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

ICLR International Conference On Learning Representations

Machine Learning for Genomics Exploration Workshop (MLGenX)

Motivation

Background

- **Predicting personalized drug response is vital for precision medicine** but limited by the scarcity of labeled clinical drug response data.
- Preclinical datasets like GDSC are widely used, but direct transfer to patient data (e.g., TCGA) suffers from **domain shift due to biological and methodological differences**.
- **Domain adaptation (DA)** has emerged as a promising strategy to bridge this gap by aligning cell line and patient data in a shared embedding space.

Challenges

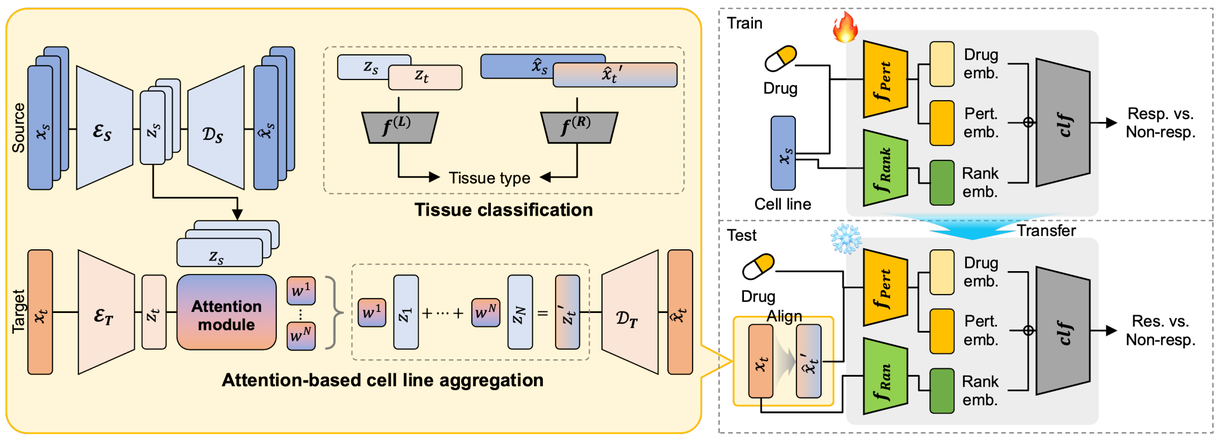
- **Tumor Heterogeneity Ignored**: Existing models do not account for tissue type and tumoral composition, which are key factors in therapy response.
- **Gene Interactions Underused**: Most DA models treat gene expression as independent features, overlooking gene-gene interaction and drug-induced perturbation signals.

Methods

THERAPI: Tumor Heterogeneity-aware Embedding for Response Adaptation and Patient Inference

(a) Step 1: Cell line-Patient Alignment

(b) Step 2: Drug response prediction



Step 1. Cell line-Patient Alignment

- **Attention-based cell line aggregation**: Model tumor heterogeneity by representing each patient tumor as a linear combination of multiple cancer cell lines.
- **Tissue classifier**: To take tissue contexts into the learning process during DA, apply MLP-based classifiers in both latent and reconstructions vectors.

Step 2. Drug Response Prediction

- **Train predictor using cell lines**: During the training of prediction, use three pre-trained models: (1) Drug-induced perturbation, (2) Rank-based representation, and (3) Attention-based molecular encoder.
- **Test predictor using patients**: After training at the source domain, the drug response predictor is applied to the target domain for patient drug response prediction.

Results

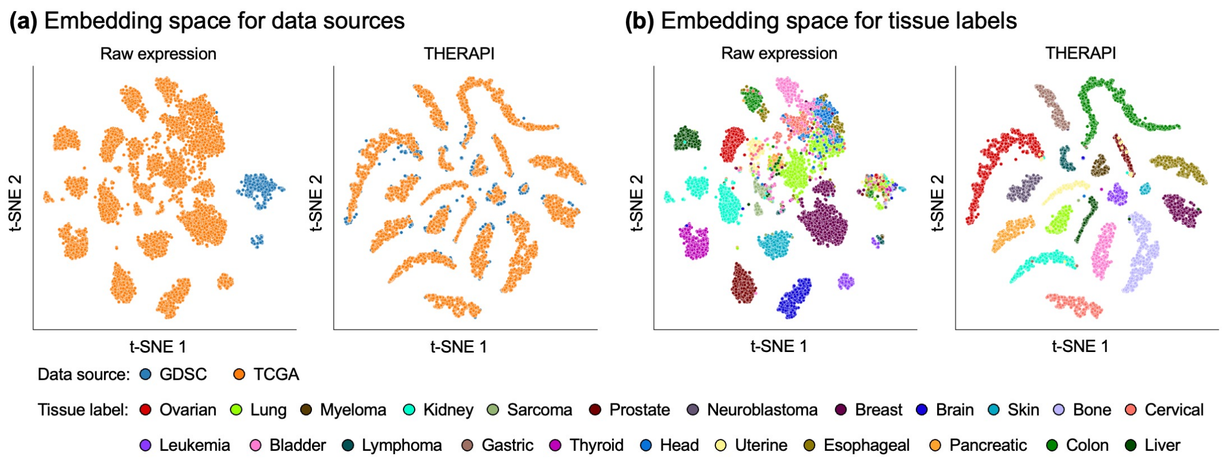
Benchmark dataset performance

- **THERAPI achieved the best performance** across most metrics. As expected, DA-based models performed better than DA-free ones.

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, with best performances in bold.

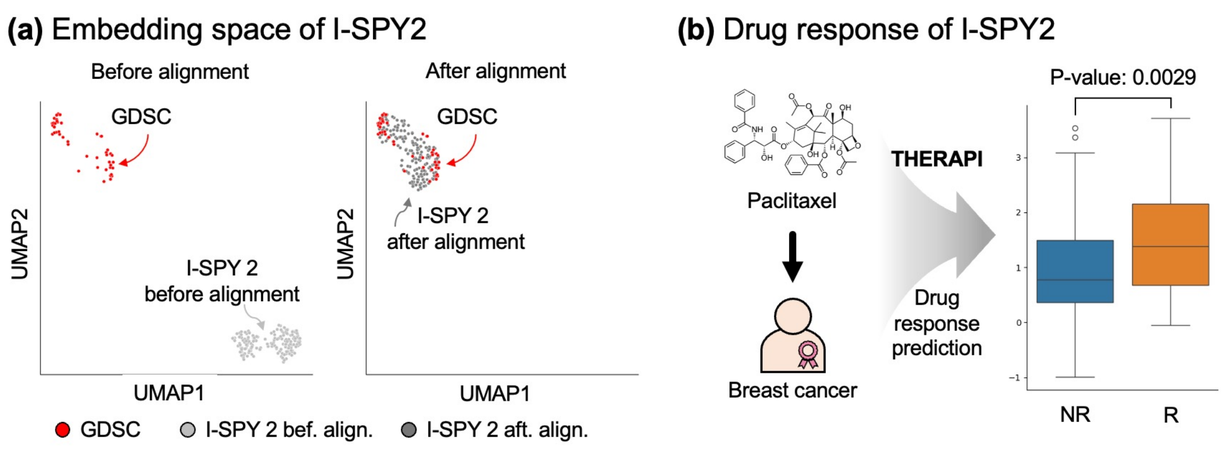
Model		AUROC	AUPRC	Accuracy	Precision	Recall	F1
DA-free (Cell-level)	DeepCDR	0.669 (0.074)	0.608 (0.058)	0.575 (0.056)	0.542 (0.044)	0.844 (0.113)	0.656 (0.046)
	CSG^{2}A	0.668 (0.053)	0.643 (0.057)	0.561 (0.060)	0.532 (0.042)	**0.857 (0.048)**	0.655 (0.032)
DA-based	CODE-AE	0.668 (0.089)	0.623 (0.067)	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)**

- Notably, **the embedding space of THERAPI not only aligned both datasets within a shared space but also clustered based on the tissue label**, outperforming other DA-based models.



External dataset performance

- Using an **external breast cancer dataset**, THERAPI constructed a unified embedding space and significantly separated responders from non-responders, **demonstrating its robustness**.



Conclusion

- THERAPI enables biologically informed domain adaptation by modeling tumor heterogeneity and gene-level interactions.
- Bridges the translational gap between preclinical and patient data through tumor-aware alignment and knowledge transfer.

Open to academic positions!

Postdoctoral research fellow @ SNU
PhD in Bioinformatics & MS in Mathematics

- E-mail: inyoung.sung@snu.ac.kr
- Linkedin: inyoung-sung