Journal Pre-proof

A disentangled generative model for improved drug response prediction in patients via sample synthesis

Kunshi Li, Bihan Shen, Fangyoumin Feng, Xueliang Li, Yue Wang, Na Feng, Zhixuan Tang, Liangxiao Ma, Hong Li

PII: S2095-1779(24)00225-9

DOI: https://doi.org/10.1016/j.jpha.2024.101128

Reference: JPHA 101128

To appear in: *Journal of Pharmaceutical Analysis*

Received Date: 10 July 2024

Revised Date: 29 September 2024

Accepted Date: 22 October 2024

Please cite this article as: K. Li, B. Shen, F. Feng, X. Li, Y. Wang, N. Feng, Z. Tang, L. Ma, H. Li, A disentangled generative model for improved drug response prediction in patients via sample synthesis, *Journal of Pharmaceutical Analysis*, https://doi.org/10.1016/j.jpha.2024.101128.

Manuscript Title:

A disentangled generative model for improved drug response prediction in patients via sample synthesis

Kunshi Li [a, 1], Bihan Shen [a, 1], Fangyoumin Feng [a], Xueliang Li [a], Yue Wang [b], Na Feng [b], Zhixuan Tang [a], Liangxiao Ma [a], Hong Li [a, *]

[a] CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, 200031, China.

[b] Molecular Pathology Laboratory, National Center for Liver Cancer, Eastern Hepatobiliary Surgery Hospital, Shanghai, 201800, China.

* Corresponding author.

E-mail addresses: lihong01@sinh.ac.cn (H. Li)

1 Both authors contributed equally to this work.

A Disentangled generative model for improved drug response prediction in patients via sample synthesis

1. Introduction

The prediction of personalized drug responses from molecular omics is important for precision medicine [1]. Large-scale omics data coupled with high-throughput drug sensitivities experiments have generated pharmacogenomic landscapes for many preclinical cancer models, such as cell lines, organoids and patient-derived tumor xenograft (PDX) mouse models. Various computational methods for drug response prediction have been developed and achieved satisfactory accuracy by cross-validation in one large preclinical dataset [1-9]. However, the application of these methods to predict the therapeutic effects of patients often results in obvious performance decreases [8]. The clinical applicability of prediction methods trained on preclinical cancer models remains an important challenge.

The discrepancies between preclinical cancer models and patient tumors stem from systematic differences in their biological characteristics, such as their microenvironments and cellular heterogeneity, as well as differences in experimental technologies, such as drug sensitivity measurements and gene expression profiling. To address these discrepancies, initial endeavors focused on aligning the transcriptional profiles of cell lines with those of patient tumors using statistical frameworks. Subsequently, drug response was predicted through regression, machine learning or deep learning methodologies. For example, combat, which is the commonly used batch-effect correction method, is utilized to homogenize cell line and patient datasets [3, 10]; Patient response estimation corrected by interpolation of subspace embeddings (PRECISE) [11] extracts consensus features by subspace alignment via principal component analysis (PCA); Celligner [12] identifies correlated variability that is more prevalent in either the tumor data or the cell line data by contrastive PCA; and Tumor response assessment by nonlinear subspace alignment of cell lines and tumors (TRANSACT) [13] captures the biological signals common to both preclinical models and tumors by nonlinear kernel PCA.

Leveraging the rapid advancements in artificial intelligence (AI) algorithms, recent studies have used domain adaptation (DA) techniques [14-16] and few-short learning technologies [17] to enhance the process of transferring knowledge from preclinical

models to patient tumors. Using large labeled preclinical pharmacogenomic datasets as the source domain and unlabeled clinical datasets as the target domain, the core aim of DA methods is to extract shared components from two domains. Adversarial inductive transfer learning (AITL) [14] employs adversarial inductive DA to adapt not only the input space but also the output space; Task uncertainty guided domain adaptation (TUGDA) [15] is a multi-task DA framework using task uncertainty and covariate-shift to improve the robustness of drug response prediction; and the context-aware deconfounding autoencoder (CODE-AE) [16] is a context-aware deconfounding autoencoder that can distinguish domain shared signals from uninteresting confounders. The classical batch-effect correction methods, PCA-based alignment methods, and most DA methods, such as AITL or TUGDA, require molecular data from both preclinical models and patient tumors during training. Moreover, AITL [14] and Transfer of Cell Line Response Prediction (TCRP) [17] require drug response labels from patients during training, limiting their practical applicability. CODE-AE implements a pretraining scheme utilizing unlabeled patient data, enabling the training process without the need for labeled patients. This feature enhances its suitability for a wider range of real-world application scenarios [16] .
Although significant advances in DA methods for drug response prediction have been reported, several potential shortcomings remain. Firstly, existing DA methods overlook features within domain-private components that may encode tumor microenvironment-related information. Secondly, due to the high heterogeneity of tumors and inadequate training datasets, existing methods may yield suboptimal predictions when the test data are outside the distribution of training data.
A promising way to address these bottlenecks is the utility of synthetically generated data for training, which is being increasingly popular in computer science [18-24]. Typically, it is necessary to disentangle the input features and remix them subsequently at sample or feature level to generate synthetic samples [18, 22, 25, 26]. In bioinformatics and medicine, data synthesis methods have also been utilized in practical applications such as medical image analysis and drug discovery [23]. Gao et al. [24] demonstrated the effectiveness of synthesized X-ray images in hip imaging and lesion segmentations. Kadurin et al. [27] applied a generative adversarial autoencoder, using a latent neuron for the growth inhibition percentage of tumor cells post-treatment to generate candidate molecular which has potential applications in cancer therapy. Polykovskiy et al. [28] proposed an entangled conditional adversarial

autoencoder for de novo drug discovery, reporting efficient *in vitro* activity of a newly discovered molecule. These applications utilized high-quality synthetic data to enhance model robustness, addressed challenges associated with limited annotated datasets, which could also eliminate privacy and ethical concerns from real data. They emphasized the potential and importance of synthetic data in advancing medical research and drug development.

Based on this, we proposed a novel disentangled synthesis transfer network (DiSyn) to improve the generalizability of drug response prediction. Its core idea is to separate features related to drug response and features private for different domains and then synthesize new samples to increase the data size and improve the prediction of label-lacked target domains. Three datasets involving cancer patients and tumor-bearing mice were used to evaluate the performance of DiSyn. The benchmark results showed that DiSyn achieved significant improvement compared to eighteen baseline methods. The contributions of decoupling and synthetic techniques were further demonstrated by visualizing latent features and performing an ablation study. We also applied DiSyn to estimate the drug response of breast cancer patients and explored drug sensitivity-related biomarkers based on the predicted results.

2. Methods

2.1 Problem formulation

First, we introduce the problem formulation used in the article. A ***domain*** includes two components [29]: a feature space $\mathcal{X}$ and a marginal probability distribution $P(X)$, where $X = \{x_1, \dots, x_n\} \in \mathcal{X}$. Given a specific domain $\mathcal{D} = \{\mathcal{X}, P(X)\}$, a ***task*** also includes two components: a label space $\mathcal{Y}$ and an objective predictive function $f(\cdot)$. In this article, we refer to the labeled dataset from the source domain as $D^S = \{(x_i^s, y_i^s)\}_{N_s}$, where $x_i^s \in \mathcal{X}^s$ is denoted as the ith instance of N_s gene expression profiles from the cell line and $y_i^s \in \mathcal{Y}^s$ is its drug response label. Similarly, we denote the unlabeled dataset from the target domain as $D^T = \{x_i^t\}_{N_t}$, where $x_i^t \in \mathcal{X}^T$ is the ith instance of N_T gene expression profiles from patients.

2.2 DiSyn architecture

DiSyn aims to construct drug response prediction models from source domain(well-labeled) and apply them to target domain(label-lacked), specifically from cancer cell

lines to clinical patients. The core strategy of DiSyn involves isolating drug response-related features from the input and enhancing prediction accuracy in target domains by generating and incorporating synthetic samples. These processes mutually iterate during training, ensuring that the model continually improves its performance in domains with limited or no labels (Fig. 1A). We refer to the factors that are relevant for specific drug responses as drug-specific features and those that are not relevant for specific drug responses as unspecific features [30, 31].

2.2.1 Task-specific training with pretraining-based DA

DiSyn takes gene expression profiles from the source domain (cancer cell lines) and target domain (e.g., patient tumors) as input, using a domain separation network [32] as a backbone to perform DA. Input samples from each domain are partitioned into two subspaces: features that are private within each domain and features that are shared across domains. The loss function can be defined as

$$L_{DA} = L_{recon} + \alpha\, L_{difference} - \beta\, L_{similarity} \quad (1.)$$

where α and β are hyperparameters, L_{recon} is used for the reconstruction of the input data, $L_{difference}$ is an orthogonality constraint to push the common and private features apart, and $L_{similarity}$ is designed to encourage the common features from two domains to be as close as possible. $L_{similarity}$ can be either the maximum mean discrepancy [33] (MMD) or adversarial loss [34] (ADV). The detailed implementations are described in the supplementary methods.

Next, DiSyn used a drug-specific training phase to learn the response information specific to each drug. The activation of $E_{drug\_specific}$ is trained to predict the drug response to a given the cell-line input. The loss is defined by

$$L_{response} = -\sum_{i=1}^{N_s} y_i \log \hat{y}_i \quad (2.)$$

where y_i and $\hat{y}_i$ are the real and predicted drug response labels, respectively, of sample x_i^s in the source domain. Through this process, DiSyn obtains drug-specific features from the encoder $\boldsymbol{E}_{drug\_specific}$.

2.2.2 Disentanglement and synthesis iteratively update the task-specific encoder

In the disentanglement step, $\boldsymbol{E}_{drug\_specific}$ is kept fixed to train two encoders

$\boldsymbol{E}^{s}_{drug\_unspecific}$ and $\boldsymbol{E}^{t}_{drug\_unspecific}$, which extract information unrelated to specific drug responses from the source and target domains. The loss function can be denoted as

$$L = L_{recon} + \sigma\, L_{difference} \tag{3.}$$

where σ is a hyperparameter and L_{recon} and $L_{difference}$ are consistent with earlier definitions. This step obtains unspecific information from encoders $\boldsymbol{E}^{s}_{drug\_unspecific}$ and $\boldsymbol{E}^{t}_{drug\_unspecific}$.

After dividing the input space into drug-specific features and nonspecific features, we can employ attribute swapping and decoupling synthesis to generate synthetic data $\left(x_i^{syn},\ y_i^{syn}\right)$. As shown in Fig. 1B, the drug response labels and drug-specific features of synthetic data originate from the source domain, while the unspecific features originate from the target domain.

$$x_i^{syn} = \boldsymbol{D}\{\mathrm{concat}(\left[\boldsymbol{E}_{drug\_specific}(x_i^{s}), \boldsymbol{E}^{t}_{drug\_unspecific}(x_j^{t})\right])\} \tag{4.}$$

$$y_i^{syn} = y_i^{s} \tag{5.}$$

where $(x_i^{s},\ y_i^{s})$ is a labeled cell line sample for the specific drug and x_j^{t} is an unlabeled target sample. $\boldsymbol{D}$ is the decoder, and $\mathrm{concat}(\cdot)$ is the vector concatenation operation.

Then, we can further update the unspecific feature extractors with synthetic data. The loss function is composed of three parts as follows:

$$L = L_{recon} + \sigma\, L_{difference} - \lambda * L_{adv\_class} \tag{6.}$$

$$L_{recon} = \frac{1}{N_s}\sum_{\mathrm{i}=1}^{N_s}\|x_i - \hat{x}_i\| + \frac{1}{N_{syn}}\sum_{\mathrm{i}=1}^{N_{syn}}\|x_i - \hat{x}_i\| \tag{7.}$$

$$L_{difference} = \left\|trans(H_c^{s}) \cdot H_p^{s}\right\|_F^2 + \left\|trans(H_c^{syn}) \cdot H_p^{\mathrm{syn}}\right\|_F^2 \tag{8.}$$

$$L_{adv\_class} = -\sum_{i=1}^{N_s+N_{syn}} y_i \log \hat{y}_i \tag{9.}$$

where σ and λ are hyperparameters, N_s and N_{syn} are the number of samples in the source domain and synthetic data, y_i is the ground truth drug response label of sample x_i, $\hat{y}_i$ is the corresponding predicted label, $\|\cdot\|_F^2$ is the squared Frobenius norm, and $trans(\cdot)$ is the matrix transposition. H_c^{s} and H_p^{s} are matrices whose rows are drug-specific and drug-unspecfic latent features of source samples obtained

from the encoders $\boldsymbol{E}_{drug\_specific}$ and $\boldsymbol{E}^{s}_{drug\_unspecific}$. Similarly, H_c^{syn} and H_p^{syn} are matrices with latent features of synthetic samples.

The synthetic data generated in the disentanglement step can be further involved in the task-specific training stage, which slightly modifies the loss function $L_{respnose}$ as follows:

$$L_{respnose} = -\sum_{i=1}^{N_s+N_{syn}} y_i \log \hat{y}_i \qquad (10.)$$

Source code that produced the findings of the study, including all main and supplemental figures, is available at https://github.com/LiHongCSBLab/DiSyn.

2.3 Experiment details

2.3.1 Datasets

We used the following datasets in our experiments (Please refer to Table S1 for more information):

• The dataset from the Genomics of Drug Sensitivity in Cancer (GDSC) [35] consisted of gene expression profiles of 966 cancer cell lines from various cancer types experimented with 282 drugs.

• The dataset from The Cancer Genome Atlas(TCGA) [36] comprised transcriptomic information from a total of 10237 patients with various types of cancer. Part of these individuals received drug treatments, among which 16 drugs were in common with GDSC, and their therapeutic responses were documented using the Response Evaluation Criteria in Solid Tumors (RECIST).

• The dataset from the Novartis Institutes for Biomedical Research PDX Encyclopedia (NIBR PDXE) [37] involved 176 samples from different cancer types. Some of these were tested with 8 drugs that were also used in the GDSC.

• The dataset from I-SPY 2 [38] consisted of 988 transcriptomic profiles from patients diagnosed with breast cancer. A portion of them received treatment with Paclitaxel.

The gene expression profiles from GDSC [35] and tumor profiles from target datasets were used for pretraining. The expression profile of each gene was scaled by Max-Min normalization. We employed a published method [39] to select the top 1000 highly variable genes (HVGs) among the samples. HVGs from cell lines and patients were merged as input variables. The drug response AUC values of the cell lines were

obtained from the GDSC and were converted to binary labels by taking the median of each drug as the cutoff. The labeled GDSC cell lines were used as the source domain. Another three datasets (TCGA [36], I-SPY 2 [38] and NIBR PDXE [37]) were taken as the target domain, and their response labels were reserved for test datasets. Considering the sample size, we selected 23 drugs for testing, with 16 drugs from TCGA, eight drugs from PDX, and one drug from I-SPY 2 [38]. Patients from TCGA and I-SPY 2 data were partitioned into two subgroups based on RECIST: positive samples that had a partial or complete response and negative samples that had progressive or stable disease [40]. PDX mice were partitioned using the same criteria based on the modified RECIST criteria [37].

The original GDSC, PDXE, TCGA and I-SPY 2 are all publicly available datasets. GDSC were downloaded from the portal website (https://www.cancerrxgene.org/). PDXE data were downloaded as supplementary information of the original paper [37]. TCGA data were obtained from the official website (https://portal.gdc.cancer.gov/). And I-SPY 2 data were downloaded from the GEO website (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE194040) and supplemental material of the original paper [38].

2.3.2 Training and evaluation

We investigated the classification performance of DiSyn in three publicly available datasets: TCGA [36], I-SPY 2 [38], and NIBR PDXE [37] datasets. First, DiSyn was trained using a labeled cell line and unlabeled data from each target dataset to build separate classification models for each drug. To assess the robustness of DiSyn, we conducted 5 independent repeated experiments with different random seeds. In each experiment, we employed a five-fold cross-validation and reported the average performance of all the test folds. Source samples (cell lines from the GDSC) were partitioned into five folds in a stratified fashion by drug sensitivity for cross-validation. With each runtime, we trained the model using 4/5 of the folds, while the remaining 1/5 was utilized as the validation set to search for the optimal hyperparameters (Table S2). The area under the receiver operating characteristic curve (AUROC) on the validation set was used as the stopping criterion because it is stable to changes in label distributions [41] and has been evaluated across various classification methods [15, 16]. For models with optimal parameters, we reported the average performance of all testing folds for multiple metrics, including AUROC, area

under the precision–recall curve (AUPRC), accuracy (ACC), F1 score and average precision score (APS).

2.3.3 Baseline methods

Baseline methods refer to a set of established methods developed by previous researchers that serve as a point of comparison for evaluating the performance of our newly developed method DiSyn. Specifically, in the TCGA dataset, we compared DiSyn with various baseline methods for cross-dataset drug response prediction, including four classical machine learning approaches (random forest [42] , support vector machine (SVM) [43] , elastic net [44] and adaptive boosting (AdaBoost) [45]), nine deep learning models without considering domain differences (clinical response prediction using deep neural network (CRDNN) [3], DrugCell [7], prediction of anticancer compound sensitivity with multimodal attention-based neural networks (PaccMann) [5], variational autoencoder with elastic net strategy (VAEN) [2], multi-omics late integration method based on deep neural networks (MOLI) [4], four twin graph neural networks for drug response prediction and a similarity augmentation module (TGSA) [6] variants with different omics data as input), and five state-of-the-art methods addressing domain differences (AITL, TUGDA, PRECISE, TRANSACT, CODE-AE).

All baseline methods were retrained using the GDSC cell line dataset with drug response labels. The unlabeled target datasets were also used for unsupervised pretraining. Due to the inductive hypothesis of AITL [14], the drug response labels of the test dataset were also used for supervised training. To ensure fairness, the setting of five experiment replicates and five-fold split settings for cross-validation were consistent across all baseline models and DiSyn.

For classical machine learning approaches, we used the Python package scikit-learn to select the optimal parameters and train the models. For other methods, the training codes were downloaded from GitHub. For TRANSACT, we used the optimal hyperparameters mentioned in the article, and for most other models that have hyperparameter ranges in their projects, we employed these ranges for grid search to obtain the best hyperparameters. The specific parameter range we used for the grid search can be found in Table S2.

2.3.4 Cell culture

The breast cell line (MDA-MB-231, obtained from American Type Culture Collection, Virginia, USA) was maintained in complete medium supplemented with 12% fetal bovine serum (FBS, Shanghai XP Biomed Ltd, Shanghai, China), 100 units/mL penicillin (New Cell & Molecular Biotech Co., Ltd, Jiangsu, China), and 100 mg/mL streptomycin (New Cell & Molecular Biotech Co., Ltd, Jiangsu, China) at 37 °C, 5% CO2 saturated humidity.

2.3.5 Cell viability assay

The cell viability was detected by Cell Counting Kit-8 (CCK-8) assays according to manufacturer's instructions. Briefly, all cells were seeded in 96-wells plates at a density of 2000 cells per well. The corresponding concentrations of 5-Fluorouracil (5-FU, Sigma-Aldrich, St. Louis, MO, USA) and Gefitinib (Sigma-Aldrich, St. Louis, MO, USA) were added to continue the incubation for 6, 12, 24, 48, 72, 96 h, respectively. After treatment, the medium was removed and replaced with 110 μL of 10% CCK-8 solution (New Cell & Molecular Biotech Co., Ltd, Jiangsu, China) in serum-free medium, and the cells were incubated at 37°C, 5% CO2 saturated humidity for 1.5 h. Then we detected an optical density 450 (OD450) of each well using Microplate Reader (BioTek, Biotek Winooski, Vermont, USA).

3. Results

3.1 DiSyn workflow

Inspired by the domain separation network (DSN) [32], DiSyn adapts a novel disentangled transfer network that contains one common encoder, two private encoders, one decoder, and two adversarial discriminators (Fig. 1B). The unlabeled gene expression profiles from the source and target domains are inputted into the network for pretraining, which allows the common encoder to extract features shared between two domains. Next, task-specific training of the network is performed using the labeled samples from the source domain to update the common encoder to a drug-specific encoder that captures drug response-related features. Then, the drug-specific features from the source domain and unspecific features from the target domain are combined by the decoder to synthesize new labeled samples, which are further involved in the disentangling process (Fig. 1C). By iterating between disentanglement

and synthesis, DiSyn progressively extracts effective features and enhances its predictive performance.

To evaluate the effectiveness of DiSyn on drug response, the model was implemented on GDSC, which is a large-scale pharmacogenomic dataset of human cancer cell lines, and three target datasets: TCGA [36] (patients, sixteen drugs), I-SPY 2 [38] (patients, one drug) and NIBR PDXE [37] (mice, eight drugs) (Table S1). The expression profiles of 966 cell lines from the GDSC dataset [35] and corresponding samples from the target dataset [46] were used for unsupervised pretraining. Then, cell lines with drug response labels were used as the source domain to predict labels of target domain samples from another three datasets.

3.2 DiSyn captures drug-specific features associated with therapeutic responses

Taking GDSC dataset and TCGA cohorts as examples, we used t-distributed stochastic neighbor embedding (*t*-SNE) plots to visualize the low-dimensional representations of the feature spaces of the source and target samples during DiSyn training. At the beginning, the original gene expression profiles were clearly separated according to the *t*-SNE plot (Fig. 2A); such difference has also been observed in many studies [10, 12, 16]. After DiSyn pretraining, the latent features encoded from the shared encoder exhibited substantial overlap between the TCGA and GDSC samples (Fig. 2B), confirming the efficiency of DiSyn in domain alignment. Then, the drug-specific encoder and unspecific encoders were derived from the first fine-tuning and disentanglement training phase. Taking paclitaxel as an example, the latent features obtained from the paclitaxel-specific encoder separated responders and non-responders better than the latent features obtained from the drug-unspecific encoders. (Figs. 2C and D, distance = 9.78 and distance = 4.76. The distance was calculated based on the centroids of the responders and nonresponders respectively.) Moreover, the addition of synthetic data enlarged the gap between responders and non-responders by improving the ability of the paclitaxel-specific encoder (Fig. 2E, distance = 26.58). The latent features encoded from the nonspecific drug encoders clearly indicate the sample domains instead of the drug response. (Figs. 2D and F) The results indicated that DiSyn gradually learned the shared drug-specific features between the source and target domains over the course of training.

3.3 DiSyn enhances the prediction accuracy in clinical settings

To evaluate the model's transferability in the clinical context, we conducted comprehensive experiments using patient data from TCGA as the target domain. Comparisons between DiSyn and 18 baseline models were conducted across 16 drugs using five evaluation metrics: AUROC, AUPRC, APS, ACC and F1 score (Figs. 3A-E, and Tables S2-7). The results indicated that DiSyn outperformed the baseline models across most of the metrics. We ranked all methods based on their performance on each drug. The average rankings of DiSyn across the 16 drugs were 1.19, 2.81, 2.63, 2.50, and 4.75 for the AUROC, AUPRC, APS, ACC and F1 scores, respectively (Fig. 3F). These rankings were significantly better than those of all baseline models ($P < 0.05$). When the AUROC was used as the evaluation index, DiSyn achieved the best results, ranking first for 14 (87.5%) drugs. The average AUROC of DiSyn was 0.775, representing a 15.36% improvement compared to the best baseline model (Fig. 3A). For five independent experiments with different random seeds, the variations in the evaluation metrics are small, suggesting that DiSyn could offer robust predictions via hyperparameter optimization.

It is worth mentioning that variations exist among different drugs regardless of the prediction method employed. The DiSyn predictions consistently demonstrated superior performance for vinorelbine, tamoxifen, bleomycin, etoposide, and vinblastine, with multiple metric values exceeding 0.8 (Figs. 3A-E). The effects of temozolomide and sorafenib are difficult to predict by any method. DiSyn, along with other state-of-the-art methods addressing domain differences (AITL, TUGDA, PRECISE, TRANSACT, and CODE-AE), obtained high AUROC and ACC but low AUPRC, APS and F1, while other deep learning methods without considering domain differences (CRDNN, DrugCell, PaccMann, VAEN, MOLI, and TGSA) obtained high AUPRC, APS and F1 but low AUROC and ACC. The reasons are multifaceted, potentially involving imbalanced sample sizes in the target domain or disparities in cancer types between the labeled source and target samples.

We further evaluated DiSyn using another independent dataset, I-SPY 2 [38], in which two drugs, paclitaxel and MK-2206, were labeled in the GDSC cell line dataset. DiSyn was used to construct prediction models for paclitaxel and MK-2206. For 179 patients treated with paclitaxel, DiSyn's prediction achieved an improvement of

10.06% in the average AUROC score compared to the second-ranked baseline method (Fig. S1). The other 60 patients were treated with a combination of paclitaxel and MK-2206. We hypothesized that if the single-drug models predict a patient to be effective for both paclitaxel and MK-2206, this patient will be more likely to be truly effective when treated with combination therapy. According to this hypothesis, when patients who ranked in the top 20, 30, or 40 according to the two single-drug models were considered to be predicted as responders to combination therapy, the true positive rates were 100% (4/4), 87.5% (7/8), and 66.7% (8/12), respectively. This finding suggested that single-drug models built with DiSyn could also provide an insight into drug combinations.

3.4 DiSyn improves the prediction accuracy of PDX models

We utilized the NIBR PDXE data [37] as a target domain to evaluate the transferability of DiSyn to patient-derived xenograft models. Eight drugs shared in the PDX and GDSC datasets were selected for evaluation. We compared DiSyn with eight representative baseline models (Fig. 4, and Tables S8-12). Consistent with the results of the previous experiments, DiSyn showed a convincing predictive ability to transfer drug response information from cancer cell lines to patient-derived tumors, obtaining a 5.44% improvement in the average AUROC compared to that of the best baseline model across the eight drugs tested. With AUPRC and APS as metrics, DiSyn also demonstrated notable improvements of 12.17% and 10.73%, respectively.

The overall improvement of DiSyn in the PDX dataset was slightly lower than that in the TCGA cohort, and DiSyn did not outperform baseline models consistently. The reason may be that the pretraining stage of DiSyn used the expression profiles of 176 mice and 10237 patients when predicting the PDX and TCGA datasets, respectively. The relatively small sample size of the PDX model may have impeded the learning capacity of the DiSyn pretraining step. Accumulations of larger-scale cancer data may further improve DiSyn's performance.

3.5 Ablation study

To further investigate the performance impact of different model components, we evaluated the performance of DiSyn with different variants, including different regularization strategies, different sample sizes and different numbers of iterations of

data synthesis. The distribution of AUROC values for 16 drugs in the TCGA dataset was used as the comparative index. Consistent with prior research [32], the adversarial regularization method(ADV) obtained a larger performance gain than the maximum mean discrepancy (MMD) regularization or the base models without adding the domain similarity loss (Fig. 5A).

In addition, the impacts of different synthetic data sizes and numbers of iterations on performance were evaluated. As illustrated in Fig. 5B, there was an evident improvement in terms of the AUROC performance, particularly when the synthetic data initially participated compared to the training without data synthesis ($P < 0.001$). A one-fold synthetic dataset means adding the same size of synthetic samples as the labeled cell lines for a given drug. Adding one-fold synthetic samples for one-round, two-round or three-round iterative training resulted in average AUROC increases of 19.65%, 21.78% or 14.17%, respectively. The best performance was achieved by conducting at least two iterations on single-fold or double-fold synthetic samples or just one iteration on triple-fold synthetic samples (Fig. 5B). These results indicate that, compared to the performance gained from different regularization strategies, the application of synthetic data has a more pronounced impact on the model's performance, which notably enhances the predictive capabilities.

3.6 Estimated atlas of drug responses for human breast cancer

Large-scale measurements of drug responses in human patients are extremely difficult. With the effective prediction method DiSyn, we can estimate drug response for any patient with expression profiles. The data of 1,089 breast cancer patients in the TCGA dataset were used as an example to generate an estimated response atlas for 16 drugs. The predictive response values were ranked for each drug, and the samples with the top 5% or bottom 5% of the predictive values were regarded as responders or nonresponders, respectively. A total of 46.8% of patients responded to at least one drug. Different drugs are effective for distinct patients, with some overlap, which further demonstrates the importance of personalized treatment (Fig. 6A). The top 3 drugs with the highest proportions of responders among patients with breast cancer (BCRs) were 5-FU, Docetaxel and Doxorubicin. All of these drugs are FDA-approved for breast cancer, indicating that the prediction of DiSyn is reliable.

To determine the association between the predicted drug response and gene

expression, 5-FU was used for further analysis. We identified 385 differentially expressed genes (DEGs) (adjusted *P* value < 0.1) between responsive and nonresponsive patients. These genes could be used to divide breast cancer patients into two groups with different responses to 5-FU (Fig. 6B). DEGs were significantly enriched in pathways of interest, such as focal adhesion and epithelial cell differentiation, and significantly overlapped with genes whose expression was upregulated when epidermal growth factor receptor (EGFR), transforming growth factor beta (TGFB) or rapidly accelerated fibrosarcoma (RAF) was overexpressed in epithelial cell lines of breast cancer (adjusted *P* value < 0.0001). Compared with responders, nonresponders had stronger EGFR pathway activity and greater EGFR gene expression (Figs. 6C and D, $P < 0.0001$). This led to the hypothesis that inhibiting the EGFR pathway could improve the therapeutic effects of 5-FU in these original nonresponders. Similar results and hypotheses have been reported in a previous experimental study. EGFR is significantly elevated in 5-FU-resistant tumors compared with normal tissues and 5-FU-sensitive tumors, and targeting EGFR sensitizes 5-FU-resistant cancer cells [47]. This consistent example demonstrated that DiSyn could be used to identify drug sensitivity-related markers and provide clues for potential combination therapy.

To evaluate the therapeutic effects of the drug combination in breast cancer, we selected the EGFR high-expressing MDA-MB-231 cell line for *in vitro* viability assays (see "Methods" section for details). The results demonstrated that the combination of Gefitinib (an EGFR inhibitor) and 5-FU inhibited cell viability with greater efficacy compared to either agent alone (Fig. 6E), suggesting the effectiveness of our methods in identifying potential drug combinations. Notably, MDA-MB-231 is also a typical triple-negative breast cancer (TNBC) cell line, which is a subtype with the worst prognosis and limited treatment options in breast cancer cases [48]. These findings provide a promising strategy for TNBC therapy, warranting further studies to fully explore its potential.

4. Discussion

In this study, we developed a new deep learning method, DiSyn, to enhance the generalization performance of drug response prediction beyond cell line datasets. The accuracy and robustness of DiSyn were demonstrated by a comprehensive benchmark

using 3 test datasets against 18 baseline models. Additionally, we estimated the therapeutic effects of 16 drugs for more than 1,000 human breast tumors, which is valuable for better understanding the heterogeneity of cancer treatment.

The most important improvement of DiSyn is training using a mix of real and synthetic data. DiSyn disentangled drug-specific and unspecific features from the input data. Then, utilizing the drug-unspecific representations from the target domain, the model could generate synthetic data, which provide examples and variations that might not be explicitly represented in the original dataset, to be further involved in model training. Inspired by Cao et al. [22], our model alternates between disentanglement and drug-specific training stages, progressively enhancing the predictive performance throughout these iterations. The initial motivation behind this iterative process is that an effective feature extractor can generate more accurate synthetic data, and accurate synthetic data can further promote extractor training. The ablation study demonstrated that DiSyn could not only transfer knowledge within the decoupled subspace but also leverage synthetic data to further promote transfer capability. Through the innovative iterative approach, DiSyn achieved a significant performance improvement compared to 18 baseline methods.

Previously, we and others have explored the effects of different omics data on model performance, such as copy number variations, genomic mutations and gene expression [8, 9, 49]. The expression profiles usually exhibit similar or better performance than other omics; therefore, we did not discuss other omics methods in this study. Like most deep learning methods [50], DiSyn training relies on hyperparameter optimization. Using the AUROC as the optimal criterion results in a high AUROC, as expected, but does not guarantee the performance of other metrics. It is recommended that algorithm developers select the most relevant evaluation indicator to set the stopping criterion for optimization. Additionally, the complex architecture of DiSyn limits its interpretability. Although we used *t*-SNE to visualize latent features and used breast cancer as an example to identify candidate biomarkers of 5-FU, these results are preliminary and require more rigorous investigations.

In conclusion, DiSyn offers a powerful and reliable framework for drug response prediction from preclinical cancer models to clinical applications. Furthermore, the application of DiSyn to large patient cohorts may reveal new biomarkers related to drug responses.

5. Conclusion

Our study presents DiSyn, a deep learning model that achieves remarkable improvements in drug response prediction for cancer. By integrating disentangled feature learning and data synthesis, DiSyn outperforms existing methods, demonstrating its robustness and potential for clinical application. The model's success in predicting drug responses across diverse datasets underscores its reliability and the promise of artificial intelligence in advancing personalized medicine.

References

1. F. Feng, B. Shen, X. Mou, et al., Large-scale pharmacogenomic studies and drug response prediction for personalized cancer medicine, J. Genet. Genomics 48 (2021) 540-551.

2. P. Jia, R. Hu, G. Pei, et al., Deep generative neural network for accurate drug response imputation, Nat. Commun. 12 (2021), 1740.

3. T. Sakellaropoulos, K. Vougas, S. Narang, et al., A deep learning framework for predicting response to therapy in cancer, Cell Rep. 29 (2019) 3367-3373.

4. S.-N. Hossein, Z. Olga, C.C. Colins, et al., MOLI: Multi-omics late integration with deep neural networks for drug response prediction, Bioinformmatics 35 (2019) i501-i509.

5. M. Manica, A. Oskooei, J. Born, et al., Toward explainable anticancer compound sensitivity prediction via multimodal attention-based convolutional encoders, Mol. Pharm. 16 (2019) 4797-4806.

6. Y. Zhu, Z. Ouyang, W. Chen, et al., TGSA: Protein-protein association-based twin graph neural networks for drug response prediction with similarity augmentation, Bioinformatics 38 (2022) 461-468.

7. B.M. Kuenzi, J. Park, S.H. Fong, et al., Predicting drug response and synergy using a deep learning model of human cancer cells, Cancer Cell 38 (2020) 672-684.

8. B. Shen, F. Feng, K. Li, et al., A systematic assessment of deep learning methods for drug response prediction: From *in vitro* to clinical applications, Brief. Bioinform. 24 (2023), bbac605.

9. J.C. Costello, L.M. Heiser, E. Georgii, et al., A community effort to assess and improve drug sensitivity prediction algorithms, Nat. Biotechnol. 32 (2014) 1202–1212.

10. P. Geeleher, N.J. Cox, R.S. Huang, Clinical drug response can be predicted using baseline gene expression levels and *in vitro* drug sensitivity in cell lines, Genome Biol. 15 (2014), R47.

11. S. Mourragui, M. Loog, M.A. van de Wiel, et al., PRECISE: A domain adaptation approach to transfer predictors of drug response from pre-clinical models to tumors, Bioinformatics 35 (2019) i510-i519.

12. A. Warren, Y. Chen, A. Jones, et al., Global computational alignment of tumor and cell line transcriptional profiles, Nat. Commun. 12 (2021), 22.

13. S.M.C. Mourragui, M. Loog, D.J. Vis, et al., Predicting patient response with models trained on cell lines and patient-derived xenografts by nonlinear transfer learning, Proc. Natl. Acad.

Sci. U. S. A. 118 (2021), e2106682118.

14. H. Sharifi-Noghabi, S. Peng, O. Zolotareva, et al., AITL: Adversarial Inductive Transfer Learning with input and output space adaptation for pharmacogenomics, Bioinformatics 36 (2020) i380-i388.

15. R. Peres da Silva, C. Suphavilai, N. Nagarajan, TUGDA: Task uncertainty guided domain adaptation for robust generalization of cancer drug response prediction from *in vitro* to *in vivo* settings, Bioinformatics 37 (2021) i76-i83.

16. D. He, Q. Liu, Y. Wu, et al., A context-aware deconfounding autoencoder for robust prediction of personalized clinical drug response from cell-line compound screening, Nat. Mach. Intell. 4 (2022) 879-892.

17. J. Ma, S.H. Fong, Y. Luo, et al., Few-shot learning creates predictive models of drug response that translate from high-throughput screens to individual patients, Nat. Cancer 2 (2021) 233-244.

18. S. Lee, S. Cho, S. Im, DRANet: Disentangling representation and adaptation networks for unsupervised cross-domain adaptation, 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA, IEEE, (2021) 15247-15256.

19. W. Tranheden, V. Olsson, J. Pinto, et al., DACS: Domain adaptation via cross-domain mixed sampling, 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). January 3-8, 2021, Waikoloa, HI, USA. IEEE, (2021) 1378-1388.

20. S. Sankaranarayanan, Y. Balaji, A. Jain, et al., Learning from synthetic data: Addressing domain shift for semantic segmentation, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, (2018) 3752-3761.

21. Y. Chen, W. Li, X. Chen, et al., Learning semantic segmentation from synthetic data: A geometrically guided input-output adaptation approach, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 15-20, 2019, Long Beach, CA, USA. IEEE, (2019) 1841-1850.

22. J. Cao, O. Katzir, P. Jiang, et al., DiDA: Iterative boosting of disentangled synthesis and domain adaptation, 2021 11th International Conference on Information Technology in Medicine and Education (ITME). November 19-21, 2021, Wuyishan, Fujian, China. IEEE, (2021) 201-208.

23. S.I. Nikolenko, Synthetic data outside computer vision. S.I. Nikolenko, Synthetic data for

deep learning, Springer Optimization and Its Applications, Vol 174, Springer, Cham, 2021, pp. 217-226.

24. C. Gao, B.D. Killeen, Y. Hu, et al., Synthetic data accelerates the development of generalizable learning-based algorithms for X-ray image analysis, Nat. Mach. Intell. 5 (2023) 294-308.

25. X. Peng, Z. Huang, X. Sun, et al., Domain agnostic learning with disentangled representations, Proceedings of the 36th International Conference on Machine Learning (ICML). June 9-15, 2019, Long Beach, California, USA. PMLR, (2019) 5102-5112.

26. W. Young Jhoo, J.P. Heo, Collaborative learning with disentangled features for zero-shot domain adaptation, 2021 IEEE/CVF International Conference on Computer Vision (ICCV). October 10-17, 2021, Montreal, QC, Canada. IEEE, (2021) 8876-8885.

27. A. Kadurin, A. Aliper, A. Kazennov, et al., The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology, Oncotarget 8 (2017) 10883-10890.

28. D. Polykovskiy, A. Zhebrak, D. Vetrov, et al., Entangled conditional adversarial autoencoder for *de novo* drug discovery, Mol. Pharm. 15 (2018) 4398-4405.

29. S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (2010) 1345-1359.

30. N. Hadad, L. Wolf, M. Shahar, A two-step disentanglement method, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 18-23, 2018, Salt Lake City, UT, USA. IEEE, (2018) 772-780.

31. M. Mathieu, J. Zhao, P. Sprechmann, et al., Disentangling factors of variation in deep representations using adversarial training, 30th Conference on Neural Information Processing Systems. December 5-10, 2016, Barcelona, Spain. NeurIPS, (2016).

32. K. Bousmalis, G. Trigeorgis, N. Silberman, et al., Domain separation networks, 30th Conference on Neural Information Processing Systems. December 5-10, 2016, Barcelona, Spain. NeurIPS, (2016) 343 - 351.

33. A. Gretton, K.M. Borgwardt, M.J. Rasch, et al., A kernel two-sample test, J. Mach. Learn. Res. 13 (2012) 723-773.

34. Y. Ganin, V. Lempitsky, Unsupervised domain adaptation by backpropagation, 32nd International Conference on Machine Learning. July 6-11, 2015, Lille, France. PMLR, (2015) 1180-1189.

35. M.J. Garnett, E.J. Edelman, S.J. Heidorn, et al., Systematic identification of genomic

markers of drug sensitivity in cancer cells, Nature 483 (2012) 570-575.

36. C. Hutter, J.C. Zenklusen, The cancer genome atlas: Creating lasting value beyond its data, Cell 173 (2018) 283-285.

37. H. Gao, J.M. Korn, S. Ferretti, et al., High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response, Nat. Med. 21 (2015) 1318-1325.

38. D.M. Wolf, C. Yau, J. Wulfkuhle, et al., Redefining breast cancer subtypes to guide treatment prioritization and maximize response: Predictive biomarkers across 10 cancer therapies, Cancer Cell 40 (2022) 609-623.

39. E.A. Clayton, T.A. Pujol, J.F. McDonald, et al., Leveraging TCGA gene expression data to build predictive models for cancer drug response, BMC Bioinformatics 21 (2020), 364.

40. Z. Ding, S. Zu, J. Gu, Evaluating the molecule-based prediction of clinical drug responses in cancer, Bioinformatics 32 (2016) 2891-2895.

41. T. Fawcett, An introduction to ROC analysis, Pattern Recognit. Lett. 27 (2006) 861-874.

42. L. Breiman, Random Forests, Mach. Learning 45 (2001) 5–32.

43. C. Cortes, V. Vapnik, Support-vector network, Mach. Learn 20 (1995) 273–297.

44. H. Zou, T. Hastie, Regularization and Variable Selection Via the Elastic Net, J. R. Stat. Soc. B 2 (2005) 301–320.

45. Y. Freund, R.E. Schapire, A desicion-theoretic generalization of on-line learning and an application to boosting, J. Comput. Syst. Sci. 1 (1997) 119-139.

46. J. Chen, L. Zhang, A survey and systematic assessment of computational methods for drug response prediction, Brief. Bioinform. 22 (2021) 232-246.

47. S. Gao, S. Ren, Y. Liu, et al., Targeting EGFR sensitizes *5-Fu*-resistant colon cancer cells through modification of the lncRNA-FGD5-AS1-miR-330-3p-Hexokinase 2 axis, Mol. Ther. Oncolytics 23 (2021) 14-25.

48. P. Zagami, L.A. Carey, Triple negative breast cancer: Pitfalls and progress, NPJ Breast Cancer 8 (2022), 95.

49. K.A. Hoadley, C. Yau, T. Hinoue, et al., Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer, Cell 173 (2018) 291-304.

50. Y. Chen, L. Zhang, How much can deep learning improve prediction of the responses to drugs in cancer cell lines? Brief. Bioinform. 23 (2022), bbab378.

Figure Captions

Fig. 1. Architecture illustration of disentangled synthesis transfer network (DiSyn). (A) Illustration of DiSyn's workflow. Drug response-related features are separated from the input and then synthetic samples are utilized to help improving prediction accuracy in label-lacked target domains. These two processes mutually iterate in the subsequent training. (B) Architecture of DiSyn. The model comprises drug-specific encoder $\boldsymbol{E}_{drug-specific}$ sharing across domains, drug unspecific encoder on both source domain $\boldsymbol{E}^{\boldsymbol{s}}_{drug-unspecific}$ and target domain $\boldsymbol{E}^{\boldsymbol{t}}_{drug-unspecific}$ respectively, one decoder, and two adversarial discriminators. The training process consists of pretraining, drug-specific training, and disentanglement stages. For more details, please refer to the DiSyn Architecture section. (C) Illustration of the data synthesis process. We specifically showcased the data synthesis process during the model training. The drug-specific features from source samples and the drug-unspecific features from target samples are used for recombination to generate novel synthetic samples, and the label of the synthetic samples was derived from the source samples.

Fig. 2. Visualization of the original profiles and latent features encoded from disentangled synthesis transfer network (DiSyn). (A) T-distributed stochastic neighbor embedding (*t*-SNE) results generated from the original gene expression features. GDSC: Genomics of Drug Sensitivity in Cancer, TCGA: The Cancer Genome Atlas. (B) Latent features encoded by the pretrained encoder. (C) Latent features encoded by the paclitaxel-specific encoder without synthetic samples. (D) Latent features encoded by unspecific encoders without synthetic samples. (E) Latent features encoded by the paclitaxel-specific encoder with synthetic samples. (F) Latent features encoded by unspecific encoders with synthetic samples.

Fig. 3. Performance comparison of disentangled synthesis transfer network (DiSyn) and baseline models on The Cancer Genome Atlas (TCGA) patient data. (A) Area under the receiver operating characteristic curve (AUROC) score for 16 drugs of DiSyn and baseline models. (B) area under the precision–recall curve (AUPRC). (C) Accuracy (ACC). (D) F1 score. (E) Average precision score (APS). (F) Heatmap showing the average rank of 16 drugs (Tables S3-7). CODE-AE: context-aware deconfounding autoencoder [16], AITL: Adversarial inductive transfer learning [14],

TRANSACT: Tumor response assessment by nonlinear subspace alignment of cell lines and tumors [13], PRECISE: Patient response estimation corrected by interpolation of subspace embeddings [11], AdaBoost: Adaptive Boosting [45], TUGDA: Task uncertainty guided domain adaptation [15], SVM: support vector machine [43], PaccMann: prediction of anticancer compound sensitivity with multimodal attention-based neural networks [5], TGSA: twin graph neural networks for drug response prediction and a similarity augmentation module [6], TGSA_EXP: TGSA variants with only gene expression input, TGSA_CNV: TGSA variants with only gene copy number variation input, TGSA_MUT: TGSA variants with only gene mutation input, CRDNN: clinical response prediction using deep neural network [3], VAEN: variational autoencoder with elastic net strategy [2], MOLI: multi-omics late integration method based on deep neural networks [4].

Fig. 4. Performance comparison of DiSyn and baseline models on PDX mouse data. **(A-H)** are the area under the receiver operating characteristic curve (AUROC) results for Tamoxifen, Ruxolitinib, LGK974, Trametinib, Erlotinib, Paclitaxel, 5-Fluorouracil, and Gemcitabine. Eight representative baseline models were selected to reduce the computational time. The detailed results of the other evaluation metrics are presented in Table S8. CODE-AE: context-aware deconfounding autoencoder [16], TRANSACT: Tumor response assessment by nonlinear subspace alignment of cell lines and tumors [13], PRECISE: Patient response estimation corrected by interpolation of subspace embeddings [11], AdaBoost: Adaptive Boosting [45], TUGDA: Task uncertainty guided domain adaptation [15], SVM: support vector machine [43].

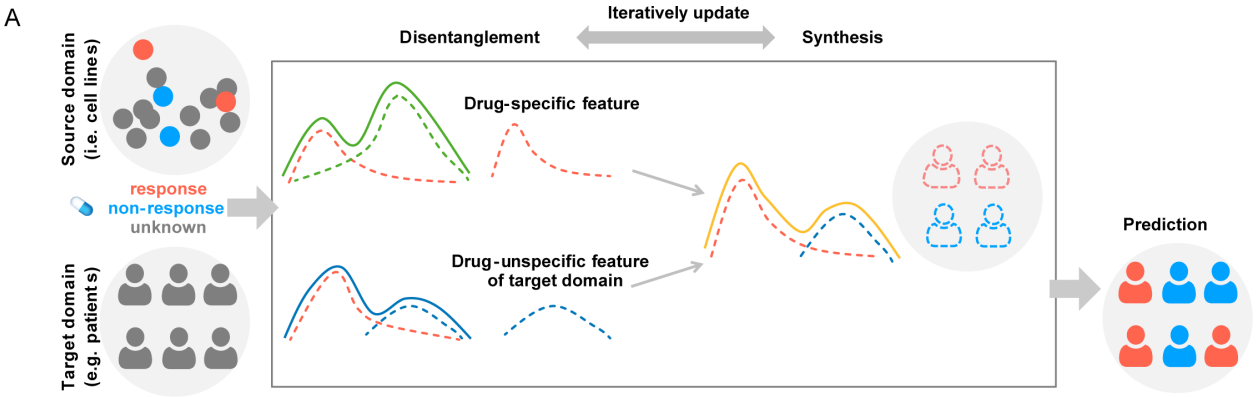
Fig. 5. Ablation studies were conducted on 16 drugs from TCGA dataset to assess the effects of various model components and the inclusion of synthetic data on prediction performance. Specifically, we analyzed how these factors influenced the area under the receiver operating characteristic curve (AUROC) values. In each figure, a single point represents the 5-fold-average AUROC for one drug. (**A**) Distribution of AUROC values under different regularization methods: BASE (base models without domain

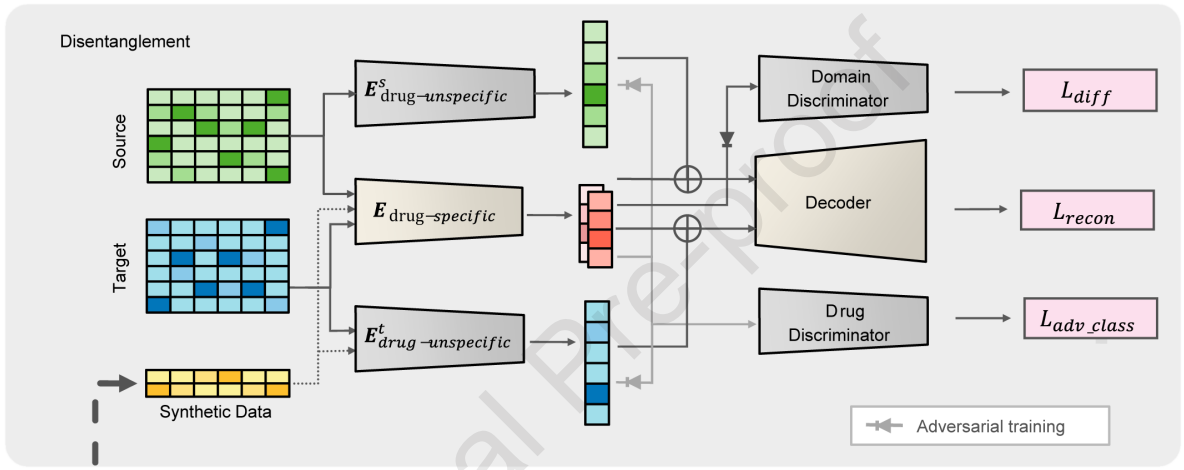
similarity loss), MMD (maximum mean discrepancy regularization), and ADV (adversarial regularization). (**B**) Distribution of AUROC values when varying the folds of synthetic data and the number of iterations during training. Asterisks indicate the level of statistical significance according to the Wilcoxon test: *$P < 0.05$, **$P < 0.01$.

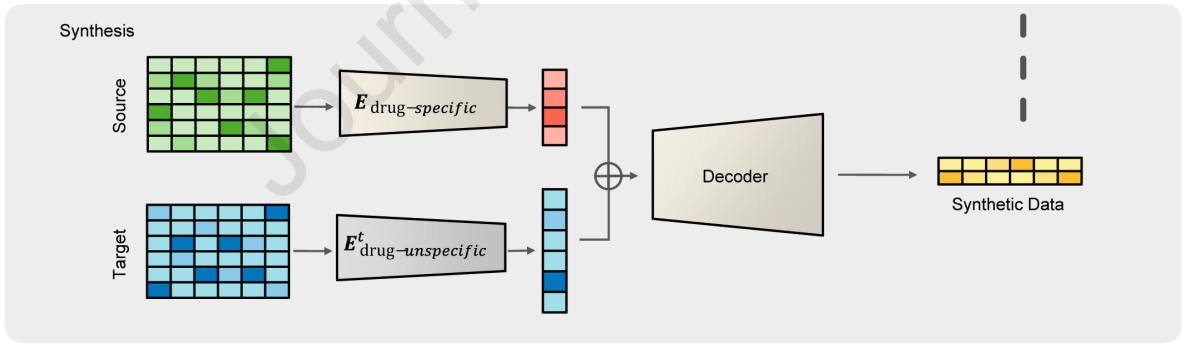
Fig. 6. Predicted drug responses of breast cancer patients. (A) Number of patients predicted to be responsive to each drug, with a pie chart representing the proportion of patients responsive to different numbers of drugs. (B) Expression of differentially expressed genes (DEGs) between breast cancer patients who were responsive (R) or non-responsive (NR) to 5-Fluorouracil (5-FU). p-values were calculated using the Wilcoxon test and adjusted for multiple tests. Each row represents a gene, and each column represents a patient. (C) Enrichment plot of DEGs in the epidermal growth factor receptor (EGFR) pathway, with the EGFR-related gene set obtained from the "EGFR_UP.V1_UP" term in the MSigDB database. The enrichment score was calculated using Gene Set Enrichment Analysis (GSEA). (D) Boxplot comparing EGFR expression levels between 5-FU responders and non-responders. Asterisks indicate the level of statistical significance according to the Wilcoxon test: **$P < 0.01$. (E) Cell Counting Kit-8 (CCK-8) assays of breast cells treated with 5-FU and Gefitinib individually or together. The abscissa is the different treatment time, and the ordinate represents the cell viability after treatment. Two-way ANOVA was used to analyze the data (****$P < 0.0001$). The data are shown as mean ± SD. CK: Control.

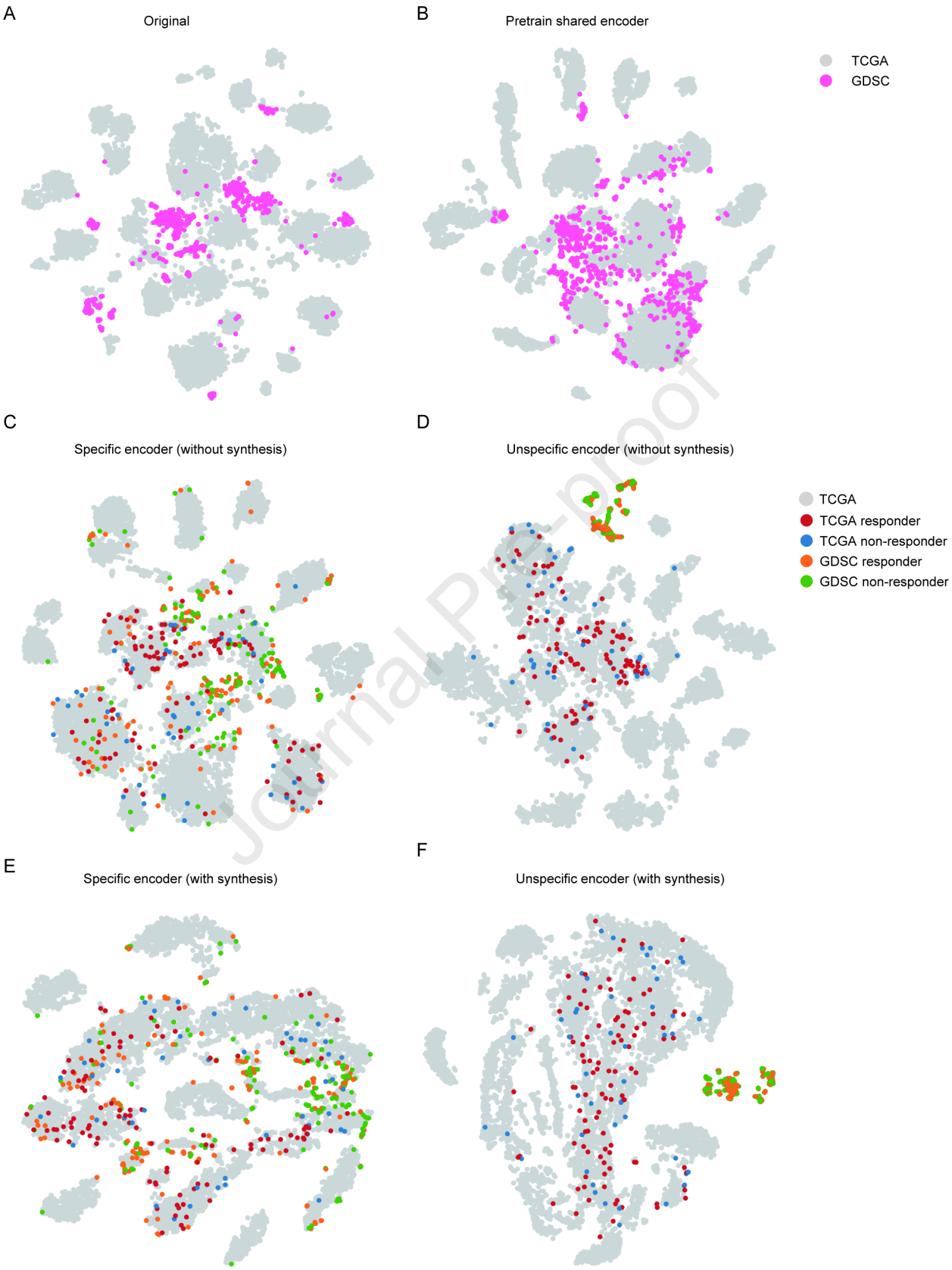
Acknowledgments

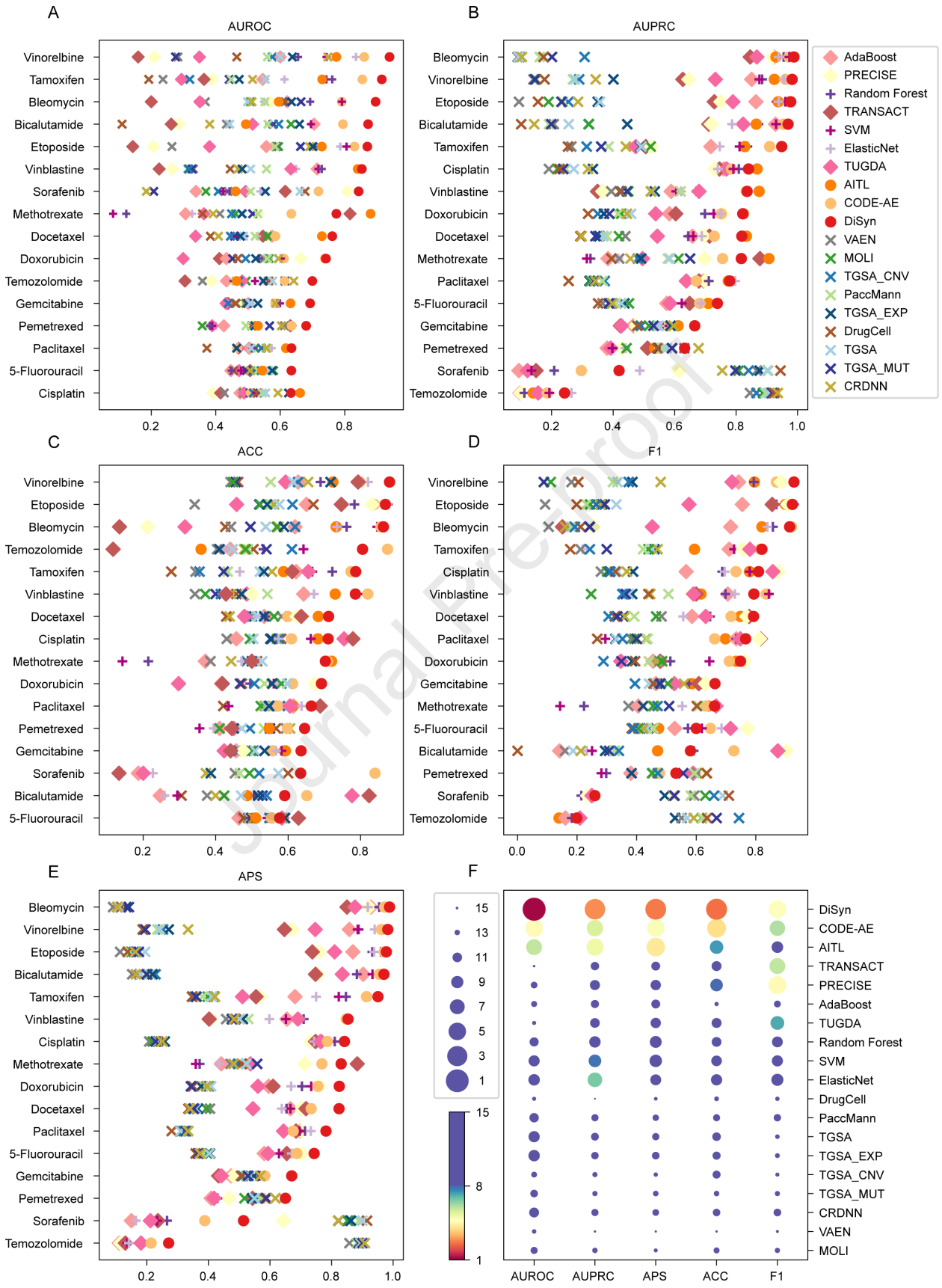
This work was supported by the National Natural Science Foundation of China (Grant Nos.: 32170680, and T2122018), the Natural Science Foundation of Shanghai (Grant No.: 21ZR1476000), and the CAS Youth Innovation Promotion Association (Grant No.: Y2022076).

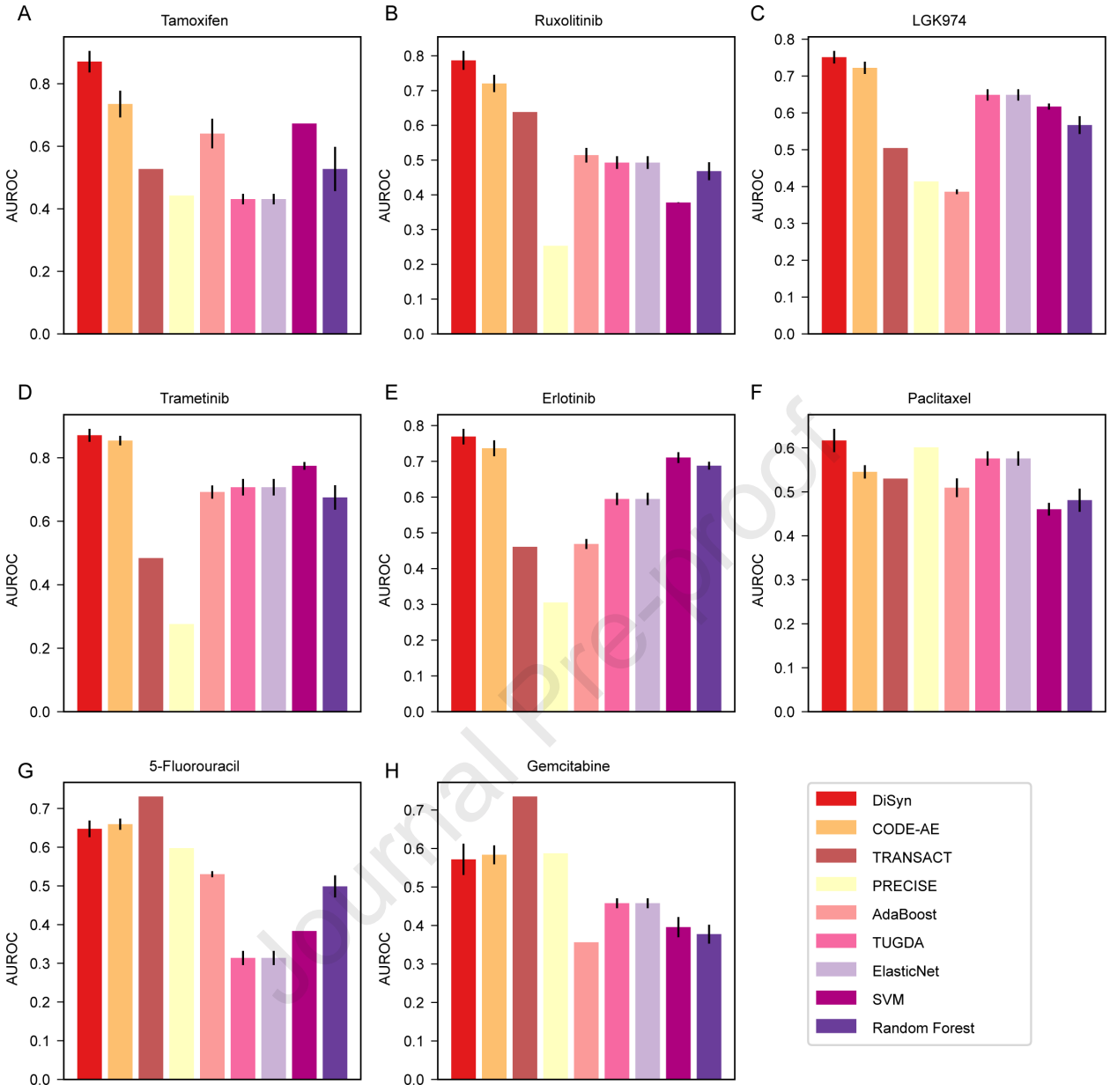


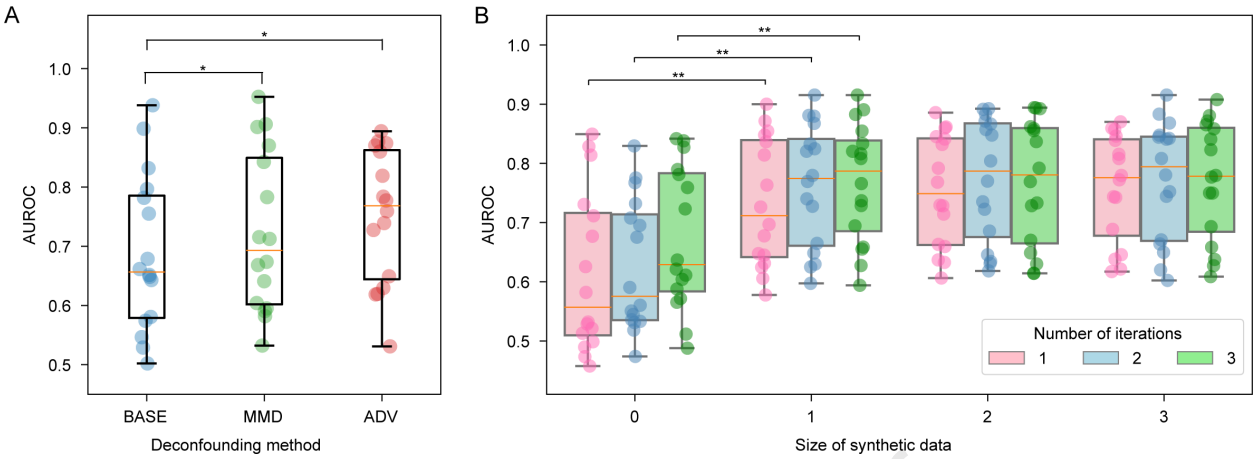


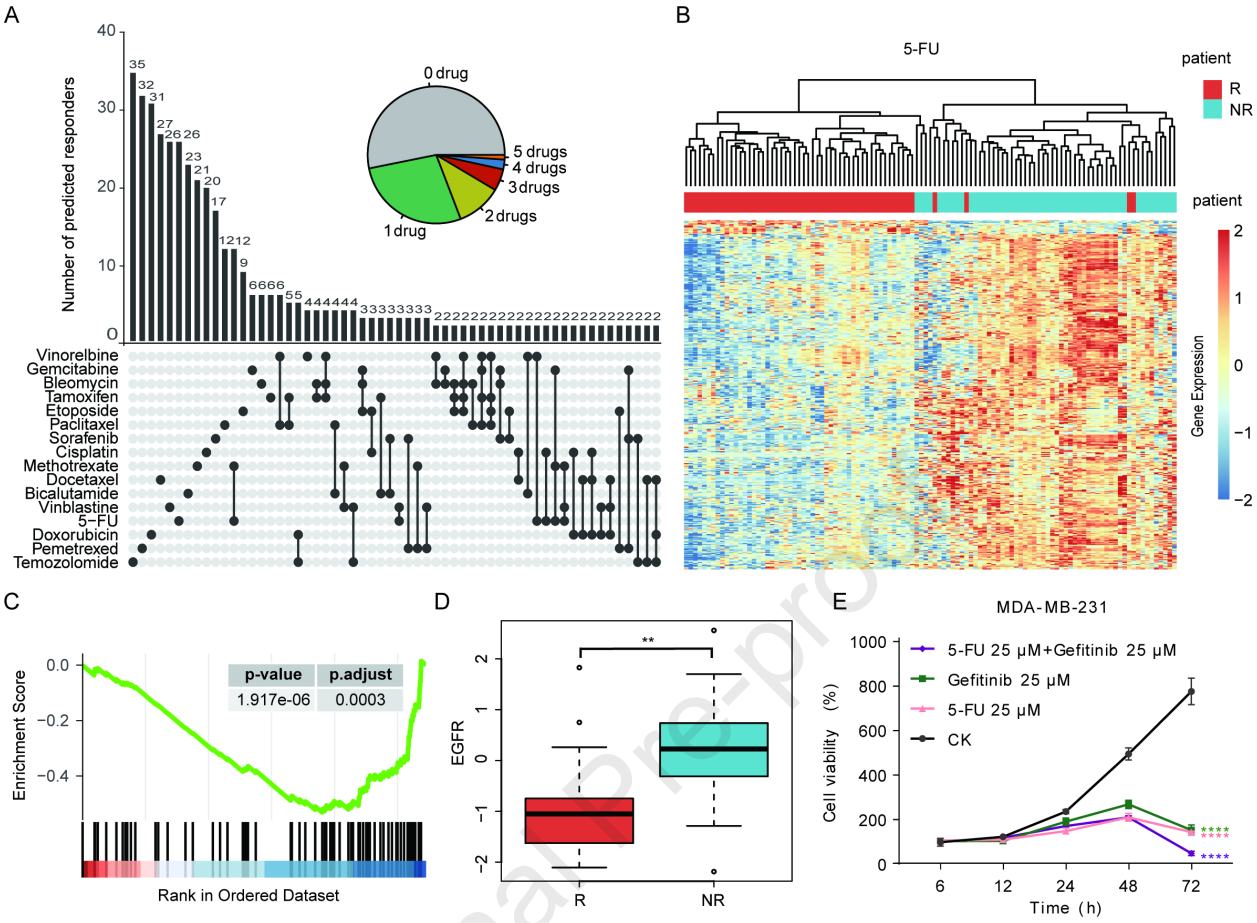


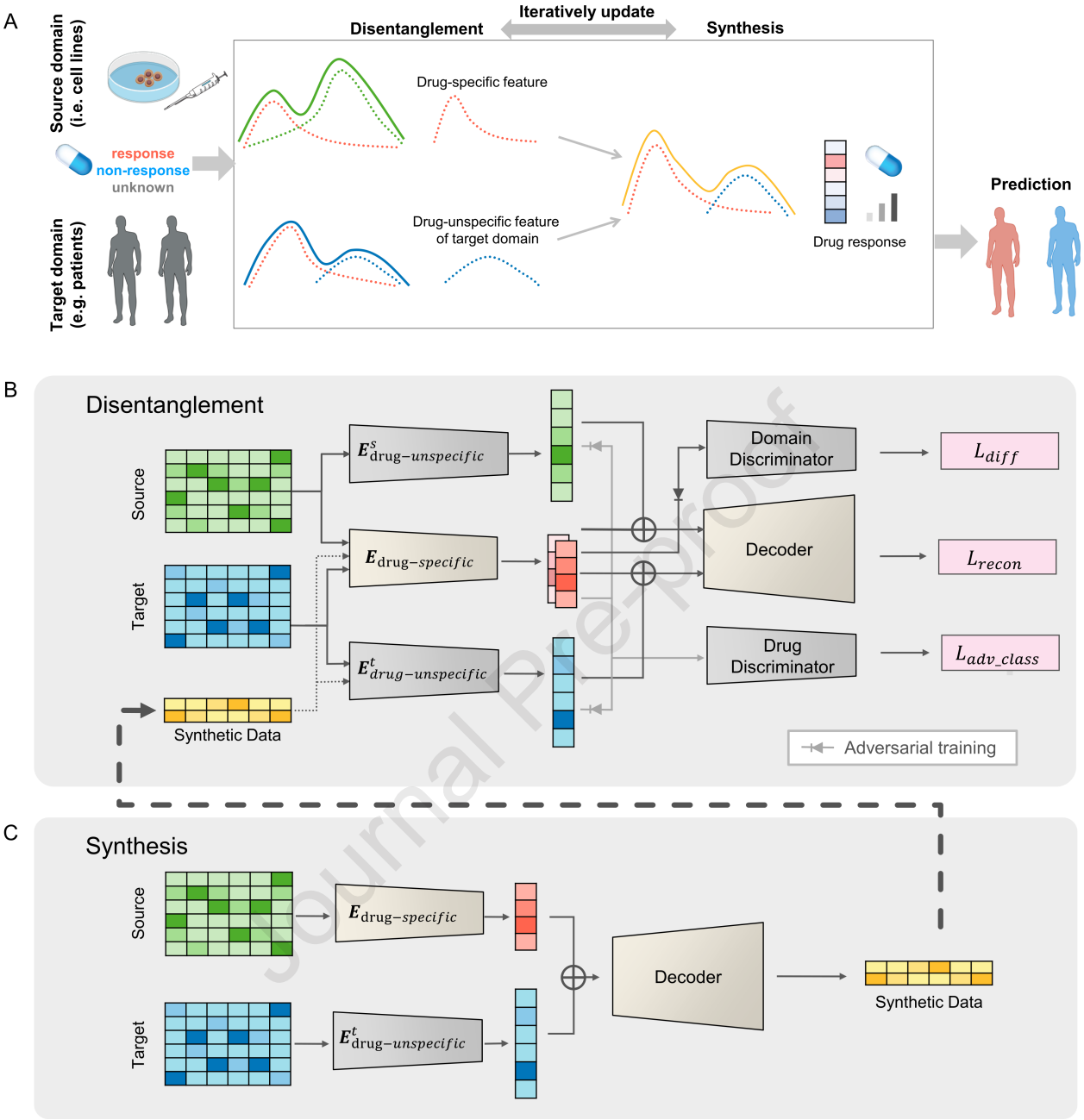












Highlights

- A novel disentangled synthesis transfer network (DiSyn) for in-vivo drug response prediction.
- Achieving state-of-the-art performance in drug response prediction by effectively generalizing preclinical data to patients.
- The application on patients reveals its potential in identifying biomarkers and optimizing therapeutic strategies.

Competing interests

The authors declare no competing interests.