Appendix

1. Details of TransDRP

1-1. Positive/negative instance construction

Strategies to yield positive and negative instances often dominate the performance of contrastive learning (Zhang et al. 2022). Unlike image and text data, it is unwise to construct positive instances by performing regular data augmentations on cell line or patient samples, as even minor perturbations may lead to significant bioproperty changes for genomic profiles. Moreover, sampling strategies based on supervised contrastive learning (Khosla et al. 2020) are also inadequate, due to the limited number of samples for some cancer categories, which cannot guarantee their inclusion in each mini-batch and prevents contrastive learning from separating different cancer categories. Therefore, we bring in the concept of prototype networks to predefine both positive and negative instances.

1-2. Implementation details

In the domain representation extractor, both the encoders and decoders are implemented as the 3-layer fully-connected network with batch normalization, the pre-training epoch τ_a is set to 200, which is determined based on experimental experience. In the multi-label response classifier, the co-occurrence rate is computed as the conditional probability $P(y_i|y_j)$, representing the likelihood of label y_i occurring when label y_j appears in the training set. And we apply a threshold of 0.1 to binarize the probabilities and obtain $\mathcal{E}$. The decoder $\boldsymbol{g}_a$ is defined as a 3-layer graph attention network (GAT) with two heads and mean aggregation. In the domain adversarial training, the discriminator $\boldsymbol{\varphi}$ is a 3-layer MLP, $\boldsymbol{\Gamma}$ is a simple bilinear function $\sigma(a^T W b)$ that estimates similarities between two vectors a and b, where W indicate a learnable matrix and σ is a sigmoid function. In the transferring drug response prediction, the adaptation epoch τ_b is set to 300, and the coefficients α and β are both fixed to 0.3 because they produced the lowest loss during the adaptation phase. At last, we utilized the Adam with a learning rate of 0.0001 to optimize the entire model, and for those missing source domain labels, we omitted to compute their loss during training.

2. Details of Results

2-1. Datasets

According to the treatment progression of cancer patients, TCGA database classifies clinical drug response into four types: complete response, partial response, stable disease, and clinical progressive disease. These response types are accessible through an open-source interface: UCSC Xena (https://xenabrowser.net/). Here, we binarize response types to obtain labels for the testing set, with complete and partial responses classified as sensitivity, and stable and progressive diseases classified as resistance. Data details of the 9 TCGA drugs used for target domain testing are listed in Table S1. In addition, GDSC dataset (i.e., drug responses for cell lines) used for pre-training of multi-label response classifier is provided in Table S2.

Table S1: Annotated samples of the 9 TCGA drugs

Drug name	TCGA samples	Pos : Neg
5-Fluorouracil	220	1.44
Cisplatin	417	1.54
Cyclophosphamide	170	7.09
Docetaxel	159	1.83
Doxorubicin	159	1.65
Etoposide	96	1.82
Gemcitabine	230	0.64
Paclitaxel	222	1.55
Temozolomide	134	0.27

"Pos:Neg" denotes the proportion of positive and negative labels in samples.

Table S2: Annotated samples of the 9 GDSC drugs

Drug name	GDSC samples	Pos : Neg
5-Fluorouracil	702	0.96
Cisplatin	702	1.24
Cyclophosphamide	578	1.28
Docetaxel	701	1.04
Doxorubicin	707	0.96
Etoposide	652	1.07
Gemcitabine	712	1.03
Paclitaxel	633	1.15
Temozolomide	694	1.37

2-2. Baselines

Pipelines of six UDA baselines are presented below:

- Celligner (Warren et al. 2021) aligned gene expression in patients with that in cell lines using an unsupervised way, which corrects batch discrepancies in scRNA-seq data to remove systematic differences between domain features, and then uses the aligned features for transfer DRP via an elastic network.
- Velodrome (Sharifi-Noghabi et al. 2021) designed two neural networks: a shared feature extractor and a domain-specific predictor, and incorporates a domain consistency loss and a feature alignment loss to transfer neural networks in the target domain.
- scDEAL (Chen et al. 2022) utilized a pre-trained autoencoder to extract embeddings from Bulk-seq and scRNA-seq samples, and used maximum mean discrepancy to enforce embedding alignment between two domains during the Bulk-to-scRNA DRP. Here, we deploy this model for cell line-to-patient DRP.
- AITL (Sharifi-Noghabi et al. 2020) took gene expression of patients and cell lines as the input, employing an adversarial domain network with multi-task learning to address data discrepancies and predict multiple drug responses simultaneously.

- CODE-AE (He et al. 2022) built upon a domain separation network, which constructs neural networks to separate gene expression from cell lines and patients into private and shared features. By ensuring orthogonal feature separation, it reduced confounding factors between domains in the DRP.
- WISER (Shubham et al. 2024) modelled genomic profiles as a combination of discrete drug representations to handle various drug responses, and then used weak supervision and subset selection on unlabeled patient data to improve the generalization of DRP classifier trained on cell line data.

Pipelines of four cell line baselines are presented below:

- tCNNs (Liu et al. 2019) employed the CNNs to predict cancer drug responses (CDR), and the SMILES sequences of drugs and genomic mutation data of cell lines are used for the input features.
- DeepCDR (Liu et al. 2020) integrated multi-omics profiles of cell lines and chemical structures of drugs, and then developed a hybrid GNN to predict CDRs.
- GraphCDR (Liu et al. 2022) accomplished a GNN-based framework based on expression profiles of cell lines, chemical structures of drugs, and their known responses for CDR prediction.
- Bi-GNN (Hostallero, Li, and Emad 2022) predicted CDRs through a graph representation learning method that incorporates information regarding the sensitivity and resistance of cell lines.

2-3. Model performance of other metrics

Other metrics (accuracy and f1-score) of TransDRP and UDA models for 9 clinical drugs on the TCGA dataset, are provided in below Figure S1.

2-4. Computational efficiency

The 5-CV running time of each method for all 9 TCGA drug data within the 300 UDA epochs, excluding pretraining phases, is presented in Table S3. All methods are deployed on a unified experimental platform, i.e., Intel CoreTM i7-13700@3.6GHz CPU, NVIDIA GTX 3080Ti GPU, 32GB RAM, Linux system, Python 3.11 and PyTorch Geometric 2.2.

Table S3: Running time of methods (sum of 9 drugs)

Methods	UDA epochs	Running time (s)
Celligner	300	2443.61
Velodrome	300	2650.33
scDEAL	300	3347.29
AITL	300	1520.10
CODE-AE	300	3784.42
WISER	300	978.77
TransDRP	300	762.05

2-5. Comparison with cell line-oriented models

Performance (e.g., AUC, AUPR, Accuracy, F1-score) of TransDRP and cell line-oriented baselines for 9 drugs on the TCGA dataset, are provided in below Table S4.

References

Chen, J.; Wang, X.; Ma, A.; Wang, Q.-E.; Liu, B.; Li, L.; Xu, D.; and Ma, Q. 2022. Deep transfer learning of cancer drug responses by integrating bulk and single-cell RNA-seq data. *Nature Communications*, 13(1): 6494.

He, D.; Liu, Q.; Wu, Y.; and Xie, L. 2022. 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.

Hostallero, D. E.; Li, Y.; and Emad, A. 2022. Looking at the BiG picture: incorporating bipartite graphs in drug response prediction. *Bioinformatics*, 38(14): 3609–3620.

Khosla, P.; Teterwak, P.; Wang, Y.; Isola, P.; Maschinot, A.; Liu, C.; and Krishnan, D. 2020. Supervised contrastive learning. *Advances in neural information processing systems*, 33: 18661–18673.

Liu, P.; Li, H.; Li, S.; and Leung, K.-S. 2019. Improving prediction of phenotypic drug response on cancer cell lines using deep convolutional network. *BMC bioinformatics*, 20: 1–14.

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

Liu, X.; Song, C.; Huang, F.; Fu, H.; and Zhang, W. 2022. GraphCDR: a graph neural network method with contrastive learning for cancer drug response prediction. *Briefings in Bioinformatics*, 23(1): bbab457.

Sharifi-Noghabi, H.; Harjandi, P. A.; Zolotareva, O.; Collins, C. C.; and Ester, M. 2021. Out-of-distribution generalization from labelled and unlabelled gene expression data for drug response prediction. *Nature Machine Intelligence*, 3(11): 962–972.

Sharifi-Noghabi, H.; Peng, S.; Zolotareva, O.; Collins, C. C.; and Ester, M. 2020. AITL: adversarial inductive transfer learning with input and output space adaptation for pharmacogenomics. *Bioinformatics*, 36(Supplement_1): i380–i388.

Shubham, K.; Jayagopal, A.; Danish, S. M.; AP, P.; and Rajan, V. 2024. WISER: Weak supervISion and supErvised Representation learning to improve drug response prediction in cancer. *arXiv preprint arXiv:2405.04078*.

Warren, A.; Chen, Y.; Jones, A.; Shibue, T.; Hahn, W.; Boehm, J.; Vazquez, F.; Tsherniak, A.; and McFarland, J. 2021. Global computational alignment of tumor and cell line transcriptional profiles. *Nature Communications*, 12(1): 22–22.

Zhang, Z.; Zhao, Y.; Chen, M.; and He, X. 2022. Label Anchored Contrastive Learning for Language Understanding. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1437–1449.

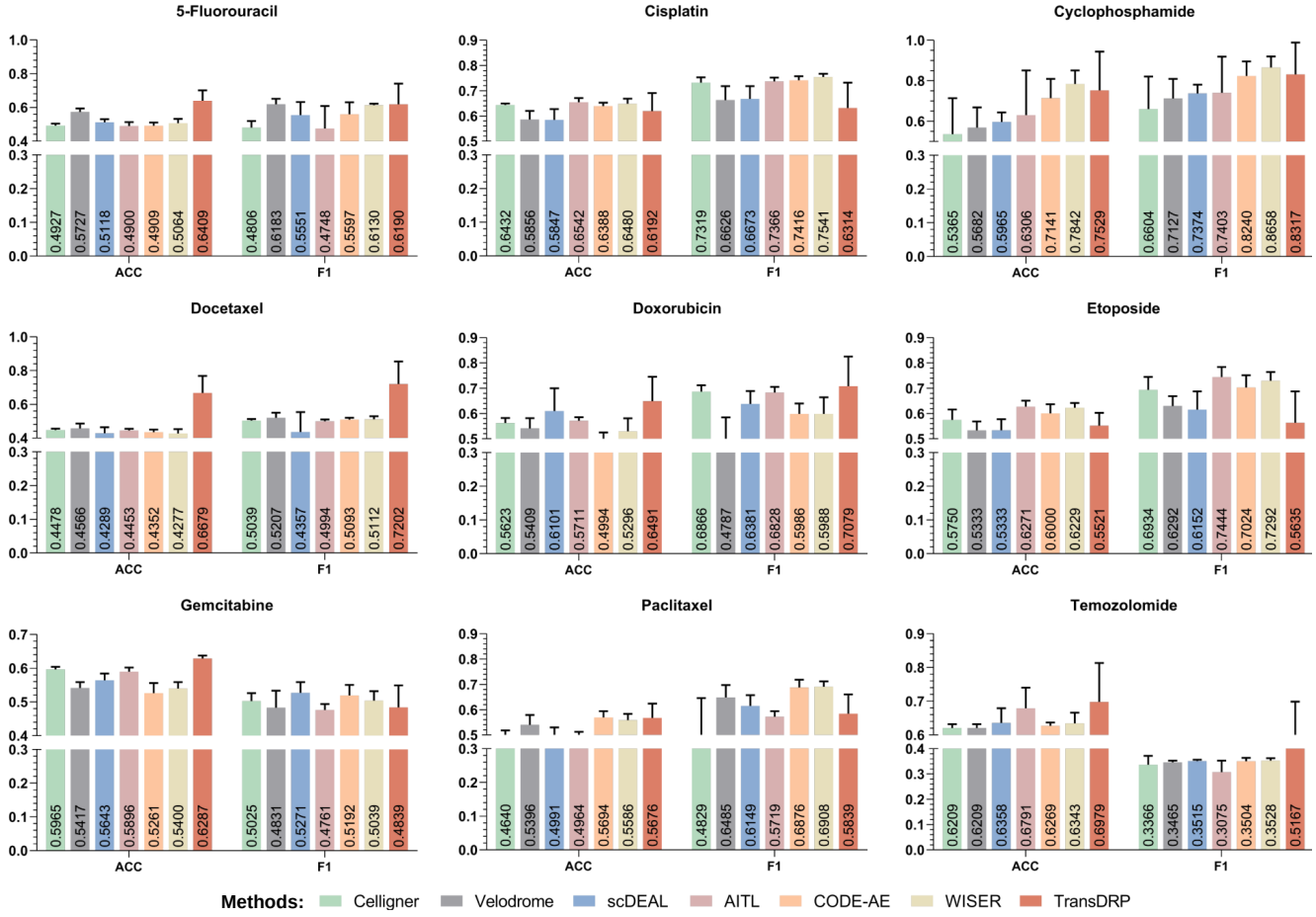

Figure S1: Accuracy (ACC) and f1-score (F1) of all methods on the TCGA dataset for 9 clinical drugs.

Table S4: Performance of TransDRP and cell line baselines on the TCGA dataset for 9 clinical drugs.

Drug name	t-CNNs		DeepCDR		GraphCDR		Bi-GNN		TransDRP	
	AUC	AUPR	AUC	AUPR	AUC	AUPR	AUC	AUPR	AUC	AUPR
5-Fluorouracil	0.6154	0.6898	0.4260	0.4668	0.6781	0.6337	0.6929	0.6977	0.7545	0.8233
Cisplatin	0.5095	0.6194	0.6311	6230	0.5580	0.6407	0.5237	0.6804	0.6815	0.7245
Cyclophosphamid	0.4711	0.7771	0.4695	0.6121	0.5660	0.7255	0.5583	0.7683	0.6717	0.8363
Docetaxel	0.4100	0.6457	0.5686	0.6402	0.3841	0.6663	0.4194	0.6324	0.6849	0.7627
Doxorubicin	0.4338	0.5288	0.4330	0.4464	0.4784	0.5029	0.4647	0.4608	0.6556	0.7118
Etoposide	0.5646	0.6616	0.5359	0.6173	0.5910	0.6476	0.5801	0.6058	0.6544	0.7247
Gemcitabine	0.5930	0.4736	0.6145	0.4593	0.5490	0.4957	0.6119	0.4772	0.6115	0.4808
Paclitaxel	0.4636	0.5750	0.4718	0.6121	0.4644	0.5919	0.5318	0.5992	0.6051	0.6945
Temozolomide	0.4074	0.5146	0.4420	0.4913	0.5712	0.5276	0.5318	0.5003	0.6689	0.5676

Note that, each value reports the average result of 5-CV experiments.