, 2022.

Francesco Iorio, Theo A Knijnenburg, Daniel J Vis, Graham R Bignell, Michael P Menden, Michael Schubert, Nanne Aben, Emanuel Gonçalves, Syd Barthorpe, Howard Lightfoot, et al. A landscape of pharmacogenomic interactions in cancer. *Cell*, 166(3):740–754, 2016.

Juyeon Kim, Sung-Hye Park, and Hyunju Lee. Pancdr: precise medicine prediction using an adversarial network for cancer drug response. *Briefings in Bioinformatics*, 25(2):bbae088, 2024.

Jin-Ku Lee, Zhaoqi Liu, Jason K Sa, Sang Shin, Jiguang Wang, Mykola Bordyuh, Hee Jin Cho, Oliver Elliott, Timothy Chu, Seung Won Choi, et al. Pharmacogenomic landscape of patient-derived tumor cells informs precision oncology therapy. *Nature genetics*, 50(10):1399–1411, 2018.

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.

Tianyu Liu, Kexing Li, Yuge Wang, Hongyu Li, and Hongyu Zhao. Evaluating the utilities of foundation models in single-cell data analysis. *bioRxiv*, pp. 2023–09, 2023.

Łukasz Maziarka, Tomasz Danel, Sławomir Mucha, Krzysztof Rataj, Jacek Tabor, and Stanisław Jastrzębski. Molecule attention transformer. *arXiv preprint arXiv:2002.08264*, 2020.

Soufiane Mourragui, Marco Loog, Mark A Van De Wiel, Marcel JT Reinders, and Lodewyk FA Wessels. Precise: a domain adaptation approach to transfer predictors of drug response from pre-clinical models to tumors. *Bioinformatics*, 35(14):i510–i519, 2019.

Bihan Shen, Fangyoumin Feng, Kunshi Li, Ping Lin, Liangxiao Ma, and Hong Li. A systematic assessment of deep learning methods for drug response prediction: from in vitro to clinical applications. *Briefings in Bioinformatics*, 24(1):bbac605, 2023.

Christina V Theodoris, Ling Xiao, Anant Chopra, Mark D Chaffin, Zeina R Al Sayed, Matthew C Hill, Helene Mantineo, Elizabeth M Brydon, Zexian Zeng, X Shirley Liu, et al. Transfer learning enables predictions in network biology. *Nature*, 618(7965):616–624, 2023.

John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120, 2013.

Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. A discriminative feature learning approach for deep face recognition. In *Computer vision–ECCV 2016: 14th European conference, amsterdam, the netherlands, October 11–14, 2016, proceedings, part VII 14*, pp. 499–515. Springer, 2016.

Denise M Wolf, Christina Yau, Julia Wulfkuhle, Lamorna Brown-Swigart, Rosa I Gallagher, Pei Rong Evelyn Lee, Zelos Zhu, Mark J Magbanua, Rosalyn Sayaman, Nicholas O'Grady, et al. Redefining breast cancer subtypes to guide treatment prioritization and maximize response: Predictive biomarkers across 10 cancer therapies. *Cancer Cell*, 40(6):609–623, 2022.

Fan Yang, Wenchuan Wang, Fang Wang, Yuan Fang, Duyu Tang, Junzhou Huang, Hui Lu, and Jianhua Yao. scbert as a large-scale pretrained deep language model for cell type annotation of single-cell rna-seq data. *Nature Machine Intelligence*, 4(10):852–866, 2022.

A DETAILS IN METHODOLOGY OF THERAPI

A.1 ALIGNMENT STEP

A.1.1 ATTENTION-BASED CELL LINE AGGREGATION

A central objective of our approach is to capture tumor heterogeneity by allowing each patient tumor sample to draw information from multiple cell lines. Let $X_S \in \mathbb{R}^{N_s \times d}$ and $X_T \in \mathbb{R}^{N_t \times d}$ denote the source (cell line) and target (tumor) expression datasets, respectively, with N_s and N_t samples and d genes per sample. Each expression profile x has an associated tissue label τ.

We employ two domain-specific encoders, $\mathcal{E}_S$ and $\mathcal{E}_T$, mapping input vectors into latent representations:

$$z_s^i = \mathcal{E}_S(x_s^i), \quad z_t = \mathcal{E}_T(x_t),$$

where $z_s^i \in \mathbb{R}^{d_z}$ and $z_t \in \mathbb{R}^{d_z}$. The attention module uses these latent representations to compute source weights:

$$w_{(t,s)}^i = \frac{\exp\left(\langle Q z_t,\, K z_s^i \rangle\right)}{\sum_{j=1}^{N_s} \exp\left(\langle Q z_t,\, K z_s^j \rangle\right)},$$

where Q and K are learnable projection matrices of size $d_z \times d_k$, and $\langle \cdot, \cdot \rangle$ denotes the dot product. These weights indicate how relevant each cell line is for reconstructing a given tumor sample. The aggregated target representation is then

$$z_t' = \sum_{i=1}^{N_s} w_{(t,s)}^i z_s^i,$$

capturing tumor heterogeneity through a combination of different cell lines. Next, $\mathcal{D}_S$ and $\mathcal{D}_T$ reconstruct the original samples:

$$\hat{x}_s = \mathcal{D}_S(z_s^i), \quad \hat{x}_t = \mathcal{D}_T(z_t').$$

We define the reconstruction loss $\mathcal{L}_{\text{rec}}$ as the mean squared error between the original and reconstructed samples, ensuring that both source and target encoders and decoders learn consistent latent representations:

$$\mathcal{L}_{\text{rec}} = \mathbb{E}_{x_s \sim X_S}\left[\|\hat{x}_s - x_s\|^2\right] + \mathbb{E}_{x_t \sim X_T}\left[\|\hat{x}_t - x_t\|^2\right].$$

By learning appropriate attention weights and minimizing $\mathcal{L}_{\text{rec}}$, the model aligns cell line features with patient tumor profiles in a way that accounts for tissue-specific heterogeneity.

A.1.2 TISSUE LABEL CLASSIFIERS

To further incorporate tissue information in our domain adaptation, we introduce tissue classification at two representation levels: the latent space (z) and the reconstructed expression space (x'). An MLP-based tissue classifier f_S is trained in both domains. In the source domain, we classify

$$f_S^{(L)}(z_s^i) \quad \text{and} \quad f_S^{(R)}(\hat{x}_s),$$

where $\hat{x}_s = \mathcal{D}_S(z_s^i)$. For the target domain, the aggregated latent representation $\sum_i w_{(t,s)}^i z_s^i$ and its reconstruction $\sum_i w_{(t,s)}^i x_s^i$ are used:

$$f_S^{(L)}\Big(\sum_i w_{(t,s)}^i z_s^i\Big) \quad \text{and} \quad f_S^{(R)}\Big(\sum_i w_{(t,s)}^i x_s^i\Big).$$

We define the overall classification loss as the sum of latent and reconstruction classification losses:

$$\mathcal{L}_{\text{class}} = \mathcal{L}^{(L)} + \mathcal{L}^{(R)},$$

where each term can be a standard cross-entropy loss.

To promote tighter clustering of tissue labels within the latent space, we incorporate a center loss $\mathcal{L}_{\text{center}}$, which pulls samples of the same tissue type closer together:

$$\mathcal{L}_{\text{center}} = \sum_{k=1}^{K} \sum_{z \in \mathcal{C}_k} \|z - c_k\|^2,$$

where K is the number of tissue types, $\mathcal{C}_k$ is the set of latent vectors with tissue label k, and c_k is the learnable center for tissue k.

Finally, we combine these objectives into the total alignment loss:

$$\mathcal{L}_{\text{align}} = \alpha\mathcal{L}_{\text{rec}} + \beta\mathcal{L}_{\text{class}} + \gamma\mathcal{L}_{\text{center}},$$

where α, β, and γ are hyperparameters. Balancing these terms ensures the model captures relevant gene expression patterns, tissue-specific distinctions, and tumor heterogeneity while bridging the gap between cell line and patient tumor domains.

A.2 Drug response prediction step

The model is first trained on drug-perturbed and rank-based representations derived from pre-trained models in the source domain. After training, the model is applied to the target domain to predict drug response in patient samples.

A.2.1 Training in the source domain

During training, the drug response predictor is constructed using three pre-trained models, each capturing different aspects of drug-gene interaction:

- Drug perturbation representation using CSG^{2}A (Bang et al., 2024), which models the gene expression changes induced by drug treatment: $z_{\text{pert}} = f_{\text{pert}}(x_s, d)$ where x_s represents the gene expression profile of a source domain sample, and d represents the drug molecular structure.
- Rank-based gene interaction representation using Geneformer (Theodoris et al., 2023), which encodes gene-gene interactions in a rank-based manner:
$$z_{\text{rank}} = f_{\text{rank}}(x_s).$$
- Molecular representation model using molecular attention transformer (MAT, Maziarka et al. (2020)), which encodes the drug molecular structure using a transformer-based embedding:
$$z_{\text{mol}} = f_{\text{mol}}(d).$$

The final representation for drug response prediction is obtained by concatenating these three embeddings:

$$z = [z_{\text{pert}}; z_{\text{rank}}; z_{\text{mol}}]$$

where z is the combined input to the MLP. The drug response predictor is defined as:

$$\hat{y} = f_{\text{MLP}}(z)$$

where $\hat{y}$ represents the predicted probability of drug response. Since the drug response is binary, the model is optimized using the binary cross-entropy (BCE) loss:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log \hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i)\right]$$

where y_i is the true drug response label for sample i, and N is the number of training samples.

A.2.2 Inference in the target domain

After training in the source domain, the drug response predictor is applied to the target domain for patient drug response prediction. The key difference in the inference step is that the patient input x_t is first transformed into an aligned representation x'_t using the attention-based cell lineage aggregation module during the alignment step:

$$x'_t = \sum_{i=1}^{N_s} w^i_{(t,s)} x^i_s$$

where $w^i_{(t,s)}$ represents the learned attention weights mapping the target sample x_t to the source cell lines x^i_s. The aligned representation x'_t is then passed through the trained drug response model to predict the final response $\hat{y}$.

B EXPERIMENTAL SETTING

B.1 DATA

For patient drug response prediction, we mainly utilized the GDSC-TCGA dataset, built using GDSC as the source domain and TCGA as the target domain. Both datasets contain gene expression profiles utilized to 978 LINCS (Library of Integrated Network-based Cellular Signatures) landmark genes and include tissue-type annotations for 23 distinct tissue types. All gene expression data are metricized by the standard transcripts per million bases for each gene, followed by log transformation.

The GDSC dataset consists of 673 cancer cell lines and 174 unique drugs, forming a total of 112,533 cell line-drug pairs. The drug response values in GDSC are represented using IC50 (half-maximal inhibitory concentration) values and we used these values in a binary form. To obtain binary drug response labels, we binarize IC50 values by computing the mean IC50 per drug across all cell lines. Samples with IC50 lower than the median are labeled as responders (1), while those with IC50 higher than the median are labeled as non-responders (0). The GDSC data can be accessed at https://www.cancerrxgene.org/.

The TCGA dataset consists of 8,400 patient samples, including 8,042 unlabeled and 358 labeled patients. The unlabeled patients are used only in the alignment step to adapt the source and target domains. In contrast, the 358 labeled patients, who were treated with 21 unique drugs, are used only in the test step for drug response prediction, forming a total of 383 patient-drug pairs. The TCGA data can be accessed at https://portal.gdc.cancer.gov/.

B.2 PERFORMANCE EVALUATION

We perform 10-fold cross-validation with a train:validation:test split of 8:1:1. The model performance is evaluated using six metrics:

Area Under the Receiver Operating Characteristic Curve (AUROC)

$$\text{AUROC} = \int_0^1 \text{TPR}(t)\, d\text{FPR}(t)$$

where TPR (True Positive Rate) and FPR (False Positive Rate) are computed at different thresholds.

Area Under the Precision-Recall Curve (AUPRC)

$$\text{AUPRC} = \int_0^1 \text{Precision}(r)\, d\text{Recall}(r)$$

where precision and recall vary based on the classification threshold.

Accuracy

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

where TP, TN, FP, and FN represent the true positives, true negatives, false positives, and false negatives, respectively.

Precision

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1-score

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The final reported performance is the average metric score across the 10 folds.

B.3 HYPERPARAMETERS

The model consists of two steps: alignment and drug response prediction step. Tabels 3 and 4 are the hyperparameters for each step.

Table 3: Hyperparameters for alignment

Component	Hyperparameter
Autoencoder for GDSC	$978 \rightarrow 256 \rightarrow 128 \rightarrow 256 \rightarrow 978$
Autoencoder for TCGA	$978 \rightarrow 128 \rightarrow 256 \rightarrow 978$
Latent Tissue Classifier	$128 \rightarrow 128 \rightarrow 32 \rightarrow 23$
Expression Tissue Classifier	$978 \rightarrow 512 \rightarrow 128 \rightarrow 32 \rightarrow 23$
Batch Size	128
Learning Rate	1×10^{-3}
Loss Weights	$\alpha = 0.2, \beta = 0.8, \gamma = 0.4$
Epochs	199

Table 4: Hyperparameters for drug response prediction

Component	Architecture/Hyperparameter
Perturbation Embedding	$978 \rightarrow 256$
Rank-Based Embedding	$978 \rightarrow 256$
Molecular Embedding	$978 \rightarrow 256$
Concatenation & Prediction	$768 \rightarrow 128 \rightarrow 1$
Dropout Rate	0.1
Batch Size	512
Learning Rate	1×10^{-3}
Training Strategy	Early stopping with patience 10 steps

In both steps, we fixed the random seed to 42 and used torch.optim.Adam as the optimizer.

C RELATED WORK

C.1 TRANSCRIPTOMIC FOUNDATION MODELS FOR TRANSFER LEARNING

Recent advances in single-cell sequencing have generated extensive multimodal datasets, propelling the development of foundation models (FMs) in this domain (Liu et al., 2023). While FMs have excelled in DNA analysis and biomedical NLP, their application to single-cell data remains in its infancy. Existing single-cell FMs, such as scBERT (Yang et al., 2022), scGPT (Cui et al., 2024), and GeneFormer (Theodoris et al., 2023), primarily focus on cell-type annotation and gene function prediction. GeneFormer distinguishes itself by the novel employment of a rank-based positional encoding scheme, which facilitates more fluent correction of batch effects—a common challenge in single-cell data analysis. This approach contrasts with conventional methods and contributes to improved integration and analysis across diverse datasets.

In the realm of drug response prediction, chemical perturbation can be defined at multiple scales—gene, cell line, and patient—with each scale capturing different aspects of cellular behavior. Addressing the gap between these scales, CSG^2A (Bang et al., 2024) introduces a condition-specific gene–gene attention mechanism. This perturbation-based modeling approach dynamically learns gene interactions under specific input conditions and transfers this detailed gene-level knowledge to the cell line level.

Together, models like GeneFormer and CSG^2A highlight the evolving capabilities of transcriptome foundation models, bridging the gap between gene-level perturbations and higher-level cellular responses, which is crucial for advancing precision medicine.

C.2 DRUG RESPONSE PREDICTION MODELS

Drug response prediction models mainly focus on accurately predicting the cell-scale drug response, mainly due to the large-scale available data provided by the GDSC project. However the essential aim is to transfer this knowledge to patient level to enable precision medicine. We first describe the cell-scale drug response models, then provide the pioneering domain adaptation-based models that aim to translate cell-scale modeling to patient data through alignment techniques.

C.2.1 DOMAIN ADAPTATION-FREE (CELL-LEVEL) DRUG RESPONSE MODELS

DeepCDR (Liu et al., 2020) DeepCDR (Precision Medicine Prediction using an Adversarial Network for Cancer Drug Response) is a hybrid graph convolutional network (GCN) designed for cancer drug response (CDR) prediction by integrating multi-omics profiles of cancer cells with intrinsic chemical structures of drugs. DeepCDR automatically learns latent representations of topological structures among atoms and bonds. The model consists of a uniform GCN and multiple subnetworks, capturing the relationships between genomic, transcriptomic, and chemical data.

CSG^{2}A (Bang et al., 2024) CSG^{2}A is a condition-specific gene–gene attention (CSG^{2}A) network designed to bridge the gap between transcriptomic-level perturbation data (LINCS L1000) and cell-line drug response datasets (GDSC). The model employs a transfer learning strategy, where it is pretrained on LINCS L1000 and fine-tuned on GDSC while preserving gene interaction networks. CSG^{2}A dynamically learns condition-specific gene-gene interactions, incorporating both biological priors and data-driven learning.

C.2.2 Domain Adaptation-based models

CODE-AE (He et al., 2022) CODE-AE is a context-aware deconfounding autoencoder that mitigates data heterogeneity and distribution shift between preclinical and clinical drug response data. It learns a shared embedding space from unlabeled data and refines it through supervised training on labeled cell-line data. To enhance domain alignment, CODE-AE employs regularization techniques, with adversarial learning (CODE-AE-ADV) demonstrating the best performance. During inference, drug responses in patients are predicted using the aligned embedding, improving generalization to clinical data.

PANCDR (Kim et al., 2024) PANCDR (Precision Medicine Prediction using an Adversarial Network for Cancer Drug Response) is a domain adaptation-based model designed to improve preclinical-to-clinical generalization in drug response prediction. It consists of two sub-models: an adversarial model that minimizes domain discrepancy between preclinical and clinical datasets, and a CDR prediction model that extracts drug-response features. PANCDR is trained on both preclinical and unlabeled clinical data, enabling it to generalize effectively to external test sets.

D Additional Results

D.1 Embedding space analysis

The Figure 3 shows the embedding space generated from the latent representations of each model, with colored by data type (Figure 3a) and colored by tissue type (Figure 3b). The results indicate that our model effectively aligns the two data types while also preserving tissue-specific information. CSG^{2}A and CODE-AE were not included because their latent representations vary depending on the drug combination, even for the same sample.

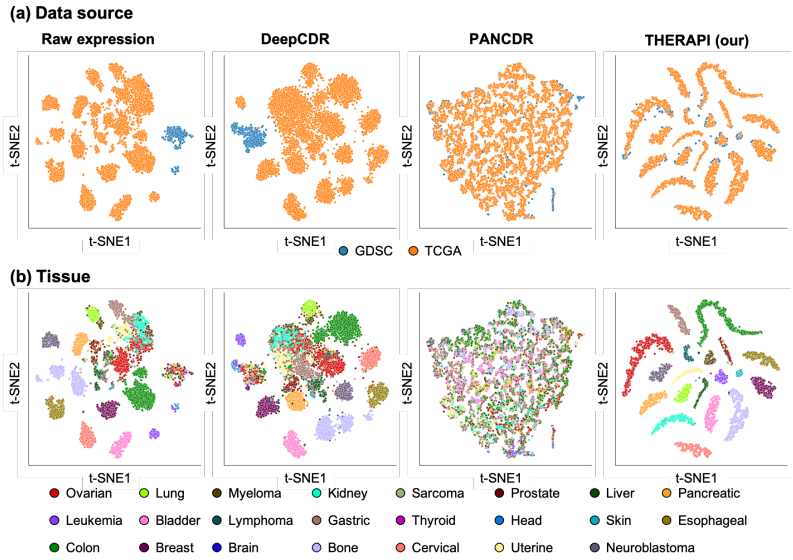

Figure 3: Embedding Space of GDSC and TCGA. t-SNE visualization of the embedding space for GDSC and TCGA, colored by (a) data type and (b) tissue type. The figure presents results for different models in the following order: raw gene expression values (left), DeepCDR, PANCDR, and THERAPI (ours, right).

D.2 External Dataset embedding space

The I-SPY 2 dataset (Wolf et al., 2022) [1], includes 988 breast cancer tumor patients with expression profiling by array data. Among these, 178 patients received Paclitaxel, with drug response labeled based on pathological complete response (pCR) status: a value of 1 indicating complete response (148 patients), and 0 indicating failure to achieve complete response (30 patients). We aligned the 810 patients who did not receive paclitaxel with breast cancer samples from GDSC and then evaluated the alignment of the 178 paclitaxel-treated patients. As shown in the Figure 4, before alignment, the two datasets exhibited distinct embedding distributions, whereas after alignment, they formed a shared embedding space with a similar distribution.

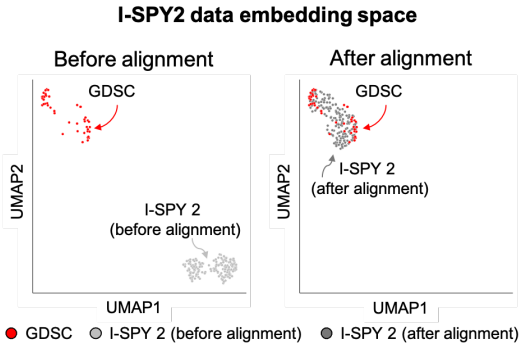

Figure 4: Embedding space of I-SPY2 dataset. Embedding space visualization before and after alignment of GDSC breast samples (red circles) and I-SPY2 samples (light gray circles before alignment and dark gray circles after alignment) using UMAP.

[1]Downloaded from GEO https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE194040