Briefings in Bioinformatics, 2026, 27, bbag337
https://doi.org/10.1093/bib/bbag337
Published: 24 June 2026
Problem Solving Protocol

OXFORD

scTumorDrug: predicting cell-type-specific drug responses for heterogeneous tumors

Qiang Zhang‡, Qing Zhang‡, Chunming Guo*, Cheng Peng*

Yunnan Key Laboratory of Cell Metabolism and Diseases, Center for Life Sciences, School of Life Sciences, Yunnan University, Easter Outer Ring Road, Chenggong District, Kunming 650500, China

*Corresponding authors. Chunming Guo and Cheng Peng, Yunnan Key Laboratory of Cell Metabolism and Diseases, Center for Life Sciences, School of Life Sciences, Yunnan University, Kunming 650500, China. E-mail: chunmingguo@ynu.edu.cn; chengpeng@ynu.edu.cn
‡Qiang Zhang and Qing Zhang contributed equally to this work.

Abstract

It is important to predict cell-type-specific drug responses within the heterogeneous tumors for precision medicine. The single-cell RNA sequencing (scRNA-seq) technique together with drug responses data provides opportunities for this mission. The previous methods were mainly evaluated in datasets derived from cancer cell lines, lacking direct validations on the real tumors derived from mouse models and clinical human samples. In this work, we integrated the labeled scRNA-seq, bulk RNA, and drug response data to develop the computational tool scTumorDrug to predict cell-type-specific drug responses for heterogeneous tumors. Overall, scTumorDrug achieved accurate predictions in public datasets derived from cell lines, mouse models, and clinical human samples, and outperformed previous methods in the selected datasets. Since there is no available experimental validation on drug responses from tumor subpopulations, we established the mouse bladder tumor model to investigate the cell-type-specific drug responses for the small molecule JQ1 by performing single nucleus RNA sequencing and spatial transcriptomics. The mouse model showed that the JQ1 treatment delayed the tumor progression though it did not completely eliminate the bladder tumors. The scTumorDrug found that the Mki67+ urothelial subpopulation was sensitive to JQ1, which was cross-validated by spatial transcriptomics. This observation also provided an explanation for JQ1 efficacy. To summarize, the scTumorDrug can be used in broad application scenarios and we provided cell-type-specific drug response prediction and validation.

Keywords drug response, tumor, single cell, spatial transcriptomics, JQ1

Introduction

Tumor heterogeneity plays key roles in tumor evolution and treatment resistance, in which distinct cell subpopulations may respond differently to drugs [1–3]. Thus, it is important to utilize advanced technologies to develop precision treatment strategies for patients. The single-cell RNA sequencing (scRNA-seq) technique can reveal heterogeneous gene expression across cancer subpopulations under specific drug treatments [4]. However, existing drug response prediction methods developed for bulk data are difficult to apply directly to large-scale and highly complex single-cell data [5–9]. Therefore, there is an urgent need to develop computational methods for inferring cancer drug responses at the single-cell level.

Currently, studies leveraging large-scale drug sensitivity screening databases have accumulated extensive bulk RNA-seq data and drug response information from cancer cell lines (CCLs), enabling the development of various drug response prediction models. For instance, the Precily model can predict drug responses from bulk data but cannot be directly applied to single-cell data [9]. Recently, several studies integrated drug-related bulk data with scRNA-seq data to predict drug responses in scRNA-seq. For example, SCAD and scDEAL employ adversarial domain adaptation and transfer learning, respectively, to learn drug response patterns from bulk RNA-seq and transfer them to scRNA-seq data [10, 11]. Beyondcell calculates drug sensitivity scores for individual cells based on drug signature enrichment analysis, thereby identifying therapeutically distinct cell clusters [12]. scPDS uses a transformer-based deep learning model to predict single-cell drug sensitivity by converting gene expression into pathway activity scores [13]. scIDUC integrates large-scale cell line and single-cell transcriptomic data via canonical correlation analysis (CCA) and applies a regression model to infer drug responses at single-cell resolution [14]. Single-cell drug response (scDR) employs differential expression analysis and scoring-based integration to predict drug response directly from scRNA-seq data using pre-identified drug-response genes [15]. Although these methods have achieved preliminary success in drug response prediction, they generally rely solely on labeled bulk data as the source domain for model training before transferring knowledge to unlabeled single-cell data. As a result, their predictive performance remains limited at single-cell resolution.

Received: January 15, 2026. **Revised:** April 30, 2026. **Accepted:** May 27, 2026

To address this challenge, we utilized the labeled single-cell and bulk RNA data from CCLs to develop a computational tool [6, 16–18], called scTumorDrug in this work, to predict cell-type-specific drug responses at single-cell resolution. We systematically validated the performance of scTumorDrug in different application scenarios, including cell lines, mouse models, and clinical human data. The model scTumorDrug achieved accurate predictions in CCLs. In solid tumors, the scTumorDrug accurately distinguished cisplatin-sensitive and -resistant tumors in both mouse bladder cancer and clinical human bladder cancer samples. Overall, the scTumorDrug performed better than the existing methods, including SCAD, scDEAL, and Beyondcell, in the selected datasets. Based on the successful application in public datasets, we generated bladder cancer mouse model and used the scTumorDrug to investigate the roles of drug JQ1 in treating bladder cancer by performing single nucleus RNA sequencing (snRNA-seq) and spatial transcriptomics. The scTumorDrug predicted that the Mki67+ urothelial subpopulation was sensitive to JQ1, and this prediction was further validated. Collectively, the scTumorDrug can be applied to broad scenarios to predict cell-type-specific drug responses in heterogeneous tumors, helping to advance precision medicine.

Materials and methods

Public datasets

The scTumorDrug training included two categories of cancer bulk gene expression data and corresponding drug response data. These data were downloaded from the Cancer Dependency Map (DepMap, https://depmap.org/portal/), the Cancer Cell Line Encyclopedia, and the Cancer Therapeutics Response Portal v2 [5, 16–19]. Gene expression and drug response data were downloaded from the Genomics of Drug Sensitivity in Cancer (GDSC) (https://www.cancerrxgene.org/) [6]. The CCL scRNA data were downloaded from National Center for Biotechnology Center (NCBI) with accession number GSE157220 (cancer type composition shown in Supplementary Fig. S1) [19]. The scRNA-seq data for I-BET-treated Mixed Lineage Leukemia-AF9 (MA9) leukemic cells and Bortezomib-treated MCF7 breast cancer cell were downloaded from GSE110894 and GSE114461 [20, 21], respectively. The scRNA-seq data for cisplatin-sensitive/resistant mouse bladder cancer tumors and cisplatin- and gemcitabine-sensitive/resistant human bladder cancer tumors were downloaded from GSE192575 [22]. The scRNA-seq data for chemonaive and chemotreated (carboplatin–paclitaxel) human ovarian tumors were downloaded from GSE201047 [23].

Data processing

Following the scIDUC method, we performed $\mathbf{log_2\,(TPM+1)}$ transformation on the raw bulk gene expression data. For each drug-cell line pair, we normalized the area under the dose-response curve (AUC) by dividing it by the tested dose range to generate the normalized AUC (nAUC). Cell line names in both the downloaded bulk gene expression and drug response data were harmonized using the Cellosaurus accession numbers with the prefix "CVCL" [24, 25], as described in scIDUC. Since each drug was only screened in a subset of cell lines, we calculated the proportion of missing values per drug and excluded those with coverage below 40% across all cell lines. For the single-cell data, the GSE157220 dataset, already in Counts Per Million (CPM) format, was further $\log_2$-transformed. The remaining single-cell datasets were uniformly normalized using $\log_2(\mathrm{CPM}+1)$.

The drug response-relevant genes (nDRGs) were derived from scIDUC, which applied the limma package to fit linear models on bulk RNA data and ranked genes by using the B-statistics to determine the strength of their association with drug response. Each drug corresponded to a ranked nDRG list. When training a prediction model for a specific drug, the top 1500 ranked genes were selected and then intersected with genes from the bulk gene expression data, CCL scRNA data, and the tumor scRNA data, ensuring a consistent gene set across all datasets.

We converted continuous nAUC values into binary labels for drug response prediction. Specifically, cell lines were first ranked in descending order based on their nAUC values to generate the nAUC cell line curve, where the *x*-axis represents cell lines and the *y*-axis represents nAUC values. The strategy proposed in scDEAL was used to separately determine the nAUC cutoff for linear and non-linear trend, since scDEAL validated the utility of this strategy. For curves exhibiting a linear trend (Pearson correlation >0.9 for the fitted regression line), the sensitive/resistant cutoff was set as the median nAUC across all cell lines. For non-linear curves, the cutoff was defined as the nAUC value of the boundary point with the largest vertical distance to the line connecting the points with the maximum and minimum nAUC values. For each drug, cell lines with nAUC below the cutoff were labeled as drug-sensitive (label 0), and those above the cutoff were labeled as drug-resistant (label 1). To ensure data balance, drugs with fewer than 20% of cell lines in either the sensitive or resistant category were excluded. The Pearson correlations and cutoff diagrams for the drugs used were provided in Supplementary Figs S2 and S3. However, it should be noted that other methods can be used to define drug sensitivity [10, 15], except the nAUC cutoff strategy used in this work.

Model training of scTumorDrug

The overall scTumorDrug framework consisted of three parts: training a shared encoder (SE) to extract latent representations between bulk RNA and target scRNA data; using the dimensionality-reduced representations from both bulk RNA and CCL scRNA data, along with their corresponding drug response labels, to train a multilayer perceptron (MLP) as a drug response predictor; and finally, applying the trained SE and MLP to the target scRNA data to predict drug responses.

First, the SE was optimized through a multitask learning paradigm integrated with domain adversarial training. This paradigm not only performed two reconstruction tasks: bulk decoder (BD) and single-cell decoder (SD), where the zero-inflated negative binomial (ZINB) loss was used in single-cell reconstruction to handle data sparsity, but also utilized the domain classifier (DC) to force the SE to learn domain-invariant features through a gradient reversal layer (GRL), thereby effectively aligning the feature distributions between bulk data (X_b) and the target single-cell data (X_s). The reconstruction loss functions of the bulk data and the target tumor scRNA data $\mathbf{loss_{Rec\_}}_b$ and $\mathbf{loss_{Rec\_}}_s$ were defined as:

$$\begin{cases} \mathrm{loss}_{Rec\_b} = \mathbf{MSEloss}\left(\mathbf{BD}\left(E_b\right), X_b\right) \\ \mathbf{loss_{Rec\_}}_s = \mathbf{ZINBloss}\left(\mathbf{SD}\left(E_s\right), X_s\right) \end{cases} \tag{1}$$

$$\mathbf{MSEloss} = \frac{1}{N \cdot n} \sum_{j=1}^{N} \sum_{i=1}^{n} \left(x_{ji} - y_{ji}\right)^2 \tag{2}$$

where $E_b = SE\left(X_b\right)$ and $E_s = SE\left(X_s\right)$. **SE**, **BD**, and **SD** represent the shared encoder, bulk decoder, and single-cell decoder, respectively,

while E_b and E_c denote the dimensionality-reduced bulk RNA and tumor scRNA, respectively. The Mean Squared Error (MSE) loss function was used to minimize the difference between the predicted values and the true values, in which x_{ij} represents the true value of the j-th feature of the i-th sample, and y_{ij} represents the corresponding predicted value from the model.

We employed the ZINBloss function [26] to model the count distribution of scRNA data by considering the overdispersion and zero-inflation characteristics of gene expression:

$$\mathbf{ZINBloss} = \frac{1}{N \bullet n} \sum_{i=1}^{N} \sum_{j=1}^{n} L_{\mathbf{ZINB}}\left(m_{ij}, d_{ij}, p_{ij}; x_{ij}\right) \tag{3}$$

$$L_{\mathbf{ZINB}}\left(m, d, p; x\right) = \begin{cases} -\log\left[p + (1-p) \cdot NB\left(m, d; x\right)\right], if\ x = 0 \\ -\log\left[(1-p) \bullet NB\left(m, d; x\right)\right], \quad if\ x > 0 \end{cases} \tag{4}$$

$$NB\left(m, d; x\right) = \begin{cases} \left(\frac{d}{d+m}\right)^{d}, & if\ x = 0 \\ \frac{\Gamma(x+d)}{\Gamma(x+1)\Gamma(d)} \left(\frac{d}{d+m}\right)^{(d+x)} \left(\frac{m}{d}\right)^{x}, & if\ x > 0 \end{cases} \tag{5}$$

where x_{ij} denotes the observed count of gene j in cell i, m_{ij} is the scaled mean parameter of the negative binomial distribution, d_{ij} is the dispersion parameter, p_{ij} is the zero-inflation probability for data comprising N cells and n genes, and Γ represents the Gamma function. For observed zero counts, the loss function distinguishes between technical zeros (caused by sequencing dropouts) and biological zeros (where the gene is genuinely not expressed) through the zero-inflation probability p. For positive counts, it describes the over-dispersion phenomenon of count data using the two key parameters of the negative binomial distribution: the mean, m, which represents gene expression abundance, and the dispersion parameter, d, which captures intercellular heterogeneity. The Gamma function extends the factorial to the continuous domain, thereby enabling the negative binomial distribution to utilize a continuous dispersion parameter that flexibly captures the over-dispersion characteristics of gene expression counts.

We used the Binary Cross-Entropy (BCE) loss function in the adversarial domain adaptation:

$$\mathbf{loss}_{\mathbf{Adv}} = \mathbf{BCEloss}\left(\mathbf{DC}\left(\mathbf{GRL}\left(E_b\right)\right), 0\right) + \mathbf{BCEloss}\left(\mathbf{DC}\left(\mathbf{GRL}\left(E_{sc}\right)\right), 1\right) \tag{6}$$

$$\mathbf{BCEloss}\left(p, y_i\right) = -\frac{1}{N} \sum_{i=1}^{N} y_i \cdot \log(p) + \left(1 - y_i\right) \cdot \log\left(1 - p\right) \tag{7}$$

where **GRL** and **DC** represent the gradient reversal layer and the domain classifier, respectively, and the variable i denotes the sample sizes for either bulk RNA or target scRNA data. For data with a sample size of N, y_i (0 or 1) is the label of the i-th sample, and p is the predicted probability for the sample.

The drug response predictor (MLP) was trained using the dimensionality-reduced features of the bulk and CCL scRNA (X_{Ls}) data from the SE, along with their corresponding labels. The labels for CCL scRNA (L_{Ls}) were derived from the bulk drug response labels (L_b). Then the Cross-Entropy Loss (CSEloss) was used in the binary classification for the bulk and CCL scRNA data:

$$\begin{cases} \mathbf{loss}_{C\_b} = \mathbf{CSEloss}\left(M_b, L_b\right) \\ \mathbf{loss}_{C\_Ls} = \mathbf{CSEloss}\left(M_{Ls}, L_{Ls}\right) \end{cases} \tag{8}$$

$$\mathbf{CSEloss} = -\frac{1}{N} \sum_{i=1}^{N} \log\left(\frac{\exp\left(z_{i,y_i}\right)}{\exp\left(z_{i,0}\right) + \exp\left(z_{i,1}\right)}\right) \tag{9}$$

where $\mathbf{loss}_{C\_b}$ and $\mathbf{loss}_{C\_Ls}$ represent losses for the bulk and CCL scRNA data, $M_b = \mathbf{MLP}\left(\mathbf{SE}\left(X_b\right)\right)$ and $M_{Ls} = \mathbf{MLP}\left(\mathbf{SE}\left(X_{Ls}\right)\right)$ denote the output matrices derived from MLP predictor for the bulk and CCL scRNA data, respectively. Given data with a batch size of N, each sample is represented by two logit values $z_{i,0}$ and $z_{i,1}$, which indicate the raw scores for the two classes. Here, y_i is the true class index of the i-th sample.

In the model training process, the total loss function was finally defined as the following formula:

$$\begin{aligned} \mathbf{loss}_{\mathbf{All}} = {} & w_1 \mathbf{loss}_{\mathbf{Rec\_}b} + w_2 \mathbf{loss}_{\mathbf{Rec\_}s} + w_3 \mathbf{loss}_{C\_b} \\ & + w_4 \mathbf{loss}_{C\_Ls} + w_5 \mathbf{loss}_{Adv} \end{aligned} \tag{10}$$

where the weight coefficients w_1, w_2, w_3, w_4, and w_5 were set to 0.1, 0.1, 1, 1, and 0.01 in this work, respectively. The training process utilized the Adam optimizer with a learning rate of 0.005, and no batch size was predefined. The model was trained for a total of 300 epochs.

Model prediction of scTumorDrug

Upon completion of model training, the SE and the predictor MLP were used in drug response prediction. Specifically, the dimensionality-reduced data of the target tumor scRNA E_{sc} were input into the MLP for prediction:

$$M_{sc} = \mathbf{MLP}\left(E_{sc}\right) \tag{11}$$

Then, the argmax function was used to convert the 2D matrix M_{sc} into the 1D binary labels L_{SC}:

$$L_{SC} = argmax\left(M_{SC}, \mathbf{dim} = 1\right) \tag{12}$$

where the output 0 denotes a drug-sensitive cell and output 1 denotes a drug-resistant cell.

We employed the integrated gradients attribution method to quantify both the magnitude and direction of gene influence on the predicted tolerance probability, in which the signed importance score was assigned to each input feature to represent its contribution to the model output. By using the all-zero expression vector as the uninformative baseline, a straight-line path was constructed from the baseline to the actual expression profile for each cell. At each step, we computed the gradient of the tolerance probability with respect to each gene's expression, and these stepwise gradients were averaged and multiplied by the difference between the gene's expression and the baseline, yielding the attribution score for the gene in the given cell. For each gene, this procedure was repeated for all cells, and the final importance score was obtained by averaging its attribution values across cells.

Model evaluation and comparison

Model performance was assessed using the five-fold cross validation. The CCL scRNA data and their corresponding labels were randomly partitioned into five non-overlapping subsets of equal size using a random seed. In each validation round, four subsets were used as the training set and the remaining one as the test set. This process was repeated five times, ensuring that each subset was used exactly once as the test set. The final model performance was evaluated based on the results from the five test folds (Supplementary Fig. S4).

In the model comparison, the single-cell expression data were normalized by $\log_2(\mathrm{CPM} + 1)$ for SCAD. This model output a sensitivity

score ranging from 0 to 1 for each cell. Based on its training paradigm, the labels 1 and 0 indicated sensitivity and resistance, respectively. Consequently, a cell population with a score >0.5 were classified as sensitive, and that with a score <0.5 was classified as resistant. The SCAD manual lists the optional values for the hyperparameters, but without giving the recommended ones. Then we used different combinations to evaluate the hyperparameter values for the tested datasets, and selected the overall better one for model comparison in this work (Supplementary Fig. S5). For drug response prediction using scDEAL, the parameters recommended for the corresponding drug were used. scDEAL directly provided a binary classification label (0 or 1) for each cell, where 1 denotes sensitive and 0 denotes resistant. The Beyondcell analysis was performed by directly inputting the Seurat object. Its output included a Beyondcell Sensitivity Score (BSC) for each cell against every compound in its drug database. A higher BSC indicates greater predicted sensitivity of a cell to a specific drug. To visualize these results, the BSC distribution was plotted as a violin plot.

Proteasome score calculation

To evaluate the overall activity level of the proteasome complex in individual cells, we selected 14 core proteasome subunit genes (*PSMA1–PSMA7* and *PSMB1–PSMB7*) as the feature set. By using the AddModuleScore function in the Seurat package [27], we calculated the Proteasome Score for each cell based on the expression profiles of these genes.

Copy number variation score calculation

To quantify genomic instability in tumor cells and distinguish malignant cells, we performed a somatic copy number variation (CNV) analysis. We employed the R package infercnv (v1.10.1) [28], using cancer-associated fibroblasts (CAFs) as a stable reference cell population, to infer CNV profiles in the remaining cells. Specifically, we calculated a CNV score for each cell based on the extent of CNV alterations, which effectively reflects the magnitude of genome-wide copy number changes. We used the non-parametric Wilcoxon rank-sum test to assess the statistical significance of CNV score differences between tumor and normal cells.

Gene Ontology enrichment analysis

We first identified significantly differentially expressed genes using the following criteria: adjusted *P*-value (adj.P.Val) <.05 and absolute log fold change |logFC| >0.5. Subsequently, Gene Ontology (GO) enrichment analysis was performed on these genes using the ClusterProfiler package (v4.1.0) [29]. Significantly enriched biological pathways were defined as those with statistical significance (adj.P.Val <.05) and were visualized using bubble plots.

Small molecule experiments of the N-Butyl-N-(4-hydroxybutyl) nitrosamine (BBN) model

All animal studies were performed according to protocols and guidelines reviewed and approved by the Institutional Animal Care and Use Committee at Yunnan University (approval numbers YNU20220176 and YNU20230525). C57BL/6 J mice (JAX#000664) were obtained from the Jackson Laboratory.

Following the method established by Beachy [30], 8-week-old male C57BL/6J mice were administered drinking water supplemented with 0.1% BBN (Sigma). The water bottles were light-protected, and the solution was refreshed twice weekly. BBN treatment was continued for 5.5 months. Between 3.5 and 5.5 months of BBN exposure, mice were treated with Dimethyl sulfoxide (DMSO) or JQ1 (5 mg/kg) i.p. twice a week. At the experimental endpoint, mice were euthanized, and bladder samples were collected for pathological evaluation and T staging. T stage was assessed according to the 8th edition American Joint Committee on Cancer criteria. T1, Tumor invasion of the lamina propria (subepithelial connective tissue). T2, Tumor invasion of the muscularis propria. T3, Tumor invasion of perivesical tissue. T4, Tumor invasion into any of the following: prostate, uterus or vagina, pelvic wall, or abdominal wall.

Hematoxylin and eosin staining

Deparaffinized sections were sequentially stained as follows: immersed in hematoxylin (Servicebio, G1005–1) for 5 min to stain nuclei, rinsed in water, placed in differentiation solution for 10 s, and then in bluing solution for 15 s, with thorough water rinsing between each solution. Subsequently, sections were sequentially stained in Eosin (Servicebio, G1005–2) for 1 min to stain the cytoplasm, followed by brief immersion in 95% ethanol, two changes of 100% ethanol (1 min each), and finally two changes of xylene (2 min each). Sections were then mounted using a xylene-based mounting medium and allowed to dry.

Formalin-Fixed Paraffin-Embedded (FFPE) sample preparation for single nucleus RNA sequencing and spatial transcriptomics

Freshly harvested mouse bladder tissues were fixed in 4% paraformaldehyde (ALDRICH, 441244) for 24 h at 4°C. The fixed tissues were then dehydrated through a graded ethanol series: 75%, 85%, and 95% ethanol, followed by twice in absolute ethanol (Sangon Biotech, A500737), for 30 min each step. Subsequently, the samples were cleared by immersion in xylene (DAMAO, 1330-20-7) twice, for 30 min each. Finally, the tissues were infiltrated with paraffin wax (CITOTEST) at 56°C twice, for 1 h each, and then embedded in paraffin blocks. The solidified blocks were trimmed and sectioned at a thickness of 25 μm using a microtome (LEICA, RM2245) for subsequent snRNA-seq, in which each group contained three different mouse samples. The paraffin blocks were also sectioned at 5 μm for spatial transcriptomics, with each chip containing three different mouse samples. The snRNA-seq [31] and spatial transcriptomics (spRandom-seq) [32] on FFPE samples were performed by M20 Genomics. The read alignment and count matrix generation in spatial transcriptomics were also performed by M20 Genomics.

Single nucleus RNA sequencing data processing

Reads were processed using the VITAseer v1.2 pipeline with default and recommended parameters. First, primer sequences and capture adaptor sequences were trimmed in raw sequencing data. For each Read1, we extracted Unique Molecular Identifier (UMI) (8 nts) and cell-specific barcode (20 nts) and merged sequenced barcodes that can be uniquely assigned to the same accepted barcode with a Hamming distance of 1 nts or less. Read2 was used to generate the gene expression matrix by the STARsolo module in STAR(2.7.10a) [33]. The valid nuclei were identified by STARsolo.

Single-nucleus and spatial transcriptomics data analysis

The preprocessed single-nucleus and spatial transcriptomics data were analyzed using Seurat v5 (v5.1.0) [27]. For the raw gene expression matrix, the NormalizeData function was used to normalize gene counts per cell, the FindVariableFeatures function was used to identify the highly variable genes, and the ScaleData function was used to scale gene expression values. The dimensionality reduction was conducted in principal component space using the RunPCA function. For the bladder cancer dataset, data integration was carried out using the IntegrateLayers function with the CCA method to correct batch effects while preserving biological variation. Other datasets were processed without applying the IntegrateLayers function. Finally, cell clustering was performed using the first 30 CCA dimensions from the integrated data, and the FindNeighbors function was used to construct a cell nearest-neighbor graph, followed by application of the FindClusters function implementing the Louvain algorithm for unsupervised clustering.

Results

Overview of scTumorDrug framework

scTumorDrug is a deep learning framework that predicts drug responses at single-cell resolution (Fig. 1). By employing cross-domain feature learning and multitask collaborative training, the model transfers the known gene expression signatures of drug responses from labeled bulk RNA and scRNA-seq derived from CCLs to the unlabeled tumor scRNA-seq data. This dual-supervision on both bulk RNA and scRNA enables the model to learn drug response patterns from two kinds of RNA-seq data, significantly enhancing prediction accuracy for heterogeneous tumor cell subpopulations.

In the model training module (Fig. 1A), we first selected the top 1500 genes from predefined gene expression signatures as nDRGs. The shared gene set across bulk RNA-seq, CCL scRNA, and target tumor scRNA data was input to an SE to learn the unified and comparable low-dimensional representations for these different kinds of gene expression data, and then the scDecoder and BulkDecoder were used to reconstruct bulk RNA and scRNA data, respectively, to preserve the intrinsic data structure. To address the high sparsity and zero-inflation characteristics of scRNA-seq data, the decoder reconstruction was performed using a ZINBLoss. To mitigate distributional discrepancies between bulk RNA and scRNA-seq data, the model further utilized an adversarial domain adaptation, in which a DC with a GRL forces the SE to learn domain-invariant representations, thereby effectively bridging the distribution gap between bulk RNA and scRNA data. We also utilized an MLP as the drug response predictor for the labeled data, in which the drug sensitivity labels from both bulk RNA and CCL scRNA data were used for model training to enhance the model generalizability and robustness in heterogeneous tumor cells. In this work, the SE was the fully connected networks with dimensions ranged from 1500, 1024, 512 to 256, and the dimensions of the scDecoder and BulkDecoder ranged from 256, 512, 1024 to 1500. The dimensions of the domain adaption were set from 256, 128, 64 to 1, and those of the MLP were set from 256, 128, 32 to 2.

As for the prediction module, the scTumorDrug used the SE to extract latent representations from tumor scRNA data, and used the trained MLP to output the drug response, sensitive or resistant, for each cell (Fig. 1B). These single-cell responses were further used to calculate the drug responses for each cell population. Thus, this computational pipeline transferred the knowledge from large-scale labeled data to the unlabeled tumor scRNA-seq data in predicting drug responses for different kinds of cell populations of heterogeneous tumors.

scTumorDrug achieved good drug response prediction in cancer cell lines

We evaluated the performance of scTumorDrug using the publicly available datasets, in which three existing methods, SCAD, scDEAL, and Beyondcell, were selected to perform comparisons. The scPDS, scIDUC, and scDR were excluded from comparisons due to either unavailable training data required in the model or the unavailable model code. In the dataset derived from MA9 leukemic cells, there exist four treatment conditions: two sensitive states (P1: DMSO and P2: I-BET 400 nM) and two resistant states (R1: I-BET RESISTANT and R2: I-BET RESISTANT WITHDRAWAL) (Fig. 2A) [20]. I-BET151 and I-BET-762 belong to the class of BET bromodomain inhibitors, which competitively bind to the bromodomains to block the recognition and binding of the BET family proteins to acetylated histones, thereby leading to the transcriptional suppression of key oncogenes and pro-inflammatory genes [34, 35]. The scTumorDrug was used to predict the responses to I-BET151 and I-BET-762, and the predicted results were consistent with the original study (Fig. 2B). We further compared the performances among scTumorDrug, SCAD, scDEAL, and Beyondcell. The scDEAL was excluded from the I-BET151 comparison since it does not include I-BET151 data, and the SCAD was excluded from the I-BET-762 comparison for the same reason. As shown in Fig. 2C, the scTumorDrug performed better than SCAD and Beyondcell in predicting cell line response to I-BET151. The scTumorDrug predicted sensitive cell proportions of 80.4% (P1), 73.9% (P2), 15.4% (R1), and 15.3% (R2), consistent with the labeled sensitive (P1 and P2) and resistant (R1 and R2) states. However, SCAD predicted sensitive cell proportions of 83.5% and 85.2% for R1 and R2, respectively, contradicting the experimentally validated resistant states. The Beyondcell reported median BSC scores of 0.58 (P1), 0.49 (P2), 0.50 (R1), and 0.58 (R2), in which higher BSC scores indicate greater sensitivity. These results showed that SCAD and Beyondcell could not distinguish well between sensitive and resistant samples. For I-BET-762, scTumorDrug predicted sensitive cell proportions of 71.9% (P1), 76.3% (P2), 24.8% (R1), and 26.1% (R2). The results from scDEAL were 96.2%, 86.3%, 9.6%, and 1.5%, respectively, and Beyondcell median BSC scores were 0.42, 0.42, 0.35, and 0.33. All three methods distinguished sensitive from resistant samples, with scDEAL showing the largest inter-group difference, followed by scTumorDrug (Fig. 2D).

We next evaluated scTumorDrug on MCF7 breast cancer cells. In the original experiment, MCF7 cells were treated and sampled with Bortezomib (500 nM) under four conditions: before treatment (t0), 12 h treatment (t12), 48 h treatment (t48), and after 72 h of treatment followed by drug washout and 24 h recovery (t96) (Fig. 2E, left panel) [21]. Bortezomib, a proteasome inhibitor, exerts its effect through decreased proteasome activity, which can be assessed using expression levels of relevant genes [36–39]. We therefore calculated the proteasome activity scores for MCF7 cells under the four conditions. The results indicated that the cells exhibited increased activity at t12

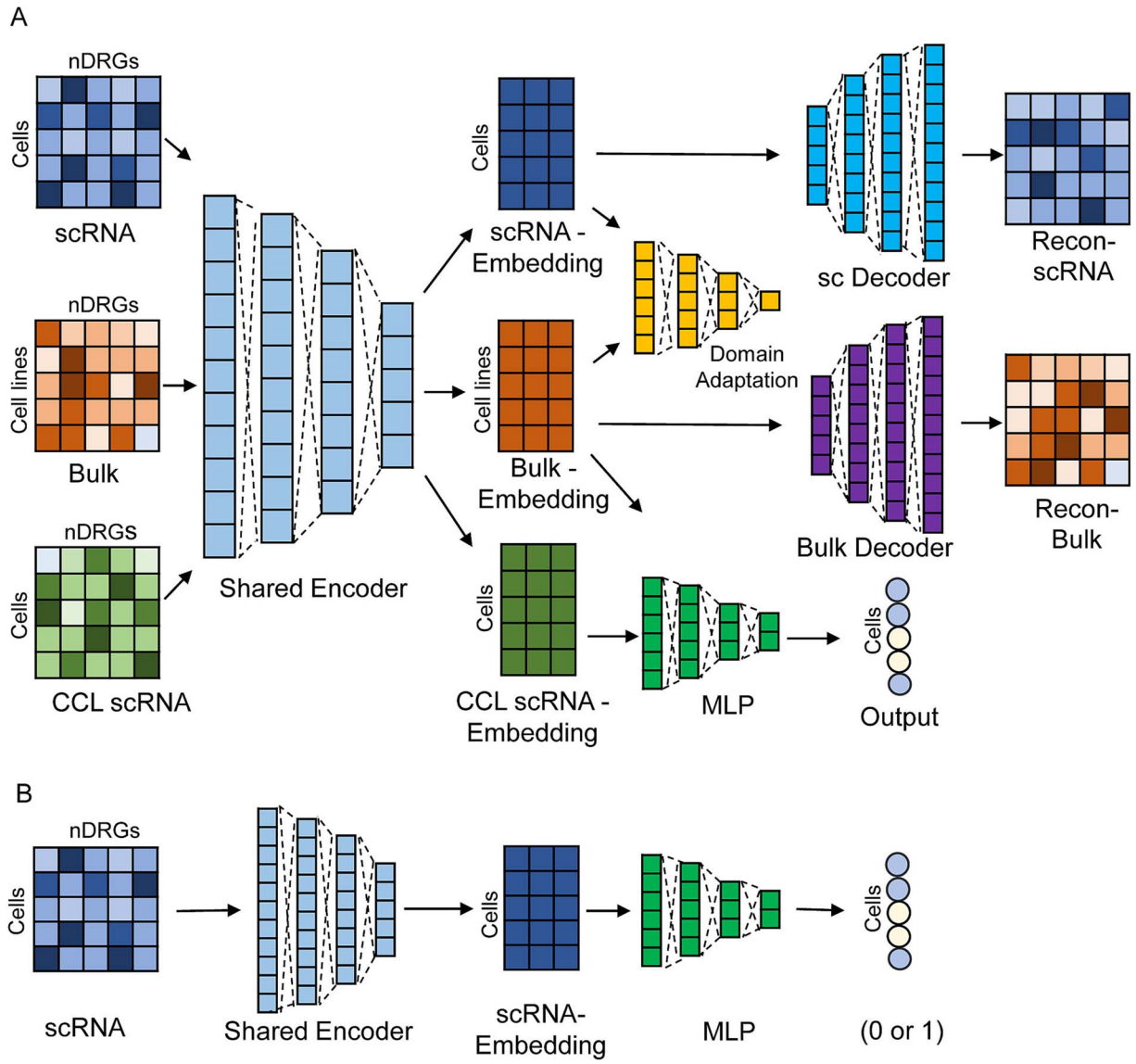

Figure 1 The scTumorDrug framework; (A) model training framework of scTumorDrug; (B) model prediction pipeline; CCL scRNA: scRNA-seq data derived from cancer cell lines; bulk: bulk RNA-seq data derived from CCL; scRNA: scRNA-seq data derived from tumors for prediction; recon: reconstruction.

and t48, and then decreased activity at t96 (Fig. 2E, right panel), consistent with the drug treatment procedure. Expectedly, scTumorDrug accurately predicted Bortezomib drug responses consistent with the original study, in which the cells at t0 and t96 were dominant sensitive cells and cells at t12 and t48 were mainly resistant cells (Fig. 2F). We subsequently compared the scTumorDrug predictions to those from SCAD, scDEAL, and Beyondcell. Notably, the predictions from SCAD and scDEAL did not match the experimental design and the observed trend in proteasome activity, whereas scTumorDrug and Beyondcell were consistent with this trend (Fig. 2G).

Collectively, in the cell line datasets, scTumorDrug demonstrated accurate predictions for all three tested drugs; Beyondcell and scDEAL achieved prediction power for two drugs and one drug, respectively, and SCAD only achieved good predictions in some but not all cell lines for each drug. These results demonstrated the applicability of scTumorDrug in predicting drug responses in CCLs.

scTumorDrug distinguished between sensitive and resistant tumors derived from mouse and human

We then evaluated the scTumorDrug performance on mouse and human tumors, which was almost neglected in previous methods. Wang *et al.* generated scRNA-seq data for drug-sensitive and drug-resistant bladder tumors from mouse and human, respectively, in which cisplatin was used in the mouse models and cisplatin–gemcitabine combination was used in human treatment [22]. Since bladder tumors are mainly composed of urothelial cells, we focused on predicting the drug responses of this cell population in the following analyses. In the mouse tumor samples, we identified eight distinct cell populations through clustering and annotation, including a major urothelial cell population (Fig. 3A and Supplementary Fig. S6). Using scTumorDrug, SCAD, scDEAL, and Beyondcell, we predicted the cisplatin response of urothelial cells in the mouse data. The

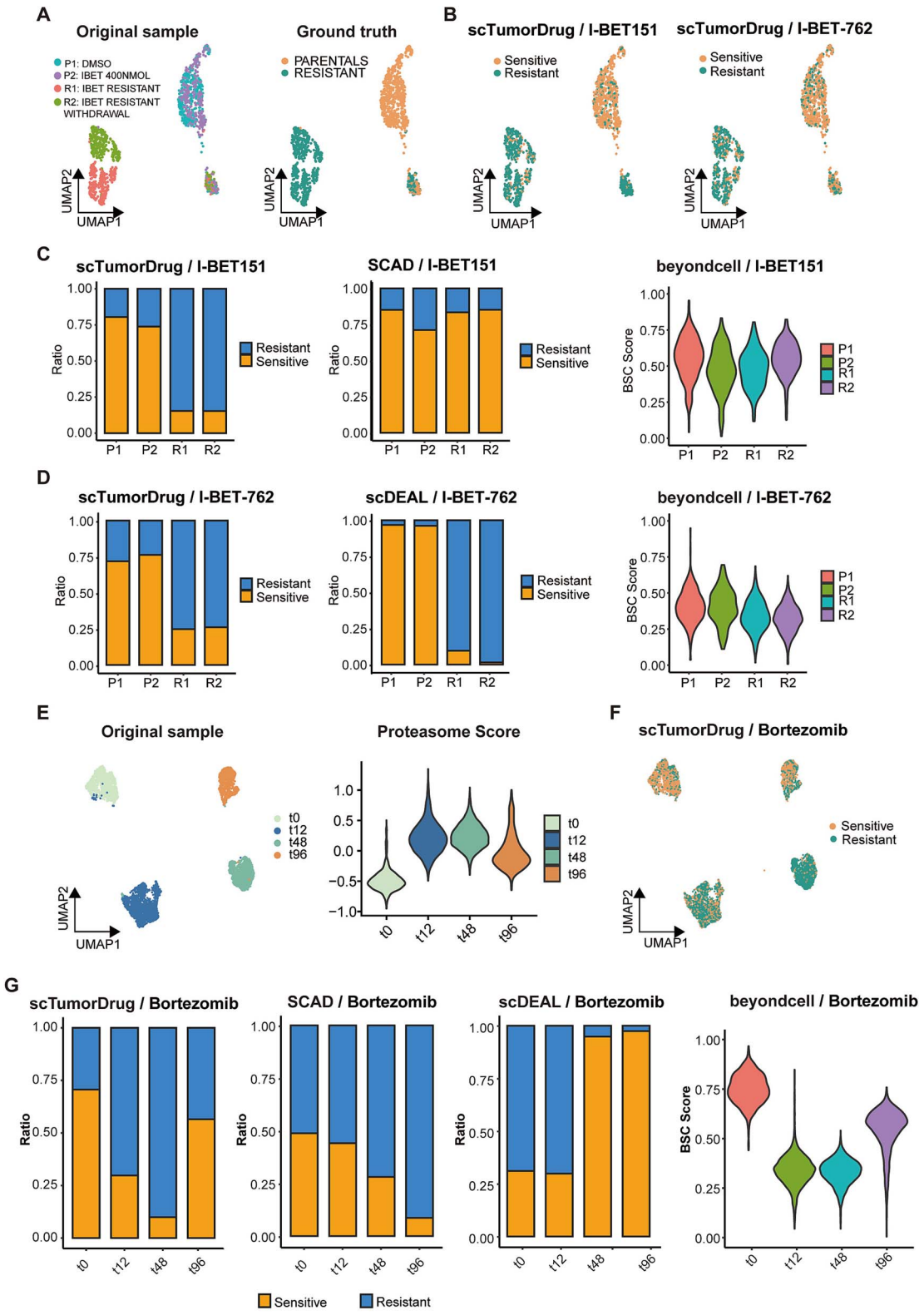

Figure 2 Evaluations on scTumorDrug using scRNA-seq data derived from CCLs; (A) UMAP visualization of treatment conditions (left) and the ground-truth drug responses (right) in the MA9 leukemic cells; (B) UMAP visualization of predicted drug responses for I-BET151 (left) and I-BET-762 (right); (C) comparison in predicting I-BET151 sensitivity; (D) comparison in predicting I-BET-762 sensitivity; (E) UMAP visualization of treatment conditions (left) and corresponding proteasome scores (right) in the MCF7 cell line; (F) UMAP visualization of predicted drug response in MCF7 cells to bortezomib; (g) comparison in predicting bortezomib sensitivity; from left to right: proportions of sensitive and resistant cells predicted by scTumorDrug, SCAD, and scDEAL; violin plot of the BSC score computed by Beyondcell; in (C), (D), and (G), scTumorDrug, SCAD, and/or scDEAL show the predicted proportions of sensitive and resistant cells, respectively, while Beyondcell shows the violin plot of the BSC score, with higher BSC value indicating greater drug sensitivity.

result for the overall urothelial cells is shown in Fig. 3B, with subpopulation-specific predictions detailed in Supplementary Fig. S7. The scTumorDrug predicted that 92.0% of urothelial cells were sensitive to cisplatin in sensitive tumor and 4.0% of urothelial cells were sensitive in resistant tumors across all urothelial cells, consistent with the experimental data. The SCAD yielded 86.7% and 0.03%, scDEAL yielded 62.3% and 27.5%, and Beyondcell produced median BSC scores of 0.66 and 0.48 in sensitive and resistant tumors, respectively. Thus, all four models were able to distinguish between sensitive and resistant tumors in this case.

In the human sample, the sensitive samples were untreated, while the resistant samples had undergone cisplatin–gemcitabine combination therapy and were clinically assessed as having progressive disease. Cell clustering and annotation identified 10 distinct clusters, including the major urothelial cell population (Fig. 3C and Supplementary Fig. S8). Using scTumorDrug, SCAD, scDEAL, and Beyondcell, we predicted the cisplatin response of urothelial cells in the human data. The result for the overall urothelial cells is shown in Fig. 3D, with subpopulation-specific predictions detailed in Supplementary Fig. S9. For all urothelial cells, scTumorDrug predicted 44.1% cisplatin-sensitive cells in sensitive tumors versus 10.7% in resistant tumors. The prediction results were 31.2% versus 18.4% for SCAD, 52.3% versus 65.4% for scDEAL, and median BSC scores of 0.41 versus 0.60 for Beyondcell. The differences in predicted sensitivity between sensitive and resistant tumors were 0.334 for scTumorDrug, 0.182 for SCAD, −0.131 for scDEAL, and −0.19 for Beyondcell. These results indicate that scTumorDrug and SCAD could predict the cisplatin-resistant human bladder tumors, while scDEAL and Beyondcell produced predictions contrary to the clinical observations. We further evaluated the four models for predicting gemcitabine responses in the same urothelial cell population (Fig. 3E) and subpopulations (Supplementary Fig. S10). The scTumorDrug predicted 58.9% gemcitabine-sensitive cells in sensitive tumors versus 5.7% in resistant tumors, consistent with its clinical observations. The corresponding predictions were 44.5% versus 16.0% for SCAD. In addition, scDEAL and Beyondcell produced inverted predictions. It should be noted that the response prediction for combinatorial drugs was not performed in this work since it is out of the scope of our model.

We then evaluated the scTumorDrug on the dataset derived from human ovarian tumor. In this dataset, the samples derived from three human patients at different stages, the chemonaive (no treatment) and chemotreated (carboplatin and paclitaxel treatment), were collected, and the CD45-negative cells were sorted to generate scRNA-seq data. The primary ovarian tumors were used for prediction in this work, but the scRNA-seq data derived from Patient 2 were excluded from prediction due to limited number of ovarian tumor cells. Then five distinct clusters were identified via clustering and annotation by referring to the original work [23], including epithelial cells, matrix CAF, vascular CAF, $STAR^+$ CAF, and endothelial cells (Fig. 4A–C). As shown in Fig. 4D, the epithelial cells, the major cancer cell type in ovarian tumor, exhibited decreased proportion in Patient 1 after treatment, whereas there was no obvious change on epithelial cell proportion in Patient 3, consistent with the original report on the tumor cell changes for Patient 1 and Patient 3 [23]. We then used scTumorDrug, SCAD, scDEAL, and Beyondcell to predict the paclitaxel response for epithelial cells. The carboplatin was excluded from prediction since this drug is not in the drug library in this work. As shown in Fig. 4E, scTumorDrug predicted that the epithelial cells in Patient 1 and Patient 3 were sensitive and resistant to paclitaxel treatment, respectively, consistent with the cell proportion analysis. The SCAD, scDEAL, and Beyondcell generated the similar predictions, but with lower proportion difference of sensitive cells between Patient 1 and Patient 3.

To summarize, scTumorDrug well predicted the drug responses in both bladder and ovarian tumors. SCAD could predict the correct sensitive/resistant trend overall, but it could not well distinguish sensitive and resistant tumor cells in some cases. The scDEAL and Beyondcell methods produced predictions contrary to experimental and clinical observations in human bladder tumors. Considering the fact that scTumorDrug also performed better in CCLs, the results from cell lines and solid tumors collectively demonstrated the superiority of scTumorDrug in discriminating sensitive and resistant tumor cells.

scTumorDrug aided in exploring the cell-type-specific role of JQ1 in bladder tumor

In public mouse models and human samples, we used the scTumorDrug and other methods to predict cell-type-specific drug responses except for the overall response (Supplementary Figs S7, S9, and S10). However, we could not directly validate the drug responses of these subpopulations since there are no experimental validations currently. In CCLs, we validated that scTumorDrug well distinguished sensitive from resistant cancer cells for the BET inhibitor iBET. BET proteins (bromodomain and extra-terminal domain proteins) are key epigenetic readers, and members such as BRD4 recognize histone acetylation marks and recruit transcription complexes to drive oncogene expression [40–43]. These results together made us to investigate the cell-type-specific drug responses for the small molecule JQ1 [44–46], a well-characterized BET (BRD4) inhibitor, in treating bladder tumors in a mouse model.

First, we established a BBN-induced mouse bladder tumor model. Between 3.5 and 5.5 months of BBN induction, the mice were administered either DMSO or JQ1 (Fig. 5A), and bladder samples were collected at the end of the experiment for pathological evaluation (Fig. 5B). The result showed that JQ1 treatment significantly reduced bladder weight in mice (Fig. 5C). Hematoxylin and eosin stain (H&E) staining revealed that bladder tumors in the DMSO group began to infiltrate the muscle layer, whereas in the JQ1 group, infiltration was limited to the lamina propria (Fig. 5D). We further statistically analyzed the T stage of the bladder tumors in each group, and the results indicated that 41.67% of mice in the DMSO group progressed to the later T2 stage, whereas the proportion of T2 stage in the JQ1-treated group decreased to 8.33% (Fig. 5E). These results demonstrated that JQ1 delayed the progression of muscle-invasive bladder tumors in mice, although it did not completely eliminate the tumors in our experimental design.

We then performed snRNA-seq on mouse bladder tumor samples treated with DMSO and JQ1 (Supplementary Fig. S11). We identified 10 cell clusters, including urothelial cells (Fig. 5F and G and Supplementary Fig. S12). The scTumorDrug predicted that the urothelial cells were only relatively sensitive to JQ1 in the DMSO-treated samples, consistent with the observation that the JQ1 treatment did not completely eliminate the bladder tumor in BBN model. We further clustered urothelial cells into three subpopulations: Mki67+ proliferating uro, basal uro, and luminal uro, in which abbreviation uro denotes urothelial cells (Fig. 6A and Supplementary Fig. S13a and b). Then we used scTumorDrug to predict JQ1 response for these three subpopulations in DMSO-treated samples, and found that Mki67+ proliferating uro exhibited the highest proportion of sensitive cells (63.3%) (Fig. 6B). Quantification revealed that Mki67+

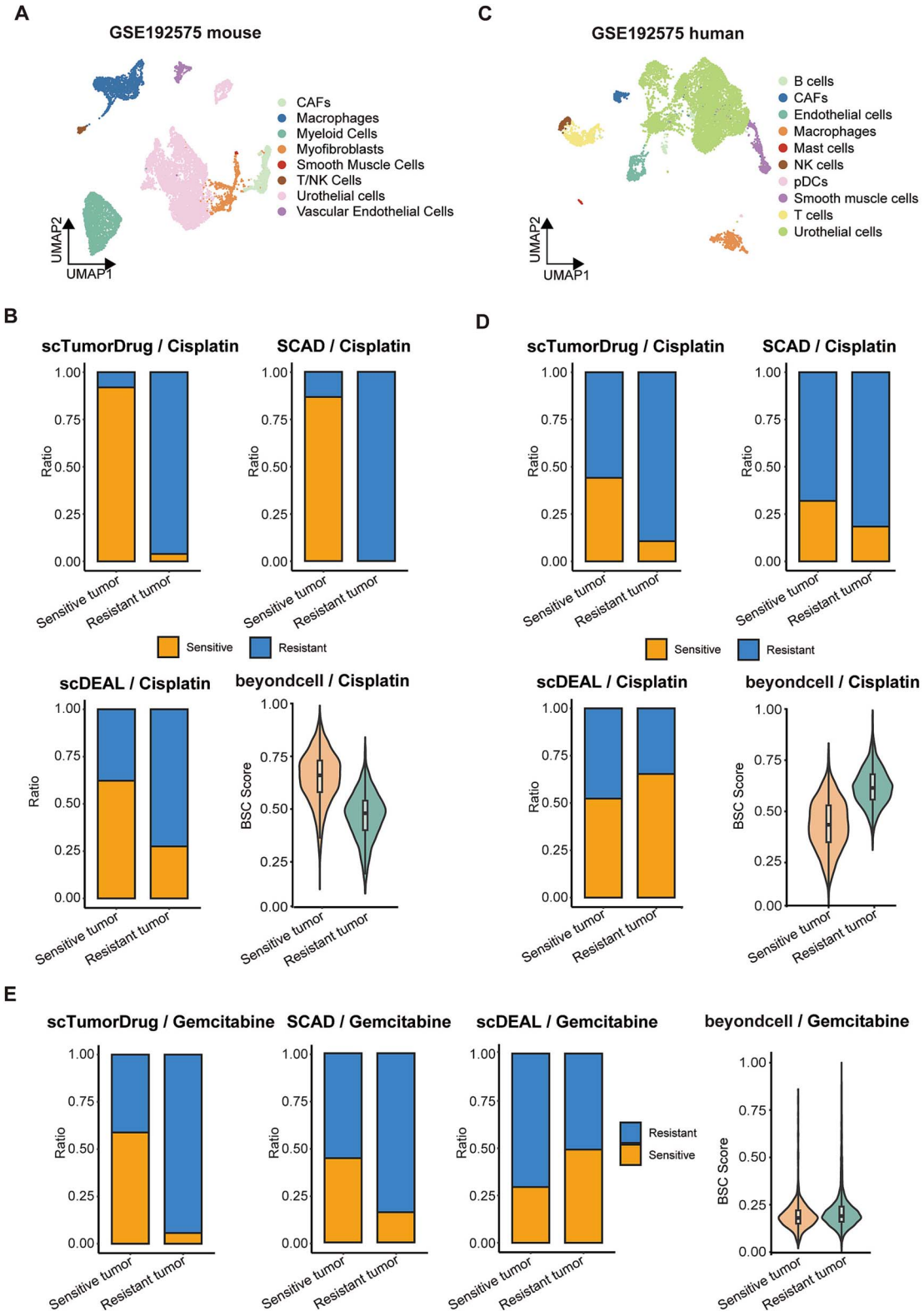

Figure 3 Evaluations on scTumorDrug using scRNA-seq data derived from mouse and human bladder tumors; (A) UMAP plot of cell populations in mouse scRNA-seq data; (B) comparison in predicting cisplatin sensitivity for urothelial cells in mouse data; (C) UMAP plot of cell populations in human scRNA-seq data; (D and E) comparisons in predicting cisplatin sensitivity and gemcitabine sensitivity for urothelial cells in human data, respectively.

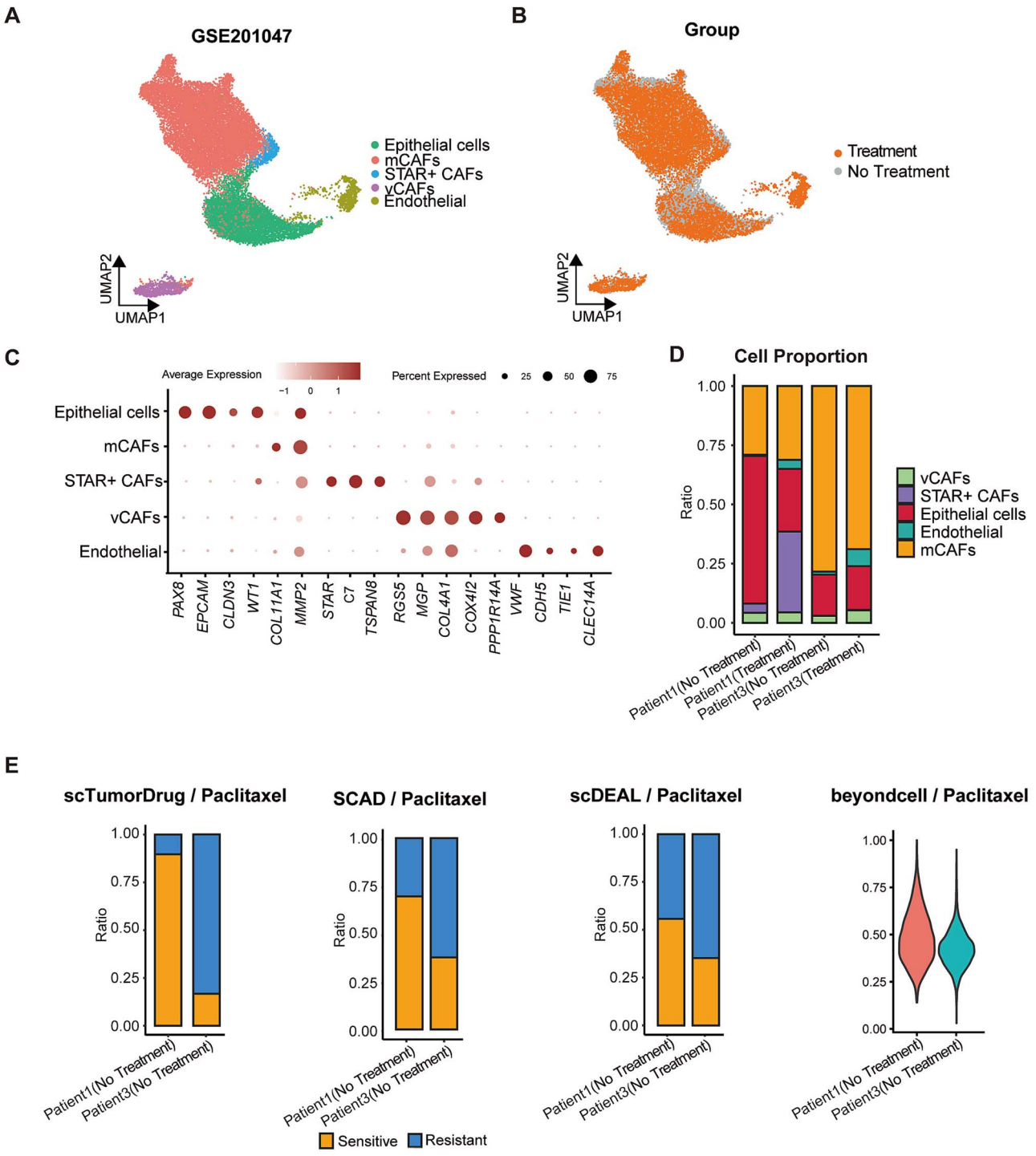

Figure 4 Evaluations on scTumorDrug using scRNA-seq data derived from human ovarian tumors; (A) UMAP plot of cell populations in human scRNA-seq data; (B) grouped plots of single-cell data from chemonaive (no treatment) and chemotreated (treatment) human ovarian tumors; (C) annotated bubble chart of single-cell data for the ovarian tumors; (D) the cell proportions in Patient 1 and Patient 3 before and after treatment; (E) comparison in predicting paclitaxel sensitivity and resistance in epithelial cells.

proliferating urothelial cells comprised 12.0% of cells in DMSO samples, but this proportion decreased to 9.4% upon JQ1 treatment (Fig. 6C). GO functional enrichment analysis revealed downregulation of cell proliferation, cell cycle, and cell migration pathways in JQ1-treated Mki67+ proliferating uro (Supplementary Fig. S13c). Additionally, inferCNV analysis showed significantly lower CNV scores in this subpopulation, implying reduced malignancy in JQ1 samples ($p = 3.2 \times 10^{-11}$) (Supplementary Fig. S13d). These analyses agreed with our predicted sensitivity in Mki67+ proliferating uro. We used SCAD and Beyondcell to predict drug responses for the three urothelial cell subpopulations, and scDEAL tool was excluded from calculation as JQ1 was not included in its database. In SCAD, the

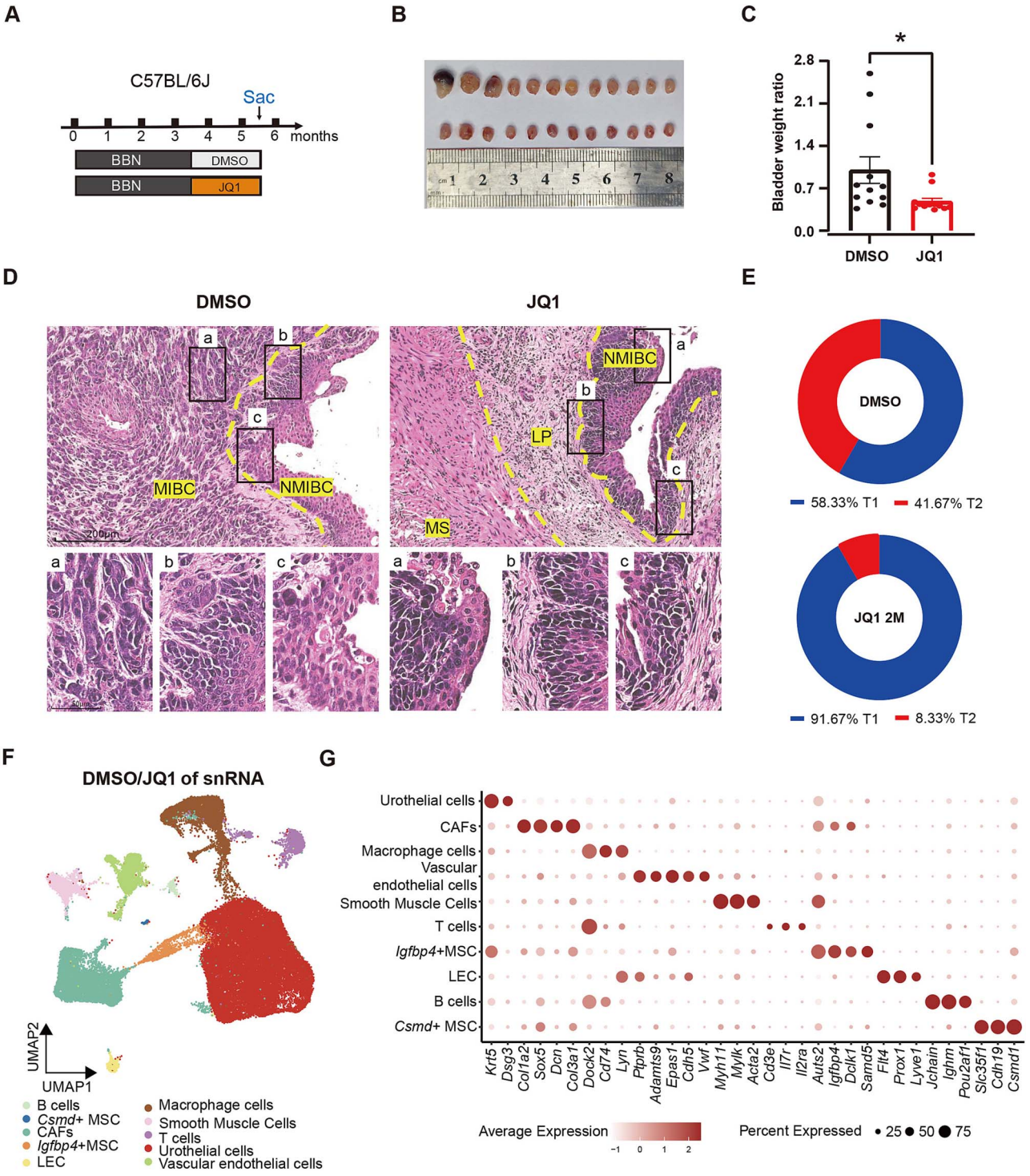

Figure 5 JQ1 delays the progression of bladder cancer in mouse model; (A) treatment strategy of JQ1 in the BBN-induced bladder cancer model; (B) representative images of collected mouse bladders for the DMSO (up) and JQ1 (down) groups; (C) statistical comparison of bladder weight between DMSO and JQ1 groups; $^{*}P < .05$; (D) representative H&E staining images showing the pathological progression of bladder cancer in the DMSO and JQ1 groups; NMIBC: non-muscle invasive bladder carcinoma; MIBC: muscle-invasive bladder carcinoma; LP: lamina propria; MS: smooth muscle; low-magnification scale bar: 200 μm; high-magnification scale bar: 50 μm; (E) the proportion of mice at each bladder cancer T stage; DMSO group (mice, $n = 12$), JQ1 group (mice, $n = 12$); (F) and (G) UMAP plot and annotated bubble chart of snRNA-seq data from mouse bladder tumors treated with DMSO and JQ1.

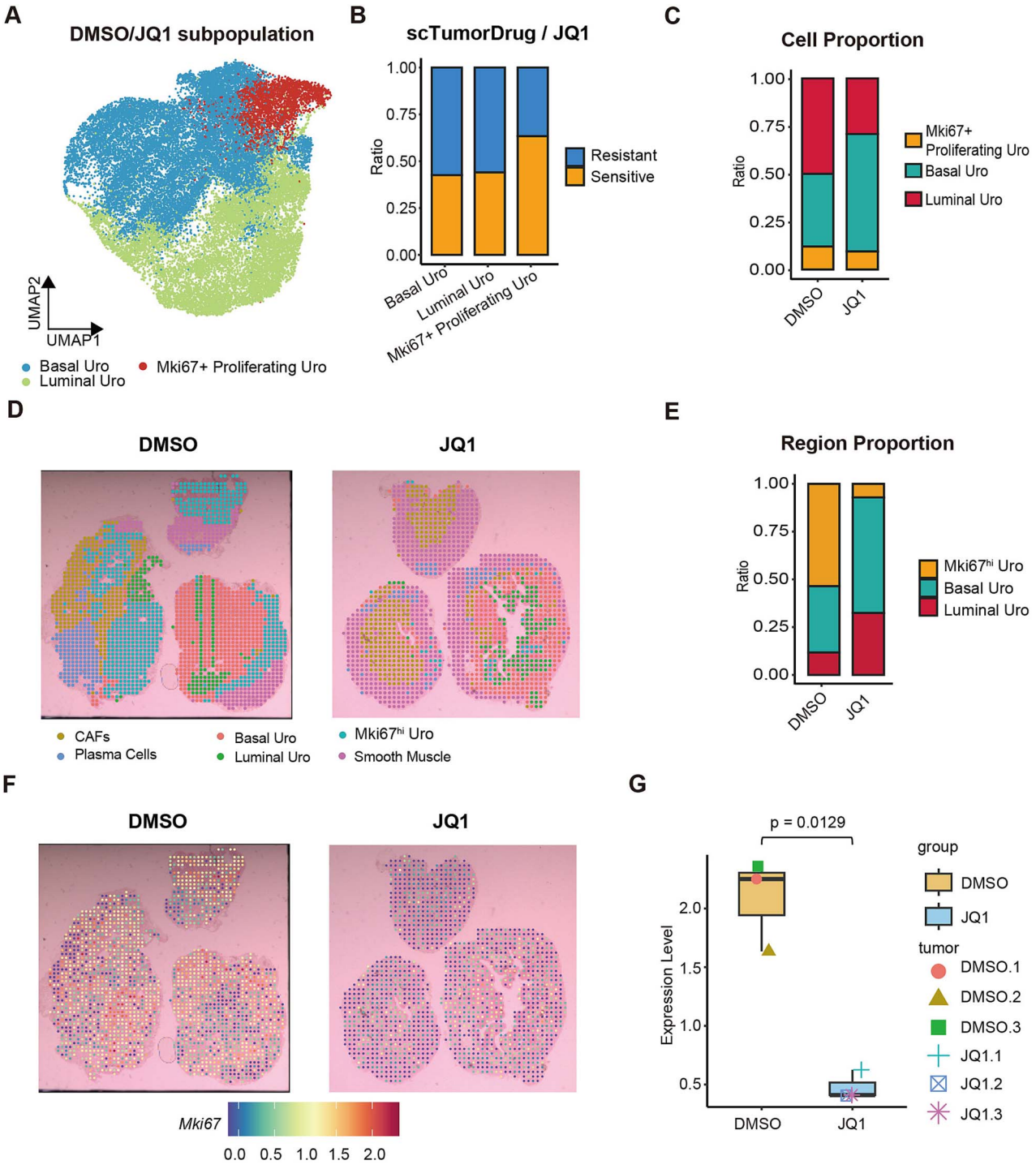

Figure 6 The prediction and validation on the sensitivity of Mki67+ urothelial cells to JQ1 treatment; (A) UMAP plot of urothelial cell subpopulations in snRNA-seq data from DMSO and JQ1 groups; (B) the scTumorDrug predicted proportions of sensitive and resistant cells in the DMSO-treated urothelial cell subpopulation; (C) the cellular proportions in DMSO and JQ1 snRNA-seq datasets; (D) the spatial regions derived from mouse bladder tumors treated with DMSO and JQ1, respectively; (E) proportions of spatial regions in DMSO and JQ1 groups; (G) spatial expressions of gene *Mki67* in DMSO and JQ1 groups; (G) comparison of *Mki67* expression levels between DMSO and JQ1 samples.

Mki67+ proliferating uro subpopulation was mainly predicted as JQ1 resistance though it exhibited higher sensitivity than other two subpopulations, while Beyondcell yielded predictions consistent with our experiments (Supplementary Fig. S14).

To further validate these findings, we performed spatial transcriptomics on the tissue sections of DMSO and JQ1 samples. We identified six clusters from spatial transcriptomics, including three clusters mainly marked by urothelial cell subpopulations: Mki67^{hi} uro, basal

uro, and luminal uro (Fig. 6D and Supplementary Fig. S13e). Consistent with snRNA-seq data, the proportion of Mki67^{hi} uro decreased from 53.5% to 7.1% upon JQ1 treatment (Fig. 6E). Meanwhile, we found that the overall expression level of the *Mki67* gene significantly decreased in JQ1 samples (Fig. 6F and G).

Considering that the Mki67+ urothelial cells are highly proliferative tumor cells, the results derived from snRNA-seq, spatial transcriptomics, and scTumorDrug collectively indicated that JQ1 reduced Mki67+ urothelial cells in the bladder tumors, which was the potential mechanism why JQ1 delayed the tumor progression to T2 stage in the BBN-induced mouse bladder tumor model. Furthermore, the scTumorDrug could aid in the discovery of sensitive subpopulations.

Discussion

In this study, we developed the computational framework scTumorDrug to predict drug responses at single-cell resolution, which has the potential to dissect the tumor heterogeneity in drug treatment for precision therapy. The performance improvement of scTumorDrug lies in its innovative learning paradigm. Inspired by the SCAD that learns the representation from both source and target domains, scTumorDrug utilized the adversarial domain adaption to align feature distributions between bulk and single-cell data, ensuring domain invariance of learned knowledge. But unlike SCAD, scDEAL and other methods that simply transfer knowledge learned from bulk data to single-cell domain, scTumorDrug constructs a dual supervision system that directly learns drug response patterns at the single-cell level by introducing labeled single-cell data. Furthermore, scTumorDrug added both single-cell and bulk data reconstructions in the total loss function to make the learned representation better fit the dual supervision. Our experimental results show that this learning approach captures subtle response differences among cell subpopulations with greater precision, and outperforms the traditional transfer learning methods. To demonstrate the role of dual supervision, we trained the model by using only bulk RNA-seq data, and the results showed that the scTumorDrug derived from dual supervision performed much better than the model trained from only bulk RNA-seq data (Supplementary Fig. S15). In addition, the scTumorDrug can also detect the important genes in this dual-supervision prediction by using gradient backpropagation technique. For example, we identified gene *PTPRC* as high contributing one to drug response prediction in the MA9 I-BET treatment (Supplementary Fig. S16), consistent with the previous work that this gene plays important roles in leukemia prognosis and treatment response [47]. However, it should be noted that the genes contributing to drug response prediction cannot completely represent functional analysis, and further validations on these genes are still needed. Collectively, the dual supervision technique can utilize the efficacy of both bulk RNA-seq and scRNA-seq data, enabling the model to work accurately on both CCL and solid tumor datasets.

Successful predictions on bladder and ovarian tumors derived from human patients provide hints for the applicability of scTumorDrug in clinical tumors. Previous works mainly used drug sensitivity data derived from CCLs, and they were not well evaluated by mouse models and clinical samples. In this work, we compared these methods in the real solid tumor data. Though SCAD, scDEAL, and Beyondcell worked on some datasets to some extent, only scTumorDrug accurately distinguished drug-sensitive from drug-resistant tumors in both mouse and human bladder tumors. These results demonstrate that scTumorDrug learns conserved and generalizable features associated with drug response.

The scTumorDrug can dissect the tumor heterogeneity at subpopulation resolution in drug responses. In this work, we investigated the cell-type-specific responses to JQ1 treatment in mouse bladder tumors by performing snRNA-seq and spatial transcriptomics on a BBN-induced mouse model. The scTumorDrug successfully identified cellular response heterogeneity among urothelial subpopulations, and also predicted higher JQ1 sensitivity in Mki67+ proliferating urothelial. The snRNA-seq and spatial transcriptomics cross-validated this drug sensitivity prediction at the subpopulation level in a solid tumor. This refined resolution enables researchers to move beyond overall tumor response and identify the sensitive or resistant subpopulations, providing a basis for exploring drug efficacy.

We also evaluated the hyperparameter sensitivity in scTumorDrug. For the weights in the total loss function in Equation 10, we trained the model with different weight coefficients. The results showed that scTumorDrug performances were robust when coefficients slightly varied, but the coefficients used in this work performed relatively better overall (Supplementary Fig. S17). We also selected different numbers of top nDRGs as model input, and found that the prediction accuracy generally increased when the nDRG number increased at the beginning. Overall, the performance was relatively better between 1000 and 2000 genes (Supplementary Fig. S18), and we selected the top 1500 genes in this work.

There are also limitations in this work. First, scTumorDrug performance depends on the scale and quality of labeled single-cell datasets, which remain relatively scarce. Since the complexity of different tumor types can change, the limited datasets in the real tumors hinder the comprehensive evaluation on the model performance. As more public data become available in the future, more model validations can be further conducted to account for different kinds of tumors. Second, we predicted the cisplatin-sensitive and gemcitabine-sensitive urothelial cell subpopulations, but there is no direct evidence to support these cell-type-specific drug responses in the mouse and human bladder tumor samples due to the lack of validation experiments. This limitation also inspires us to further investigate the cell-type-specific roles of JQ1 in mouse bladder tumor. Third, we mainly focused on drug responses on primary tumor cells in this work. However, primary tumor cell sensitivity cannot completely represent treatment efficacy or prognosis in some cases since there exist complicated intercellular communications, tumor relapse, and/or metastasis in real tumors. This kind of prediction is out of the scope of our work due to limited data. Further works are needed to deal with the cellular communications and other issues [48–50].

Conclusion

In this work, we provided a computational tool scTumorDrug for single-cell drug response prediction through its innovative dual supervision, which expands the computational toolkit for precision medicine in cancer treatment. We also used the scTumorDrug to aid the investigation of cell-type-specific responses to JQ1 treatment in mouse bladder tumor using snRNA-seq and spatial transcriptomics, and found that the Mki67+ proliferating urothelial cells are sensitive to JQ1, revealing the mechanism underlying the observation that JQ1 delayed the bladder tumor progression stage. With the accumulation of more single-cell data on drug responses, the scTumorDrug can be further evaluated or improved in the future.

Key Points

- We developed the software scTumorDrug to predict single-cell drug responses for heterogeneous tumors by integrating bulk and single-cell RNA data.
- The scTumorDrug worked well on public tumor datasets derived from mouse models and/or clinical human patients, outperforming the existing methods in the selected datasets.
- We established a mouse bladder cancer model to investigate the cell-type-specific drug responses of JQ1 at the subpopulation level by using scTumorDrug, and we also evaluated the prediction performance.
- We performed single-cell RNA sequencing and spatial transcriptomics to cross-validate the finding that the Mki67+ proliferating urothelial subpopulation is sensitive to JQ1 in a mouse bladder cancer model.

Acknowledgments

We thank the M20 Genomics team for the FFPE snRNA-seq and spatial transriptomics technologies and the support of sequencing data preprocessing. We thank Dr Maosen Ye in Kunming Institute of Zoology, Chinese Academy of Sciences, for providing the in-house Shiny application to select individual mouse sample from spatial transcriptome for *Mki67* expression statistics.

Author contributions

Qiang Zhang developed the software and performed data analysis, and Cheng Peng supervised this process. Qing Zhang conducted animal and molecular experiments, and Chunming Guo supervised this process. Chunming Guo and Cheng Peng conceived and administrated the project. All authors read and approved the manuscript.

Competing interests

All authors declared no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China [32170662, 82460142], and Guangdong Hybribio Biotech Funding [No. H20230314, H20230311, and H20230313].

Supplementary material

Supplementary material is available at *Briefings in Bioinformatics* online.

Data availability

The single cell RNA sequencing and spatial transcriptome data generated in this study have been deposited in the National Genomics Data Center (https://ngdc.cncb.ac.cn) at the BioProject PRJCA052916, in which the sequencing reads are under the GSA accession number CRA034798, and the preprocessed data and staining images in spatial transcriptomics are under OMIX accession number OMIX013647.

The source code of scTumorDrug is freely available at Github (https://github.com/CPenglab/scTumorDrug).

References

1. Jamal-Hanjani M, Quezada SA, Larkin J *et al.* Translational implications of tumor heterogeneity. *Clin Cancer Res* 2015;**21**:1258–66. https://doi.org/10.1158/1078-0432.CCR-14-1429
2. Schmidt F, Efferth T. Tumor heterogeneity, single-cell sequencing, and drug resistance. *Pharmaceuticals* 2016;**9**:33. https://doi.org/10.3390/ph9020033
3. McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell* 2017;**168**:613–28. https://doi.org/10.1016/j.cell.2017.01.018
4. Rambow F, Rogiers A, Marin-Bejar O *et al.* Toward minimal residual disease-directed therapy in melanoma. *Cell* 2018;**174**:843–55.e819. https://doi.org/10.1016/j.cell.2018.06.025
5. Barretina J, Caponigro G, Stransky N *et al.* The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;**483**:603–7. https://doi.org/10.1038/nature11003
6. Yang W, Soares J, Greninger P *et al.* Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* 2012;**41**:D955–61. https://doi.org/10.1093/nar/gks1111
7. Verjans ET, Doijen J, Luyten W *et al.* Three-dimensional cell culture models for anticancer drug screening: worth the effort? *J Cell Physiol* 2018;**233**:2993–3003. https://doi.org/10.1002/jcp.26052
8. Schirle M, Jenkins JL. Identifying compound efficacy targets in phenotypic drug discovery. *Drug Discov Today* 2016;**21**:82–9. https://doi.org/10.1016/j.drudis.2015.08.001
9. Chawla S, Rockstroh A, Lehman M *et al.* Gene expression based inference of cancer drug sensitivity. *Nat Commun* 2022;**13**:5680. https://doi.org/10.1038/s41467-022-33291-z
10. Zheng Z, Chen J, Chen X *et al.* Enabling single-cell drug response annotations from bulk RNA-Seq using SCAD. *Adv Sci* 2023;**10**:2204113. https://doi.org/10.1002/advs.202204113
11. Chen J, Wang X, Ma A *et al.* Deep transfer learning of cancer drug responses by integrating bulk and single-cell RNA-seq data. *Nat Commun* 2022;**13**:6494. https://doi.org/10.1038/s41467-022-34277-7
12. Fustero-Torre C, Jiménez-Santos MJ, García-Martín S *et al.* Beyondcell: targeting cancer therapeutic heterogeneity in single-cell RNA-seq data. *Genome Med* 2021;**13**:187. https://doi.org/10.1186/s13073-021-01001-x
13. Yao Y, Xu Y, Zhang Y *et al.* Single cell inference of cancer drug response using pathway-based transformer network. *Small Methods* 2025;**9**:2400991. https://doi.org/10.1002/smtd.202400991
14. Zhang W, Maeser D, Lee A *et al.* Integration of pan-cancer cell line and single-cell transcriptomic profiles enables inference of therapeutic vulnerabilities in heterogeneous tumors. *Cancer Res* 2024;**84**:2021–33. https://doi.org/10.1158/0008-5472.CAN-23-3005
15. Lei W, Yuan M, Long M *et al.* scDR: predicting drug response at single-cell resolution. *Genes* 2023;**14**:268. https://doi.org/10.3390/genes14020268
16. Rees MG, Seashore-Ludlow B, Cheah JH *et al.* Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol* 2016;**12**:109–16. https://doi.org/10.1038/nchembio.1986
17. Basu A, Bodycombe NE, Cheah JH *et al.* An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* 2013;**154**:1151–61. https://doi.org/10.1016/j.cell.2013.08.003

18. Seashore-Ludlow B, Rees MG, Cheah JH *et al.* Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discov* 2015;**5**:1210–23. https://doi.org/10.1158/2159-8290.CD-15-0235
19. Kinker GS, Greenwald AC, Tal R *et al.* Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat Genet* 2020;**52**:1208–18. https://doi.org/10.1038/s41588-020-00726-6
20. Bell CC, Fennell KA, Chan Y-C *et al.* Targeting enhancer switching overcomes non-genetic drug resistance in acute myeloid leukaemia. *Nat Commun* 2019;**10**:2723. https://doi.org/10.1038/s41467-019-10652-9
21. Ben-David U, Siranosian B, Ha G *et al.* Genetic and transcriptional evolution alters cancer cell line drug response. *Nature* 2018;**560**:325–30. https://doi.org/10.1038/s41586-018-0409-3
22. Wang M, Chen X, Tan P *et al.* Acquired semi-squamatization during chemotherapy suggests differentiation as a therapeutic strategy for bladder cancer. *Cancer Cell* 2022;**40**:1044–59. https://doi.org/10.1016/j.ccell.2022.08.010
23. Loret N, Vandamme N, De Coninck J *et al.* Distinct transcriptional programs in ascitic and solid cancer cells induce different responses to chemotherapy in high-grade serous ovarian cancer. *Mol Cancer Res* 2022;**20**:1532–47. https://doi.org/10.1158/1541-7786.MCR-21-0565
24. Ling A, Gruener RF, Fessler J *et al.* More than fishing for a cure: the promises and pitfalls of high throughput cancer cell line screens. *Pharmacol Ther* 2018;**191**:178–89. https://doi.org/10.1016/j.pharmthera.2018.06.014
25. Bairoch A. The cellosaurus, a cell-line knowledge resource. *J Biomol Tech* 2018;**29**:25–38. https://doi.org/10.7171/jbt.18-2902-002
26. Eraslan G, Simon LM, Mircea M *et al.* Single-cell RNA-seq denoising using a deep count autoencoder. *Nat Commun* 2019;**10**:390. https://doi.org/10.1038/s41467-018-07931-2
27. Hao Y, Stuart T, Kowalski MH *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* 2024;**42**:293–304. https://doi.org/10.1038/s41587-023-01767-y
28. Patel AP, Tirosh I, Trombetta JJ *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 2014;**344**:1396–401. https://doi.org/10.1126/science.1254257
29. Wu T, Hu E, Xu S *et al.* clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb)* 2021;**2**:100141. https://doi.org/10.1016/j.xinn.2021.100141
30. Shin K, Lim A, Odegaard JI *et al.* Cellular origin of bladder neoplasia and tissue dynamics of its progression to invasive carcinoma. *Nat Cell Biol* 2014;**16**:469–78. https://doi.org/10.1038/ncb2956
31. Xu Z, Zhang T, Chen H *et al.* High-throughput single nucleus total RNA sequencing of formalin-fixed paraffin-embedded tissues by snRandom-seq. *Nat Commun* 2023;**14**:2734. https://doi.org/10.1038/s41467-023-38409-5
32. Liao Y, Chen J, Zhang S *et al.* Spatial total RNA sequencing of formalin-fixed paraffin embedded tissues by spRandom-seq. *bioRxiv* 2025;2025.02.19.638981. https://doi.org/10.1101/2025.02.19.638981
33. Le DT, Durham JN, Smith KN *et al.* Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* 2017;**357**:409–13. https://doi.org/10.1126/science.aan6733
34. Dawson MA, Prinjha RK, Dittmann A *et al.* Chung C-w, Hopf C, Savitski MM: inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature* 2011;**478**:529–33. https://doi.org/10.1038/nature10509
35. Mirguet O, Gosmini R, Toum J *et al.* Discovery of epigenetic regulator I-BET762: lead optimization to afford a clinical candidate inhibitor of the BET bromodomains. *J Med Chem* 2013;**56**:7501–15. https://doi.org/10.1021/jm401088k
36. Csizmar C, Kim D, Sachs Z. The role of the proteasome in AML. *Blood Cancer J* 2016;**6**:e503. https://doi.org/10.1038/bcj.2016.112
37. Ghansah H, Debreceni IB, Fejes Z *et al.* The proteasome inhibitor bortezomib induces apoptosis and activation in gel-filtered human platelets. *Int J Mol Sci* 2021;**22**:8955. https://doi.org/10.3390/ijms22168955
38. Britton M, Lucas MM, Downey SL *et al.* Selective inhibitor of proteasome's caspase-like sites sensitizes cells to specific inhibition of chymotrypsin-like sites. *Chem Biol* 2009;**16**:1278–89. https://doi.org/10.1016/j.chembiol.2009.11.015
39. Maseda D, Meister S, Neubert K *et al.* Proteasome inhibition drastically but reversibly impairs murine lymphocyte development. *Cell Death Differ* 2008;**15**:600–12. https://doi.org/10.1038/sj.cdd.4402297
40. Shi J, Vakoc CR. The mechanisms behind the therapeutic activity of BET bromodomain inhibition. *Mol Cell* 2014;**54**:728–36. https://doi.org/10.1016/j.molcel.2014.05.016
41. Dhalluin C, Carlson JE, Zeng L *et al.* Structure and ligand of a histone acetyltransferase bromodomain. *Nature* 1999;**399**:491–6. https://doi.org/10.1038/20974
42. Jiang YW, Veschambre P, Erdjument-Bromage H *et al.* Mammalian mediator of transcriptional regulation and its possible role as an end-point of signal transduction pathways. *Proc Natl Acad Sci U S A* 1998;**95**:8538–43. https://doi.org/10.1073/pnas.95.15.8538
43. Yang Z, Yik JHN, Chen R *et al.* Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol Cell* 2005;**19**:535–45. https://doi.org/10.1016/j.molcel.2005.06.029
44. Filippakopoulos P, Qi J, Picaud S *et al.* Selective inhibition of BET bromodomains. *Nature* 2010;**468**:1067–73. https://doi.org/10.1038/nature09504
45. French CA, Ramirez C, Kolmakova J *et al.* BRD–NUT oncoproteins: a family of closely related nuclear proteins that block epithelial differentiation and maintain the growth of carcinoma cells. *Oncogene* 2008;**27**:2237–42. https://doi.org/10.1038/sj.onc.1210852
46. Mertz JA, Conery AR, Bryant BM *et al.* Targeting MYC dependence in cancer by inhibiting BET bromodomains. *Proc Natl Acad Sci U S A* 2011;**108**:16669–74. https://doi.org/10.1073/pnas.1108190108
47. Drube J, Ernst T, Pfirrmann M *et al.* PTPRG and PTPRC modulate nilotinib response in chronic myeloid leukemia cells. *Oncotarget* 2018;**9**:9442–55. https://doi.org/10.18632/oncotarget.24253
48. Ren X, Zhang L, Zhang Y *et al.* Insights gained from single-cell analysis of immune cells in the tumor microenvironment. *Annu Rev Immunol* 2021;**39**:583–609. https://doi.org/10.1146/annurev-immunol-110519-071134
49. Golo M, Newman PLH, Kempe D *et al.* Mechanoimmunology in the solid tumor microenvironment. *Biochem Soc Trans* 2024;**52**:1489–502. https://doi.org/10.1042/BST20231427
50. Frankiw L, Li G. The tumor microenvironment's role in malignant progression and treatment response. *Cancer Lett* 2022;**548**:215888. https://doi.org/10.1016/j.canlet.2022.215888

Briefings in Bioinformatics, 2026, **27**, bbag337
https://doi.org/10.1093/bib/bbag337
Problem Solving Protocol