Gene expression

A Survey and Systematic Assessment of Computational Methods for Drug Response Prediction

Jinyu Chen and Louxin Zhang*

Department of Mathematics and Computational Biology Programme, National University of Singapore, Singapore 119076

* Corresponding author: Louxin Zhang, Tel: +65-65166579; Email: matzlx@nus.edu.sg

Abstract

Drug response prediction arises from both basic and clinical research of personalized therapy, as well as drug discovery for cancer and other diseases. With gene expression profiles and other omics data being available for over 1000 cancer cell lines and tissues, different machine learning approaches have been applied to solve drug response prediction problems. These methods appear in a body of literature and have been evaluated on different datasets with only one or two accuracy metrics. We systematically assessed 17 representative methods for drug response prediction, which have been developed in the past five years, on four large public datasets in nine metrics. This study provides insights and lessons for future research into drug response prediction.

Keywords: Drug response prediction; regression model; low rank matrix factorization based approach; Bayesian inference; deep learning; benchmark

Jinyu Chen is currently a postdoctoral research fellow at the Department of Mathematics, National University of Singapore. She completed her PhD at the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China. Her research interests include integrative analysis of omics data and matrix-factorization methods in bioinformatics.

Louxin Zhang is currently a professor at the Department of Mathematics and the Computational Biology Programme, National University of Singapore. He received his PhD in computer science from the University of Waterloo, Canada. His research interests include bioinformatics in comparative and evolutionary genomics, cancer genomics, cellular protein networks and discrete algorithms.

© The Author 2019.

1 Introduction

Cancer is still an incurable disease. Recent cancer genome studies have suggested that each cancer patient owns a unique profile of gene mutations and that there is no silver bullet for defeating all the cancers. Therefore, how to use the genomic profiles of cancer patients to design a personalized treatment is vital for effective control of disease progression [1, 2]. This requires accurate prediction of the patients' response to target-specific drugs and therapies.

Given the paucity of patients' molecular profiles and their responses to drugs, large-scale drug screening experiments on cultured and genome-wide profiled cancer cell lines have been used to study drug response [3, 4]. The US National Cancer Institute (NCI) 60 human tumor cell lines screen database provides drug screening data for thousands of drugs in 60 tumor cell lines, for which mRNA, microRNA expression, DNA methylation and mutation profiles are also available [3, 5]. In 2012, another two large-scale pharmacogenomic datasets were made publicly available: the Cancer Cell Line Encyclopedia (CCLE) [6] and the Genomics of Drug Sensitivity in Cancer (GDSC) [7]. The CCLE contains drug sensitivity data from over 11,000 experiments involving 24 drugs and 504 cell lines and both the gene mutational and expression profiles for about 1000 cell lines. The GDSC includes the data generated from about 200,000 drug response experiments that tested 266 drugs on nearly 1000 human cancer cell lines.

With a large amount of drug response data for cell lines being available, various machine learning approaches have been applied to design computational methods for drug response prediction in recent years. These prediction methods, their interpretability and the importance of the integration of multiple omics profiles for performance improvement are well documented in two survey articles [8, 9]. Additionally, key methods developed before 2014 were evaluated in [10] and [11]. Since 2015, a number of new approaches have been developed for drug response prediction, including regression methods with or without kernels, [12–16], Bayesian inference methods [17–20], matrix-factorization-based methods [21–23] and other miscellaneous (deep) machine learning methods [24–28].

In this study, we aim to systematically assess 17 representative computational methods that have been developed since 2015 (Table 1) and the classic regression approaches on four public datasets: GDSC, CCLE, NCI-60 and the Cancer Therapeutics Response Portal (CTRP) dataset [29]. These methods have often been evaluated by the developers on the datasets that were obtained from the aforementioned datasets in one or two metrics. It remains therefore unclear which are among the best methods. We evaluated these state-of-the-art methods on the same datasets in nine metrics. On the basis of these analyses, we draw lessons and insights that could hopefully benefit future research into drug response prediction.

2 Brief overview of drug response prediction and machine learning methods

2.1 The problems of drug response prediction

Drug response in cell lines are often measured in terms of either the half maximal inhibitory concentration (IC_{50} or GI_{50}). Both IC_{50} and GI_{50} represent how much of a particular drug is needed to inhibit a cell activity and the cell growth by 50%, respectively. Another quantitative measure for drug response is the area under the concentration response curve (AUC) [30]. For IC_{50}, GI_{50} or AUC, the lower the value is, the more effective is the drug. The response to m drugs in n cell lines are given as an $n \times m$ real matrix, in which the (i, j)-th entry is the IC_{50}, GI_{50} or AUC value of the j-th drug in the i-th cell line.

Drug response prediction is a supervised learning problem: given a training drug response dataset $R_{n\times m}$ with n cell lines and m drugs, a response function $r(x, y)$ is learned from $R_{n\times m}$ and then used to compute the response to (i) a known drug (in the training dataset) in a new cell line, (ii) a new drug in a known cell line or (iii) a new drug in a new cell line. In this study, we assess different methods for the Type-i drug response prediction. The Type-iii drug response prediction is believed to be more challenging than the first two.

2.2 Integration of omics data

The rationale for the predictability of drug response is that the response to drugs that have similar chemical properties in a cell line are highly related and cell lines that have similar genomic profiles also have similar drug response to a drug. Good methods for drug response prediction rely therefore on extra omics data other than the drug response matrix. The omics data that are frequently used for the purpose include:

- Gene expression profiles, which characterize the expression level of a set of genes in a cell line. Currently, both microarray and RNA-Seq gene expression datasets are available in the aforementioned datasets.
- MicroRNA expression profiles. MicroRNAs are non-coding RNA molecules consisting of about 20 nucleotides. They function in RNA silencing and post-transcriptional regulation of gene expression.
- Copy number alteration/aberration (CNA) profiles. Copy number alterations are changes in copy number for genes in a (tumor) cell line.
- Mutational profiles. These refer to a list of mutation types (e.g., base substitutions, short insertions and deletions) and mutation counts in a (tumor) cell line.
- DNA methylation profiles. DNA methylation is tissue-specific and dynamic. The DNA methylation profiles characterize the patterns of aberrant DNA methylation occurring in a (tumor) cell line.
- Gene similarity profiles. These comprise curated data characterizing functionally and structurally similarities of a set of genes, which are often computed from genomic and/or proteomic profiles.
- Cellular protein networks. These could be cell-specific protein–protein interaction networks, gene regulatory networks, or signaling pathways.
- Protein expression profiles. These characterize the expression levels of proteins in a cell line. Protein expression levels can be measured on the reverse phase protein lysate microarray or the global mass spectrometry platform.
- Drug features. There are two main types of drug features. One is in formation on the drug's primary targets or target pathways, which are very useful for drug classification. Another is the molecular structural information. Each drug can be characterized by different 1-dimensional and 2-dimensional structural properties, such as the number of atoms, the presence of chemical substructures, etc.

By concatenating different omics datasets and encoding non-numeric entries into numbers, we can represent multiple omics profiles of n cell lines into a real-valued matrix $C_{n\times s}$, in which each column corresponds to a feature in one of the omics profiles. Similarly, we can also represent multiple profiles of m drugs by a real-valued matrix D of m rows and integrate the relationships among k genes such as a gene similarity profile and a cellular protein network into a $k \times k$ matrix G. Thus, mathematically, the training dataset for drug response prediction includes a cell line–drug response matrix $R_{n\times m}$ and at most three auxiliary profile matrices: $C_{n\times s}$, D and G.

Table 1 summarizes which omics data are used by each of the methods to be evaluated in this work.

Table 1. Summary of key types of omics data that can be integrated in the 17 assessed methods (Section 2.4–2.7) in addition to drug response data. GE: Gene expression; CNA: Copy number aberration; MUT: Mutational profile; PN: Cellular protein network; DM: DNA methylation; PE: Protein expression; GS: Gene similarity profile; DF: Drug features. FS: Feature selection, indicating whether a method has a feature selection function or not.

		Omics Data Type							
	FS	GE	CNA	MUT	PN	DM	PE	GS	DF
BMTMKL [11]		√	√	√		√	√		
CaDRReS [18]		√							
CDRscan [24]				√					√
cwKBMF [17]		√						√	√
DualNets [23]		√							√
ENet	√	√	√	√		√	√		
GENet [14]	√	√			√				
KRL [27]		√							
KRR		√							
MACAU [22]		√			√				√
MERGE [20]	√	√	√	√		√			
MVLR [19]	√	√						√	√
pairwiseMKL [26]		√							√
RF	√	√							
RWEN [12]	√	√							
SRMF [21]		√							√
TANDEM [15]	√	√	√	√					

2.3 Single-task vs multi-task learning

A linear regression model assumes a linear relationship between the drug response level and the genomic profile of cell lines. In other words, the linear regression approach learns a linear function $r(x)$ that maps the profiles x of cell lines to the drug response levels. It often identifies good bio-markers. However, it has relatively low prediction accuracy particularly when there are limited samples for a drug, which is also confirmed by this study.

Recently, many methods have been developed to predict the response of multiple drugs in a single model [11, 17, 19–21, 26]. In this so-called multi-task learning framework [31], sharing information on different drugs (tasks) often enables us to obtain a prediction method with high accuracy. It is easy to find out which methods are developed in the multi-task learning framework.

In the rest of this section, we groups different methods into four groups depending on their underlying mathematical principle.

2.4 Linear regression methods and their generalizations

Different regression methods have been developed for drug response prediction. They use the drug response vector $R_{n\times 1}$ of a drug d in the n cell lines to learn a linear response function $r_d(\mathbf{x}) = \mathbf{w}\cdot\mathbf{x}$ with the coefficient vector $\mathbf{w} = (w_1, w_2, \cdots, w_s)$, where $\mathbf{x} = (x_1, x_2, \cdots, x_s)^T$ represents the profile vector of a cell line. For a new cell line with the profile vector $\mathbf{c}$, $r_d(\mathbf{c})$ is output as the response to the drug in the cell line.

A linear regression method computes the coefficient vector $\mathbf{w}$ of the function $r_d(\mathbf{x})$ by solving the following optimization problem:

$$\min_{\mathbf{w}} \frac{1}{n}\|B_{n\times n}(R_{n\times 1} - C_{n\times s}\mathbf{w}^T)\|_2^2 + \lambda_1\|\mathbf{w}\|_1 + \lambda_2\mathbf{w}M\mathbf{w}^T, \quad (1)$$

where $\|\cdot\|_1$ and $\|\cdot\|_2$ are, respectively, the L_1-and L_2-norm defined for real vectors; M is either the $s\times s$ identity matrix or an $s\times s$ symmetric semi-definite matrix derived from an input profile matrix; T denotes the transpose operator for matrices; and $B_{n\times n}$ is a diagonal matrix with n parametric diagonal entries b_j. The free parameters λ_i's and b_j's are set differently for different methods.

Elastic net (ENet)

In this method, $M = \mathbf{I}_{s\times s}$ and $B_{n\times n} = I_{n\times n}$ (i.e., $b_i = 1$) in Eqn. (1) [32]. Thus, the third term of the equation simply becomes the L_2-norm of the coefficient vector $\mathbf{w}$. Additionally, the ENet uses a 5- or 10-fold cross-validation procedure to determine the values of λ_1 and λ_2.

The implementation of the ENet available in the R package has the time complexity $O(s^2n + s^3)$, where s and n are the number of features and cell lines in the input cell line profile matrix, respectively.

Generalized ENet (GENet)

In this method, $B_{n\times n} = I_{n\times n}$ in Eqn. (1). We assessed two versions of this model: **GENet-Lap** and **GENet-NLap** in which M is set to the Laplacian matrix and the normalized Laplacian matrix of a gene interaction network [14], respectively.

In [14], the authors proposed to solve the optimization problem in Eqn. (1) through cyclic coordinate descent [33] by changing one coefficient w_k at a time, while keeping the values of other coefficients w_j ($j \neq k$) unchanged. For this computational method, the time complexity of each iteration is $O(nms^3)$.

Response-weighted ENet (RWENet)

This approach uses an iterative procedure to set (b_j) and compute the coefficient vector $\mathbf{w}$ in Eqn. (1). Here, the drug response data $R_{n\times 1} = (r_i)_{i=1}^n$ is assumed to follow a normal distribution $N(\mu, \sigma^2)$ with the parameters being estimated as [12]:

$$\mu = \text{median}(R_{n\times 1}),$$
$$\sigma = 1.4826 \times \text{median}_{1\le i\le n}\left(|r_i - \text{median}(R_{n\times 1})|\right).$$

Under this assumption, the drug response value is less than ℓ (resp. larger than γ) with a probability of 0.05, where $\ell = \mu - 1.65\times\sigma$ and $\gamma = \mu + 1.65\times\sigma$.

In the RWENet method, b_j is initially set to 1, that is, $b_j^{(0)} = 1$ ($1 \le j \le n$). Thus in the first step, ENet is simply called to compute the coefficient vector $\mathbf{w}^{(0)}$. In the t-th iteration step, $t \ge 1$, $b_j^{(t)}$ are set as follows, depending on whether or not $r_j \le \ell$ with the current coefficient vector $\mathbf{w}^{(t-1)}$:

$$b_j^{(t)} = \begin{cases} \exp\left(1 + |r_j - \mathbf{c}_j\left(\mathbf{w}^{(t-1)}\right)^T|\right), & r_j \le \ell, \\ \delta, & \text{otherwise}, \end{cases} \quad (2)$$

where δ is a small fixed constant close to 0 and $\mathbf{c}_j$ is the j-th row of $C_{n\times s}$. The purpose of setting (b_j) in this way is to find a coefficient vector that has a good prediction for cell lines with low drug responses, which are sensitive to the drug. The algorithm continues until the difference between the input drug responses and the predicted values are small enough. This variant is called the RWENet-left method. The time complexity of the RWENet-left method is $O(ms^3 + mns^2 + ms)$ for each iteration.

Symmetrically, the method can be modified for predicting drug responses for cell lines that are resistant to a drug through setting $r_j \ge \gamma$ in Eqn. (2), as well as for predicting the responses for cell lines that are either sensitive or resistant to a drug. These two variants of RWENet are called the RWENet-right and RWENet-both methods [12], respectively.

MERGE (mutation, expression hubs, known regulators, genomic CNV, and methylation)

Lee *et al.* implemented a linear regression method by incorporating multi-omics prior information about genes for all the drugs simultaneously [20]. They used cell line profiles $C_{n\times s} = (c_{ij})$ for inference and gene profiles

$D_{s\times k} = (d_{ij})$ for assigning scores for the s genes. Lee *et al.* inferred the regression coefficient matrix $W_{s\times m} = (w_{ij})$ by solving the following optimization problem:

$$\min_{w_{jt},v_\ell} \sum_{j,t=1}^{s,m}\sum_{i=1}^{n}(r_{it}-c_{ij}w_{jt})^2 + \sum_{j,t=1}^{s,m}\left(\lambda - \sum_{\ell=1}^{k} d_{j\ell}v_\ell\right) w_{jt}^2$$
$$- m\sum_{j=1}^{s}\log\left(\lambda - \sum_{\ell=1}^{k} d_{j\ell}v_\ell\right)$$
$$\text{s.t. } \lambda - \sum_{\ell=1}^{k} d_{j\ell}v_\ell > 0, \text{ for } j = 1, 2, \cdots, m.$$

Here, λ is a hyperparameter to be determined on the basis of the data and can be considered as the upper bound of the gene scores $\sum_{\ell=1}^{k} d_{i\ell}v_\ell$. Thus, the second term is used to penalize genes with lower scores.

Kernel ridge regression (KRR)
This method first maps a cell line vector v to a vector $f(v)$ in a high-dimensional feature space and then computes the coefficient column vector $\mathbf{w}$ in the image vector space [27, 34]. More specifically, it solves the following optimization problem:

$$\min_{\mathbf{w}} \frac{1}{n}\|R_{n\times 1} - f(C_{n\times s})\mathbf{w}\|_2^2 + \lambda \mathbf{w}^{\mathrm{T}}\mathbf{w}.$$

This problem can be solved as easily as solving Eqn. (1). The reason is that one only needs to know how to compute the inner product $f(\mathbf{u})\cdot f(\mathbf{v})$ of image vectors to solve the optimization problem. He *et al.* used the so-called Gaussian kernel function to define the inner product of the vectors $f(\mathbf{u})$ and $f(\mathbf{v})$ [27]:

$$f(\mathbf{u})\cdot f(\mathbf{v}) = \exp\left(-\frac{\|\mathbf{u}-\mathbf{v}\|_2^2}{\gamma}\right), \tag{3}$$

where γ is a free parameter.

Pairwise multiple kernel learning (pairwiseMKL)
When drug profiles are available, it is possible to map a cell line profile v_c to $f(v_c)$ in a new feature space H, map a drug profile u_d to $g(u_d)$ in another new feature space H' and then find a linear regression function in the product space $H \times H'$ to predict drug response [26].

Let k_c and k_d be the two kernel functions such that:

$$k_c(v_c, v_{c'}) = \langle f(v_c), f(v_{c'})\rangle,\ k_d(u_d, u_{d'}) = \langle g(u_d), g(u_{d'})\rangle.$$

Define $K^{(c)}_{n\times n} = (k_c(v_i, v_j))$, $K^{(d)}_{m\times m} = (k_d(u_i, u_j))$ and their Kronecker product $K^{(d)}_{m\times m} \otimes K^{(c)}_{n\times n} = \left(k_d(u_i, u_j)K^{(c)}_{n\times n}\right)$. In the Kronecker product, each drug-cell line pair (u_d, v_c) corresponds to a row and a column. Hence, $K^{(d)}_{m\times m} \otimes K^{(c)}_{n\times n}$ can be used to define a linear regression function in $H \times H'$, by which the drug response r_{uv} to a drug u in a cell line v can be predicted as:

$$r_{uv} = (w_1, w_2, \cdots, w_{nm})k,$$

where k is the column vector consisting of the kernel values between each training drug-cell line pair (u_d, v_c) and the test pair (u, v). Here, the coefficient row vector $\mathbf{w} = (w_1, w_2, \cdots, w_{nm})$ can be found by solving the following system of linear equations:

$$(K^{(d)}_{m\times m} \otimes K^{(c)}_{n\times n} + \lambda I_{nm\times nm})\mathbf{w}^T = \text{vector}(R_{n\times m}), \tag{4}$$

where vector$(X) = (x'_k)$ is the flat form of an $n \times m$ matrix X that is defined by $x'_{(j-1)n+i} = x_{ij}$ $(1 \le i \le n,\ 1 \le j \le m)$ and λ is a hyperparameter to be determined.

Cichonska et al. incorporated a few ideas into the above kernel regression approach. First, they proposed to use both multiple kernels $K^{(d)}_i$ $(1 \le i \le k_d)$ for drugs and multiple kernels $K^{(c)}_i$ $(1 \le i \le k_c)$ for cell lines. In this way, $K^{(d)}_{m\times m} \otimes K^{(c)}_{n\times n}$ in Eqn. (4) is replaced by a weighted combination of the pairwise Kronecker products:

$$K_\omega = \sum_{i=1}^{k_d}\sum_{j=1}^{k_c} \omega_{ij} K^{(d)}_i \otimes K^{(c)}_i. \tag{5}$$

Second, the authors applied the central kernel alignment technique [35] to compute the kernel mixture weight matrix $\omega = (\omega_{ij})$ in Eqn. (5).

Let 1_{nm} be the matrix of size nm in which each entry is 1 and $C = I_{nm\times nm} - \frac{1}{nm}1_{nm}$, which satisfies that $C^2 = C$. In the central kernel alignment approach, the kernel function K_ω in Eqn. (5) is normalized into $\hat{K}_\omega = CK_\omega C$. Let

$$\hat{R} = C \times \text{vector}(R_{n\times m})\text{vector}(R_{n\times m})^T C. \tag{6}$$

The weight matrix (ω_{ij}) is computed by maximizing the similarity between $\hat{K}_\omega$ and $\hat{R}$:

$$\max_{(\omega_{ij})} \frac{\langle \hat{K}_\omega, \hat{R}\rangle_F}{\|\hat{K}_\omega\|_F}, \text{subject to : } \sum_{i,j}\omega_{ij}^2 = 1, \omega_{ij} \ge 0. \tag{7}$$

Third, Cichonska et al. proposed to apply a Gaussian response kernel to each drug response r_{ij} such that r_{ij} is replaced by a vector of size nm that corresponds to a histogram vector of the Gaussian distribution centered around r_{ij} ([36]). In this way, vector$(R_{n\times m})$ in Eqn. (6) is replaced by an $nm \times nm$ matrix.

Lastly, one disadvantage of the pairwise multiple kernel regression approach is its high time complexity. Cichonska et al. used the bilinearity of the Kronecker product of matrices to simplify the time complexity of solving the optimization problems in Eqn. (4) and (7). In their implementation of the conjugate gradient method that iteratively solves the optimization problems, each iteration takes $O(k_d k_c nm(n+m))$ time.

2.5 Bayesian inference methods

Bayesian inference is a statistic method in which Bayes' theorem is used to update the probability for a model as more data becomes available. Bayesian inference methods for drug response predication often take multi-task learning approach. They attempt to predict all the responses $r(i, j)$ $(1 \le i \le n;\ 1 \le j \le m)$ through jointly learning a single model from the training response matrix $R_{n\times m}$ and auxiliary cell line and drug profile matrices.

The Bayesian multi-task multi-kernel learning method (BMTMKL)
This multi-task learning method had the best performance in the 2014 NCI-DREAM drug sensitivity prediction challenge [11].

For a matrix X, we use $X[i, *]$ and $X[*, j]$ to denote its i-th row and j-th column, respectively. We can rewrite the response matrix $R_{n\times m}$ as follows:

$$R_{n\times m} = (R[*, 1], R[*, 2], \cdots, R[*, m]),$$

where $R[*, j]$ is the responses to drug j in the cell lines. Costello *et al.* proposed to decompose $R[*, j]$ into a linear combination of p components

$Y_j^{(k)} (k = 1, 2, \cdots, p)$ as follows [11]:

$$R[*, j] \approx \sum_{k=1}^{p} e_k Y_j^{(k)} + \begin{bmatrix} b_j & b_j & \cdots & b_j \end{bmatrix}^T \tag{8}$$

for each $j = 1, 2, \cdots, m$; equivalently:

$$R_{n\times m} \approx \mathbf{e} \begin{bmatrix} Y_1^{(1)} & Y_2^{(1)} & \cdots & Y_m^{(1)} \\ Y_1^{(2)} & Y_2^{(2)} & \cdots & Y_m^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ Y_1^{(p)} & Y_2^{(p)} & \cdots & Y_m^{(p)} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{bmatrix} \begin{bmatrix} b_1 & b_2 & \cdots & b_m \end{bmatrix},$$

where $\mathbf{e} = [e_1 \ e_2 \ \cdots \ e_p] \in \mathbb{R}^{1\times p}$ is common to all drugs, and both $Y_j^{(i)} \in \mathbb{R}^{n\times 1}$ and b_j depend only on drug j for each pair of i and j.

One may consider that there are p different profiles of cell lines (such as the DNA methylation profiles or gene expression profiles) and $Y_j^{(i)}$ is the response component of drug j that is inferred from the i-th cell line profile for each $j = 1, 2, \cdots, m$ and $i = 1, 2, \cdots, p$. Costello *et al.* assumed the following connection between $Y_j^{(i)}$ and cell line profiles $K_j^{(i)}$:

$$Y_j^{(i)} = K_j^{(i)} \begin{bmatrix} a_{1j} & a_{2j} & \cdots & a_{nj} \end{bmatrix}^T \tag{9}$$

for each drug j, where $K_j^{(i)}$ is the $n \times n$ similarity matrix that is computed via the Gaussian kernel function in Eqn. (3) or other similarity function.

Costello *et al.* inferred drug responses by using Bayesian inference. They assumed that each e_i of the parameter vector $\mathbf{e}$ follows a normal distribution $\mathcal{N}(0, \sigma_i^{-1})$ such that σ_i follows a Gamma distribution:

$$\sigma_i \sim \mathcal{G}\left(\alpha_\sigma, \beta_\sigma\right).$$

Similarly, for each drug $j = 1, 2, \cdots, m$, each b_j is sampled from a similar distribution:

$$b_j \sim \mathcal{N}(0, \eta_j^{-1}), \quad \eta_j \sim \mathcal{G}\left(\alpha_\eta, \beta_\eta\right).$$

Each component of the vector $(a_{1j}, a_{2j}, \cdots, a_{nj})$ is independently sampled from the following distribution:

$$a_{ij} \sim \mathcal{N}(0, \delta_{ij}^{-1}), \quad \delta_{ij} \sim \mathcal{G}\left(\alpha_\delta, \beta_\delta\right).$$

Costello *et al.* further assumed that the response components $Y_j^{(i)}$ $(i = 1, 2, \cdots, p)$ are sampled from the following multivariate normal distributions:

$$Y_j^{(i)} \sim \mathcal{N}\left(K_j^{(i)} \begin{bmatrix} a_{1j} & a_{2j} & \cdots & a_{nj} \end{bmatrix}^T, \epsilon_j^{-1} \mathbf{I}_{n\times n}\right),$$
$$\epsilon_j \sim \mathcal{G}(\alpha_\epsilon, \beta_\epsilon)$$

and

$$R[*, j] \sim \mathcal{N}\left(\sum_{k=1}^{p} e_k Y_j^{(k)} + (b_j)_{n\times 1}, \zeta_j^{-1} \mathbf{I}_{n\times n}\right),$$
$$\zeta_j \sim \mathcal{G}\left(\alpha_\zeta, \beta_\zeta\right).$$

In summary, this is a very general stochastic model involving the following 10 parameters for prior Gamma distributions:

$$\alpha_\sigma, \beta_\sigma, \alpha_\eta, \beta_\eta, \alpha_\delta, \beta_\delta, \alpha_\epsilon, \beta_\epsilon, \alpha_\zeta, \beta_\zeta. \tag{10}$$

Costello *et al.* applied their method to an instance where the training dataset involved 28 drugs and 35 cell lines and the test data included 18 cell lines. They set the number of components (p) to 22 by considering different combinations of input cell line profiles. It is worth pointing out that they used a form that is more general than Eqn. (9) so that the responses data do not necessarily involve the same set of cell lines for different drugs.

Component-wise kernel Bayesian matrix factorization (cwKBMF)
This was developed on the basis of a matrix factorization that was more general than Eqn. (8)–(9) [17]. Ammad-ud-din et al. took drug profile into account through a drug similarity matrix.

Multi-view multi-task linear regression (MVLR)
This method approaches the linear regression model via Bayesian inference. For m drugs targeting one or more common pathways, a linear regression model assumes the following relationship between the drug responses $R_{n\times m}$ and cell line profiles $C_{n\times s}$:

$$R_{n\times m} \approx C_{n\times s} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_s \end{bmatrix} [a_1, a_2, \cdots, a_m]. \tag{11}$$

In [19], the cell line profile was divided into p parts, i.e.,

$$C_{n\times s} = \begin{bmatrix} C_{n\times s_1}^{(1)} & C_{n\times s_2}^{(2)} & \cdots & C_{n\times s_p}^{(p)} \end{bmatrix},$$

where $\sum_{1\le i\le p} s_i = s$. The rationale of this approach is that each $C_{n\times s_i}^{(i)}$ represents a data view, generated from related features; the corresponding w_i's form the regression coefficient vector for each data view; and a_i is a parameter being set for each drug i. Note that if $p = 1$, $s_1 = s$, Eqn. (11) is equivalent to the relationship between $R[*, j]$ and $C_{n\times s}$ on which the linear regression model Eqn. (1) is based for each j.

Ammad-ud-din *et al.* inferred the parameters (w_j) $(1 \le j \le s)$ and a_i $(1 \le i \le m)$ by assuming the following distributions [19] :

$$a_i \sim \text{Dir}(\alpha), \quad i = 1, 2. \cdots, m;$$
$$w_k \sim \text{Cauchy}(0, \beta_k \gamma_v), \quad \sum_{i=1}^{v-1} s_i + 1 \le k \le \sum_{i=1}^{v} s_i;$$
$$\beta_k \sim \text{Cauchy}_+(0, \epsilon), \ \gamma_v \sim \text{Dir}(\delta),$$

and

$$R[*, j] \sim \mathcal{N}\left(a_j \left(\sum_{k=1}^{s} c_{ik} w_k\right)_{n\times 1}, \sigma_j \mathbf{I}_{n\times n}\right),$$
$$\sigma_j \sim \text{Cauchy}_+(0, \tau),$$

in which there are four free parameters $(\alpha, \epsilon, \delta, \tau)$ for the Dirichlet prior and Cauchy distribution.

MACAU
This is a Bayesian matrix factorization learning method [37]. Yang et al. used this method on drug target information and perturbed pathway information for drug response prediction [22]. The drug response matrix $R_{n\times m}$ is assume to has the following factorization:

$$R_{n\times m} = C_{n\times s} M_{s\times \ell}^{(c)} M_{\ell\times t}^{(d)} S_{t\times m},$$

where t is the number of drug targets, s is the number of perturbed pathways, ℓ is a hyperparameter of the method. Here, $M_{s\times \ell}^{(c)}$ and $M_{\ell\times t}^{(d)}$ are two so-called link matrices that project the side information about cell lines and drugs into the latent matrix factors of the drug response matrix $R_{n\times m}$, respectively.

Macau uses Gibbs sampling to sample both the latent vectors and the link matrices. Having the time complexity $O(st)$, it can support high-dimensional side information when conjugate gradient based noise injection sampler is employed [37].

2.6 Matrix-factorization based methods

Predicting cancer drug response using a recommender system (CaDRRes)

This method was developed under the assumption that the drug response function can be approximated by the sum of a few low-rank matrices. More specifically, the response matrix $R_{n\times m}$ for m drugs and n cell lines has the following decomposition [18]:

$$R_{n\times m} \approx \mu(R)\begin{bmatrix}1 & 1 & \cdots & 1\\ 1 & 1 & \cdots & 1\\ \vdots & \vdots & \ddots & \vdots\\ 1 & 1 & \cdots & 1\end{bmatrix} + \begin{bmatrix}a_1\\ a_2\\ \vdots\\ a_n\end{bmatrix}\begin{bmatrix}1 & 1 & \cdots & 1\end{bmatrix} + \begin{bmatrix}1\\ 1\\ \vdots\\ 1\end{bmatrix}\begin{bmatrix}b_1 & b_2 & \cdots & b_m\end{bmatrix} + K_{n\times n}U_{n\times f}W_{f\times m}, \quad (12)$$

where K is a similarity matrix of the cell lines, f is a small positive integer to be determined, and $\mu(R)$ is the average value of entries in $R_{n\times m}$.

Let $\bar{R}_{n\times m}((a_i),(b_j),U,W)$ denote the right-hand side of Eqn. (12). Suphavilai *et al.* solved the following minimization problem by using the gradient descent method to predict drug responses [18]:

$$\min_{(a_i),(b_j),U,W} \|R_{n\times m} - \bar{R}_{n\times m}((a_i),(b_j),U,W)\|_F^2,$$

where K is calculated via the Pearson's correlation of gene expression profiles for cell lines and $f = 10$.

DualNets

Zhang *et al.* developed this simple but efficient network interpolation method for drug response prediction [23]. They assumed that the response to a known drug d in a new cell line is a weighted combination of the responses of the neighboring cell lines c_i $(i = 1, 2, \cdots, n)$ to drug d as given below:

$$r_{cd} = \sum_{i=1}^{n} w_i r_{c_i d}, \quad (13)$$

where $r_{c_i d}$ is the response level to drug d in cell line c_i and the coefficient w_i is defined as a negative exponential function. Let $s(c, c_i)$ be the gene expression correlation between the two cell lines c and c_i. Then w_i is given as:

$$w_i = \frac{\exp\left(-\frac{(1-s(c,c_i))^2}{2\sigma^2}\right)}{\sum_{j=1}^{n}\exp\left(-\frac{(1-s(c,c_j))^2}{2\sigma^2}\right)}, \quad (14)$$

where σ is the decay rate parameter.

Similarly, the response value of a known cell line to a new drug could be inferred by using drug similarity.

The similarity-regularized matrix factorization (SRMF) method

This method simply assumes that the drug response matrix $R_{n\times m}$ is approximately a product of two low-rank matrices $U_{n\times k}$ (a compression of the cell line profile matrix) and $V_{m\times k}$ (a compression of the drug profile matrix). Wang *et al.* used a cell line similarity matrix S_{cell} and a drug similarity matrix S_{drug} to compute the two low-rank matrices [21] as follows:

$$S_{\text{cell}} \approx U_{n\times k} \times U_{n\times k}^T, \quad S_{\text{drug}} \approx V_{m\times k} \times V_{m\times k}^T.$$

In this way, Wang *et al.* proposed to compute U and V by solving the following optimization problem:

$$\begin{aligned}\min_{U,V} \|B\circ(R-UV^T)\|_F^2 + \lambda_1\|S_{\text{cell}} - UU^T\|_F^2 \\ + \lambda_2\|S_{\text{drug}} - VV^T\|_F^2 + \lambda_3\left(\|U\|_F^2 + \|V\|_F^2\right)\end{aligned} \quad (15)$$

where $B_{n\times m} = (b_{ij})$ is the indicator matrix such that $b_{ij} = 0$ if the drug response value r_{ij} is missing and 1 otherwise, and the operator "$\circ$" is the entry-by-entry matrix product used to mask off the missing entries in R.

2.7 Miscellaneous methods

Random forest with gene filtering (RF-g) and with both gene and sample filtering (RF-gs)

Random forest is a regression approach that captures nonlinear relationships between predictors and response variables [38]. It first uses bootstrapping to split the data into several sub-datasets, fits a decision tree to each of them, and finally averages the results. In order to improve the accuracy of the drug response predictions, Riddick *et al.* incorporated the procedures of filtering gene signatures and removing cell line outliers in the random forest framework [39].

Kernel rank learning (KRL)

This method focuses on accurately recommending the k most sensitive drugs for each cell line [27]. It therefore proposed to use the normalized discounted cumulative gain (DCG) defined in Eqn. (23) to construct the loss function instead of the square of the loss between observed and predicted values:

$$\min_W \sum_{i=1}^{n}(1 - \text{NDCG}(C_{n\times s}[i,*]W, R_{n\times m}[i,*])) + \lambda\sum_{i=1,j=1}^{s,m} w_{ij}^2,$$

where W is a real $s\times m$ matrix. The drugs are recommended on the basis of the predicted rank vector xW for the cell line with the profile x.

TANDEM

When applied to multiple omics data of cell lines, the ENet or other regression methods make drug response prediction mainly on the basis of gene expression profile. To take advantage of information contained in other profiles, Aben et al. developed a two-stage approach to predict drug response, named TANDEM [15]. In this approach, an ENet model is first used to predict drug response using mutation, CNA, cancer type and methylation profiles and then the second ENet model is used to adjust the prediction made in the first stage using the gene expression profile.

Cancer drug response profile scan (CDRscan)

CDRscan is a deep convolutional neural network approach for drug response prediction [24]. CDRscan employs a two-stage convolution architecture, in which the genomic mutational profiles of cell lines and the molecular profiles of drugs are processed separately in Stage 1 and then merged by simulating the 'docking' process in Stage 2.

3 Evaluation Metrics

In this study, we chose nine widely used metrics to evaluate the 17 methods on the basis of the difference between the observed responses, $\mathbf{y} = (y_i)_{i=1}^n$, and the predicated responses, $\widehat{\mathbf{y}} = (\widehat{y}_i)_{i=1}^n$, of n cell lines to a drug.

Root mean square error (RMSE)
The basic RMSE attempts to capture the deviation between the true and predicated responses for a method and is defined as:

$$\mathrm{RMSE}\,(\mathbf{y},\widehat{\mathbf{y}}) = \sqrt{\frac{\sum_i (y_i-\widehat{y}_i)^2}{n}}. \tag{16}$$

To facilitate the comparison of two prediction methods with different scales, one may use the following normalized RMSE:

$$\mathrm{NRMSE}(\mathbf{y},\widehat{\mathbf{y}}) = \frac{\mathrm{RMSE}(\mathbf{y},\widehat{\mathbf{y}})}{\max_i y_i - \min_i y_i}. \tag{17}$$

For a drug, one may be concerned only with the sensitive cell lines only or the resistant cell line only. For these purposes, RMSE is further modified as [12]:

$$\text{L-RMSE}(\mathbf{y},\widehat{\mathbf{y}}) = \sqrt{\frac{\sum_i (y_i-\widehat{y}_i)^2\cdot \mathbf{1}\{y_i\le \ell\}}{\sum_i \mathbf{1}\{y_i\le \ell\}}} \tag{18}$$

$$\text{R-RMSE}(\mathbf{y},\widehat{\mathbf{y}}) = \sqrt{\frac{\sum_i (y_i-\widehat{y}_i)^2\cdot \mathbf{1}\{y_i\ge \gamma\}}{\sum_i \mathbf{1}\{y_i\ge \gamma\}}} \tag{19}$$

$$\text{LR-RMSE}(\mathbf{y},\widehat{\mathbf{y}}) = \sqrt{\frac{\sum_i (y_i-\widehat{y}_i)^2\cdot \mathbf{1}\{(y_i\le \ell)\mathrm{or}(y_i\ge \gamma)\}}{\sum_i \mathbf{1}\{(y_i\le \ell)\mathrm{or}(y_i\ge \gamma)\}}} \tag{20}$$

where $\mathbf{1}\{\}$ denotes the indicator function,

$$\ell = \mathrm{median}(\mathbf{y}) - 1.65\times\sigma,$$

$$\gamma = \mathrm{median}(\mathbf{y}) + 1.65\times\sigma,$$

$$\sigma = 1.4826\times \mathrm{median}_{1<i<n}(|y_i-\mathrm{median}(\mathbf{y})|).$$

Note that ℓ and γ are the left and right cutoff points, respectively, and are computed by the lower and upper limit of the 90% confidence interval for the data following a Gaussian distribution with standard deviation σ.

Pearson correlation coefficient (PCC)
Let $\mu(X)$ denote the mean of a random variable. The PCC of $\mathbf{y}$ and $\widehat{\mathbf{y}}$ is defined as:

$$\mathrm{PCC}(\mathbf{y},\widehat{\mathbf{y}}) = \frac{\sum_i \left(\widehat{y}_i-\mu(\widehat{\mathbf{y}})\right)\left(y_i-\mu(\mathbf{y})\right)}{\sqrt{\sum_i \left(\widehat{y}_i-\mu(\widehat{\mathbf{y}})\right)^2}\sqrt{\sum_i \left(y_i-\mu(\mathbf{y})\right)^2}}. \tag{21}$$

Spearman correlation coefficient (SCC)
Assume that $\mathbf{y} = (y_i)_{i=1}^n$ is sorted into $y_{i_1}\ge y_{i_2}\ge\cdots\ge y_{i_n}$ in decreasing order. The rank of the response y_{i_k}, $r(y_{i_k})$, is defined to be k. The integral vector $\mathbf{r(y)} = (r(y_i))_{i=1}^n$ is called the rank vector of $\mathbf{y}$. The rank vector of $\widehat{\mathbf{y}}$ can be defined similarly. The SCC of $\mathbf{y}$ and $\widehat{\mathbf{y}}$ is then defined as:

$$\mathrm{SCC}\big(\mathbf{y},\widehat{\mathbf{y}}\big) = \mathrm{PCC}\big(\mathbf{r(y)},\mathbf{r}(\widehat{\mathbf{y}})\big). \tag{22}$$

Normalized DCG (NDCG)
This is a widely used score for ranking quality [18, 40]. The DCG for the response prediction made by a method is defined as:

$$\mathrm{DCG}\big(\mathbf{y},\widehat{\mathbf{y}}\big) = \sum_{i=1}^n \frac{2^{-y_i}-1}{\log_2\left(r(-\widehat{y}_i)+1\right)}.$$

The numerator gives more weight to high sensitivity score whereas the denominator gives preference to high ranks. The normalized DCG is further defined as:

$$\mathrm{NDCG}\big(\mathbf{y},\widehat{\mathbf{y}}\big) = \frac{\mathrm{DCG}(\mathbf{y},\widehat{\mathbf{y}})}{\mathrm{DCG}(\mathbf{y},\mathbf{y})} \tag{23}$$

Normalized weighted probabilistic c-index (NWPC)
In the 2014 NCI-DREAM drug sensitivity prediction challenge [11], NWPC was used to evaluate the performance of a method. It measures the agreement between a predicted ranked list and the true list, defined as follows:

$$\text{pc-index}\big(\mathbf{s},\mathbf{r}(\widehat{\mathbf{s}})\big) = \frac{2}{n(n-1)}\sum_{i<j}\mathrm{hp}\big(s_i,s_j,r(\widehat{s}_i),r(\widehat{s}_j),\sigma(\mathbf{s})\big),$$

where $\sigma(X)$ is the standard deviation of a random variable X and the function $\mathrm{hp}\big(s_i,s_j,r(\widehat{s}_i),r(\widehat{s}_j),\sigma(\mathbf{s})\big)$ is defined as

$$\mathrm{hp}\big(s_i,s_j,r(\widehat{s}_i),r(\widehat{s}_j),\sigma(\mathbf{s})\big) = \begin{cases} \frac{1}{2}\Big(1+\mathrm{erf}\big(\frac{s_i-s_j}{2\sigma(\mathbf{s})}\big)\Big), & r(\widehat{s}_i)<r(\widehat{s}_j),\\ 0.5, & r(\widehat{s}_i)=r(\widehat{s}_j),\\ \frac{1}{2}\Big(1+\mathrm{erf}\big(\frac{s_j-s_i}{2\sigma(\mathbf{s})}\big)\Big), & r(\widehat{s}_i)>r(\widehat{s}_j),\end{cases}$$

and $\mathrm{erf}(a) = \frac{2}{\sqrt{\pi}}\int_0^a e^{-t^2}dt$.

For a method, the pc-index is calculated separately for each drug d, denoted $\mathrm{pc}_d = \text{pc-index}\Big(\mathbf{s^{(d)}},\mathbf{r}(\widehat{\mathbf{s}}^{\mathbf{(d)}})\Big)$ where $\mathbf{s^{(d)}},\widehat{\mathbf{s}}^{\mathbf{(d)}}$ are the true and predicted score vectors for the drug d, respectively. The final score of a method M is calculated as the weighted average of the pc-index scores across all the drugs (i.e. the wpc-index) as follows:

$$\text{wpc-index}(M) = \frac{\sum_d w_d\cdot \mathrm{pc}_d}{\sum_d w_d}$$

The weight w_d is calculated from the empirical null distribution of pc_d. For each drug d, (i) produce a random ranking of n items, denoted $\tilde{\mathbf{r}}^{\mathbf{(d)}}$, and calculate the $\text{pc-index}\Big(\mathbf{s^{(d)}},\tilde{\mathbf{r}}^{\mathbf{(d)}}\Big)$; (ii) repeat this procedure 10,000 times to construct an empirical null distribution with a median of μ_d and a standard deviation of σ_d; (iii) compute the pc-index with the standard ranking $\mathbf{r}(\mathbf{s^{(d)}})$, $\mathrm{pc}_d^* = \text{pc-index}\Big(\mathbf{s^{(d)}},\mathbf{r}(\mathbf{s^{(d)}})\Big)$; (iv) drug weight $w_d = \frac{\mathrm{pc}_d^*-\mu_d}{\sigma_d}$.

In order to scale the range of the wpc-index into $[0,1]$, use the transformation:

$$\mathrm{NWPC} = \frac{\mathrm{wpc}-\mathrm{wpc}_{\min}}{\mathrm{wpc}_{\max}-\mathrm{wpc}_{\min}} \tag{24}$$

where $\mathrm{wpc}_{\max},\mathrm{wpc}_{\min}$ are obtained from the real ranking vector $\mathbf{r}(\mathbf{s^{(d)}})$ and its inverse, respectively, for each drug when computing pc_d.

Unlike the previous eight metrics, the wpc-index measures the overall performance for all the predicted drugs. Therefore, only one value is given for each method.

Lastly, we point out that a smaller value indicates good performance for the first five metrics introduced above, whereas a value close to 1 suggests good performance for the last four metrics.

4 Test datasets

We assessed the selected methods with four public drug response datasets that were respectively downloaded from:

- The GDSC website (https://www.cancerrxgene.org/downloads, accessed 1 Feb. 2019),
- the CCLE website (https://portals.broadinstitute.org/ccle/data, accessed 1 Aug. 2019),
- the NCI-60 (version 2.6.0) [5] and
- The CTRP website (https://portals.broadinstitute.org/ctrp/, accessed 1 Aug. 2019).

The curated datasets (Tables S1–S44) are available on the first author's website: https://github.com/Jinyu2019/Suppl-data-BBpaper.

4.1 The GDSC data

Genomic datasets

The GDSC genomic profiles include the gene expression, CNA, mutation, and methylation datasets. These datasets were preprocessed by (i) removing genes with little variation in expression, (ii) extracting both mutations and CNAs for the cancer genes listed in the COSMIC database (https://cancer.sanger.ac.uk/cosmic/download, version 88), and (iii) keeping CpG probes with variation in methylation levels, respectively. In the end, we obtained:

- A gene expression profile $E_{734\times 8046}$,
- A DNA methylation profile $D_{734\times 8473}$,
- A mutation profile $M_{734\times 636}$, and
- A CNA profile $C_{734\times 694}$

for 734 cancer cell lines (Tables S1–S4), of which annotations are provided in Table S5.

Drug response dataset

The level of drug response was measured in both IC_{50} and AUC in the GDSC database. After removing drugs of which the responses are missing for 80% or more cell lines, we obtained two drug response matrices (Tables S6 and S7) for 734 tumor cell lines and 201 drugs. Table S6 summarizes the common logarithm of the IC_{50} drug response readouts, whereas Table S7 summarizes the AUC readouts.

Protein–protein interaction network

A protein–protein interaction (PPI) network on 7,136 proteins (Table S8) was derived by extracting interactions relations between them in the Pathway Commons database (http://www.pathwaycommons.org/, accessed 11 March 2019) [41].

Drug groups and drug similarity matrix

The structural information on the 201 drugs include 1,444 1-dimensional, 2-dimensional descriptors and nine types of molecular fingerprints in PubChem [42] (Table S9). The structural information on the drugs were further processed with the PaDEL-Descriptor [43] to construct a drug similarity network, named $S^{201\times 201}$ (Table S10).

The 201 drugs were divided into 23 groups (Table S12), each containing drugs targeting the same pathway, according to the information available in the GDSC database.

Drug-targeted pathways

We obtained the drug-targeted pathways via the data preprocessing method used in [17]. We downloaded 1,329 canonical pathways and 189 oncogenic signature gene sets in C2:CP and C6 collections in the MSigDB database (http://software.broadinstitute.org/ gsea/msigdb/index.jsp, accessed 1 Feb 2019) [44]. We further extracted only those pathways that contained one or more primary targets of the 201 drugs, leading to 71 drug-targeted pathways. All the genes that were not in any extracted pathways formed a single group named "others". We obtained 72 gene subsets in total (Table S11).

4.2 The CCLE data

The CCLE data contain six genomic profiles: gene expression, microRNA expression, protein expression, DNA methylation, mutation and CNA. These six profiles were preprocessed in the same way as we preprocessed the GDSC data. The final CCLE data for 385 cancer cell lines consisted of:

- an expression profile $E_{385\times 17726}$ of 17,726 genes,
- an expression profile $O_{385\times 734}$ of 734 microRNAs,
- an expression profile $P_{385\times 214}$ of 214 proteins,
- a methylation profile $D_{385\times 4507}$ of 4,507 DNA methylation sites,
- a mutational profile $M_{385\times 444}$ of 444 cancer genes, and
- a CNA profile $C_{385\times 700}$ of 700 cancer genes,

which can be found in Tables S13–S18.

The CCLE drug response datasets contain both IC_{50} and AUC readouts for 23 drugs and 385 cell lines (Tables S19–S20). Other CCLE datasets used for assessment include drug-similarity matrix, drug group information, a protein–protein interaction network, drug structural information, cell-line tissue information and gene groups (Tables S21–S26).

4.3 The NCI-60 data

The NCI-60 data were downloaded via calling a R program named rcellminer [45]. The NCI-60 genomic profiles were preprocessed in the same way as we preprocessed the GDSC and CCLE data. The final data for 59 cancer cell lines included:

- an expression profile $E_{59\times 18077}$ of 18,977 genes,
- an expression profile $O_{59\times 410}$ of 410 microRNAs,
- an expression profiles $P_{59\times 162}$ of 162 proteins,
- a methylation profiles $D_{59\times 11330}$ of 11,330 DNA methylation sites,
- a mutation profiles $M_{59\times 443}$, and
- a CNA profile $C_{59\times 394}$ of 394 cancer genes,

which are found in Tables S27–S32.

The NCI-60 drug response datasets contain only GI_{50} readouts for 215 drugs and 59 cell lines (Table S33), where GI_{50} is the concentration for 50% of maximal inhibition of cell proliferation. Other CCLE datasets used for assessment include a protein–protein interaction network, drug-similarity matrix, drug structural information, cell-line tissue information and gene group information, drug group information (Tables S34–S39). .

4.4 The Cancer Therapeutics Response Portal (CTRP) dataset

The CTRP data (version 2.1) include gene expression profile (Table S40), AUC readouts for 63 drugs in 720 caner cell lines (Table S41), cell line tissue information (Table S42) and drug group information (Table 43) [29].

5 Assessment methods

Using the GDSC, CCLE, NCI-60 and CTPR datasets, we assessed the following 13 representative methods developed in the past five years in the nine metrics (Eqn. (16)–(24)):

- CaDRReS (`https://github.com/CSB5/CaDRReS`),
- CDRscan (`https://github.com/summatic/CDRScan`),
- cwKBMF (`https://research.cs.aalto.fi/pml/`),
- DualNets (coded in R by the first author,)
- GENet-Lap and GENet-NLap (available in R),
- KRL (`https://github.com/BorgwardtLab`),
- MACAU (`https://github.com/saezlab`),
- MERGE (`http://merge.cs.washington.edu/`),
- MVLR (`https://github.com/suleimank/mvlr`),
- pairwiseMKL (`https://github.com/aalto-ics-kepaco`),
- RWEN-left (resp. right, and both) (`https://github.com/kiwtir/RWEN`),
- SRMF (`https://github.com/linwang1982/SRMF`),
- TANDEM (`http://ccb.nki.nl/software/tandem`),

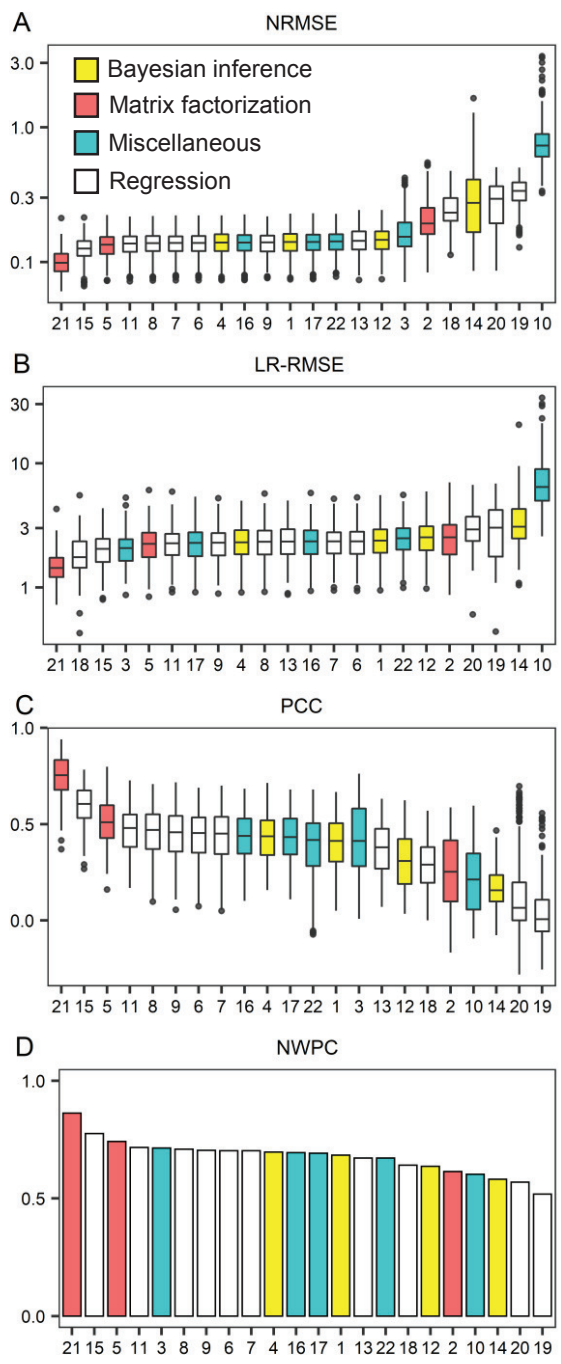

Fig. 1. The performance of the 17 methods on the GDSC IC_{50} dataset. Box plots were drawn for the normalized root mean square error (NRMSE) (A), the left-right (LR)-RMSE (B), the Pearson correlation coefficient (PCC) (C) and the normalized weighted probabilistic c-index (NWPC) (D), where the methods were listed from the best (leftmost) to the worst. Each box plot was drawn on the basis of the scores of 10 repeated cross validation tests for the following methods: BMTMKL (1), CaDRReS (2), CDRscan (3), cwKBMF (4), DualNets (5), ENet (6), M-ENet (7), GENet-Lap (8), GENet-NLap (9), KRL (10), KRR (11), MACAU (12), MERGE (13), MVLR (14), pairwiseMKL (15), RF-g (16), RF-gs (17), RWENet-both (18), RWENet-left (19), RWENet-right (20), SRMF (21), TANDEM (22).

as well as four classical approaches: BMTMKL (`https://github.com/mehmetgonen/bmtmkl`), ENet, RF-g and RF-gs, which are available in R.

Table 2. The run time of the 17 methods on the GDSC data. The CDRscan was assessed on a GPU machine, whereas the rest were assessed on a PC cluster with 6 cores and 45GB RAM. The time taken by parameter selection was included into the run time for each method. GNet-Lap and GNet-NLap are two variants of GNet; M-ENet is a variant of ENet that integrated multiple genomic profiles of cell lines; RWENet-left, RWENet-right and RWENet-both are three variants of RWENet;

Method	Time (min)	Method	Time (min)
Regression		**Bayesian inference**	
MERGE	45	MACAU	18
KRR	66	cwKBMF	110
ENet	91	M-BMTMKL	163
M-ENet	169	MVLR	1289
pairwiseMKL	974	**Matrix-factorization**	
RWENet-right	993	DualNets	43
RWENet-left	1299	CaDRRes	50
RWENet-both	1565	SRMF	1425
GENet-NLap	3125		
GENet-Lap	3168		
Miscellaneous			
TANDEM	142		
CDRscan	865		
RF-g	1392		
RF-gs	2088		
KRL	3153		

The auxiliary omics data used by the methods are provided in Table 1. The input data were prepared in the format recommended by the developer for each method.

We used the five-cross validation approach to assess the 17 methods. All the cell lines were randomly partitioned into five equal parts, four of which were used for training and the other part was used for test. The partition was done under the constraint that for each tissue type, the cell lines in the training data were kept in the same proportion as in the test data. The predicted response level to the drugs in all cell lines were then compared against the observed values in each metric. This assessment process was repeated 10 times to obtain a robust evaluation for each dataset.

To find out the extend to which the prediction accuracy increases through integrating multiple genomic profiles other than gene expression, we ran ENet and BMTMKL in two settings: (i) gene expression profile only and (ii) multiple profiles including gene expression, mutational, CNA, and DNA methylation profiles. The multi-view version of them are denoted M-Enet and M-BMTMKL in Table 2, respectively.

A method that has low accuracy in general may perform extremely well when applied to the cell lines from a tissue or the drugs within group. To find out whether this is true or not, we tested the methods on the datasets containing cell lines derived from only one tissue. We assessed the methods on the following tissue specific datasets:

- GDSC: breast, central nervous, haematopoietic and lymphoid, lung and skin;
- CCLE: breast, haematopoietic and lymphoid, lung, ovary and skin;
- NCI-60: melanoma and lung.

We also tested whether a method trained in a dataset still performs well on another drug response dataset. To this purpose, we trained the methods on the GDSC data and test them on the CCLE and CTRP datasets, respectively.

Lastly, the one-sided paired Wilcoxon Sign-Rank test ([46, page 338]) was used to determine whether or not a method was significantly better than another in each metric. The heat map of the -$\log_{10}$(p-values) of the tests is summarized for pairwise method comparison.

The CDRscan method was validated on a PC cluster empowered with a GPU (Nvidia Tesla V100-SXM2). All other methods were validated on a PC cluster with 6 cores and 45GB RAM. The computational time of different methods are summarized in Table 2.

6 Results and discussions

The test results on the GDSC data are summarized in Figures 1–3 (and Supplementary file 1: Figures S1–S3) for the IC_{50} drug response readouts and in Supplementary file 2: Figures S9–S11 for the AUC readouts. The results on the CCLE data are summarized in Supplementary file 3: Figures S17–S19 for the IC_{50} readouts and in Supplementary file 4: Figures S25–S27 for the AUC readouts. The results on the NCI-60 data are summarized in Supplementary file 5: Figures S33–S35 for the GI_{50} readouts.

The performance of the methods on the GDSC tissue-specific data are summarized in Supplementary file 1: Figures S4–S8 and Supplementary file 2: Figures S12–S16 for the IC_{50} and AUC readouts, respectively. The performance of the methods on the CCLE tissue-specific data are summarized in Supplementary file 3: Figures S20–S24 and Supplementary file 4: Figures S28–S32 for the IC_{50} and AUC readouts, respectively. The performance of the methods on the NCI-60 tissue-specific data with the GI_{50} readouts are summarized in Supplementary file 5: Figures S36–S37.

The performance of the methods when trained on the GDSC data and tested on the CCLE data are summarized in Supplementary file 6: Figures S38–S44 for the IC_{50} readouts and Supplementary file 7: Figures S45–S51 for the AUC readouts. The performance of the methods when trained on the GDSC AUC data and tested on the CTRP AUC data are summarized in Supplementary file 8: Figures S52–S58.

6.1 Pros and cons of the methods

Different methods have quite different time complexity (Table 2). The fastest methods such as MACAU, MERGE and DualNets took less than an hour on average, including the parameter selection time. The slowest methods include GELNet, KRL, SRMF and RWENet and took over 20 hours.

Our studies suggest that the assessed methods can be divided into three groups in terms of prediction accuracy: good methods (DualNets (Method 5 in the figures), KRR (11), pairwiseMKL (15) and SRMF (21)), unstable methods (BMTMKL, CDRscan, cwKDMF, ENet and M-ENet, GENet-Lap and GENet-NLap, MACAU, MERGE, RF and TANDEM), and poor methods (CaDRReS (2), KRL(10), MVLR (14) and RWENet (18)) (Figure 1, Figures S1, S9, S17, S25 and S33). The four good methods outperformed constantly the other methods almost on every dataset in each metric, whereas the poor methods had the lowest accuracy almost on every test.

The accuracy of each unstable method fluctuated dramatically across datasets and evaluation metrics. For instance, on the GDSC dataset, the deep learning method CDRscan had the forth or fifth best performance if LR-RMSE, L-RMSE, R-RMSE or NDCG was used for evaluation; in contrast, it had the 7-th worst performance if each of the RMSE and NRMSE was used.

When trained on the GDSC data and tested on the CCLE or CTRP data, the four good methods did not retain a clear edge in performance (Figures S38–S58).

Our results show that the matrix-factorization-based approach, represented by SRMF, performed better than the regression models (with or without kernel extension), whereas the regression models performed better than the different Bayesian inference methods. This has the following implications.

First, the drug response is likely a complicated non-linear function of the cell line profiles and the drug profiles. When predicting the response to a known drug in a new cell line, as one of the best methods, DualNets assumes that the drug response level is linear to the drug profile but non-linear to the cell line profile (Eqn. (13)–(14)), whereas ENet and GENet and other regression methods consider the drug response function to be linear to the cell line profiles (see Eqn. (1)).

SRMF, another best method, assumes that the drug response matrix is a product of the low-rank proximity of the cell line similarity matrix and the low-rank proximity of the drug similarity matrix, which are both non-linear to cell line profiles and drug profiles, respectively.

Second, Bayesian inference is usually powerful for solving machine learning problems. Possible causes for the ineffectiveness of the assessed Bayesian inference methods may include the distribution of drug responses may not be as simple as a normal distribution, the parameter space in each method may be too large to compute an optimal set of parameter values and the parameter space is so large that the methods may overfit the training dataset. Our Shapiro-Wilk normality test indicates that most drug response dataset do not follow a normal distribution.

6.2 Trade-off between feature selection and prediction accuracy

Feature selection is an important function for methods that are developed for drug response prediction. It allows to identify valuable biomarkers and thus enables the follow-up biological studies into drugs. The methods with this function include ENet and its generalization GEnet, MERGE, MVLR, RF, RWEN and TANDEM (Table 1). ENet and GENet are the best among these methods. They were fast (Table 2) and had slightly lower accuracy than the most accurate methods DualNets, KRR, pairwiseMKL and SRMF in our test (Figure 1, Supplementary file 1: Figure S2, Supplementary file 2: Figure S9, Supplementary file 3: Figure S17, Supplementary file 4: Figure S25 and Supplementary file 5: Figure S33).

6.3 Protein–protein interaction networks are informative

GENet is the only model that uses explicitly cellular protein–protein interaction networks for drug response prediction. GENet-Lap and GENet-NLap were the two versions developed from ENet through incorporating the (normalized) Laplacian matrix of a protein–protein interaction network into the last regularization term in Eqn. (1). Interestingly, the two methods outperformed ENet for seven out of the nine metrics (Figure 1–2, see also Figure S1–S2). This fact may indicate that drug response prediction can benefit from the integration of both protein–protein interaction networks and gene expression profiles with the drug response data.

6.4 Mutation, CNV and methylation profiles of cell lines and drug grouping add little to drug response prediction

We assessed ENet and BMTMKL in two settings: (a) only gene expression profile was used and (b) the gene expression, mutation, CNA and methylation profiles were integrated. We did not observe clear improvement in accuracy when multiple genomic profiles of cell lines were used (Figure 1–2 and Figure S1–S2).

MVLR is the only method that used the information on drug-targeted pathways. It builds a linear regression model for each group of drugs that have the same target pathway. Its predictive performance varied from group to group (Figure 3). Both Figures 1 and 3 suggest that MVLR performed better only on four out of 23 drug groups than without group information.

6.5 The performance of a method varies a lot under different measurements

We used nine performance metrics in this study. RMSE and NRMSE measure the performance of a method on all the data points, whereas

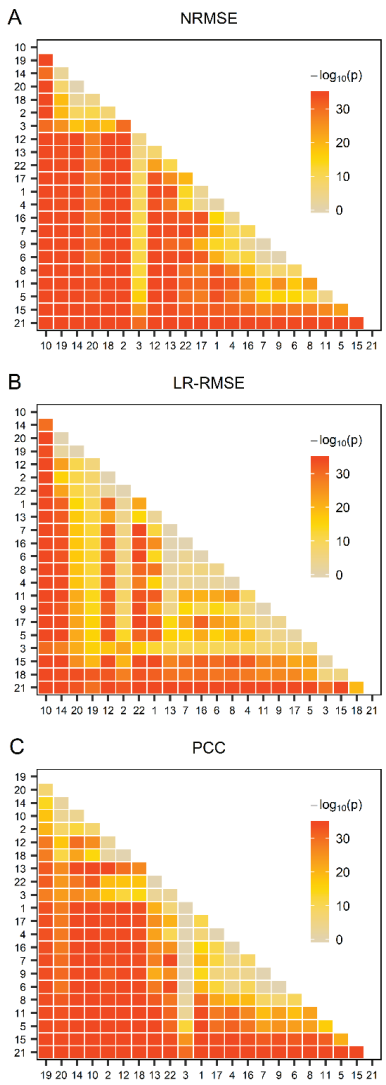

Fig. 2. The comparison of the performance of the methods on the GDSC IC_{50} dataset. The one-sided paired Wilcoxon Sign-Rank test was conducted on the data appearing in Figure 1 (A–C). The heatmap was drawn from the negative common logarithm of the p-value for the hypothesis that a method performs better than another. BMTMKL (1), CaDRReS (2), CDRscan (3), cwKBMF (4), DualNets (5), ENet (6), M-ENet (7), GENet-Lap (8), GENet-NLap (9), KRL (10), KRR (11), MACAU (12), MERGE (13), MVLR (14), pairwiseMKL (15), RF-g (16), RF-gs (17), RWENet-both (18), RWENet-left (19), RWENet-right (20), SRMF (21), TANDEM (22).

L-RMSE, R-RMSE and LR-RMSE focus on the performance at the head, tail or both sides of the drug response data distribution. SCC, NDCG and NWPC consider all the data points but measure the rank coherency between the predicted and observed values.

Our study suggests that the performance can be very sensitive to the metric selected for assessment for unstable methods. An unstable method could significantly outperform another method for a metric, but perform poorly when comparing other metrics. For instance, cwKBMF significantly outperformed RF-gs, with P-value 10^{-3}, 10^{-4} and 10^{-6} in the RMSE, NRMSE and SCC, respectively (Figures 2 and S2); in contrast, RF-gs significantly outperformed cwKBMF, with P-value 10^{-10}, 10^{-9} and 10^{-2} in L-RMSE, LR-RMSE and NDCG; and there was no significant difference between cwkBMF and RF-gs for R-RMSE and PCC.

Taken together, the two facts suggest that three or four different metrics such as NRMSE, PCC and NDCG may be used for a robust assessment of a method for drug response prediction.

7 Conclusion

We assessed four classical methods and 13 representative computational methods that have recently been developed for drug response prediction using public cell line–drug response data and nine metrics. This work provides lessons worth considering ahead of future research.

First, the study reconfirms that the gene expression profiles of cell lines provide crucial information for predicating the responses to a drug in cell lines [10, 11]. How to generate and use the time point series expression profiles of cell lines is an important issue for drug response prediction in future preclinical and clinical research.

It is commonly accepted that omics data other than gene expression profiles of cell lines often add slight improvements for drug response prediction [9, 10]. However, the excellent performance of DualNets suggests that smart integration of other omics data, particularly drug structure data and protein interaction networks, with gene expression profiles should not be ignored in the study of drug response prediction.

Second, SRMF learns drug response functions from the similarity matrices of cell lines and drugs by an iterative approach. This may suggest deep neural network and other AI approaches are likely to be powerful for drug response prediction. Given that the deep learning method CDRscan did not perform better than the top four methods mentioned above, it is non-trivial to apply deep learning approach to drug response prediction. Another research question is how to improve the performance of Bayesian inference based methods for drug response prediction.

Third, interpretability and reproducibility are two important factors for drug response models in translation medicine research [11, 47, 48]. For pan-cancer drug response data, different metrics give different pictures of the evaluated methods' performance. Linear regression models and their generalizations are good at feature selection and thus has advantage in interpretability. Unfortunately, these methods have low accuracy. Therefore, how to balance predication accuracy and interpretability is an important issue. Aben et al. has made a promising progress toward this goal

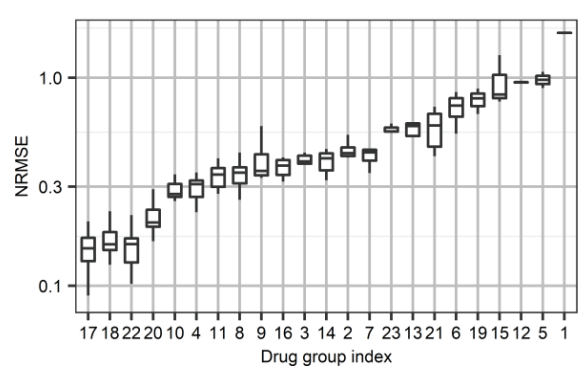

Fig. 3. The distribution of the NRMSE scores of the MVLR method when trained and tested within drug groups. All drugs were divided into 23 groups depending on their signaling pathway target. For each group, MVLR learnt a model and made prediction. The drug groups are indexed alphabetically: ABL signaling (1), Apoptosis regulation (2), Cell cycle (3), Chromatin histone acetylation (4), Chromatin histone methylation (5), Chromatin other (6), Cytoskeleton (7), DNA replication (8), EGFR signaling (9), ERK MAPK signaling (10), Genome integrity (11), Hormone-related (12), IGFR signaling (13), JNK and p38 signaling (14), Metabolism (15), Mitosis(16), other drugs (17), Kinases (18), P53 pathway (19), PI3K/MTOR signaling (20), Protein stability and degradation (21), RTK signaling (22), WNT signaling (23).

in [15]. In future, computational biologists may focus on cancer-specific drug response data to design robust prediction methods with interpretable features [49].

Key Points

- The computational methods for drug response prediction can be grouped into three classes: (a) linear regression models (with or without kernel extension), (b) Bayesian inference methods and (c) matrix factorization via nonlinear optimization methods. In our test, the matrix factorization approach outperformed the linear regression models, whereas linear regression models outperformed different Bayesian inference methods. It remains unclear how to improve the performance of Bayesian inference methods for drug response prediction.
- SRMF with the best performance learns drug response functions from the similarity matrices of cell lines and drugs by an iterative approach, revealing that the drug response functions of cell lines are non-linear and complicated. This may suggest deep neural network and other AI approaches are likely to be able to achieve high accuracy for drug response prediction. On the other hand, our analyses also suggest that it is far from trivial to design a deep learning method that surpasses the top regression or matrix factorization-based methods for drug response prediction.
- The mutation, CNV and methylation profiles of cell lines and drug grouping seem to add little to drug response prediction when they are used with the existing methods. It is worth studying how to use wisely these omics data.
- The performance of a method is sensitive to the choice of measurements. One method could significantly outperform another method for one metric, but performs poorly when comparing other metrics. Three or more metrics that are defined for assessing different aspects of drug response predication, such as the combination of NRMSE, PCC and NDCG, should be used for evaluation of a new method.
- Four large datasets of cell line genomic profiles and drug profiles were curated for benchmarking for future study into drug response prediction.

Supplementary documents

Supplementary tables are available at
`https://github.com/Jinyu2019/Suppl-data-BBpaper`

Funding

This work was supported by Singapore National Research Fund [Grant: NRF2016NRF-NSFC001-026].

Acknowledgements

The authors would like to thank Rudi Alberts for help in testing CDRscan and Lin Wang for useful discussion on matrix factorization based methods. They also thank anonymous reviewers for their useful comments on the first version of the manuscript, which improve significantly our work.

References

1. Dugger SA, Platt A, Goldstein DB. Drug development in the era of precision medicine. Nat Rev Drug Discov. 2018;17(3):183–196.
2. Roses AD. Pharmacogenetics in drug discovery and development: a translational perspective. Nat Rev Drug Discov. 2008;7(10):807–817.
3. Alley MC, Scudiero DA, Monks A, et al. Feasibility of drug screening with panels of human tumor cell lines using a microculture tetrazolium assay. Cancer Res. 1988;48(3):589–601.
4. Weinstein JN. Drug discovery: cell lines battle cancer. Nature. 2012;483(7391):544–545.
5. Shoemaker RH. The NCI-60 human tumour cell line anticancer drug screen. Nat Rev Cancer. 2006;6(10):813–823.
6. Barretina J, Caponigro G, Stransky N, et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012;483(7391):603–607.
7. Garnett MJ, Edelman EJ, Heidorn SJ, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature. 2012;483(7391):570–575.
8. Azuaje F. Computational models for predicting drug responses in cancer research. Brief Bioinform. 2016;18(5):820–829.
9. Ali M, Aittokallio T. Machine learning and feature selection for drug response prediction in precision oncology applications. Biophys Rev. 2019;11(1):31–39.
10. Jang IS, Neto EC, Guinney J, et al. Systematic assessment of analytical methods for drug sensitivity prediction from cancer cell line data. In: Proc. Pac Symp Biocomput. World Scientific; 2014. p. 63–74.
11. Costello JC, Heiser LM, Georgii E, et al. A community effort to assess and improve drug sensitivity prediction algorithms. Nat Biotechnol. 2014;32(12):1202–1212.
12. Basu A, Mitra R, Liu H, et al. RWEN: response-weighted elastic net for prediction of chemosensitivity of cancer cell lines. Bioinformatics. 2018;34(19):3332–3339.
13. Li Q, Shi R, Liang F. Drug sensitivity prediction with high-dimensional mixture regression. PLoS One. 2019;14(2):e0212108.
14. Sokolov A, Carlin DE, Paull EO, et al. Pathway-based genomics prediction using generalized elastic net. PLoS Comput Biol. 2016;12(3):e1004790.
15. Aben N, Vis DJ, Michaut M, et al. TANDEM: a two-stage approach to maximize interpretability of drug response models based on multiple molecular data types. Bioinformatics. 2016;32(17):413–420.
16. Yuan H, Paskov I, Paskov H, et al. Multitask learning improves prediction of cancer drug sensitivity. Sci Rep. 2016;6:31619.
17. Ammad-ud-din M, Khan SA, Malani D, et al. Drug response prediction by inferring pathway-response associations with kernelized Bayesian matrix factorization. Bioinformatics. 2016;32(17):455–463.
18. Supahvilai C, Bertrand D, Nagarajan N. Predicting cancer drug response using a recommender system. Bioinformatics. 2018;34(22):3907–3914.
19. Ammad-ud-din M, Khan SA, Wennerberg K, et al. Systematic identification of feature combinations for predicting drug response with Bayesian multi-view multi-task linear regression. Bioinformatics. 2017;33(14):359–368.
20. Lee SI, Celik S, Logsdon BA, et al. A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. Nat Commun. 2018;9(1):42.
21. Wang L, Li X, Zhang L, et al. Improved anticancer drug response prediction in cell lines using matrix factorization with similarity regularization. BMC Cancer. 2017;17(1):513.
22. Yang M, Simm J, Lam CC, et al. Linking drug target and pathway activation for effective therapy using multi-task learning. Scientific reports. 2018;8(1):8322.

23. Zhang N, Wang H, Fang Y, et al. Predicting anticancer drug responses using a dual-layer integrated cell line-drug network model. PLoS Comput Biol. 2015;11(9):e1004498.
24. Chang Y, Park H, Yang HJ, et al. Cancer drug response profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature. Sci Rep. 2018;8(1):8857.
25. Rahman R, Dhruba SR, Ghosh S, et al. Functional random forest with applications in dose-response predictions. Sci Rep. 2019;9(1):1628.
26. Cichonska A, Pahikkala T, Szedmak S, et al. Learning with multiple pairwise kernels for drug bioactivity prediction. Bioinformatics. 2018;34(13):509–518.
27. He X, Folkman L, Borgwardt K. Kernelized rank learning for personalized drug recommendation. Bioinformatics. 2018;34(16):2808–2816.
28. Su R, Liu X, Xiao G, et al. Meta-GDBP: a high-level stacked regression model to improve anticancer drug response prediction. Brief Bioinform. 2019;bbz022.
29. Rees MG, Seashore-Ludlow B, Cheah JH, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. Nature Chem Biol. 2016;12(2):109.
30. Neubig RR, Spedding M, Kenakin T, et al. International union of pharmacology committee on receptor nomenclature and drug classification. XXXVIII. Update on terms and symbols in quantitative pharmacology. Pharmacol Rev. 2003;55(4):597–606.
31. Caruana R. Multitask learning. Machine learning. 1997;28(1):41–75.
32. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Statist Soc B. 2005;67(2):301–320.
33. Friedman J, Hastie T, Höfling H, et al. Pathwise coordinate optimization. Annals Applied Stat. 2007;1(2):302–332.
34. Murphy KP. Machine learning: a probabilistic perspective. MIT press, MA, USA; 2012.
35. Cortes C, Mohri M, Rostamizadeh A. Algorithms for learning kernels based on centered alignment. J Machine Learning Res. 2012;13(March):795–828.
36. Shawe-Taylor J, Cristianini N, et al. Kernel Methods for Pattern Analysis. Cambridge university press; 2004.
37. Simm J, Arany A, Zakeri P, et al. Macau: Scalable Bayesian factorization with high-dimensional side information using MCMC. In: 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE; 2017. p. 1–6.
38. Breiman L. Random forests. Machine Learning. 2001;45(1):5–32.
39. Riddick G, Song H, Ahn S, et al. Predicting in vitro drug sensitivity using random forests. Bioinformatics. 2010;27(2):220–224.
40. Järvelin K, Kekäläinen J. Cumulated gain-based evaluation of IR techniques. ACM Transactions on Information Systems. 2002;20(4):422–446.
41. Cerami EG, Gross BE, Demir E, et al. Pathway Commons, a web resource for biological pathway data. Nucleic Acids Res. 2010;39(Suppl 1):D685–D690.
42. Gindulyte A, Shoemaker BA, Yu B, et al. PubChem 2019 update: improved access to chemical data. Nucleic Acids Res. 2018;47(D1):D1102–D1109.
43. Yap CW. PaDEL-Descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem. 2011;32(7):1466–1474.
44. Liberzon A, Birger C, Thorvaldsdottir H, et al. The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417–425.
45. Luna A, Rajapakse VN, Sousa FG, et al.. rcellminer: exploring molecular profiles and drug response of the NCI-60 cell lines in R. Oxford University Press; 2015.
46. Rosner B. Fundamentals of Biostatistics. Boston: Brooks/Cole, Cengage Learning, USA; 2011.
47. Fey D, Halasz M, Dreidax D, et al. Signaling pathway models as biomarkers: patient-specific simulations of JNK activity predict the survival of neuroblastoma patients. Sci Signal. 2015;8(408):ra130.
48. Zhang F, Wang M, Xi J, et al. A novel heterogeneous network-based method for drug response prediction in cancer cell lines. Sci Rep. 2018;8(1):3355.
49. Fowles JS, Brown KC, Hess AM, et al. Intra-and interspecies gene expression models for predicting drug response in canine osteosarcoma. BMC Bioinformatics. 2016;17(1):93.