Transferring Cell-level Drug Response to Patient via Tumor Heterogeneity-Aware Alignment and Gene-level Foundational Models

Inyoung Sung [1] **Dongmin Bang** [1 2] **Sun Kim** [1 2] **Sangseon Lee** [3]

Abstract

Prediction of patient-level drug response is critical for precision oncology but remains limited by the scarcity of labeled clinical data. While machine learning models trained on cancer cell lines offer a scalable alternative, biological differences introduce domain shifts that hinder direct translation to patient tumors. We present THERAPI (Tumor Heterogeneity-aware Embedding for Response Adaptation and Patient Inference), a deep learning framework designed to bridge this gap. THERAPI aligns patient and cell line transcriptomes via an attention-based aggregation of cell lines guided by tissue context, enabling tumor heterogeneity-aware modeling. For drug response prediction, it further transfers gene-level knowledge from foundation models based on drug-induced perturbations and rank-based representations. Our approach outperforms nine baselines in predicting patient drug responses and generalizes to external cohorts, while providing interpretable insights into tumor heterogeneity and clinical outcomes. These results highlight the promise of biological context-aware domain adaptation and gene-level knowledge integration for robust, interpretable drug response prediction 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 dataset like The Cancer Genome Atlas (TCGA, Weinstein et al. (2013)) forces many studies to rely on preclinical dataset 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 fundamental 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 patients in TCGA. To address this challenge, domain adaptation (DA)-based techniques have been explored to match the distributions across different datasets (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

[*]Equal contribution [1]Seoul National University, Seoul, Republic of Korea [2]AIGENDRUG Co., Ltd., Seoul, Republic of Korea [3]Inha University, Incheon, Republic of Korea. Correspondence to: Sangseon Lee <ss.lee@inha.ac.kr>.

Proceedings of the ICML 2025 Workshop on Multi-modal Foundation Models and Large Language Models for Life Sciences, Vancouver, Canada. 2025.

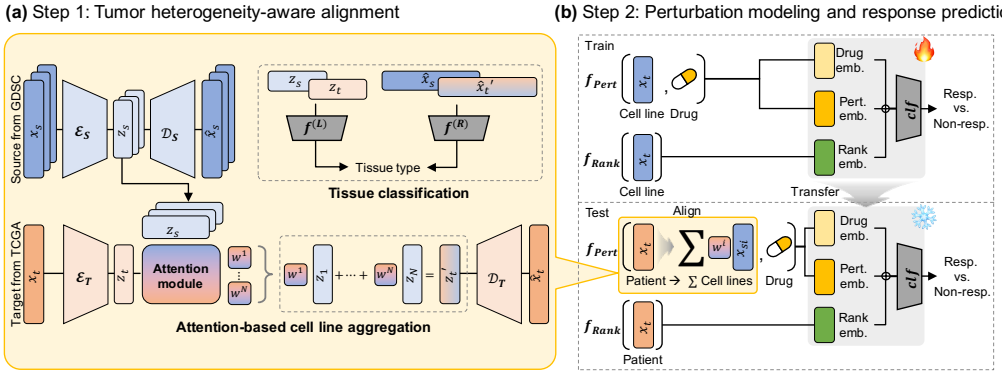

Figure 1: Overview of the THERAPI. (a) Alignment of preclinical (GDSC) and clinical (TCGA) transcriptomes into a shared embedding space. Patient samples are represented as attention-weighted sums of cell lines. (b) Drug response prediction using two modules: a drug-induced perturbation module and a rank-based expression module. The aligned patient representation is passed to both modules to generate the final prediction.

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 perturbations and rank-based representations derived from foundation models, enabling more comprehensive modeling of gene expression without the constraints of low-dimensional latent spaces. 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

Our objective is to predict patient drug response by modeling the biological characteristics of tumors and gene-gene interactions. As shown in Fig. 1, our model THERAPI consists of two steps: Step 1 performs alignment between preclinical and clinical transcriptomes via attention-based aggregation, bringing them into a shared representation space (Section 2.1); Step 2 predicts drug response by integrating perturbation- and rank-based modules for modeling gene-level knowledge (Section 2.2). For a detailed description of the THERAPI architecture, refer to Appendix A.

2.1. 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 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 Fig. 1(a), THERAPI uses two unique features: Attention-based cell line aggregation and Tissue label classifiers. The aggregation module learns to align preclinical and clinical transcriptomes while minimizing reconstruction error within a shared representation space. Tissue label classifiers further inject the tissue information via tissue label classification and an additional loss term.

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, respectively. 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

Table 1: Performance comparison with baseline models on GDSC-TCGA dataset. Models are grouped into cell line-trained (i.e., DA-free) and DA-based categories based on whether domain adaptation was applied. Mean and standard deviation of 10-fold CV are provided. Best performance and its comparable results are marked in bold, and second-best are underlined.

	Model	AUROC	AUPRC	Accuracy	Precision	F1
Cell line-trained (DA-free)	Random Forest	0.506 (0.011)	0.486 (0.006)	0.490 (0.013)	0.486 (0.006)	0.649 (0.004)
	SVM	0.662 (0.015)	0.589 (0.012)	0.660 (0.015)	0.632 (0.016)	0.669 (0.020)
	AdaBoost	0.535 (0.036)	0.504 (0.022)	0.534 (0.041)	0.535 (0.059)	0.528 (0.090)
	XGBoost	0.520 (0.034)	0.494 (0.018)	0.509 (0.037)	0.498 (0.026)	0.621 (0.027)
	DeepCDR (Liu et al., 2020)	0.669 (0.074)	0.608 (0.058)	0.575 (0.056)	0.542 (0.044)	0.656 (0.046)
	DEERS (Koras et al., 2021)	0.511 (0.116)	0.543 (0.077)	0.537 (0.045)	0.517 (0.194)	0.388 (0.261)
	CSG^{2}A (Bang et al., 2024)	0.668 (0.053)	0.643 (0.057)	0.561 (0.060)	0.532 (0.042)	0.655 (0.032)
DA-based	CODE-AE (He et al., 2022)	0.668 (0.089)	0.623 (0.067)	0.628 (0.082)	0.643 (0.092)	0.551 (0.144)
	PANCDR (Kim et al., 2024)	0.714 (0.029)	0.687 (0.025)	0.638 (0.049)	0.609 (0.052)	0.663 (0.032)
	THERAPI (ours)	**0.775 (0.034)**	**0.710 (0.024)**	**0.716 (0.039)**	**0.713 (0.051)**	**0.703 (0.051)**

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 (Wen et al., 2016) $\mathcal{L}_{\text{center}}$ to construct a compact space according to the tissue labels in the latent embedding space.

The overall loss in the alignment step 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.

2.2. 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 (Fig. 1(b)-Train). Then, this model is applied to the target domain to predict drug response in patient samples (Fig. 1(b)-Test).

During training of the drug response predictor on the source domain, the input representation is constructed using two foundational models: CSG^{2}A generates post-treatment and drug embeddings that reflect drug-gene interaction (Bang et al., 2024), while Geneformer generates pre-treatment embedding that capture gene-gene interaction (Theodoris et al., 2023). These embeddings are concatenated to form the final input to a MLP, which is trained with binary cross-entropy loss to predict drug response.

After training in the source domain, the drug response predictor is applied to the target domain for patient drug response prediction. A key aspect of the inference step is that the patient input x_t is first mapped into a shared embedding space as $x'_t = \sum_{i=1}^{N_s} w^i_{(t,s)} x^i_s$ using the attention-based aggregation module trained during the alignment step. The transformed patient embedding x'_t is then fed into the trained drug response predictor to estimate patient-specific response. We note that the drug response-labeled patient data are not used in the alignment step to prevent data leakage.

2.3. Experimental Setting

We used GDSC as the source domain and TCGA as the target domain for patient drug response prediction. The GDSC dataset includes 673 cell lines and 174 drugs, totaling 112,533 cell line–drug pairs. TCGA contains 8,400 patient samples, with 383 drug-treated pairs spanning 21 drugs. Model performance was evaluated using AUROC, AUPRC, accuracy, precision, and F1-score, averaged over 10-fold cross-validation. Details are provided in Appendix B.

3. Results

3.1. Benchmark Dataset Performances

We evaluated THERAPI on the task of translating drug response predictions from GDSC to TCGA data. Our benchmark included seven DA-free models (DeepCDR (Liu et al., 2020), DEERS (Koras et al., 2021), CSG^{2}A (Bang et al., 2024), and classical ML models) and two DA-based models (CODE-AE (He et al., 2022) and PANCDR (Kim et al., 2024)). As shown in Table 1, THERAPI achieved the best performance across all evaluation metrics. Consistent with previous findings, DA-based models outperformed DA-free ones, underscoring the importance of DA.

Among DA-based approaches, THERAPI showed the highest accuracy, which we attribute to its use of tissue labels and tumor heterogeneity information for biologically informed alignment. As illustrated in Fig. 2, THERAPI preserved tissue structure in the shared embedding space, unlike PANCDR or raw expression embeddings. This biological consistency may explain its superior performance in cross-domain generalization.

Ablation studies To investigate the contribution of each component in THERAPI, we performed an ablation study by removing the alignment module, the rank-based representation, and the perturbation-based representation. The full model achieved the best overall performance across all

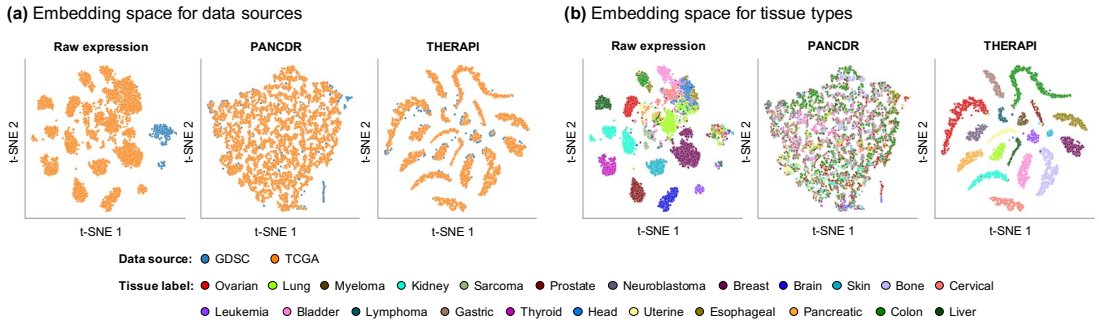

Figure 2: Embedding space of data sources and tissue types. t-SNE plots of GDSC and TCGA data, colored by (a) data source and (b) tissue type. In each panel, from left to right: raw expression data, PANCDR, and THERAPI-aligned embeddings.

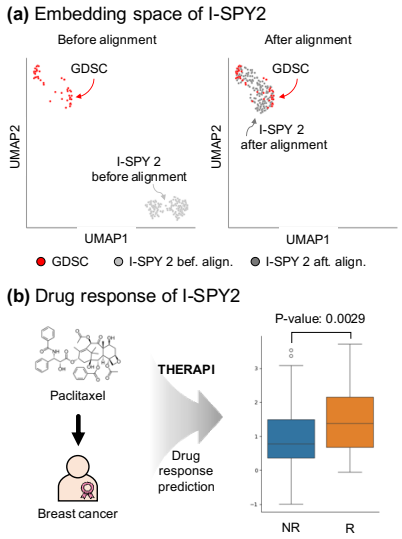

Figure 3: External validation on the I-SPY 2 dataset. (a) UMAP plots before (left) and after (right) alignment, showing improved domain integration. (b) THERAPI significantly distinguished responders (R) and non-responders (NR) to paclitaxel (p = 0.0029).

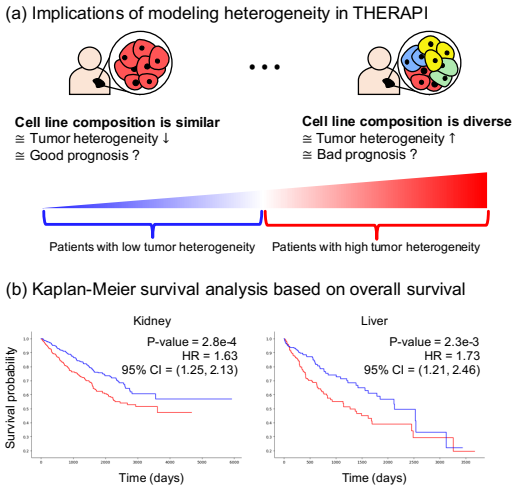

Figure 4: Entropy-based interpretation of tumor heterogeneity. (a) Tumors represented as attention-weighted mixtures of cell lines; entropy captures heterogeneity. (b) Kaplan–Meier curves are presented for kidney and liver cancers.

ablation settings (Appendix C). We observed that adding the alignment module improved F1 scores, suggesting effective reduced domain discrepancy. Among the two representation modules, the perturbation-based module contributed the most to AUROC, highlighting the importance of modeling drug-induced gene perturbations. These results support the complementary roles of each module in THERAPI.

3.2. 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, the alignment step successfully integrated the I-SPY 2 and GDSC samples into a shared space (Fig. 3(a)), and THERAPI significantly distinguished responders and non-responders to paclitaxel treatment (p = 0.0029, Fig. 3(b)). We further evaluated THERAPI on a colorectal cancer dataset (Marisa et al., 2013), with results highlighting its robustness and potential for drug recommendation (Appendix D).

3.3. Entropy-Based Interpretation of THERAPI

In the alignment step, THERAPI models each patient tumor as a linear combination of cancer cell lines, a strategy supported by prior studies in translational research (Salvadores et al., 2020; Hsu et al., 2024). We interpret this formulation as an indicator of tumor heterogeneity, where greater diversity in cell line contributions indicates higher heterogeneity (Fig. 4(a)). To quantify heterogeneity, we computed the Shannon entropy of cell line weights and stratified patients into high- and low-entropy groups. As shown in Fig. 4(b), Kaplan–Meier analysis based on overall survival (OS) showed significantly worse prognosis for high-heterogeneity patients, notably in kidney (p = 2.8e-4) and liver (p = 2.3e-3) cancers. Results across all four clinical endpoints and sixteen tissue types are provided in Appendix E.

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

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

Feng, F., Shen, B., Mou, X., Li, Y., and Li, H. Large-scale pharmacogenomic studies and drug response prediction for personalized cancer medicine. *Journal of Genetics and Genomics*, 48(7):540–551, 2021.

Goldman, M. J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A. N., et al. Visualizing and interpreting cancer genomics data via the xena platform. *Nature biotechnology*, 38(6):675–678, 2020.

He, D., Liu, Q., Wu, Y., and Xie, L. 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.

Hsu, Y.-C., Chiu, Y.-C., Lu, T.-P., Hsiao, T.-H., and Chen, Y. Predicting drug response through tumor deconvolution by cancer cell lines. *Patterns*, 5(4), 2024.

Iorio, F., Knijnenburg, T. A., Vis, D. J., Bignell, G. R., Menden, M. P., Schubert, M., Aben, N., Gonçalves, E., Barthorpe, S., Lightfoot, H., et al. A landscape of pharmacogenomic interactions in cancer. *Cell*, 166(3):740–754, 2016.

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

Koras, K., Kizling, E., Juraeva, D., Staub, E., and Szczurek, E. Interpretable deep recommender system model for prediction of kinase inhibitor efficacy across cancer cell lines. *Scientific reports*, 11(1):15993, 2021.

Lee, J.-K., Liu, Z., Sa, J. K., Shin, S., Wang, J., Bordyuh, M., Cho, H. J., Elliott, O., Chu, T., Choi, S. W., et al. Pharmacogenomic landscape of patient-derived tumor cells informs precision oncology therapy. *Nature genetics*, 50(10):1399–1411, 2018.

Liu, Q., Hu, Z., Jiang, R., and Zhou, M. Deepcdr: a hybrid graph convolutional network for predicting cancer drug response. *Bioinformatics*, 36(Supplement_2):i911–i918, 2020.

Marisa, L., de Reyniès, A., Duval, A., Selves, J., Gaub, M. P., Vescovo, L., Etienne-Grimaldi, M.-C., Schiappa, R., Guenot, D., Ayadi, M., et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS medicine*, 10(5):e1001453, 2013.

Mourragui, S., Loog, M., Van De Wiel, M. A., Reinders, M. J., and Wessels, L. F. Precise: a domain adaptation approach to transfer predictors of drug response from pre-clinical models to tumors. *Bioinformatics*, 35(14): i510–i519, 2019.

Salvadores, M., Fuster-Tormo, F., and Supek, F. Matching cell lines with cancer type and subtype of origin via mutational, epigenomic, and transcriptomic patterns. *Science Advances*, 6(27):eaba1862, 2020.

Shen, B., Feng, F., Li, K., Lin, P., Ma, L., and Li, H. A systematic assessment of deep learning methods for drug response prediction: from in vitro to clinical applications. *Briefings in Bioinformatics*, 24(1):bbac605, 2023.

Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., Gould, J., Davis, J. F., Tubelli, A. A., Asiedu, J. K., et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*, 171 (6):1437–1452, 2017.

Theodoris, C. V., Xiao, L., Chopra, A., Chaffin, M. D., Al Sayed, Z. R., Hill, M. C., Mantineo, H., Brydon, E. M., Zeng, Z., Liu, X. S., et al. Transfer learning enables predictions in network biology. *Nature*, 618(7965):616–624, 2023.

Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J. M. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120, 2013.

Wen, Y., Zhang, K., Li, Z., and Qiao, Y. 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.

Wolf, D. M., Yau, C., Wulfkuhle, J., Brown-Swigart, L., Gallagher, R. I., Lee, P. R. E., Zhu, Z., Magbanua, M. J., Sayaman, R., O'Grady, N., 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.

A. Details in Methodology

A.1. Alignment step

Attention-based cell line aggregation by autoencoder framework A central objective of our approach is to model 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 x_s^i denotes the expression vector of the i-th sample from the source domain, x_t denotes the expression vector from the target domain, and $z_s^i, z_t \in \mathbb{R}^{d_z}$. The attention module uses these latent representations to compute source weights:

$$w_{(t,s)}^i = \frac{\exp\left(\langle Qz_t,\, Kz_s^i\rangle\right)}{\sum_{j=1}^{N_s} \exp\left(\langle Qz_t,\, Kz_s^j\rangle\right)},$$

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 shared 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 tumor heterogeneity.

Tissue label classifiers To further incorporate tissue information in the attention-based autoencoder framework, we introduce tissue classification at two representation levels: the latent space $\{z\}$ and the reconstructed expression space $\{x'\}$. A MLP-based tissue classifier f_S is trained in both domains. In the source domain, we train $f_S^{(L)}(z_s^i)$ and $f_S^{(R)}(\hat{x}_s)$, where $f_S^{(L)}$ is a tissue classifier in the latent space, $f_S^{(R)}$ is a tissue classifier in the reconstruction space, and $\hat{x}_s = \mathcal{D}_S(z_s^i)$. 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 tightly cluster tumor representations of same tissue type 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

Training in the cell line data During training, the drug response predictor is constructed using two foundational models, each capturing a distinct aspect of drug–gene or gene–gene interactions. First, we use CSG^{2}A (Bang et al., 2024) to generate drug-perturbation and drug representations. Given a gene expression-drug SMILES pair as input, it outputs two components: a perturbed gene expression embedding z_{pert} and a drug embedding z_{drug}, computed as $[z_{\text{pert}}, z_{\text{drug}}] = f_{\text{pert}}(x_s, d)$, where x_s is the pre-treatment gene expression of a source domain sample and d is the molecular structure of the drug. Next we use Geneformer (Theodoris et al., 2023) to generate rank-based gene interaction representations from gene expression: $z_{\text{rank}} = f_{\text{rank}}(x_s)$. The final input for drug response prediction is formed by concatenating the three components: $z = [z_{\text{pert}}; z_{\text{drug}}; z_{\text{rank}}]$, which is then passed into a MLP predictor f_{MLP}. The predicted drug sensitivity probability is given by $\hat{y} = f_{\text{MLP}}(z)$, where $\hat{y}$ represents the predicted probability of drug sensitivity. As the task is formulated as binary classification, 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 ground truth drug response label for sample i, and N is the number of training samples.

Inference in the patient data After training on the source domain, the drug response predictor is applied to the target domain to infer patient-specific drug responses. An important aspect of the inference step lies in how the inputs to the representation models are handled. For the rank-based representation, the patient's gene expression profile x_t is directly input into f_{rank}, following the same procedure as during training, thereby generating an embedding that reflects pre-treatment profile. For the perturbation-based representation, the patient's gene expression is first transformed into an aligned expression profile x'_t using attention-derived weights from the alignment step. Specifically, the aligned expression is computed as

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

where x^i_s is the expression profile of the i-th cell line in the source domain and $w^i_{(t,s)}$ is the corresponding attention weight for patient t. This aligned expression x'_t is then used as input to f_{pert}, which generates a perturbation-informed embedding reflecting the post-treatment profile. The resulting embeddings from all three modules are concatenated and evaluated using the drug response predictor f_{MLP} trained on source domain data, yielding the final prediction $\hat{y}_t$.

B. Experimental Setting

B.1. Datasets

For patient drug response prediction, we mainly utilized the GDSC-TCGA dataset, built using GDSC (Iorio et al., 2016) as the source domain and TCGA (Weinstein et al., 2013) as the target domain. Both datasets contain gene expression profiles utilized to 978 LINCS (Subramanian et al., 2017) 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.

B.1.1. Genomics of Drug Sensitivity in Cancer(GDSC)

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 in GDSC dataset, we converted the original scalar pIC50 values (the negative base 10 logarithm of IC50) into binary labels based on whether each value was above or below the average pIC50 for each drug across all cell lines. Samples with pIC50 values above the mean were labeled as responsive (1), while those below the mean were labeled as non-responsive (0). The GDSC dataset is publicly available at the GDSC portal (https://www.cancerrxgene.org/).

B.1.2. The Cancer Genome Atlas (TCGA)

The TCGA dataset consists of 8,400 patient samples, including 8,042 unlabeled and 358 labeled patients. The unlabeled patients were used exclusively during the alignment step to adapt the source and target domains. The 358 labeled patients were treated with 21 unique drugs, resulting in a total of 383 patient-drug pairs, and were used only in the drug response prediction step. Additionally, for survival analysis, we evaluated four clinical endpoints: overall survival (OS), progression-free interval (PFI), disease-free interval (DFI), and disease-specific survival (DSS) in the TCGA patients. The TCGA dataset is publicly available through the UCSC Xena Cancer Genome Browser (https://xena.ucsc.edu/) (Goldman et al., 2020).

B.1.3. External Datasets: I-SPY2 and GSE39582

The I-SPY2 dataset consists of expression profiling by array data derived from fresh-frozen pre-treatment breast cancer tumor samples of 988 patients (Wolf et al., 2022). Among these, 178 patients received Paclitaxel, with drug response labeled based on pathological complete response (pCR) status: a value of 1 indicating complete response, and 0 indicating failure to achieve complete response. The remaining 810 untreated patients were used in the alignment step alongside the GDSC dataset, while the 178 drug-treated patients were used in the drug response prediction step. The I-SPY2 data was obtained from the GEO database under accession number GSE194040.

The GSE39582 dataset consists of mRNA expression profiles from 519 colorectal cancer patients, including 75 with microsatellite instability (MSI) and 444 with microsatellite stability (MSS) (Marisa et al., 2013). Patient status was determined based on five microsatellite markers: if two or more markers showed instability, the sample was classified as MSI; if one or none showed instability, it was classified as MSS. Since GSE39582 does not include drug response information, we randomly partitioned the dataset into alignment and prediction subsets. Using `train_test_split` from `sklearn` with a fixed seed of 42, the dataset was split in a 7:3 ratio. As a result, 311 samples (259 MSS, 52 MSI) were used for the alignment step, and the remaining 156 samples (133 MSS, 23 MSI) were used for the drug response prediction step. The GSE39582 dataset was downloaded from the GEO database.

B.2. Evaluation metrics

To evaluate the performance of drug response prediction, we employed the following metrics commonly used for binary classification tasks:

Area Under the Receiver Operating Characteristic Curve (AUROC): Measures the model's ability to distinguish between positive and negative classes across all classification thresholds. Higher values indicate better separability.

Area Under the Precision-Recall Curve (AUPRC): Evaluates the trade-off between precision and recall, especially useful in imbalanced classification settings.

Accuracy: The proportion of correctly predicted samples among all samples,

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

Precision: The proportion of correctly predicted positive samples among all samples predicted as positive,

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

F1-score: The harmonic mean of precision and recall, balancing both metrics,

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \quad \text{where} \quad \text{Recall} = \frac{TP}{TP + FN}$$

Here, TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

B.3. Hyperparameters and experimental settings of THERAPI

These tables summarize the model hyperparameters used in the alignment step (top) and the drug response prediction step (bottom) of THERAPI. In both steps, we fixed the random seed to 42 and used `torch.optim.Adam` as the optimizer.

Hyperparameters for Alignment

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

Hyperparameters for Drug Response Prediction

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

C. Ablation study

This table reports the contributions of THERAPI's three core components: the attention-based alignment module, the rank-based gene interaction representation, and the drug-induced perturbation representation. Results are reported as mean and standard deviation from 10-fold cross-validation, with the best performances highlighted in bold.

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

D. External Dataset on Colorectal cancer

The GSE39582 dataset includes 519 colorectal cancer samples labeled as either microsatellite stable (MSS) or unstable (MSI). As it lacks drug response data, we split the dataset (7:3 ratio, seed = 42) into alignment (311 samples) and prediction (156 samples) subsets using `train_test_split` from `sklearn`.

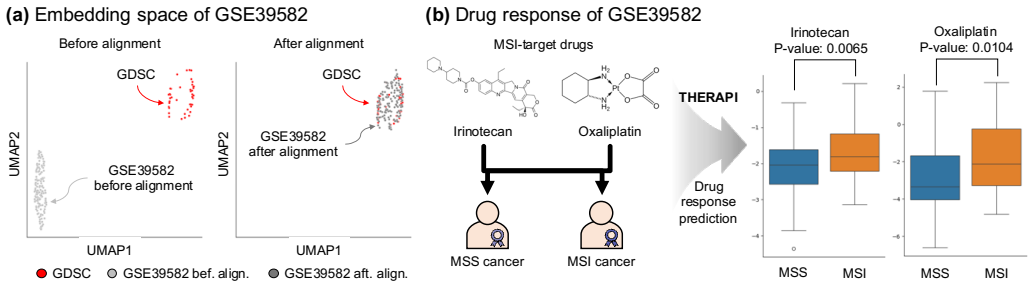

Figure 5: External validation on the colorectal cancer dataset. (a) UMAP plots before (left) and after (right) alignment, showing improved domain integration. (b) Drug response prediction for MSI-targeting drugs Irinotecan and Oxaliplatin in MSS and MSI colorectal cancer patients from the GSE39582 dataset. THERAPI successfully separated the two groups with p-values of 0.0065 (left) and 0.0104 (right), respectively.

E. Survival analysis results of THERAPI

This table reports the survival analysis results of TCGA patients across 16 tissue types using heterogeneity scores derived from THERAPI's alignment weights, evaluated across four clinical endpoints: overall survival (OS), progression-free interval (PFI), disease-free interval (DFI), and disease-specific survival (DSS). Combinations with p-values less than 0.05 are marked with an asterisk (*), and cases where the high heterogeneity group shows worse survival than the low heterogeneity group are highlighed in bold.

Clinical endpoint	Tissue label	p-value
OS	Prostate	0.11983
OS	Kidney	**0.00028** *
OS	Ovarian	0.50917
OS	Esophageal	0.85378
OS	Cervical	0.08865
OS	Breast	0.02713 *
OS	Pancreatic	0.97616
OS	Head	0.19384
OS	Thyroid	0.14876
OS	Lung	**0.00390** *
OS	Skin	0.82554
OS	Bladder	**0.00328** *
OS	Leukemia	0.37799
OS	Liver	**0.00226** *
OS	Colon	0.64275
OS	Uterine	0.31731
DSS	Prostate	**0.04437** *
DSS	Kidney	**0.00006** *
DSS	Ovarian	0.53130
DSS	Esophageal	0.21953
DSS	Cervical	0.08087
DSS	Breast	0.10834
DSS	Pancreatic	0.72395
DSS	Head	0.14643
DSS	Thyroid	**0.01508** *
DSS	Lung	**0.00643** *
DSS	Skin	0.55142
DSS	Bladder	**0.00099** *
DSS	Leukemia	1.00000
DSS	Liver	**0.00085** *
DSS	Colon	0.94083
DSS	Uterine	0.31731

Clinical endpoint	Tissue label	p-value
DFI	Prostate	0.76160
DFI	Kidney	0.41007
DFI	Ovarian	0.61714
DFI	Esophageal	**0.00885** *
DFI	Cervical	0.73464
DFI	Breast	0.90478
DFI	Pancreatic	0.60727
DFI	Head	0.44692
DFI	Thyroid	0.22566
DFI	Lung	0.49217
DFI	Skin	1.00000
DFI	Bladder	0.85834
DFI	Leukemia	1.00000
DFI	Liver	**0.01803** *
DFI	Colon	0.07087
DFI	Uterine	0.31731
PFI	Prostate	0.14422
PFI	Kidney	0.02077 *
PFI	Ovarian	0.33629
PFI	Esophageal	0.09761
PFI	Cervical	0.05734
PFI	Breast	0.62507
PFI	Pancreatic	0.99714
PFI	Head	0.10686
PFI	Thyroid	0.31057
PFI	Lung	0.55263
PFI	Skin	0.19494
PFI	Bladder	**0.03877** *
PFI	Leukemia	1.00000
PFI	Liver	**0.00108** *
PFI	Colon	0.98653
PFI	Uterine	0.69489