MetaSCDrug: Meta-Transfer Learning for Single-Cell-Level Drug Response Prediction from Transcriptome and Molecular Representations

Shuang Ge[a,b], Shuqing Sun[b], Yiming Ren[a], Huan Xu[c] and Zhixiang Ren[a,*]

[a]Peng Cheng Laboratory, Shenzhen, 518055, Guangdong Province, China
[b]Shenzhen International Graduate School, Tsinghua University, Shenzhen, 518055, Guangdong Province, China
[c]School of Public Health, Anhui University of Science and Technology, Hefei, 231131, Anhui Province, China

ARTICLE INFO

Keywords:
Drug response prediction
Meta-transfer learning
single-cell RNA sequencing
Molecular representations

ABSTRACT

Analyzing the drug response at the cellular level is crucial for identifying biomarkers and understanding the mechanisms of resistance. Although studies on the drug response of individual cells can provide novel insights into tumor heterogeneity, pharmacogenomic data related to single-cell (SC) RNA sequencing is often limited. Transfer learning provides a promising approach to translate the knowledge of drug response from bulk cell lines to SC analysis, potentially providing an effective solution to this challenge. Previous studies often use data from single drug-cell lines to pre-train specific models and adapt the models on SC datasets, which lack pharmacogenomic information from other drugs and hinder model generalization. In this work, we introduce MetaSCDrug as a unified meta pre-training framework that integrates molecular information with transcriptomic data to simultaneously modeling cellular heterogeneity in response to multiple pre-trained drugs and generalize to unseen drugs. Our model requires only one pre-training session, followed by fine-tuning on multiple single-cell datasets by few-shot learning, achieving an average of 4.58% accuracy increase in drug response prediction compared to the baselines. Furthermore, our meta pre-training strategy effectively captures transcriptome heterogeneity in the generalization of unseen drugs, achieving a 20% improvement over the model without meta pre-training. Case studies of our framework highlight its capability to identify critical genes for resistance, providing a method for exploring drug action pathways and understanding resistance mechanisms.

1. Introduction

Post-treatment resistance studies are critical in exploring the mechanisms of drug-induced resistance and in identifying potential therapeutic targets, as well as in assessing the selective pressure of therapies[1]. While prior-treatment resistance studies focus on characterizing the heterogeneity of perturbative biological signatures within the tumor, thereby facilitating personalized treatment strategies. Intra-tumor heterogeneity has become a key factor in the failure of cancer treatments, arising from a variety of factors, including the genetic diversity of cells and numerous distinct cell subpopulations within a tumor[2, 3].This diversity poses substantial challenges in predicting responses to personalized anti-cancer treatments[4]. In this context, single-cell RNA sequencing (scRNA-seq) analysis has emerged as a powerful tool, providing novel insights into the cellular composition of tumors[5] and enabling the identification of distinct cellular subpopulations that contribute to resistance[6, 7]. The analysis of scRNA-seq holds great promise for enhancing drug response prediction, tailoring therapies to individual patients, and facilitating personalized oncology[8].

However, single-cell-level (SC-level) drug response studies currently face several challenges, mainly due to the limited pharmacogenomic data related to comprehensive scRNA-seq datasets and the complexity of data analysis. However, numerous studies have successfully used drug response profiling in cancer cell lines. These cell lines have been developed to represent a wide range of cancer types originating in the primary tumor, effectively preserving tissue characteristics and the immune microenvironment[4]. Numerous studies have demonstrated that pharmacogenomic profiling of bulk cell lines has become a platform for biomarker discovery and for predicting in vitro drug responses to a variety of targeted therapies[9, 10]. In this context, the potential of bulk cell lines pre-training followed by transfer learning on SC-level analysis emerges as a compelling approach to overcome these obstacles. A few recent studies, such as DREEP[11], scDEAL[12], and SCAD[13] attempt to leverage publicly available drug-cell line sensitivity profiles to predict drug response at the cellular level. However, previous studies have only used transcriptomic information to pre-train a single drug-specific model, which lack pharmacogenomic information for other drugs and hinder model generalization.

Leveraging both pharmaceutical and genomic information during pre-training could be a solution for generalization to multiple drug-cell line datasets. The intuitive notion that similar molecules exhibit similar properties is empirically substantiated, especially when structural similarities suggest a common interaction with biological targets[14]. By incorporating molecular information into the training phase, the model is able to model cellular heterogeneity in response to multiple drugs by only one pre-training session[15]. This approach not only improves the predictive accuracy of the models, but also extends their utility to a wider range of therapeutic contexts, providing benefits for drug repurposing [16] and the discovery of novel therapeutic targets[17]. While

*Corresponding author at: Peng Cheng Laboratory
E-mail addresses: jason.zhixiang.ren@outlook.com

this strategy has been used for patient-derived cell lines cultured in vitro, its application to single cells, where data heterogeneity is more pronounced, remains unexplored.

In this work, we adapted the idea of meta-transfer learning for few-shot learning[18] to task-specific multi-modal representations and developed a transfer learning framework called MetaSCDrug. Experimental results show that the model can facilitate generalization to multiple drugs, and case studies further demonstrate interpretability and applicability. In summary, our main contributions are as follows:

- We present a novel meta pre-training and adaptive transfer framework that integrates representations of molecular to predict the response of multiple drugs at cellular level.
- We propose a few-shot adaptive distribution alignment strategy for cross-domain migration, facilitating task-specific few-shot fine-tuning while minimizing sample size requirements.
- We evaluate the generalization capabilities of our proposed model for the first time on over 100 drugs not seen during the pre-training phase, achieving an average classification F1 score of 87.06%.

The rest of this paper is organized as follows. In Section 2, we briefly reviewed the existing work in SC-level drug response prediction that translate from bulk-level, mainly introduce methods about deep transfer learning. In Section 3, we describe the details of the data, model, strategy, implementations and evaluations. In Section 4, we evaluate the effectiveness of transferring pharmacogenomic knowledge from bulk RNA sequencing (bulk RNA-seq) to scRNA-seq. We perform an ablation study to understand the impacts of knowledge transfer from bulk RNA-seq to scRNA-seq. Additionally, we explore the model's performance on unseen drugs and conduct a case study to explore the practical applications of the model. In Section 5, we further discuss our results and discovery, and summarize in Section 6.

2. Related Work

2.1. SC-level drug response prediction

Numerous studies have employed high-throughput drug screening on cancer cell lines to determine their responsiveness to various drugs. However, most of these studies confined to the cell-line level and often overlooked the impact of cell subpopulation heterogeneity on drug sensitivity. The emergence of SC-level drug response prediction, which aims to assess drug response at the cellular level, presents an opportunity to explore tumor heterogeneity from new insights. While there have been efforts to leverage deep learning networks for SC-level drug response prediction by translating knowledge from drug-cell line interactions, the optimization of these model structures tends to be drug-specific[19], e.g. scDEALL[12] and SCAD[13]. These methods require substantial amounts of target data and extensive parameter tuning for each drug to reach an optimal state suitable for transfer learning. This requirement may limit their applicability across a wide range of drugs or facilitate high-throughput drug screening[19].

2.2. Deep transfer learning method

Unlike traditional machine learning methods, deep transfer learning can often handle situations where the distributions of source and target data differ significantly. Techniques such as adversarial discriminative domain adaptation (ADDA)[20] and domain-adaptive neural network (DaNN)[21] have been applied to predict SC-level drug responses under bulk-level pre-training. However, these methods face inherent challenges that make it difficult to optimize networks and apply to previously unobserved drugs. On the other hand, these networks often require a large number of samples to align feature distributions, which may not always be feasible. In contrast, few-shot learning strategies are designed to learn efficiently from a minimal number of examples. By learning to recognize patterns from only a few instances, the model can understand the heterogeneity of the data, potentially enhancing their performance on unseen data or tasks, making them more versatile and efficient in data-scarce scenarios.

3. Materials and methods

3.1. Data collection and processing

3.1.1. High-throughput drug responses datasets

The bulk pharmacogenomic datasets are collected from GDSC1000 project, which systematically tests cell line responses by high-throughput drug screening and provide half-maximal inhibitory concentration (IC50) values as metrics for drug sensitivity. Moleculars selected for screening are mainly anti-cancer therapeutics, covering both targeted agents and cytotoxic chemotherapies[4]. Following the approaches of previous studies such as SCAD[13], GraphDRP[22] and GraTransDRP[23], we focused on 223 drugs applied to 1018 cell lines, with drug response values in terms of IC50 normalized within a range of (0,1). In addition, the GDSC bulk RNA-seq data was downloaded from https://www.cancerrxgene.org/downloads/bulk_download.

3.1.2. Drug molecular representations

For chemical information about drugs, the notation of each molecular were downloaded from PubChem[24] in SMILES[25] format, based on the drug names and PubChem IDs provided by the GDSC. These chemical structure information was preprocessed using RDKit[26], an open-source cheminformatics toolkit, to construct molecular graphs where atoms serve as nodes and chemical bonds as edges. Atom features were constructed using five attributes: atomic symbol, atom degree (calculated from the number of bound neighbors including hydrogen), total number of hydrogens, implicit valence of the atom, and aromaticity. These features were encoded into an one-hot multidimensional feature vector following the translation rule for each SMILES string. This approach facilitates a detailed characterization of the

Table 1
Summary of the scRNA-seq datasets.

GEO access	Drug	Cell line	Cancer type	No. Res	No. Sens	N.A.	Author
GSE117872 _HN120	Cisplatin	OSCC	Oral squamous cell carcinomas	172	346	/	Sharma, et al.
GSE117872 _HN137	Cisplatin	OSCC	Oral squamous cell carcinomas	150	388	/	Sharma, et al.
GSE149383	Erlotinib	PC9	Lung cancer	617	849	/	Aissa, et al.
GSE110894	I-BET-762	leukaemic cells	Acute myeloidleukemia	670	719	/	Bell, et al.
GSE157220	Gefitinib, Vorinostat, AR-42	SCC47	Head and Neck Cancer	162	162	472	Kinker et.al
GSE157220	NVP-TAE684, Afatinib, Sorafenib	JHU006	Head and Neck Cancer	81	81	259	Kinker et.al
GSE228154	Afatinib	MDA-MB-468 cells	metastatic breast cancer	665	846	/	J. M. McFarland

molecular structure, which is crucial for the analysis of drug properties and interactions.

3.1.3. scRNA-seq datasets

For post-treatment data, we collected dataset 1-5 (the first five datasets in the Table 1) from National Center for Biotechnology Information's Gene Expression Omnibus (GEO [https://www.ncbi.nlm.nih.gov/geo/]). For prior-treatment data, we obtained the pan-cancer dataset 6 (the sixth dataset in Table 1) from Kinker et al.[27] available from the Broad Institute's single cell portal (SCP542). In their study, Kinker et al. demonstrated that EpiSen-high and EpiSen-low cells, derived from the same cell lines, exhibit differential responses to a variety of drugs[27]. Building on this foundation[13], we adopted a threshold using the 10th percentile EpiSen score to define the sensitivity of each cell to all drugs. A detailed description of each dataset is provided in Table 1. The scRNA-seq data and metadata files were all collected from previous studies. Following the preprocessing strategy used in scBERT[28], we utilized genes from the Panglao dataset as a reference to unify the dimensions across both bulk RNA-seq and all scRNA-seq datasets. The Panglao dataset[29] , which includes 16,906 genetic symbols, was downloaded from the PanglaoDB website https://panglaodb.se/. To ensure consistency in gene nomenclature, the test SC transcriptomic data files were pre-processed by revising gene symbols according to the NCBI Gene database https://www.ncbi.nlm.nih.gov/gene. This step is essential to ensure that the gene names in the dataset correspond to the symbols in PanglaoDB, allowing us to obtain gene expression profiles that are dimensionally consistent.

We performed quality control and preprocessing of the scRNA-seq data using the Python package SCANPY[30]. This involved filtering out cells that exhibited fewer than 200 genes, and genes that appeared in fewer than three cells. Then, cells with more than 10% expression of mitochondrial genes, which could suggest compromised cell integrity, were excluded. The count matrix was then normalized using the shifted logarithm method[31] to stabilize the variance across cells. Following normalization, expression values were scaled using the MinMaxScaler function from the preprocessing module in the sklearn package, which is suitable for preparing data for deep learning networks. Different from the bulk dataset, all SC datasets were labeled as sensitivity or resistance.

3.1.4. Unseen drug datasets

The unseen drug datasets, collected from GSE157220, include 108 drugs that are not present in the GDSC dataset. To maintain consistency in data processing and analysis, we applied the same preprocessing pipeline as used for dataset 6.

3.2. MetaSCDrug framework

3.2.1. Overview

In the pre-training phase, we use all available bulk cell line-drug response datasets to develop a "learn how to learn" network. This network is designed to quickly adapt when presented with a few additional SC samples. The model is specifically trained to identify commonalities and distinctions between tasks, utilizing this information to facilitate new tasks effectively. We adopted a meta-transfer learning strategy, originally designed for few-shot learning scenarios[18], and tailored its framework to better suit the requirements of SC drug response prediction. This modification ensures that the model can flexibly handle cross-domain challenges and label inconsistency issues. An overview of the MetaSCDrug framework is shown in Figure 1(a) and (b).

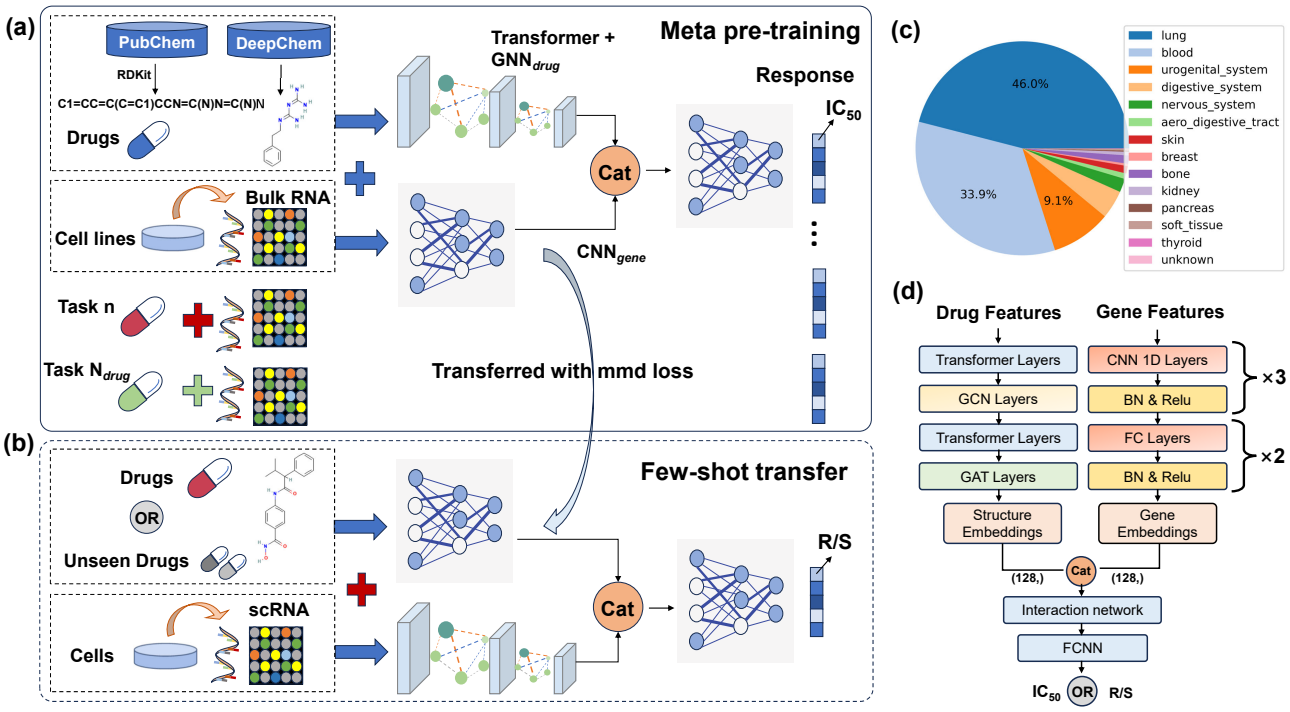

Figure 1: Description of the MetaSCDrug workflow. a. An overview of our proposed framework, including data collection and pre-processing, task-specific representation, and three modules for feature extraction and feature interaction. **b.** An overview of the few-shot transfer framework, including task-specific few-shot fine-tuning. **c.** Dataset statistic shows the distribution of bulk datasets, including 14 tissues with 58 subtissues (Figure S1). **d.** The details of each modules, including Transformer-GNN_{drug}, CNN_{gene} and task-specific interaction network.

3.2.2. Meta pre-training design

We developed a meta pre-training framework where each drug in the bulk dataset is treated as a distinct meta pre-training task, resulting in a total of 223 tasks corresponding to 223 drugs. For each task T, we further categorize the cell lines into 14 tissues as depicted in Figure 1(c), with each tissue representing a set of learning samples. During each pre-training iteration, we randomly select a dataset S corresponding to one drug and then sample a subset of 8 tissues S_i to create a smaller dataset. This subset S_i is subsequently divided into two non-overlapping groups, each containing four tissues. These groups are designated as cell lines Q and cell lines K, which serve distinct roles in the training process. For the loss function, we employ the Mean-Square Error (MSE) function L as the pre-training loss,

$$L(X,\theta) = \frac{1}{M}\Sigma_{x_i \in X}\left(p_i - \hat{p}_i(\theta)\right)^2 \tag{1}$$

The loss function L is defined as a function of the sample set X, and the model parameters θ, where p_i represents the continuous label for sample i and M is the total number of samples in X.

We first perform one-step gradient descent on the subset Q, using it to compute a preliminary update to the parameters θ. Following this initial step, we then focus on optimizing the parameters further by seeking a better solution for subset K. This is achieved through iterative refinement, where we apply the Adam gradient descent algorithm[32] to minimize the regression loss. For each training iteration, we aim to find the optimal parameters θ that result in a smaller regression loss. This involves updating the parameters over 100 iterations specifically on subset K. The entire pre-training phase comprises 1000 training iterations, allowing us to iteratively refine the model using 1000 paired subsets Q and subsets K, respectively. For each S, a loss function is defined as follows:

$$E_{S_i \in S}\left[E_{<Q,K> \in S_i}\left[L\left(K, \theta - \alpha\frac{\partial L(Q,\theta)}{\partial \theta}\right)\right]\right] \tag{2}$$

Where L is MSE loss as described in equation(1). For each training iteration, we seek the parameters θ that can achieve a smaller regression error on subset K after only one iteration of the gradient descent performed on subset Q.

3.2.3. Task-specific features

Drug encoder. For each drug d, we construct a graph-transformer network designed to learn the representations of molecular chemical structure (Figure 1(d)). The input to this network consists of graph features represented as $G_d = \left(N_d, E_d\right)$. $N_d \in \mathbb{R}^{A_d \times F_d}$ is the node matrix of the graph, where each node corresponds to an atom within the molecular structure. $E_d \in \mathbb{R}^{B_d \times D_d}$ represents the adjacency matrix of edges, detailing the connections between the nodes,

i.e., the chemical bonds. Where A_d denotes the number of atoms in the drug molecule, B_d represents the number of chemical bonds connecting these atoms. F_d corresponds to the feature embedding of the atomic properties. D_d refers to the bond endpoints, which, since each bond connects two atoms, is consistently set to 2. The combination of a Transformer architecture with a Graph Neural Network (termed Transformer-GNN_{drug}) effectively represents the chemical structures of the drugs. This approach leverages the Transformer's capability to handle long-range dependencies within sequential data, enhancing the representation of linear sequences such as SMILES strings. Simultaneously, the Graph Neural Network component can capture the intricate inter-connectivity of nodes (atoms), thus reflecting the actual spatial structures and bonding relationships in molecular.

Bulk cell lines encoder. We implemented the gene feature extraction network as an end-to-end convolutional neural network (CNN) without manual feature extraction. This approach allows exploiting the high-dimensional feature representations of deep neural networks, ensuring that the true genetic signature is fully preserved. The input to this network consists of RNA sequences of length 16, 906, which corresponds to the number of genetic symbols in the previously described Panglao dataset.

3.2.4. Few-shot transfer design

In the few-shot transfer phase, each SC dataset is treated as a new task with only a few additional samples available. Due to the distribution gap between bulk RNA-seq and scRNA-seq, we use the pre-trained weights only as initialization parameters and refine all parameters during fine-tuning. For each task, n-shot($n = 1, 5, 10$) with 10-query samples is preformed to update the parameters, and the remaining samples are used for testing. One iteration of gradient descent is computed on the shot set, followed by 100 iterations of updating the parameters on the query samples. This process is repeated 10 times for each k. We then calculated and reported the average and standard deviation of the prediction performance across all these repetitions (Figure 3 (c), (d)). The binary cross entropy (BCE) loss is employed with m ($m \in 0 \sim 1000$) iterations, activating an early stopping mechanism if the area under the curve (AUC) of the query sets reaches 0.99. In addition, we incorporate maximum mean discrepancy (MMD)[33] loss to align the hidden spatial feature distribution of the genes:

$$L_{MMD}\left(E_b\left(X_b\right), E_s\left(X_s\right)\right) = \left|\frac{1}{n}\sum_{i=1}^{n}\phi\left(x_b^i\right) - \frac{1}{m}\sum_{j=1}^{m}\phi\left(x_s^j\right)\right|_H \quad (3)$$

MMD is a measure used to determine the difference between the average embeddings of two distributions within a high-dimensional feature space known as the reproducing kernel hillbert space (RKHS) denoted by $\phi(.)$. Specifically, $X_b = \left\{x_b^i\right\}_{i=1,\ldots,n}$ and $X_s = \left\{x_s^j\right\}_{j=1,\ldots m}$ represent the embeddings of gene features from bulk cell lines n and single cells m, respectively. The intuition is that the identical distributions E_s and E_b will have similar average representations when the MMD values are close to zero. As a result, the total loss for our model is computed as the sum of the BCE loss and the MMD loss:

$$L_{all} = L_{BCE} + \lambda L_{MMD} \quad (4)$$

Where λ is introduced to balance the influence of the alignment loss. After m iterations of gradient descent, the few-shot parameter θ is adapted to the new samples:

$$\theta_{\text{few-shot}} = \theta_{\text{pre-training}} - \alpha \frac{\partial L(query, \theta)}{\partial \theta}\bigg|_{\theta - \theta_{\text{pre-training}}} \quad (5)$$

The pseudo-codes of MetaSCDrug are illustrated in Algorithm 1.

3.2.5. Model training protocols

Our deep learning model constructs an end-to-end pharmacogenomic interaction model. By extracting long-range sequence features of the transcriptome and graph features of the chemical information, the model can automatically extract interaction features to distinguish pre-existing resistant cell lines under specific drugs. As depicted in Figure 1, in the meta pre-training phase, MetaSCDrug requires three steps to obtain the prediction results. First, drug features are extracted by converting the drug SMILES into a molecular graph, which is then fed into the Transformer-GNN_{drug} module. At the same time, the expression profiles of genes are processed through the CNN_{gene} module to extract features from the gene expression. Subsequently, the outputs from both the drug and gene modules, standardized to 128 dimensions, and concatenated into a nonlinear feature convergence network designed to predict the mutual relationships, ultimately estimating the IC50 value. Within the interaction model, the hidden layers are configured with 1024 and 128 neurons, respectively. The CNN_{gene} module includes three convolutional kernels of size 8, followed by two fully connected layers that map to a 128-dimensional hidden space. The Transformer-GNN_{drug} module integrates the capabilities of graph convolutional networks (GCN)[34] and graph attention networks (GAT)[35], each following a transformer layer that serves as an invariant dimension feature extractor. During the few-shot transfer phase, the model retains the same architecture as the pre-trained model but outputs predictions using binary drug sensitivity or resistance labels. Therefore, we load the pre-trained weights up to the last fully-connected layer for task-specific fine-tuning.

To prevent overfitting and improve training efficiency, we insert the Dropout[36] layer after each convolutional layer. The ReLU function[37] is employed as a nonlinear activation function following each linear transformation. All models are optimized with the Adam optimizer, with the learning rate managed by StepLR. The initial learning rate is set to 0.0001 and the update step size is set to 100 iterations. The model environment is based on Python 3.10, running on the Pytorch framework of the GPU version.

Algorithm 1 The pre-training and fine-tunning of our proposed framework.

Input: Task distribution $p(\mathcal{T})$ and corresponding dataset $\mathcal{D}$, learning rates α, β and γ.
Output: Predict drug response value, feature extractor Θ, interaction network θ.

Randomly initialize Θ and θ.
for samples in tissues $\mathcal{S}$ **do**
  Sample a subset of 8 tissues S_i;
  Randomly separate S_i into cell line $\mathcal{Q}$ and $\mathcal{K}$ (4 tissues each) ;
  Evaluate $\mathcal{L}(\mathcal{Q}, [\Theta; \theta])$ by Eq.1;
  Optimize Θ and θ by $[\Theta; \theta] =: [\Theta; \theta] - \alpha \nabla \mathcal{L}_{\mathcal{Q}}([\Theta; \theta])$;
  while not done **do**
    Evaluate $\mathcal{L}(\mathcal{K}, [\Theta; \theta])$ by Eq.1;
    Optimize Θ and θ by $[\Theta; \theta] =: [\Theta; \theta] - \beta \nabla \mathcal{L}_{\mathcal{K}}([\Theta; \theta])$;
  end while
end for
Initialize Θ and θ by pre-trained weights for few-shot tasks.
for each task k **do**
  Randomly choose n-shot(n=1.5.10) and 10 query samples as a batch $\mathcal{T}_{(tr)}$
  while not done **do**
    Extract gene features from bulk RNA-seq($\mathcal{X}_b$) and scRNA-seq($\mathcal{X}_s$).
    Evaluate $\mathcal{L}_{MMD}$ by Eq.3.
    Calculate $\mathcal{L}(n - shot, [\Theta; \theta])$ by Eq.1.
    Optimize Θ and θ by $[\Theta; \theta] =: [\Theta; \theta] - \gamma \nabla \mathcal{L}_{n-shot}([\Theta; \theta])$;
    while not done **do**
      for samples in query **do**
        Calculate $\mathcal{L}(query, [\Theta; \theta])$ by Eq.1.
        Optimize Θ and θ by $[\Theta; \theta] =: [\Theta; \theta] - \gamma \nabla \mathcal{L}_{query}([\Theta; \theta])$;
      end for
    end while
    Calculate $\mathcal{L}_{all}$ by Eq.4
    Optimize Θ and θ by Eq. 5
  end while
end for

3.2.6. Response prediction of unseen drugs

In particular, there are scenarios where it becomes necessary to make predictions for new drugs that are not included in the training set. In this section, we explore the performance of our model on drugs that were not part of the pre-trained GDSC dataset. We use the few-shot learning approach described in Section 3.2.4, and test on SC datasets for both the bulk-level pre-trained model and a model initialized randomly. This allows us to assess the ability of our network to adapt to new pharmacological challenges.

3.2.7. Interpretability and Visualization analysis

We applied integral gradient (IG)[38] approach to quantify the contribution of input gene features to the output predictions. An IG score is calculated for each gene by integrating the gradient of the model's output with respect to the gene expression input, moving from a baseline of zero expression to the actual input expression level. The integration is approximated by a Riemann sum, which involves dividing the path from the baseline expression level to the input expression level into small segments and summing the gradient values across these segments:

$$\mathrm{IG}_{(x_i)} ::= \left(x_i - x_i'\right) \times \int_{\alpha=0}^{1} \frac{\partial F\left(x_i' + \alpha \times \left(x_i - x_i'\right)\right)}{\partial x_i} d \quad (6)$$

The importance of the i-th gene expression in the input cell x is calculated using a scaling coefficient α. The baseline expression level for gene i, denoted as x_i', is set to zero. The notation $\partial F(x_i)/\partial x_i$ refers to the gradient of the function $F(x_i)$ with respect to the i-th dimension, indicating how changes in the i-th gene expression affect the output of $F(x)$. We divided the input gene features into 50 segments ($\alpha = 1/50$) and input them into the network successively to calculate the gradient. The output of the result is an IG matrix with the same shape as the input gene expression matrix. The rows represent genes and the columns represent cells. Values are IG scores. We utilized the Wilcoxon test in SCANPY to compare the integral gradient (IG) scores between the sensitive and the resistant cells, aiming to identify significantly difference genes. Genes were selected based on several criteria: a Bonferroni-adjusted p-value of less than 0.05, a log-fold change greater than 0.1, and the condition that more than 20% of cells in either group

Table 2
Comparison of MetaSCDrug to previous studies on SC datasets

Dataset	AUC		AP		Literature
	MetaSCDrug	Baseline	MetaSCDrug	Baseline	
GSE117872_HN120	**98.66**	73.34	**99.33**	86.88	scDEAL
GSE117872_HN137	**97.32**	90.18	**98.88**	81.45	scDEAL
GSE149383	87.92	**94.94**	90.02	**94.45**	scDEAL
GSE110894	**99.00**	95.17	**98.87**	93.70	scDEAL
GSE157220_SCC47	75.60	**79.50**	75.66	**81.23**	SCAD(average)
GSE157220_JHU006	**99.41**	94.57	**99.38**	94.90	SCAD(average)
Average	**92.99**	88.41	**93.69**	88.77	

(sensitive or resistant) exhibited an IG score above 0.2. These selected genes are referred to as Critical Genes (CGs).

3.3. Model evaluation

3.3.1. Evaluation metric

In this work, we construct a regression model for the bulk dataset and a classification model for the SC dataset. For the regression model, the performance of the pre-training network was evaluated by five methods, including R^2 (coefficient of determination), root mean square error (RMSE), mean square error (MSE), Spearman correlation coefficient (SCC), and Pearson correlation coefficient (PCC). For the classification model, we used AUC and Average Percision (AP) to evaluate the performance on the scRNA-seq dataset, because these two evaluation indicators are less affected by the imbalanced data. Where AUC is the area under the receiver operating characteristic (ROC) curve and AP is the area under the precision-recall (PR) curve.

3.3.2. Baselines

To evaluate the performance of the MetaSCDrug model, we compare our method with the state-of-the-art baseline models, including the results of previous work and the other deep transfer learning models.

scDEAL[12] is a deep transfer learning framework designed to predict cancer drug responses at the cellular level by leveraging large-scale bulk cell line data. But to the best of our knowledge, this relies on training specific model on single drug-cell line data, which poses challenges in generalizing the findings to other drugs, particularly for predicting previously unobserved drugs. At the same time, it often requires a large sample of individual cells for distribution alignment.

SCAD[13] is a transfer learning model designed to infer drug sensitivities at the cellular level by leveraging bulk transcriptome profiles from large public datasets of cell lines, offering insights into drug resistance and potential drug combination strategies through the identification of drug sensitivity biomarkers. However, it suffers from the same problems as scDEAL.

ADDA[20] is a transfer learning method first implemented on natural images. The goal of ADDA is to enable a model trained on a source domain to perform well on a target domain, even when the distribution is significantly different between the two domains. It trains a feature extractor and an adversarial discriminator to maximize and minimize the discriminability of features, respectively.

DaNN[21] refers to a transfer learning model designed to improve model performance on a target domain by leveraging data from a related source domain. The core idea behind DaNN is to minimize domain discrepancy and enhance generalization ability across different but related domains.

4. Result

4.1. Data statistics and visualization

For the in vitro bulk dataset, we used 191,034 drug-cell line pairs with normalized continuous IC50 values as labels, including 1018 cell lines with 223 drugs. This was due to the absence of response data for some drug-cell line combinations. The dataset totally covers 14 tissues and 58 tissue subtypes (see Figure 1(c) and Figure S1). The normalized distribution of total gene counts in bulk cell lines and single cells (e.g. dataset GSE110894) can be observed in Figures 2(b) and (c). We found that the gene expression levels in scRNA-seq are significantly lower than those in bulk RNA-seq. This disparity is likely due to differences in sample size, sequencing techniques, and underlying biological factors, which can influence the capture and detection of RNA molecules. We further investigated the dataset's heterogeneity in the feature space. We plotted the UMAP of preprocessed bulk RNA-seq and all scRNA-seq (including prior-treatment and post-treatment cell populations) as depicted in Figure 2(a). This visualization demonstrates the overlap between the distributions of bulk RNA-seq and scRNA-seq in feature space. Figure 2(d) plots the UMAP visualizations of ground-truth drug response labels (sensitivity represents as 1 and resistance represents as 0), the predicted binary drug response label and the predicted continuous values of the model on the dataset GSE110894, which illustrates that our model can well distinguish the heterogeneity of drug responses.

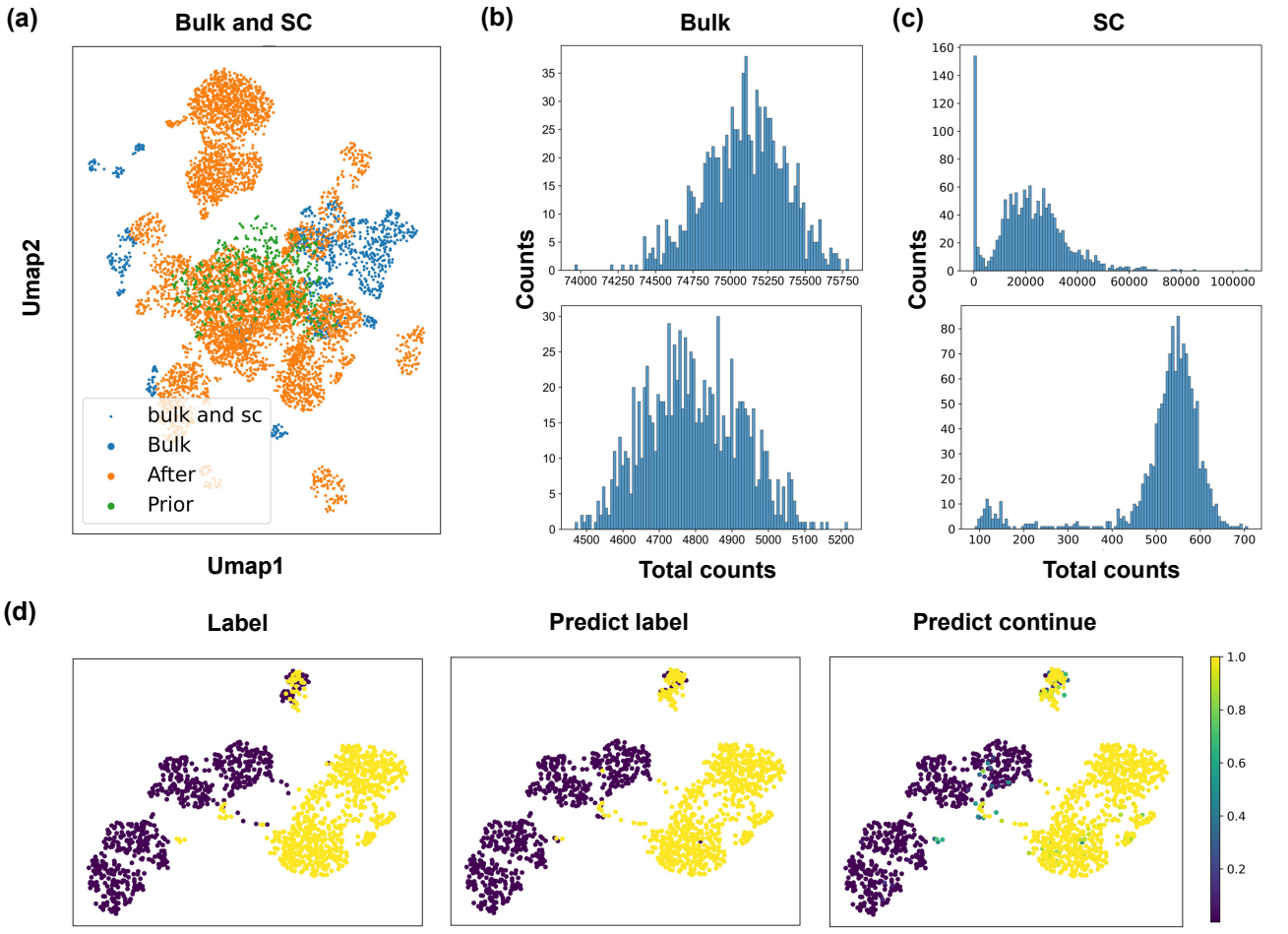

Figure 2: Description of the data distribution. **a.** The UMAP of preprocessed bulk RNA-seq and all scRNA-seq (including prior-treatment (prior) and post-treatment cell populations) demonstrates the overlap between the distributions of bulk RNA-seq and scRNA-seq in feature space. **b, c.** The normalized distribution of total gene counts in bulk cell lines and single cells (e.g. dataset GSE110894) before (up) and after (bottom) preprocessing shows that the gene expression in scRNA-seq was much lower compared to bulk RNA-seq. **d.** The ground-truth drug response labels (sensitivity represented as 1 and resistance represented as 0), predict binary labels and the predicted continuous values of the model.

4.2. In vitro prediction model evaluation of drug sensitivity

4.2.1. Meta pre-training assessment

Initially, to assess the reliability of our pre-trained model, we split the data into training, validation, and test sets in a 9:1:1 ratio. After completing 300 iterations of training, we obtained a PCC score of 0.940 and a SCC score of 0.917 on the test set.

Subsequently, in the meta pre-training stage, we evaluated the model's ability to predict continuous drug responses across all bulk datasets. MetaSCDrug achieved an R^2 score of 0.919 and a PCC score of 0.959 between the predictions and the ground truth. This performance surpasses advanced deep learning models, which recorded an R^2 score of 0.863 and a PCC of 0.929[15], demonstrating MetaSCDrug's robust performance in drug response prediction for regression tasks. As shown in Figure 3(b), the four metrics calculated for each drug indicate that the response prediction values are significantly correlated with the truth, with 87.4% of them achieving a PCC value above 0.8. To investigate the impact of transfer learning, we carried out an ablation experiment to assess model performance with and without pre-trained weights. Surprisingly, we found that the pre-trained model significantly outperformed the non-pretrained model on prior-treatment datasets. However, on the post-treatment dataset, pre-trained model only resulted in a minor improvement. Furthermore, we compare our findings with those of previous studies, as detailed in Table 2. In contrast to previous studies, MetaSCDrug shows its competitive ability to distinguish subpopulation response heterogeneity.

However, it is essential to recognize that bulk-based transfer learning is not a universal solution. Taking the GSE117872 dataset as an illustration (refer to Supplementary Figure S2), we observed substantial variations in the feature space between the same patient's cell line before and after drug intervention. Our insights from bulk RNA-seq primarily

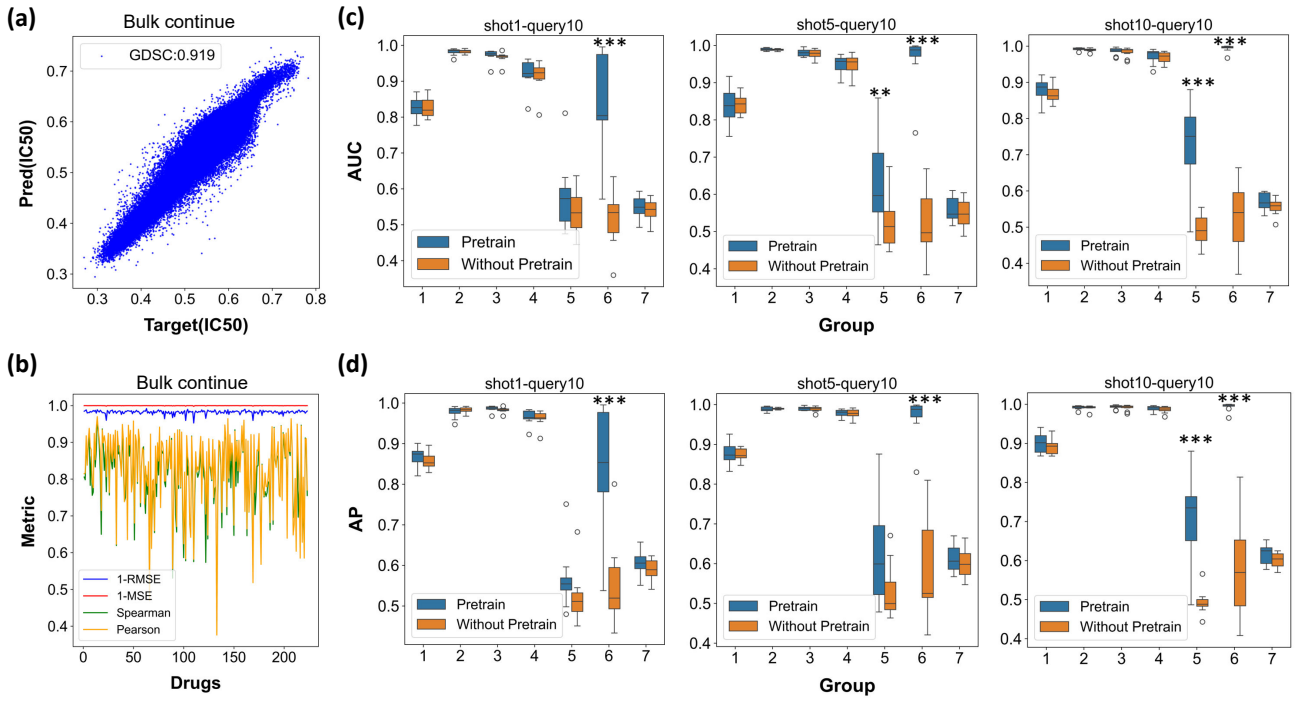

Figure 3: Assessment of pre-training and few-shot transfer. **a.** Linear scatter plots for the bulk dataset in the regression model. The closer the scatter is to the diagonal, the closer the model prediction is to the true value. The approximation of the model prediction to the true value achieves 0.919 R^2 score. **b.** Four metrics were computed for each drug on meta pre-trained bulk data (from $1 \sim 223$), including 1-RMSE, 1-MSE, Spearman and Pearson correlation coefficient. **c, d.** Box plots of the performance of MetaSCDrug and the baseline on SC datasets with n-shot ($n = 1, 5, 10$) and 10-query samples transfer learning across 10 repetitions. We use '*' to indicate the degree of significance of the difference.

focus on the "pri" feature space, which is about the transcriptome information from pre-existing resistant cell lines. The limited improvement after transitioning to the post-treatment SC dataset is most likely due to the disparity in the feature space of gene expression before and after drug selection. Specifically, bulk data lacks transcriptomic information under drug selection, and thus knowledge transfer remains limited.

4.2.2. Few-shot transfer assessment

We further compare the few-shot transfer strategy with other transfer learning methods, including Adversarial Discriminative Domain Adaptation (ADDA) and Domain-adaptive Neural Network (DaNN). These methods, widely used in computer vision, focusing on transferring knowledge from the source domain to the target domain. We benchmark various transfer learning methods to evaluate their applicability within our proposed integrated multimodal framework. The experimental results, detailed in Table 3, indicate that domain adaptation and adversarial learning methods are less effective at implementing transfer from the source domain to the target domain compared to the MetaSCDrug approach. We conjecture that this is due to the significant difference between the two sequencing techniques. However, this challenge could potentially be mitigated by incorporating few-shot labels, which could help bridge the gap between source and target domain in these complex transfer scenarios.

4.3. Unseen drug response prediction assessment

We evaluate the transferability of gene expression-molecular interaction knowledge using a prior-treatment SC dataset that includes drugs not present in the GDSC dataset. Specifically, we tested MetaSCDrug on the remaining 108 drugs in GSE157220. We compared the adaptation capabilities under both pre-trained and non-pre-trained conditions after 100 iterations. Since a dynamic threshold may not be feasible in real-world applications, we set a fixed threshold of 0.5 for both training and testing. We use F1 score to comprehensively evaluate the performance of the network. As a result, we achieved an average F1 score of 87.06% for the pre-trained group, compared to 67.03% for the non-pre-trained group, showing about a 20% improvement. Examples of drugs are listed in Table 4, and results for all drugs are provided in Supplementary Data S1.

4.4. Case Study: Identify critical genes and pathways for drug response

While MetaSCDrug achieves excellent results in predicting drug responses at cellular level, it is crucial to determine whether the model prediction results are reliable in clinical scenarios. We aim to explore the correlation between drug response and gene expression in order to explore pathways of drug action, evaluate the validity and reliability of the model and provide new methods for the discovery of possible targets. We conducted a case study for

Table 3
Comparison of MetaSCDrug to deep transfer learning methods on SC datasets

Dataset	AUC			AP		
	MetaSCDrug	ADDA	DaNN	MetaSCDrug	ADDA	DaNN
GSE117872_HN120	**98.66**±0.92	28.63±4.39	52.13±13.93	**99.33**±0.46	55.27±0.25	69.10±9.51
GSE117872_HN137	**97.32**±1.73	55.69±3.59	51.24±14.71	**98.88**±0.68	74.35±0.23	72.32±8.76
GSE149383	**87.92**±2.84	50.65±0.24	48.48±4.23	**90.02**±1.99	59.99±0.12	43.3±3.22
GSE110894	**99.00**±0.87	38.85±1.03	54.51±10.65	**98.87**±1.27	46.79±0.72	58.01±8.63
GSE157220_SCC47	**75.60**±14.07	56.99±1.53	50.13±1.09	**75.66**±14.00	52.96±0.18	50.64±1.37
GSE157220_JHU006	**99.41**±0.78	47.28±2.29	50.47±3.45	**99.38**±0.87	41.06±0.98	50.81±2.25
GSE228154	**57.03**±2.33	47.73±0.13	49.86±1.03	**61.64**±2.33	55.48±0.09	55.52±1.03

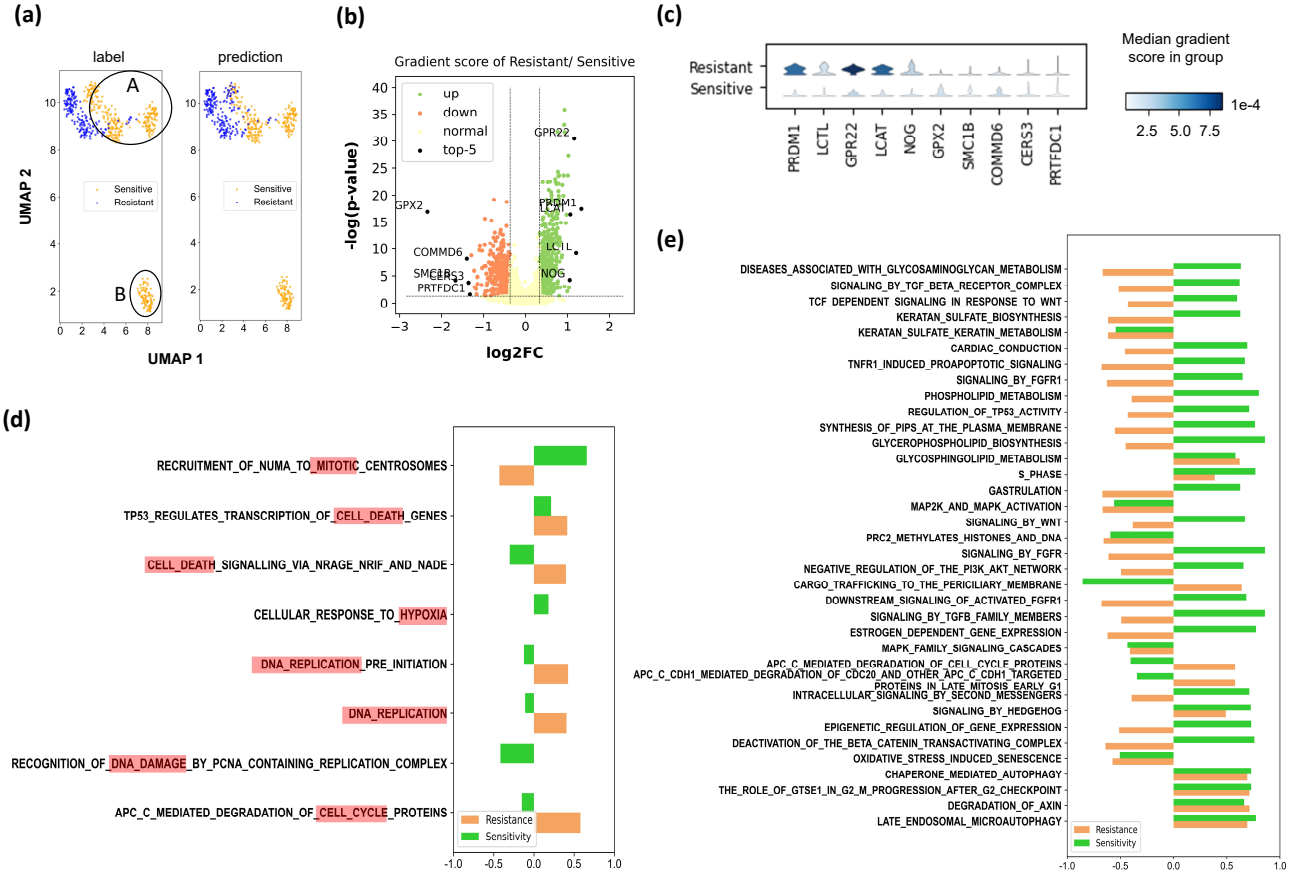

Figure 4: Case study of identifying critical genes and pathways for drug response on GSE117872_HN120. **a.** The UMAP of ground-truth (left) and predicted binary drug response (right). A and B are the different subpopulations with the same drug resistance label. **b.** The volcano map showed that the drug-resistant group significantly up-regulated (green) and down-regulated (orange) critical genes compared with the drug-sensitive group. **c.** The top 5 up- and down-regulated genes are plotted on a violin diagram. The color of each violin indicates the median gene gradient score of the resistant or sensitive groups. **d.** The pathways about cell mitotic and death, showing different or diametrically opposite correlations between the sensitivity and the resistance. **e.** 36 significant pathways of drug-resistant ($p<0.05$) and the correlation coefficient. Among them, the direction of regulation for resistance and sensitivity is opposite in 20 pathways.

oral squamous cell carcinoma (OSCC) treated with Cisplatin (dataset GSE117872_HN120). Cisplatin acts by generating DNA crosslinks via interactions with purine bases, causing DNA damage and interfering with DNA replication, leading to cell apoptosis, particularly in rapidly dividing cells such as cancer cells. Consequently, cells sensitive to Cisplatin may express genes associated with apoptosis[39].

We find that cells showing drug resistance may have large distance in the feature space, as shown in Figure 4(a) for subgroups A and B. Nonetheless, MetaSCDrug was able to

Table 4
Unseen drug response prediction on non-pre-trained and pre-trained model

Unseen drug in GDSC	Non-pre-trained (shot5)	pre-trained (shot5)
PRT062607	81.20	**100.00**
BAY 11-7082	38.19	**89.69**
ZSTK474	33.33	**97.73**
Vorinostat	58.09	**100.00**
YM201636	38.19	**97.73**
Cediranib	67.63	**93.15**

effectively distinguish resistance in cells with similar cluster spaces (subgroups A). The model achieved an AUC of 0.99, with only a few predictions failing, as indicated in Table 3 and Figure 4(a). Critical genes were filtered based on p-values <0.05, log-fold changes <0.1, and more than 20% of cells in either comparison group exhibited a significant expression difference. The results showed that of all 16,906 genes, 1,897 were up-regulated in drug-sensitive cells and 1,712 were up-regulated in drug-resistant cells, both showing significantly different attribution gradients. Now, considering both up-regulated and down-regulated genes in the resistant group compared to the sensitive group, genes with a log-fold change greater than threshold $th1$ are considered to be significantly up-regulated, while those with a log-fold change lower than threshold $th2$ are considered to be significantly down-regulated (Figure 4(b)). We analyzed the distribution of the log-fold changes for genes and determined thresholds $th1$ and $th2$ at the 5% and 95% percentile, respectively. We took the top 5 up-regulated and down-regulated genes and plotted them on a violin graph (Figure 4(c)). We implemented gene set enrichment analysis (GSEA) of the 1897 drug-sensitive genes and 1712 drug-resistant genes by Decoupler[40]. The significant enrichment pathways were filtered by p-values <0.05, and the results were presented in Figure S3, S4, and Supplementary Data S2, S3. We find that the genes sensitive to Cisplatin are significantly down-regulated on pathway "recognition of DNA damage by PCNA containing replication complex" (score=-0.416, p-value=0.038). Inefficient identification of DNA damage may lead to a delay or failure in the initiation of the DNA repair process, resulting in the accumulation of various DNA lesions such as mismatches, insertions, deletions and crosslinks[41, 42], which is one of the biological processes accociated with Cisplatin-sensitive. At the same time, we also observed the "DNA_replication" pathway was up-regulated in drug-resistant cells (score=0.405, p-value=0.078). This enhanced replicate capacity enables cancer cells to repair Cisplatin-induced DNA cross-links more effectively, contributing to resistance.

On the other hand, we also noticed that Cisplatin-resistant critical genes are significantly enriched in the pathway related to "APC-C mediated degradation of cell cycle proteins" (score=0.579, p-value=0.019), which plays a critical role during mitosis. The Anaphase-Promoting Complex or Cyclosome (APC/C) is an E3 ubiquitin ligase that regulates the cell cycle by targeting specific cell cycle proteins for degradation, thus controlling progression through mitosis and the G1 phase of the cell cycle. The effectiveness of Cisplatin largely depends on its ability to induce DNA damage that leads to cell cycle arrest and apoptosis[24]. The up-regulation of APC/C activity can lead to an accelerated transition to mitosis through the timely degradation of key cell cycle inhibitors. This rapid progression may reduce the time available for Cisplatin-induced DNA damage to be detected and repaired before cell division. This scenario potentially allows DNA-damaged cells to continue dividing, which can contribute to the development of resistance.

Specifically, we identified 36 significant pathways associated with drug resistance (p<0.05), as illustrated in Figure 4 (e). In this analysis, the correlation coefficient measures the relationship between the drug response predicted by MetaSCDrug and each pathway. A positive enrichment score indicates that a pathway is up-regulated or enriched in the population, whereas a negative enrichment score suggests that the pathway is down-regulated or depleted. Remarkably, the direction of regulation is opposite in 20 pathways compared to the drug-sensitive group. Additionally, other critical pathways, such as those involved in cell mitosis and death, are listed in Figure 4 (d). These pathways show different or even diametrically opposite correlations between sensitivity and resistance. This analysis demonstrates that MetaSCDrug can effectively identify critical genes and pathways associated with drug response, highlighting its utility in uncovering the complex biological dynamics that contribute to drug resistance.

5. Discussion

SC pharmacogenomic analysis focuses on the resistance of a particular cell population, which is beneficial for precision medicine. However, SC-level drug response studies face challenges due to limitations in sequencing technology and limited sample size. In this work, we integrate molecular information with transcriptomic data and develop a generalization model to predict cellular responses to pre-trained and unseen drugs. However, it is important to note that while MetaSCDrug has demonstrated potential in predicting the response of cancer drugs, it also has certain limitations. We infer that the limitations are mainly caused by the lack of transcriptomic knowledge of bulk cell lines under drug selection, the requirement of diverse SC drug response data, and the availability of transferable multimodal data.

Limit knowledge from bulk RNA-seq. The gain in post-treatment prediction accuracy is not significant, although some prediction errors can still be corrected, indicating that the bulk-based pre-training method still has limitations in knowledge transfer. The reason for this may be the lack of knowledge about bulk transcriptome of post-treatment, so collecting and integrating post-treatment transcriptome data could be a solution.

Clinical generalization. The applicability of our proposed model in real-world clinical settings has yet to be thoroughly assessed, primarily due to the limited diversity of clinical cell types currently presented in SC-level drug response datasets. There is potential for the model to be extended on a wider array of SC datasets, potentially enhancing its generalizability and prediction accuracy across more clinical scenarios.

Multimodal data transfer learning. Our framework is constrained by the single modality of information available in SC datasets, with a particular focus on gene expression data. This limitation hinders the development of pre-trained models that can leverage multiple modalities, such as genetic mutations and methylation patterns. Expanding the model to incorporate these data types may significantly enhance its prediction accuracy and allow for a more comprehensive understanding of drug responses at the cellular level.

6. Conclusion

In summary, we propose MetaSCDrug to predict cellular responses to multiple drugs, which overcomes the limitations of model generalization. Firstly, we introduce a meta pre-training and few-shot adaptive distribution alignment strategy for cross-domain transfer while minimizing the sample size requirement. Secondly, the proposed model can generalize to drugs not encountered during pre-training, suggesting that the model can be extended to a wider range of SC-based drug screening applications. Thirdly, attribution gradient methods can be employed to identify crucial genes and pathways under multi-modal inputs, providing a new way to understand the resistance mechanisms. Our work could potentially guide clinical precision therapy by predicting resistance based on scRNA-seq data. In the long term, MetaSCDrug may also benefit the virtual screening of anti-cancer drugs discovery.

Supplementary Materials

Supplementary material S1. Figure S1∼S4

Supplementary Data S1. Unseen drug response prediction assessment.

Supplementary Data S2. GSEA results of the 1897 drug-sensitive genes.

Supplementary Data S3. GSEA results of the 1712 drug-resistant genes.

Data availability

All datasets analyzed in this study are publicly available. Details of the data resources are described in Section 3.1.

Funding

The research is supported by the Major Key Project of PCL (PCL2021A13).

CRediT authorship contribution statement

Shuang Ge: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Shuqing Sun:** Writing – original draft, Writing – review & editing. **Huan Xu:** Methodology, Writing – original draft, Writing – review & editing. **Yiming Ren:** Writing – original draft, Writing – review & editing. **Zhixiang Ren:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relations that could have appeared to influence the work reported in this paper.

Acknowledgements

We appreciate the Peng Cheng Cloud-Brain.

References

[1] Swetha Vasudevan, Ibukun A Adejumobi, Heba Alkhatib, Sangita Roy Chowdhury, Shira Stefansky, Ariel M Rubinstein, and Nataly Kravchenko-Balasha. Drug-induced resistance and phenotypic switch in triple-negative breast cancer can be controlled via resolution and targeting of individualized signaling signatures. *Cancers*, 13(19):5009, 2021.

[2] Andriy Marusyk, Michalina Janiszewska, and Kornelia Polyak. Intratumor heterogeneity: the rosetta stone of therapy resistance. *Cancer cell*, 37(4):471–484, 2020.

[3] Suzanne Carreira, Alessandro Romanel, Jane Goodall, Emily Grist, Roberta Ferraldeschi, Susana Miranda, Davide Prandi, David Lorente, Jean-Sebastien Frenel, Carmel Pezaro, et al. Tumor clone dynamics in lethal prostate cancer. *Science translational medicine*, 6(254):254ra125–254ra125, 2014.

[4] Wanjuan Yang, Jorge Soares, Patricia Greninger, Elena J Edelman, Howard Lightfoot, Simon Forbes, Nidhi Bindal, Dave Beare, James A Smith, I Richard Thompson, et al. Genomics of drug sensitivity in cancer (gdsc): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic acids research*, 41(D1):D955–D961, 2012.

[5] Zhixiang Ren, Yiming Ren, Pengfei Liu, and Huan Xu. Cytokine expression patterns: A single-cell rna sequencing and machine learning based roadmap for cancer classification. *Computational Biology and Chemistry*, page 108025, 2024.

[6] G Gambardella, G Viscido, B Tumaini, A Isacchi, R Bosotti, and D di Bernardo. A single-cell atlas of breast cancer cell lines to study tumour heterogeneity and drug response. *bioRxiv*, pages 2021–03, 2021.

[7] Orit Rozenblatt-Rosen, Aviv Regev, Philipp Oberdoerffer, Tal Nawy, Anna Hupalowska, Jennifer E Rood, Orr Ashenberg, Ethan Cerami, Robert J Coffey, Emek Demir, et al. The human tumor atlas network: charting tumor transitions across space and time at single-cell resolution. *Cell*, 181(2):236–249, 2020.

[8] Xiaodun Li, Hayley E Francies, Maria Secrier, Juliane Perner, Ahmad Miremadi, Núria Galeano-Dalmau, William J Barendt, Laura Letchford, Genevieve M Leyden, Emma K Goffin, et al. Organoid cultures recapitulate esophageal adenocarcinoma heterogeneity providing a model for clonality studies and precision therapeutics. *Nature communications*, 9(1):2983, 2018.

[9] Jordi Barretina, Giordano Caponigro, Nicolas Stransky, Kavitha Venkatesan, Adam A Margolin, Sungjoon Kim, Christopher J Wilson, Joseph Lehár, Gregory V Kryukov, Dmitriy Sonkin, et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391):603–607, 2012.

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

[11] Simona Pellecchia, Gaetano Viscido, Melania Franchini, and Gennaro Gambardella. Predicting drug response from single-cell expression profiles of tumours. *BMC medicine*, 21(1):476, 2023.

[12] Junyi Chen, Xiaoying Wang, Anjun Ma, Qi-En Wang, Bingqiang Liu, Lang Li, Dong Xu, and Qin Ma. Deep transfer learning of cancer drug responses by integrating bulk and single-cell rna-seq data. *Nature Communications*, 13(1):6494, 2022.

[13] Zetian Zheng, Junyi Chen, Xingjian Chen, Lei Huang, Weidun Xie, Qiuzhen Lin, Xiangtao Li, and Ka-Chun Wong. Enabling single-cell drug response annotations from bulk rna-seq using scad. *Advanced Science*, 10(11):2204113, 2023.

[14] Rachel A Hodos, Brian A Kidd, Khader Shameer, Ben P Readhead, and Joel T Dudley. In silico methods for drug repurposing and pharmacology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 8(3):186–210, 2016.

[15] Yimeng Wang, Xinxin Yu, Yaxin Gu, Weihua Li, Keyun Zhu, Long Chen, Yun Tang, and Guixia Liu. Xgraphcds: An explainable deep learning model for predicting drug sensitivity from gene pathways and chemical structures. *Computers in Biology and Medicine*, 168:107746, 2024.

[16] Xiangxiang Zeng, Xiang Song, Tengfei Ma, Xiaoqin Pan, Yadi Zhou, Yuan Hou, Zheng Zhang, Kenli Li, George Karypis, and Feixiong Cheng. Repurpose open data to discover therapeutics for covid-19 using deep learning. *Journal of proteome research*, 19(11):4624–4636, 2020.

[17] Sofia D'Souza, KV Prema, and Seetharaman Balaji. Machine learning models for drug–target interactions: current knowledge and future directions. *Drug Discovery Today*, 25(4):748–756, 2020.

[18] Qianru Sun, Yaoyao Liu, Tat-Seng Chua, and Bernt Schiele. Meta-transfer learning for few-shot learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 403–412, 2019.

[19] Danielle Maeser, Weijie Zhang, Yingbo Huang, and R Stephanie Huang. A review of computational methods for predicting cancer drug response at the single-cell level through integration with bulk rnaseq data. *Current Opinion in Structural Biology*, 84:102745, 2024.

[20] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7167–7176, 2017.

[21] Muhammad Ghifary, W Bastiaan Kleijn, and Mengjie Zhang. Domain adaptive neural networks for object recognition. In *PRICAI 2014: Trends in Artificial Intelligence: 13th Pacific Rim International Conference on Artificial Intelligence, Gold Coast, QLD, Australia, December 1-5, 2014. Proceedings 13*, pages 898–904. Springer, 2014.

[22] Tuan Nguyen, Giang TT Nguyen, Thin Nguyen, and Duc-Hau Le. Graph convolutional networks for drug response prediction. *IEEE/ACM transactions on computational biology and bioinformatics*, 19(1):146–154, 2021.

[23] Thang Chu, Thuy Trang Nguyen, Bui Duong Hai, Quang Huy Nguyen, and Tuan Nguyen. Graph transformer for drug response prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 20(2):1065–1072, 2022.

[24] Sunghwan Kim, Paul A Thiessen, Evan E Bolton, Jie Chen, Gang Fu, Asta Gindulyte, Lianyi Han, Jane He, Siqian He, Benjamin A Shoemaker, et al. Pubchem substance and compound databases. *Nucleic acids research*, 44(D1):D1202–D1213, 2016.

[25] David Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36, 1988.

[26] Greg Landrum et al. Rdkit: Open-source cheminformatics and machine learning, 2021.

[27] Gabriela S Kinker, Alissa C Greenwald, Rotem Tal, Zhanna Orlova, Michael S Cuoco, James M McFarland, Allison Warren, Christopher Rodman, Jennifer A Roth, Samantha A Bender, et al. Pan-cancer single-cell rna-seq identifies recurring programs of cellular heterogeneity. *Nature genetics*, 52(11):1208–1218, 2020.

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

[29] Oscar Franzén, Li-Ming Gan, and Johan LM Björkegren. Panglaodb: a web server for exploration of mouse and human single-cell rna sequencing data. *Database*, 2019:baz046, 2019.

[30] F Alexander Wolf, Philipp Angerer, and Fabian J Theis. Scanpy: large-scale single-cell gene expression data analysis. *Genome biology*, 19:1–5, 2018.

[31] RA Dorfman. A note on the! d-method for finding variance formulae. *Biometric Bulletin*, 1938.

[32] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[33] Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012.

[34] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[35] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

[36] Srivastava Nitish. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15:1, 2014.

[37] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323. JMLR Workshop and Conference Proceedings, 2011.

[38] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks. In *International conference on machine learning*, pages 3319–3328. PMLR, 2017.

[39] Paul B Tchounwou, Shaloam Dasari, Felicite K Noubissi, Paresh Ray, and Sanjay Kumar. Advances in our understanding of the molecular mechanisms of action of cisplatin in cancer therapy. *Journal of experimental pharmacology*, pages 303–328, 2021.

[40] Pau Badia-i Mompel, Jesús Vélez Santiago, Jana Braunger, Celina Geiss, Daniel Dimitrov, Sophia Müller-Dott, Petr Taus, Aurelien Dugourd, Christian H Holland, Ricardo O Ramirez Flores, et al. decoupler: ensemble of computational methods to infer biological activities from omics data. *Bioinformatics Advances*, 2(1):vbac016, 2022.

[41] Wei-Lun Huang, Chi-Wen Luo, Chia-Lin Chou, Ching-Chieh Yang, Tzu-Ju Chen, Chien-Feng Li, and Mei-Ren Pan. High expression of ube2b as a poor prognosis factor in patients with rectal cancer following chemoradiotherapy. *Anticancer research*, 40(11):6305–6317, 2020.

[42] Errol C Friedberg, Alan R Lehmann, and Robert PP Fuchs. Trading places: how do dna polymerases switch during translesion dna synthesis? *Molecular cell*, 18(5):499–505, 2005.