

Systems biology

Kernelized rank learning for personalized drug recommendation

Xiao He^{1,2,†,*}, Lukas Folkman^{1,2,†,*} and Karsten Borgwardt^{1,2}

¹Machine Learning and Computational Biology Lab, Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland and ²Swiss Institute of Bioinformatics, Basel, Switzerland

*To whom correspondence should be addressed.

†These authors contributed equally to this work.

Associate Editor: Jonathan Wren

Received on July 7, 2017; revised on January 4, 2018; editorial decision on March 4, 2018; accepted on March 7, 2018

Abstract

Motivation: Large-scale screenings of cancer cell lines with detailed molecular profiles against libraries of pharmacological compounds are currently being performed in order to gain a better understanding of the genetic component of drug response and to enhance our ability to recommend therapies given a patient's molecular profile. These comprehensive screens differ from the clinical setting in which (i) medical records only contain the response of a patient to very few drugs, (ii) drugs are recommended by doctors based on their expert judgment and (iii) selecting the most promising therapy is often more important than accurately predicting the sensitivity to all potential drugs. Current regression models for drug sensitivity prediction fail to account for these three properties.

Results: We present a machine learning approach, named Kernelized Rank Learning (KRL), that ranks drugs based on their predicted effect per cell line (patient), circumventing the difficult problem of precisely predicting the sensitivity to the given drug. Our approach outperforms several state-of-the-art predictors in drug recommendation, particularly if the training dataset is sparse, and generalizes to patient data. Our work phrases personalized drug recommendation as a new type of machine learning problem with translational potential to the clinic.

Availability and implementation: The Python implementation of KRL and scripts for running our experiments are available at <https://github.com/BorgwardtLab/Kernelized-Rank-Learning>.

Contact: lukas.folkman@bsse.ethz.ch or xiao.he@bsse.ethz.ch

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

One of the key goals of precision medicine is the ability to suggest personalized therapies to patients based on their molecular profiles. This approach is particularly interesting for cancer treatment given the heterogeneous nature of this disease, with somatic alterations in cancer genes shown to be the determinants of a patient's response to therapy (Chapman *et al.*, 2011). For the development of targeted treatments, large collections of patients with recorded clinical outcomes and molecularly profiled tumor samples are needed. However, the cost of such a medical records knowledge bank is currently prohibitive except for a few encouraging examples (Gerstung

et al., 2017). Therefore, pre-clinical biological models, such as cultured human cancer cell lines, are a relatively inexpensive alternative approach for finding biomarkers. Recently, several large-scale drug sensitivity screens of genomically profiled cell lines ranging different cancer (sub)types have been established (Barretina *et al.*, 2012; Iorio *et al.*, 2016; Seashore-Ludlow *et al.*, 2015). One of the current challenges lies in building accurate predictive models and translating these models into the clinic.

A number of regression models (Costello *et al.*, 2014; Jang *et al.*, 2014) have been proposed to predict drug sensitivity measured by the half maximal inhibitory concentration (IC₅₀) or the area under

the fitted dose response curve (AUC). These methods typically build a predictor for each drug individually (Aben *et al.*, 2016; Barretina *et al.*, 2012; Iorio *et al.*, 2016), use multi-task learning (Ammad-ud-din *et al.*, 2016; Gönen and Margolin, 2014) or combine genomic profiles with information about the drug’s chemical structure (Ammad-ud-din *et al.*, 2014; Menden *et al.*, 2013; Zhang *et al.*, 2015). However, there are three potential problems with these approaches. Firstly, medical records knowledge banks are much sparser than cell line panels, meaning that only a few therapies are recorded for each patient compared to a comprehensive screen of tens to hundreds of drugs for each cell line. Secondly, these therapies are carefully selected by experienced doctors based on their expert judgment. Finally, the regression-based methods are not directly optimized for the prediction of the clinically relevant case in which a clinician needs to know a few most-suited drugs for a given patient.

Here, we phrase the personalized drug recommendation as a ranking problem (Section 2) and propose a method, named Kernelized Rank Learning (KRL), that is directly optimized for the clinically relevant scenario of personalized drug recommendation illustrated above. To this end, we extend the approach of Weimer *et al.* (2007) and Teo *et al.* (2010) in Section 3. The key difference between KRL and the available regression approaches is that for a given cell line (patient), KRL directly predicts the top ranking drugs rather than the exact sensitivity values for all potential drugs. Our experimental evaluation using a dataset of cancer cell lines (Section 4) shows that KRL outperforms competing approaches, particularly in the realistic scenario when only a few training examples per cell line are available. Next, using publicly available patient data and the gene expression homogenization pipeline of Geeleher *et al.* (2014, 2017), we show that even when KRL is trained on cell line models, it can produce plausible recommendations for cancer patients. Finally, we summarize our findings and suggest potential directions for future work in Section 5.

2 Problem statement

We are given $X \in \mathbb{R}^{n \times p}$, a set of n samples (cell lines or patients) where each of them is represented with a molecular profile (e.g. gene expression) with p variables. Furthermore, we are given $Y \in \mathbb{R}^{n \times m}$, a set of drug response measurements to m distinct drugs. Importantly, the majority of values in Y might be missing since the response of each individual sample is known only for a few of all potential drugs. Given X and Y as described above, and a molecular profile of a new sample, not present in X , our goal is to rank drugs based on their efficacy so that the most effective drug is ranked first. This goal is thus different from predicting the exact value of the drug response (as done by drug sensitivity prediction methods). Here we focus on the relative ordering (ranking) of a few, most effective, drugs.

From a machine learning perspective, drug recommendation can be framed into a problem of learning a weight matrix $W \in \mathbb{R}^{p \times m}$ and then recommending drugs based on the predicted ranking vector $f = xW$. Here $x \in \mathbb{R}^p$ is the molecular profile of a new sample, and $f \in \mathbb{R}^m$ is a ranking vector which entails the predicted ranking scores for each of the m distinct drugs. Finally, the ranking of the drugs is obtained as a permutation π which sorts the drugs in the decreasing order of f (thus, $\pi_1 = i$ such that $\max(f) = f_i$, that is, π_1 gives the identity of the first recommended drug, the one with the largest ranking score in f).

The weight matrix W can be viewed as containing weights of all molecular features towards each drug. Since we are interested in

relative ordering of the drugs, W needs to be learned jointly across all drugs rather than for each drug individually. Learning W can be formulated as the following minimization problem:

$$\min_W \mathcal{L}(XW, Y) + \lambda \mathcal{R}(W) \quad (1)$$

where $\mathcal{L}(XW, Y)$ is a loss function, $\mathcal{R}(W)$ is a regularization term with λ controlling the effect of the regularization (regularization is necessary to avoid over-fitting to the training set).

2.1 Normalized discounted cumulative gain (NDCG)

The key to our approach lies in employing a loss function based on the Discounted Cumulative Gain (DCG) ranking metric (Järvelin and Kekäläinen, 2002) as $\mathcal{L}(XW, Y)$ in Equation (1). Given the predicted ranking vector $f \in \mathbb{R}^m$ and the true drug response vector $y \in \mathbb{R}^m$, DCG@ k evaluates how well the order of the top k recommendations in f agrees with the order of the most effective drugs in y :

$$\text{DCG}@k(f, y) = \sum_{i=1}^k \frac{2^{y_{\pi_i}} - 1}{\log_2(1 + i)} \quad (2)$$

where π sorts f in decreasing order. DCG@ k can be intuitively interpreted as follows: the numerator gives rewards proportional to the true responses of the k recommended drugs, while the denominator penalizes the incorrect ordering of these k recommendations. The gains are maximized when the most effective k drugs are the top k recommended drugs, ranked in the correct order.

The Normalized DCG (NDCG) is more practical:

$$\text{NDCG}@k(f, y) = \frac{\text{DCG}@k(\pi, y)}{\text{DCG}@k(\sigma, y)} \quad (3)$$

where π and σ sort f and y in decreasing order, respectively. Hence, NDCG@ k ranges from 0 to 1 and is maximized (perfect recommendation) when $\pi_i = \sigma_i, 1 \leq i \leq k$.

Finally, we convert NDCG@ k , a gain function, into a loss function by taking $1 - \text{NDCG}@k(f, y)$, which allows us to rewrite Equation (1) and state the drug recommendation problem formally:

$$\min_W \sum_{i=1}^n (1 - \text{NDCG}@k(X_i W, Y_i)) + \lambda \|W\|_{Frob}^2 \quad (4)$$

where X_i (the i -th row of X) is the molecular profile of the i -th cell line and Y_i (the i -th row of Y) is the true response vector of the i -th cell line to the m drugs. $\|W\|_{Frob}^2$ is the squared Frobenius norm of W for the purpose of regularization.

3 Materials and methods

We stated the drug recommendation as minimization of $1 - \text{NDCG}@k$. Unfortunately, it turns out that $1 - \text{NDCG}@k$ is a highly non-convex and non-smooth function which makes its direct optimization difficult. Therefore, we minimize its convex upper bound derived by Weimer *et al.* (2007) instead and extend the Bundle Method for regularized Risk Minimization (BMRM) proposed by Teo *et al.* (2010) for the joint optimization of W across all drugs. Finally, Equation (4) employs a linear predictor ($X_i W$), which does not take feature interactions into account. To boost prediction performance by introducing non-linearity, we apply the well-known kernel trick (Cortes and Vapnik, 1995). We thus refer to our approach as Kernelized Rank Learning (KRL).

3.1 Convex upper bound of the loss function

Based on the work by Tsochantaridis *et al.* (2005), which describes how to find convex upper bounds of non-convex optimization problems, Weimer *et al.* (2007) showed that the following loss function is a convex upper bound of $1 - \text{NDCG}@k(f, y)$:

$$\ell(f, y) = \max_{\pi} [1 - \text{NDCG}@k(f, y) + \langle c, f_{\pi} - f \rangle] \quad (5)$$

where $c \in \mathbb{R}^m$ is a decreasing sequence, e.g. $c_i = (i + 1)^{-0.25}$ as suggested by the authors. Finding the permutation π is a linear assignment problem which can be solved with the Hungarian Marriage algorithm (Kuhn, 1955).

3.2 Kernelized rank learning (KRL)

Non-linear kernel methods have shown good performance in a drug sensitivity collaborative competition (Costello *et al.*, 2014). Therefore, we decided to kernelize the objective function defined in Equation (4).

Let $\mathbf{W} = \mathbf{X}^T \mathbf{U}$, where $\mathbf{U} \in \mathbb{R}^{n \times m}$, the objective is rewritten as:

$$\min_{\mathbf{U}} \sum_i \ell(\mathbf{X}_i \mathbf{X}_i^T \mathbf{U}, \mathbf{Y}_i) + \lambda \text{Tr}(\mathbf{U}^T \mathbf{X} \mathbf{X}^T \mathbf{U}) \quad (6)$$

where ℓ is the loss function defined in Equation (5), and Tr is the trace of a square matrix. By replacing $\mathbf{X} \mathbf{X}^T$ with a kernel matrix \mathbf{K} on \mathbf{X} , we get the KRL's kernelized objective function:

$$\min_{\mathbf{U}} \sum_i \ell(\mathbf{K}_i \mathbf{U}, \mathbf{Y}_i) + \lambda \text{Tr}(\mathbf{U}^T \mathbf{K} \mathbf{U}) \quad (7)$$

The time complexity of evaluating KRL's objective function comes down to the complexity of the linear assignment algorithm, which is $O(\tilde{m}^3)$, where \tilde{m} is the number of non-missing drug response measurements for the given sample, which is typically only a few. Moreover, the KRL loss function can be decomposed and evaluated in parallel across all samples.

Finally, to optimize Equation (7), we need the subgradient of the loss function $\mathcal{L}(\mathbf{K} \mathbf{U}, \mathbf{Y}) = \sum_i \ell(\mathbf{K}_i \mathbf{U}, \mathbf{Y}_i)$ with regards to \mathbf{U} :

$$\partial_{\mathbf{U}} \mathcal{L}(\mathbf{K} \mathbf{U}, \mathbf{Y}) = \mathbf{K} \partial_{\mathbf{F}} \mathcal{L} \quad (8)$$

where $\mathbf{F} = \mathbf{K} \mathbf{U}$ and $(\partial_{\mathbf{F}} \mathcal{L})_i = \partial_{\mathbf{F}_i} \ell(\mathbf{F}_i, \mathbf{Y}_i) = c_{\bar{\pi}^{-1}(i)} - c$ where $\bar{\pi}$ is the permutation that maximizes $\ell(f, y)$, $\bar{\pi}^{-1}$ is the inverse permutation of $\bar{\pi}$, i.e. $(f_{\bar{\pi}^{-1}})_{\bar{\pi}^{-1}(i)} = f_i$, and c is defined as in Equation (5). Supplementary Algorithm S1 details how to calculate the subgradient $\partial_{\mathbf{U}} \mathcal{L}$.

3.3 KRL optimization

The objective function in Equation (7) is convex but due to the maximization problem in ℓ expensive to evaluate. The BMRM algorithm proposed by Teo *et al.* (2010) has been shown to work well on such problems. However, the authors introduced only a single-task version of BMRM, which means it cannot be directly applied to solve Equation (7). Therefore, we extended BMRM for KRL to allow for the following differences when compared to the single-task BMRM: (i) \mathbf{U} is a matrix rather than a vector and (ii) the objective function includes regularization on a positive definite kernel matrix, $\text{Tr}(\mathbf{U}^T \mathbf{K} \mathbf{U})$, rather than an L^2 norm of a weight vector.

Algorithm 1 describes BMRM for KRL. The general idea of the BMRM algorithm is to improve the lower bound of the objective function by its linearization (the first-order Taylor approximation) at the subgradient $\partial_{\mathbf{U}} \mathcal{L}$ in different iterations. Lines 5 and 6 of Algorithm 1 require finding \mathbf{U} that minimizes the lower bound function $\mathcal{J}(\mathbf{U})$ of the objective function defined in Equation (7). The

Algorithm 1. BMRM for Kernelized Rank Learning

Data: $\mathbf{U}[0], \mathbf{K}, \mathbf{Y}, \lambda, \epsilon > 0, t = 0$

Result: $\mathbf{U}[t]$

```

1 repeat
2    $t = t + 1$ ;
3    $\mathbf{A}[t] = \partial_{\mathbf{U}} \mathcal{L}(\mathbf{K} \mathbf{U}[t - 1], \mathbf{Y})$ ;
4    $\mathbf{B}[t] = \mathcal{L}(\mathbf{K} \mathbf{U}[t - 1], \mathbf{Y}) - \langle \mathbf{U}[t - 1], \mathbf{A}[t] \rangle$ ;
5    $\mathcal{J}(\mathbf{U}) = \lambda \text{Tr}(\mathbf{U}^T \mathbf{K} \mathbf{U}) + \max_{1 \leq i \leq t} \langle \mathbf{U}, \mathbf{A}[i] \rangle + \mathbf{B}[i]$ ;
6    $\mathbf{U}[t] = \min_{\mathbf{U}} \mathcal{J}(\mathbf{U})$ ;
7    $\epsilon[t] = \min_{0 \leq i \leq t} \mathcal{L}(\mathbf{K} \mathbf{U}[i], \mathbf{Y}) + \lambda \text{Tr}(\mathbf{U}[i]^T \mathbf{K} \mathbf{U}[i]) - \mathcal{J}(\mathbf{U}[i])$ ;
8 until  $\epsilon[t] \leq \epsilon$ ;
```

primal problem involves a maximization sub-problem, which is difficult to solve. The original BMRM algorithm solves the primal problem in its dual form, which is a quadratic program. In the following we show that also in the case of KRL, with \mathbf{U} being a matrix and regularization on a kernel, the problem can be solved in its dual form.

Firstly, \mathbf{U} is decomposed into m column vectors, and the objective function is rewritten as:

$$\min_{\mathbf{U}} \mathcal{L}(\mathbf{K} \mathbf{U}, \mathbf{Y}) + \sum_{i=1}^m \lambda (\mathbf{U}^i)^T \mathbf{K} \mathbf{U}^i \quad (9)$$

where \mathbf{U}^i is the i -th column of \mathbf{U} . Next, after introducing non-negative Lagrange multipliers α , the dual function is also decomposed by column:

$$\mathbf{U}^i[t] = -\mathbf{K}^{-1} \mathbf{A}^i \alpha[t] / \lambda \quad (10)$$

$$\alpha[t] = \max_{\alpha \in \mathbb{R}^t} -\frac{1}{2\lambda} \alpha^T \left(\sum_{i=1}^m (\mathbf{A}^i)^T \mathbf{K}^{-1} \mathbf{A}^i \right) \alpha + \alpha^T \mathbf{B} \quad (11)$$

subject to $\alpha \geq 0$ and $\|\alpha\|_1 = 1$

where $\mathbf{A}^i = [\mathbf{A}^i[1], \dots, \mathbf{A}^i[t]]$, $\mathbf{B} = [\mathbf{B}[1], \dots, \mathbf{B}[t]]$, and $\alpha[t]$, $\mathbf{U}^i[t]$, $\mathbf{A}^i[t]$, and $\mathbf{B}[t]$ represent α , \mathbf{U}^i , \mathbf{A}^i , and \mathbf{B} in iteration t , respectively. This results in a quadratic program problem.

3.4 Datasets

3.4.1 Cancer cell lines

We designed, evaluated and trained our method using the Genomics of Drug Sensitivity in Cancer (GDSC) release 6 dataset with drug sensitivity measurements of 1001 cancer cell lines (ranging 31 cancer types) to 265 pharmacological compounds (Iorio *et al.*, 2016). Majority of the cell lines were characterized using gene expression, whole-exome sequencing, copy number variation and DNA methylation. We evaluated our method with all four data types individually, however, we focused our analysis on the predictions based on gene expression since it has been shown previously as the most predictive data type (Costello *et al.*, 2014). To this end, we encoded each of the 962 profiled cell lines with the RMA-normalized (robust multi-array average) basal expression of 17737 genes. Supplementary Table S1 lists the statistics regarding the number of profiled cell lines, features and missing values for all four data types.

Genomics of Drug Sensitivity in Cancer reports the drug sensitivity as the logarithm of half maximal inhibitory concentration ($\log \text{IC}_{50}$), which denotes the concentration of the compound required to inhibit the cell growth at 50%. The measured IC_{50} values are not, however, comparable between different compounds. Therefore, Iorio *et al.* (2016)

calculated drug-specific sensitivity (binarization) thresholds for each of the 265 tested drugs. These thresholds were determined using a heuristic outlier procedure (Knijnenburg *et al.*, 2016) following a previous observation that the majority of cell lines are typically resistant to a given drug (Garnett *et al.*, 2012).

We used these drug-specific sensitivity thresholds to normalize the dataset. Furthermore, for compatibility with the ranking metric (NDCG@ k), we scaled the normalized dataset into the range of $[0, v]$, where $v \in \mathbb{R}^+$ represents the strongest drug response in the dataset. Thus, the sensitivity scores used in this work were equal to $-\log \frac{IC_{50}}{thr_d} + a$, where thr_d is the sensitivity threshold of the given drug and a is the maximum normalized $\log IC_{50}$ score across all drugs.

3.4.2 Clinical trials

Recently, Geeleher *et al.* (2014) showed that ridge regression models trained on *in vitro* cell line models can predict drug response *in vivo*. Their approach was particularly successful for the docetaxel breast cancer (Chang *et al.*, 2003) and bortezomib multiple myeloma (Mulligan *et al.*, 2007) clinical trials, yielding the area under the receiver operating characteristic curve (AUROC) scores of 0.81 and 0.71, respectively. The AUROC can be interpreted as the probability that for a randomly selected responder and non-responder, the responder was predicted to respond better to the given therapy than the non-responder. In Geeleher *et al.* (2014), the ridge regression models were trained using the GDSC release 2 dataset, which comprised responses of 482 and 280 cell lines for docetaxel and bortezomib, respectively. After re-running their analysis pipeline using the GDSC release 6 containing 833 and 391 cell lines for docetaxel and bortezomib, respectively, we found that while the performance for bortezomib (AUROC of 0.70) was comparable to the GDSC release 2 results, the docetaxel model resulted in a random performance (AUROC of 0.49). We conjecture that this could be attributed to experimental errors or unsuccessful homogenization of the expression profiles for some of the GDSC release 6 cell lines. Therefore, to avoid these potential outliers during training, we used the cell lines in the intersection of the GDSC release 2 and release 6. This approach resulted in a reasonable performance for both docetaxel (394 cell lines) and bortezomib (253 cell lines) with AUROCs of 0.74 and 0.71, respectively.

We used the docetaxel (24 patients) and bortezomib (169 patients) clinical trials to test if KRL’s recommendations can generalize to cancer patients. To this end, KRL was trained using the 394 (docetaxel) and 253 (bortezomib) GDSC release 6 cell lines as explained above. We used the pipeline provided by Geeleher *et al.* (2014) to download the clinical trials datasets and homogenize the gene expression profiles of the cell lines, measured with Affymetrix HG-U219 and tumor biopsies, measured with Affymetrix HG-U95-v2 (docetaxel trial) and Affymetrix HG-U133B (bortezomib trial). The homogenization was implemented using the ComBat method (Johnson *et al.*, 2007; Leek and Storey, 2007).

3.4.3 Breast cancer patients

We employed The Cancer Genome Atlas (TCGA) breast cancer cohort (BRCA, 1098 patients) to further evaluate how well KRL’s predictions generalize to cancer patients (The Cancer Genome Atlas Network, 2012). We downloaded the Level 1 clinical annotations and Level 3 reverse phase protein array (RPPA) data from the Broad Institute TCGA Genome Data Analysis Center (2016), Firehose data run 2016_01_28 (<https://doi.org/10.7908/C11G0KM9>). We used the immunohistochemistry annotations to identify HER2-overexpressed (HER2⁺, 164 patients) and triple-negative breast

cancers (TNBCs, 116 patients), and RPPA z -scores to identify patients with normal-to-high expression of JAK2 (JAK2⁺, z -score ≥ 0 , 421 patients). Furthermore, we defined the BRCA1/2-mutant (mBRCA) patients using the list of 37 patients carrying germline loss-of-function BRCA1/2 mutations (Maxwell *et al.*, 2017). Finally, we used the pipeline proposed by Geeleher *et al.* (2017) to download and homogenize the Level 3 Illumina HiSeq RNA-seq v2 data (1 080 patients) from the Firehose data run 2015_08_21. The homogenization approach was in part based on the Remove Unwanted Variation (RUV) method (Risso *et al.*, 2014).

3.5 Evaluation

To get an unbiased estimate of KRL’s prediction performance using the GDSC cell lines dataset, we employed three-fold cross-validation, where each fold was used once for testing while the other two folds were merged for training. Finally, we reported the mean performance across the three test folds. For the experiments with subsampled training datasets, we repeated the subsampling procedure 10 times, calculated standard deviations, and estimated statistical significance of our results using paired Student t -test. For the experiments with the full training dataset, standard deviations and statistical significance were estimated from the three cross-validation folds.

To evaluate the accuracy of our approach, we quantified if the top recommendations in $f \in \mathbb{R}^m$ agree with the most effective drugs in $y \in \mathbb{R}^m$, where f and y are the predicted ranking vector and the true drug response vector, respectively. To this end, we used two different ranking metrics: Precision@ k and NDCG@ k . NDCG@ k was defined in Equation (3), Section 2.1. The main shortcoming of NDCG@ k is that there is no intuitive explanation of its values. Therefore, we used also Precision@ k , which measures precision between the k highest-ranked recommendations in f and the k most effective drugs in y :

$$\text{Precision@}k(f, y) = \frac{1}{k} \sum_{i=1}^k \sum_{j=1}^k \delta(\pi_i, \sigma_j) \quad (12)$$

where π and σ sort f and y in decreasing order, respectively, and $\delta(a, b) = 1$ if $a = b$ and 0 otherwise. The interpretation of Precision@ k is straightforward: Precision@5 = 0.4 means that out of five most effective drugs in y , two were among the top five recommendations in f .

3.6 Baselines and related work

We compared the method proposed here, Kernelized Rank Learning (KRL), with several state-of-the-art approaches for drug sensitivity prediction including the Elastic Net (EN), Kernel Ridge Regression (KRR), Random Forest (RF) regression and Kernelized Bayesian Multi-Task Learning (KBMTL). EN (Zou and Hastie, 2005) and KRR (Murphy, 2012) have been recommended as one of the best-performing algorithms in a systematic survey (Jang *et al.*, 2014). Another survey identified KBMTL (Gönen and Margolin, 2014) and RF (Breiman, 2001) as the two best-performing algorithms in a challenge-based competition on a breast cancer dataset (Costello *et al.*, 2014). KBMTL is a multi-task approach, thus instead of training a model for each task (drug) separately as done by EN, KRR and RF, it learns the relationship between different drugs during training. While there is an extension of the KBMTL method, which employs multiple kernel learning (MKL) to allow for the integration of different molecular data types, we focused on prediction using a single data type to facilitate a fair comparison of all compared

methods. Finally, we compared two implementations of the proposed method: with linear and Radial Basis Function (RBF) kernels (LKRL and KRL, respectively).

We used the implementation of EN, KRR and RF from the Python *scikit-learn* package (Pedregosa et al., 2011). For KBMTL, we used the MATLAB implementation provided by the authors (<https://github.com/mehmetgonen/bmtmkl>).

3.7 Hyper-parameter optimization

For each of the compared methods, we tuned the relevant hyper-parameters with a grid search using nested three-fold cross-validation on the training set (Supplementary Table S2). That is, we split the two training folds of the ‘outer’ cross-validation procedure into three folds for hyper-parameter optimization. To simplify the hyper-parameter optimization for the evaluation of KRL on the clinical trials and TCGA-BRCA datasets, we heuristically determined the RBF kernel width (γ) using the ‘median trick’ which sets γ to the inverse of the median squared Euclidean distance of all training samples pairs. Then, we tuned the regularization parameter λ with a line search, $\lambda \in \{1.0, 0.1, \dots, 1 \times 10^{-6}\}$.

4 Results

As described in Section 1, our method (KRL) is motivated by the clinically relevant scenario where (i) the training dataset in the form of a medical records knowledge bank is much sparser than a typical drug sensitivity screen of cultured cell lines, (ii) therapies recorded in the knowledge bank were prescribed by experienced doctors based on their expert judgment and (iii) only a few, accurate, recommendations are needed (as opposed to predicting the exact values of patient’s sensitivity to all drugs on the market). To show how KRL performs under these three requirements, we conducted the following experiments. First, for the sake of completeness, we evaluated KRL on the full GDSC dataset. Second, to address (i), we randomly subsampled the number of drugs a cell line was treated with to make the training dataset sparse. Third, to address (ii), we extended the subsampling procedure to account for doctor’s expert judgment by sampling from a predefined proportion of the most effective drugs. To address (iii) across all of the experiments, we evaluated our method with NDCG@ k and Precision@ k ranking metrics for different values of k , where k defines the number of top recommendations considered in the evaluation. We performed all experiments using the gene expression dataset (as it proved to be the most predictive data type) and then employed the other data types in the final analysis using the most realistic subsampling strategy where, for each cell line, only a few drugs were used for training. Finally, we evaluated how well KRL models trained on cell lines generalize to patient data using docetaxel and bortezomib clinical trials and TCGA breast cancer cohort.

4.1 Prediction using the full training dataset

We used the full GDSC dataset, which due to missing values included 82% of the drug sensitivity measurements, to compare the performance of KRL with related work using different evaluation ranking criteria. Figure 1 and Supplementary Figure S1 show the comparison for Precision@ k and NDCG@ k , respectively, with $k \in \{1, 3, 5, 10, 15, 20\}$. KRL yielded comparable drug recommendations (paired Student t -test, $P > 0.05$) to the second best method, KBMTL. Comparing KRL, which employs the RBF kernel, with the linear kernel implementation of our method (LKRL), KRL either outperformed ($P < 0.05$, Precision@ k for $k \in \{1, 5, 15, 20\}$) or

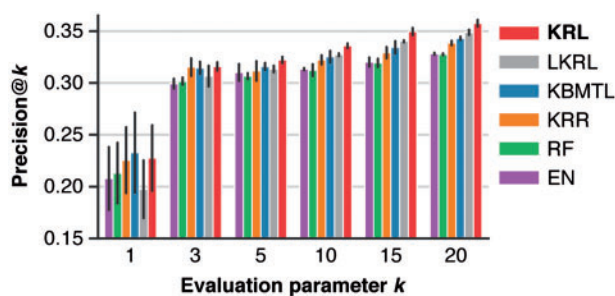


Fig. 1. Comparison of KRL with related work (see Section 3.6 for details) in terms of Precision@ k using the full training dataset for different values of the evaluation parameter k , which controls the number of predicted recommendations that are compared with the true drug ranking. The error bars show standard deviations from three cross-validation folds

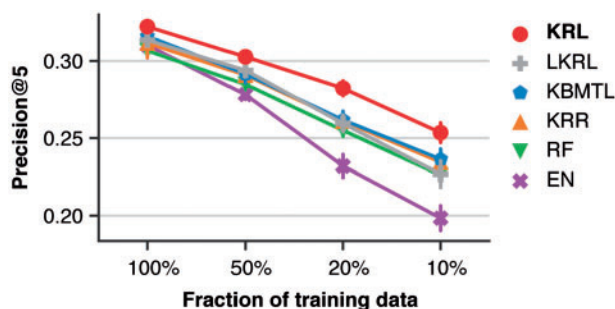


Fig. 2. Comparison of KRL with related work (see Section 3.6 for details) in terms of Precision@5 using the subsampled training datasets. The error bars show standard deviations from 10 randomly subsampled training datasets

performed comparably to LKRL. Overall, KRL’s performance was in the range of 23–36% for Precision@ k and 47–58% for NDCG@ k .

4.2 Prediction using sparse training datasets

To evaluate the compared methods in a more clinically relevant scenario, in which only a few therapies can be recorded for each patient, we subsampled the training dataset (training folds) at sampling rates of 50%, 20% and 10%. At the same time, we kept the test set (test folds) unchanged. In the case of 10% sampling, this strategy resulted in training on 22 drug responses per cell line on average.

Figure 2 and Supplementary Figure S2 show Precision@5 and NDCG@5 as a function of the relative training dataset size, respectively. As expected, the performance of all compared methods decreased as the training dataset became sparser. Nonetheless, the Precision@5 improvements of KRL compared to the second method, KBMTL, were 1.1% ($P = 8.3 \times 10^{-6}$), 2.1% ($P = 2.5 \times 10^{-8}$) and 1.7% ($P = 5.2 \times 10^{-7}$) for the sampling rates of 50%, 20% and 10%, respectively, compared to 0.6% ($P = 0.138$) for the full training dataset. Thus, KRL appears to be more robust to sparse data than the regression approach. The same trend could also be observed for NDCG@5. The Precision@5 improvements of KRL compared to the other three methods using only 10% of the training dataset were 2.7% ($P = 2.5 \times 10^{-7}$), 1.9% ($P = 2.4 \times 10^{-7}$), 2.7% ($P = 4.9 \times 10^{-8}$) and 5.5% ($P = 1.6 \times 10^{-9}$) for LKRL, KRR, RF and EN, respectively.

4.3 Prediction using sparse training datasets biased towards effective therapies

As discussed in the previous section, KRL appears to be more robust to missing data than related work. This is shown in Figure 2 where

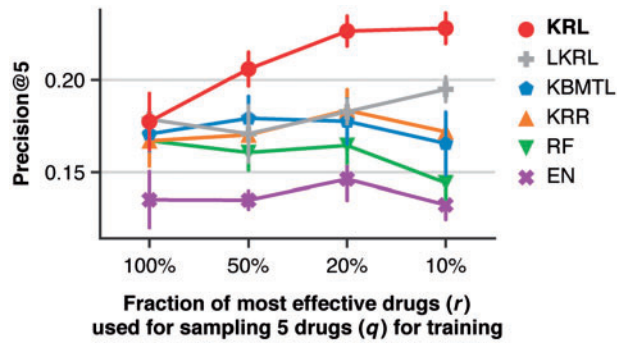


Fig. 3 Comparison of KRL with related work (see Section 3.6 for details) in terms of Precision@5 using the subsampled training datasets, keeping five drugs (q) per cell line sampled from a predefined fraction (r) of the cell line’s most effective drugs. The error bars show standard deviations from 10 randomly subsampled training datasets

random subsampling of the training dataset was used to induce sparsity. Nonetheless, it can be argued that choosing a subset of drugs for each cell line *randomly* does not simulate medical records knowledge banks realistically. This is because clinicians do not prescribe therapies randomly but rather based on their expert judgment. Thus we employed another subsampling strategy: first, we selected a predefined proportion r of the most effective drugs per cell line and second, we randomly sampled q drugs per cell line from this selection. After this procedure, there might not be enough training data left for some of the drugs, therefore, after subsampling, we removed drugs with <6 cell lines. We evaluated KRL and related work using training datasets generated with $q \in \{3, 5, 10\}$ (yielding datasets with 143–199, 202–246, and 239–265 drugs in total, respectively) and $r \in \{100\%, 50\%, 20\%, 10\%\}$ (where $r = 100\%$ is equivalent to sampling q drugs randomly). Here we analyze results with $q = 5$ for brevity. Results for $q = 3$ and $q = 10$ showed a similar trend and are summarized in Supplementary Figures S3 and S4.

Figure 3 and Supplementary Figure S3a show Precision@5 and NDCG@5, respectively, as a function of the fraction of the most effective drugs (r) used for sampling five drugs (q) per cell line. We found that as the training dataset was composed of more effective drugs, the improvement of KRL compared to the second best method, KBMTL, was gradually increasing from 0.7% to 6.2% (Precision@5) and from 0.3% to 10.4% (NDCG@5).

Both Precision@ k and NDCG@ k improvements of KRL compared to related work were statistically significant ($P < 0.05$) for $r \in \{50\%, 20\%, 10\%\}$ across a wide range of $k \in \{3, 5, 10, 15, 20\}$. KRL was also able to outperform LKRL ($P < 0.05$) across all evaluated values of r and k with the exception of $r = 100\%$, and $r = 10\%$ for $k = 1$, in which case KRL and LKRL performed comparably. Figure 4 and Supplementary Figure S4a show Precision@ k and NDCG@ k as a function of the evaluation parameter k for the case of sampling five drugs from the 20% of the most effective drugs.

4.4 Evaluation of the single top drug recommendations

We found that KRL outperformed related work in terms of both Precision@ k and NDCG@ k for a variety of settings. Arguably, the most clinically relevant setting was presented in the previous section where a few (q) drugs were sampled from a given proportion (r) of the most effective drugs. We analyzed this set of results further by looking at each cell line’s *single* top recommendation (i.e. the drug with the highest predicted ranking score) and quantifying how close this recommendation was to the most effective drug (i.e. the ground truth).

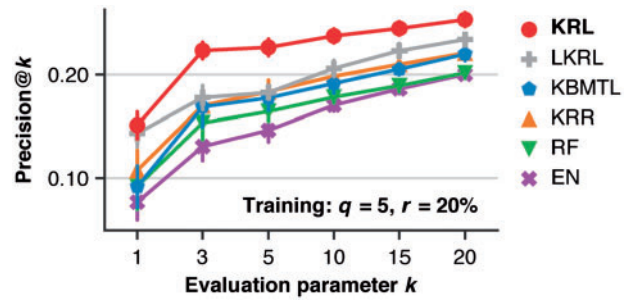


Fig. 4. Comparison of KRL with related work (see Section 3.6 for details) in terms of Precision@ k (for different values of the evaluation parameter k) using the subsampled training datasets, keeping five drugs (q) per cell line sampled from the 20% of the cell line’s most effective drugs (r). The error bars show standard deviations from 10 randomly subsampled training datasets

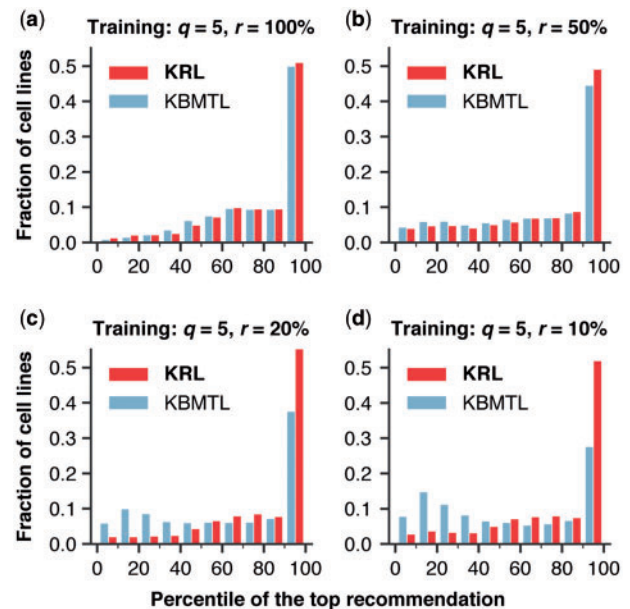


Fig. 5. Histograms comparing the distributions of percentile ranks of drugs recommended by KRL and the second best method, KBMTL, using the subsampled training datasets, keeping five drugs (q) per cell line sampled from a predefined fraction (r) of the cell line’s most effective drugs. The figure highlights that regardless of the sampling strategy, 49–56% of KRL’s recommendations were within the top 10% of drugs to which the given cell line was most sensitive

To this end, we looked at the distribution of percentile ranks of the recommended drugs.

We calculated a percentile rank of a recommended drug as the percentage of drugs to which the given cell line was *less* sensitive than it was to the recommended drug. For instance, a percentile rank of 90 means that the recommended drug was in the top 10% of the most effective drugs for the given cell line. Figure 5a–d compares histograms of percentile ranks for KRL and the second best method, KBMTL, when trained using five drugs (q) per cell line sampled from the 100%, 50%, 20% and 10% of the most effective drugs (r), respectively. The figure highlights that regardless of the sampling strategy, 49–56% of KRL’s recommendations were within the top 10% of drugs to which the given cell line was most sensitive. While KRL and KBMTL performed comparably for sampling from the full dataset (Fig. 5a, $P = 0.50$), the proportion of KRL’s recommendations with

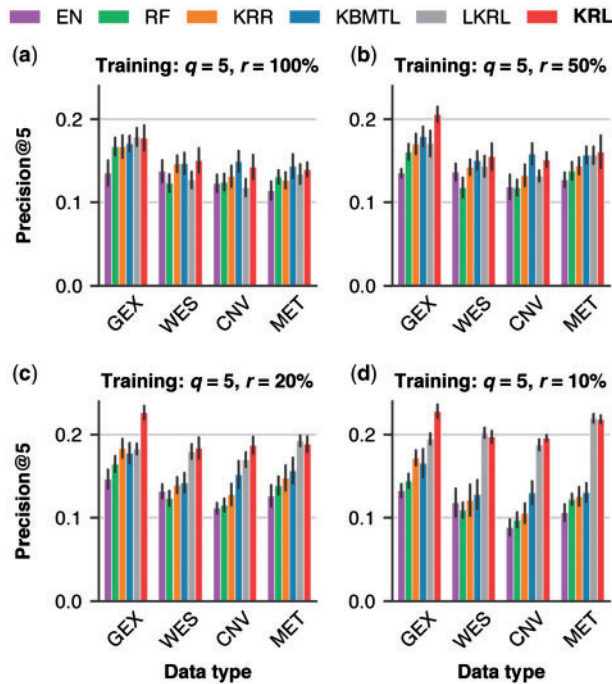


Fig. 6. Comparison of KRL with related work (see Section 3.6 for details) across the four molecular data types: gene expression (GEX), whole-exome sequencing (WES), copy number variation (CNV) and DNA methylation (MET). The six compared methods were evaluated in terms of Precision@5 using the subsampled training datasets, keeping five drugs (q) per cell line sampled from a predefined fraction (r) of the cell line's most effective drugs. The error bars show standard deviations from 10 randomly subsampled training datasets

percentile ranks > 90 was higher by 4.5%, 18.6% and 24.4% for sampling from the 50%, 20% and 10% of the most effective drugs, respectively (Fig. 5b–d, $P = 7.3 \times 10^{-5}$, 7.2×10^{-6} and 7.9×10^{-8} , respectively). Supplementary Figure S5 shows a similar trend for sampling three and 10 drugs per cell line.

4.5 Prediction with different data types

So far, all abovementioned results were based on representing a cell line with its gene expression profile. It is of interest to evaluate KRL and related work also for the other three available molecular profiles: whole-exome sequencing, copy number variation and DNA methylation (see Supplementary Table S1 for a summary of features derived from these data types). Figure 6 and Supplementary Figure S6 compare Precision@5 and NDCG@5, respectively, for the six compared methods across all four available data types. These results are based on sampling five drugs (q) from the 100%, 50%, 20% and 10% of the most effective drugs (r) per cell line (Fig. 6a–d, respectively). In agreement with the findings from a drug sensitivity collaborative competition (Costello et al., 2014), gene expression was the most predictive data type regardless of the employed method. Focusing on the other three data types, KRL yielded a comparable performance to KBMTL (the second best method) for $r = 100\%$ and 50% and achieved Precision@5 improvements of 3–9% ($P < 0.05$) for $r = 20\%$ and 10%. NDCG@5 improvements were in the range of 7–15%.

4.6 Clinical trials prediction with KRL

Recently, Geeleher et al. (2014) showed that ridge regression models trained on *in vitro* cell line models can predict drug response *in vivo*.

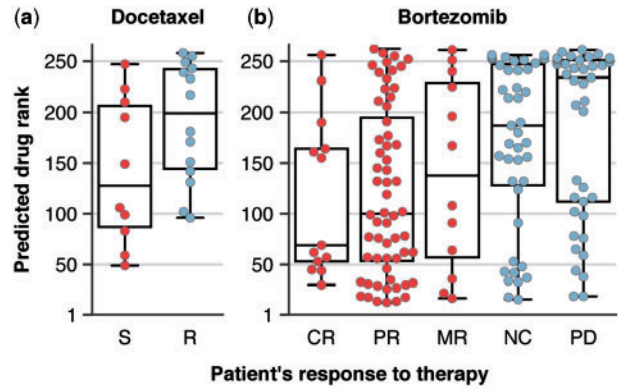


Fig. 7. Predicted drug ranks for the docetaxel (a) and bortezomib (b) clinical trials. Docetaxel trial designated patients as sensitive (S) or resistant (R). Bortezomib trial classified the response into five categories: complete response (CR), partial response (PR), minimal response (MR), no change (NC) and progressive disease (PD). Moreover, the patients were grouped as responders (CR, PR and MR) and non-responders (NC and PD)

We were interested if KRL models can also generalize to patient data. We used the pipeline provided by Geeleher et al. (2014) to download and homogenize docetaxel breast cancer (Chang et al., 2003) and bortezomib multiple myeloma (Mulligan et al., 2007) clinical trials datasets.

First, for docetaxel, there were 10 patients labeled as sensitive (S) and 14 as resistant (R). We used the predicted drug ranks to discriminate between the sensitive and resistant patients. We found that KRL had a tendency to rank docetaxel higher for the sensitive than the resistant patients (Wilcoxon rank sum test, $P = 0.04$) with AUROC of 0.71 (Fig. 7a). Second, for bortezomib, there were 169 patients and their responses were classified as complete response (CR), partial response (PR), minimal response (MR), no change (NC), or progressive disease (PD). Comparably to the ridge regression approach (Geeleher et al., 2014), the medians of the predicted drug ranks were able to sort the five categorical responses in the correct order (Fig. 7b). The bortezomib trial defined also another classification which grouped 85 CR, PR and MR patients as responders and 84 NC and PD patients as non-responders. KRL was able to rank bortezomib higher for responders than for non-responders (Wilcoxon rank sum test, $P = 2.4 \times 10^{-5}$) with AUROC of 0.68.

4.7 TCGA breast cancer cohort prediction with KRL

We showed that drug ranks predicted with KRL trained on cell lines correlate with patient response to treatment. However, the clinical trials datasets do not allow for evaluating how well KRL ranks two (or more) alternative drugs for a given patient. Therefore, we employed the TCGA-BRCA cohort (The Cancer Genome Atlas Network, 2012) and analyzed different molecular breast cancer subtypes to gauge if KRL's drug recommendations for patients are plausible.

Geeleher et al. (2017) showed how to homogenize gene expression profiles measured with RNA-seq (TCGA) and micro-arrays (GDSC), which in turn allowed them to train ridge regression models on the GDSC cell lines and use these models to 'impute' drug response of the TCGA patients. In their evaluation, they showed that HER2⁺ TCGA-BRCA patients were predicted to be more sensitive to lapatinib, a first-line therapy for HER2⁺ patients (Gomez et al., 2008), than patients without HER2 overexpression. Here we followed the idea of using molecular subtypes for evaluating KRL's

Table 1. KRL's recommendation rates for TCGA-BRCA molecular subtypes

Recommendation	HER2 ⁺ <i>n</i> = 163	mBRCA TNBC <i>n</i> = 9	JAK2 ⁺ TNBC <i>n</i> = 56
lapatinib > PARPi	0.91	0.11	—
lapatinib > veliparib	0.95	0.22	—
lapatinib > olaparib ^a	0.94	0.38	—
lapatinib > talazoparib	0.92	0.22	—
lapatinib > rucaparib	0.94	0.44	—
lapatinib > ruxolitinib	0.86	—	0.29

Note: Recommendation rates shown here (lapatinib > alternative drug) are equal to the fractions of patients for which lapatinib was ranked higher than the alternative drug.

^aOlaparib was measured twice in the GDSC dataset but KRL performed comparably for both measurements, here we reported the mean recommendation rates.

recommendations and compared lapatinib to emerging targeted therapies for TNBCs (Kalimutho *et al.*, 2015).

First, we compared lapatinib and four PARP1/2 inhibitors (PARPi), veliparib, olaparib, talazoparib and rucaparib, some of which have shown promising therapeutic response in *BRCA1/2*-mutant (mBRCA) tumors (Tutt *et al.*, 2010). Consistent with our hypothesis, we found that lapatinib was ranked higher than all four PARPi for 149 of the 163 HER2⁺ patients (91%). In contrast, lapatinib was recommended over (ranked higher than) PARPi only for one out of nine mBRCA TNBCs (11%). We re-trained KRL 10 times with randomly selected 100 drugs to confirm that the results were statistically different from a random recommendation rate of 50% (unpaired Student *t*-test, $P = 8.2 \times 10^{-14}$ for HER2⁺ and 1.7×10^{-138} for mBRCA TNBCs). Table 1 details lapatinib recommendation rates for the individual PARPi. Upon randomly selecting 100 drugs, these results were also statistically different from random ($P \ll 0.05$) except for rucaparib ($P = 0.15$), for which a recent clinical trial failed to report objective responses in breast cancer patients (Drew *et al.*, 2016).

Second, we considered ruxolitinib, a JAK1/2 inhibitor approved for the treatment of myelofibrosis, currently being investigated in several clinical trials as a treatment for TNBC (Kalimutho *et al.*, 2015). As expected, we found that lapatinib was ranked higher than ruxolitinib for 140 of the 163 HER2⁺ patients (86%), whereas it was recommended over ruxolitinib only for 16 of the 56 TNBC patients (29%) with normal-to-high JAK2 expression (JAK2⁺). We re-trained KRL 10 times with randomly selected 100 drugs to confirm that these results were statistically different from random ($P = 1.1 \times 10^{-11}$ for HER2⁺ and 7.7×10^{-8} for JAK2⁺ TNBCs).

5 Conclusion and outlook

This work phrased personalized drug recommendation as a ranking problem of choosing the most effective drugs. This follows from the observation that in a clinical setting, we are interested in providing a recommendation of a few, most effective, drugs rather than predicting the exact response (sensitivity) to all drugs available on the market. To this end, we proposed KRL, which directly optimizes a ranking loss function.

In our empirical evaluation, we aimed to simulate a clinically relevant scenario, in which only a few therapies are prescribed for each patient, by subsampling the training cell lines dataset. We found that KRL outperforms state-of-the-art predictors in drug recommendation with the improvements being largest when the

sampling was not random but biased towards effective therapies, thus simulating a medical records knowledge bank in which clinicians' prior knowledge is reflected.

In order to provide reliable and robust recommendations in a clinical setting, KRL needs to be re-trained on patient data and treatment outcomes. Unfortunately, there is currently no publicly available dataset of sufficient size to evaluate KRL in this setting directly. Nevertheless, we were able to show that even when trained on cell lines, KRL recommendations can generalize to patient data and provide plausible recommendations.

One potential direction for future work is extending KRL to learn treatment histories, interdependence and combinations of prescribed therapeutics. We envision that medical records knowledge banks which are currently being gathered by various research consortia will provide the opportunity to study these phenomena with a machine learning approach.

Funding

This work has been supported by the SNSF Starting Grant 'Significant Pattern Mining'.

Conflict of Interest: none declared.

References

- Aben, N. *et al.* (2016) TANDEM: a two-stage approach to maximize interpretability of drug response models based on multiple molecular data types. *Bioinformatics*, **32**, i413–i420.
- Ammad-ud-din, M. *et al.* (2014) Integrative and personalized QSAR analysis in cancer by kernelized bayesian matrix factorization. *J. Chem. Inform. Model.*, **54**, 2347–2359.
- Ammad-ud-din, M. *et al.* (2016) Drug response prediction by inferring pathway-response associations with kernelized bayesian matrix factorization. *Bioinformatics*, **32**, i455–i463.
- Barretina, J. *et al.* (2012) The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, **483**, 603–607.
- Breiman, L. (2001) Random forests. *Mach. Learn.*, **45**, 5–32.
- Broad Institute TCGA Genome Data Analysis Center (2016). Analysis-ready standardized TCGA data from Broad GDAC Firehose 2016_01_28 run.
- Chang, J.C. *et al.* (2003) Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet*, **362**, 362–369.
- Chapman, P.B. *et al.* (2011) Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N. Engl. J. Med.*, **364**, 2507–2516.
- Cortes, C. and Vapnik, V. (1995) Support-vector networks. *Mach. Learn.*, **20**, 273–297.
- Costello, J.C. *et al.* (2014) A community effort to assess and improve drug sensitivity prediction algorithms. *Nat. Biotechnol.*, **32**, 1202–1212.
- Drew, Y. *et al.* (2016) Phase 2 multicentre trial investigating intermittent and continuous dosing schedules of the poly(ADP-ribose) polymerase inhibitor rucaparib in germline BRCA mutation carriers with advanced ovarian and breast cancer. *Br. J. Cancer*, **114**, 723.
- Garnett, M.J. *et al.* (2012) Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, **483**, 570–575.
- Geeleher, P. *et al.* (2014) Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biol.*, **15**, R47.
- Geeleher, P. *et al.* (2017) Discovering novel pharmacogenomic biomarkers by imputing drug response in cancer patients from large genomics studies. *Genome Res.*, **27**, 1743–1751.
- Gerstung, M. *et al.* (2017) Precision oncology for acute myeloid leukemia using a knowledge bank approach. *Nat. Genet.*, **49**, 332–340.
- Gomez, H.L. *et al.* (2008) Efficacy and safety of lapatinib as first-line therapy for ErbB2-amplified locally advanced or metastatic breast cancer. *J. Clin. Oncol.*, **26**, 2999–3005.

- Gönen,M. and Margolin,A.A. (2014) Drug susceptibility prediction against a panel of drugs using kernelized bayesian multitask learning. *Bioinformatics*, **30**, i556–i563.
- Iorio,F. et al. (2016) A landscape of pharmacogenomic interactions in cancer. *Cell*, **166**, 740–754.
- Jang,I.S. et al. (2014) Systematic assessment of analytical methods for drug sensitivity prediction from cancer cell line data. In *Pacific Symposium on Biocomputing*, pp. 63–74.
- Järvelin,K. and Kekäläinen,J. (2002) Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inform. Syst.*, **20**, 422–446.
- Johnson,W.E. et al. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, **8**, 118–127.
- Kalimutho,M. et al. (2015) Targeted therapies for triple-negative breast cancer: combating a stubborn disease. *Trends Pharmacol. Sci.*, **36**, 822–846.
- Knijnenburg,T.A. et al. (2016) Logic models to predict continuous outputs based on binary inputs with an application to personalized cancer therapy. *Sci. Rep.*, **6**, 36812.
- Kuhn,H.W. (1955) The hungarian method for the assignment problem. *Nav. Res. Logist.*, **2**, 83–97.
- Leek,J.T. and Storey,J.D. (2007) Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genetics*, **3**, e161.
- Maxwell,K.N. et al. (2017) BRCA locus-specific loss of heterozygosity in germline BRCA1 and BRCA2 carriers. *Nat. Commun.*, **8**, 319.
- Menden,M.P. et al. (2013) Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS One*, **8**, e61318.
- Mulligan,G. et al. (2007) Gene expression profiling and correlation with outcome in clinical trials of the proteasome inhibitor bortezomib. *Blood*, **109**, 3177–3188.
- Murphy,K.P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, MA, USA.
- Pedregosa,F. et al. (2011) Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830.
- Risso,D. et al. (2014) Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.*, **32**, 896–902.
- Seashore-Ludlow,B. et al. (2015) Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discov.*, **5**, 1210–1223.
- Teo,C.H. et al. (2010) Bundle methods for regularized risk minimization. *J. Mach. Learn. Res.*, **11**, 311–365.
- The Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. *Nature*, **490**, 61.
- Tsochantaridis,I. et al. (2005) Large margin methods for structured and interdependent output variables. *J. Mach. Learn. Res.*, **6**(Sep), 1453–1484.
- Tutt,A. et al. (2010) Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and advanced breast cancer: a proof-of-concept trial. *Lancet*, **376**, 235–244.
- Weimer,M. et al. (2007) COFI-RANK: maximum margin matrix factorization for collaborative ranking. In *Advances in Neural Information Processing Systems 20*, p. 1593–1600.
- Zhang,N. et al. (2015) Predicting anticancer drug responses using a dual-layer integrated cell line-drug network model. *PLoS Comput. Biol.*, **11**, e1004498.
- Zou,H. and Hastie,T. (2005) Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B (Stat. Methodol.)*, **67**, 301–320.