

An overview of machine learning methods for monotherapy drug response prediction

Farzaneh Firoozbakht, Behnam Yousefi, Benno Schwikowski

▶ **To cite this version:**

Farzaneh Firoozbakht, Behnam Yousefi, Benno Schwikowski. An overview of machine learning methods for monotherapy drug response prediction. Briefings in Bioinformatics, 2021, pp.bbab408. 10.1093/bib/bbab408 . hal-03376028

HAL Id: hal-03376028

https://hal.sorbonne-universite.fr/hal-03376028

Submitted on 13 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Briefings in Bioinformatics, 00(00), 2021,1–18

https://doi.org/10.1093/bib/bbab408
Review

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbab408/6383559 by BIUS Jussieu user on 13 October 2021

An overview of machine learning methods for monotherapy drug response prediction

Farzaneh Firoozbakht, Behnam Yousefi and Benno Schwikowski

Corresponding author: Benno Schwikowski, Systems Biology Group, Department of Computational Biology, Institut Pasteur, Paris, France.
E-mail: benno.schwikowski@pasteur.fr

Abstract

For an increasing number of preclinical samples, both detailed molecular profiles and their responses to various drugs are becoming available. Efforts to understand, and predict, drug responses in a data-driven manner have led to a proliferation of machine learning (ML) methods, with the longer term ambition of predicting clinical drug responses. Here, we provide a uniquely wide and deep systematic review of the rapidly evolving literature on monotherapy drug response prediction, with a systematic characterization and classification that comprises more than 70 ML methods in 13 subclasses, their input and output data types, modes of evaluation, and code and software availability. ML experts are provided with a fundamental understanding of the biological problem, and how ML methods are configured for it. Biologists and biomedical researchers are introduced to the basic principles of applicable ML methods, and their application to the problem of drug response prediction. We also provide systematic overviews of commonly used data sources used for training and evaluation methods.

Introduction

The continued development of new high-throughput molecular profiling ('omics') technologies and their concomitant increased accessibility for biomedical applications make it increasingly attractive to study complex biological phenomena using data-driven and machine learning (ML) approaches [1]. ML is a specific subset of artificial intelligence that allows automatic learning from data, and has contributed to a wide range of genomics research [2]. One particular application of ML is drug response prediction (DRP), in which phenotypic responses of biological samples are predicted on the basis of their molecular profiles, with predictors often providing mechanistic insight.

In the context of biomedical research, DRP has, until now, largely focused on the response to anticancer drugs, possibly owing to the importance of cancer as a major public health issue, and the relative ease with which cancer cells can be cultured for *in vitro* drug screening assays. Cancer is a multifactorial and heterogeneous disease; each tumour typically has its own unique molecular characteristics, often leading to diverse responses to the same therapy for patients despite similar clinical phenotypes [3]. Hence, personalized, instead of one-size-fits-all, approaches to therapy are thought to have great potential to improve the clinical treatment of cancer [4] by extending survival and reducing side effects [5].

Unlike in other applications of ML, the establishment of systematic training datasets, e.g. in clinical trials, where a large number of drugs are evaluated for each of a large number of patients, is ethically impossible, and practically extremely expensive to approximate. The systematic datasets used for DRP therefore come mainly from preclinical models, such as cell lines [6], organoids [7] and xenograft mice [8]. In contrast to cancer cell lines, organoids and xenograft mice comprise interactions with the tumour microenvironment, and are thus considered to be better models for clinical outcomes [9–11]. Another concern with cell lines is that they may diverge from the original tumour. Yet, the observable pharmacological patterns in cell lines, which are relatively easy to culture, are still considered as rich in information about the drug mechanisms of action [6, 12, 13], and encompass clinically relevant genomic alterations [14]. The relatively good availability of large-scale, systematic, cell line

Farzaneh Firoozbakht is a researcher in the Systems Biology Group at Institut Pasteur, Paris, France.
Behnam Yousefi is a PhD student in the Systems Biology Group at Institut Pasteur, Paris, France, and Sorbonne Université, École Doctorale Complexité du vivant, Paris, France.
Benno Schwikowski is a Research Director and Group Leader of the Systems Biology Group at Institut Pasteur, Paris, France.
Submitted: 7 July 2021; **Received (in revised form):** 25 August 2021

© The Author(s) 2021. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

DRP datasets [15–18] make them the most frequently used input for ML-based DRP.

Given the omics data acquired from cell lines or tumours, along with their phenotypic responses to drugs, various ML approaches have been employed for DRP, with different inputs, models and learning processes, and different levels of complexity and interpretability of the resulting DRP models. DRP models range from established simple linear regression models to recent and more complex such as deep neural networks (DNNs). In addition, DRP methods can use biomolecular networks (between, e.g. genes and cell lines) that represent known or derived relationships.

The most recent comprehensive review of DRP methods, published in 2015 [19], naturally does not encompass important recent developments, such as recommender systems-based DRP. Other, more recent, reviews focus on specific aspects. Adam *et al*. [20] provide a broad view of methods linked to recent technologies, such as deep learning and single cell sequencing, and discuss important extensions, such as combination therapy and the prediction of drug resistance. Ali and Aittokallio [21] focus on the principled lessons from the NCI-DREAM7 community challenge [22]. Paltun *et al*. [23] discuss three specific subclasses of DRP methods based on data integration approaches. Furthermore, Chen and Zhang [24] provide a systematic performance evaluation of 17 selected DRP methods.

The present review thus intends to provide a structured, up to date, panorama of DRP methods from the past 10 years, rather than deeper technical discussions relating to specific classes of methods, which can be found in some of the above reviews and the original and ML literature. While introducing basic biological and ML concepts when needed, (i) we discuss all major databases commonly used to learn DRP ML models, together with a systematic overview of the drug response measurement that was used, numbers of drugs/samples, and URLs, (ii) we survey more than 70 DRP methods across 13 subcategories, together with a systematic overview of used inputs, outputs, evaluation approach, and code and software availability, and (iii) we discuss common approaches to evaluate DRP methods.

Databases

A number of ambitious experimental efforts have produced large *pharmacogenomic* datasets, i.e. collections of drug responses of clinical or preclinical samples, along with their molecular profiles. Box I provides an overview of commonly used preclinical drug response metrics that are used to represent the drug responses in these datasets.

Pharmacogenomic data have provided insight into the relationship between clinical and preclinical molecular profiles and responses to anticancer drugs, and drug response-predictive biomarkers. In this section, we survey the commonly used large publicly available pharmogenomic datasets.

The Cancer Genome Atlas (TCGA) [25] provides genomic, epigenomic, transcriptomic and proteomic features of over 11 000 patients across 33 cancer types. TCGA is a collaborative effort between the National Cancer Institute and the National Human Genome Research Institute started in 2006 aiming to better understand cancer biology and improve cancer treatment and prevention. In TCGA, patient drug responses are categorized into four classes of complete response, partial response, progressive disease and stable disease, based on tumour growth measured by RECIST, an imaging-based criterion to evaluate the response in solid tumours [26]. The TCGA dataset can be downloaded from Genomic Data Commons Data Portal.

Based on the hypothesis that *in vitro* drug activity on cancer cell lines can capture essential aspects of *in vivo* drug activity on corresponding tumours [12], molecular and drug sensitivity profiles across a large number of cell lines have been generated; Table 1 provides an overview. In the late 1980s, a project led by the US National Cancer Institute measured the ability of thousands of anticancer drugs to inhibit cell growth across 60 human cancer cell lines [6, 12]. The *NCI-60* panel of measurements includes gene expression (GE), copy number variation (CNV), single nucleotide polymorphism (SNP), DNA methylation (DM) and proteomics profiles of each cell line. Analysis of this data resulted in the discovery of a number of gene mutations that are highly predictive of clinical responses [27] and provided new insights into drug activity modulators and drug-target interactions.

Following the NCI-60 effort, two large collaborative datasets were established: the Cancer Cell Line Encyclopedia (CCLE) [16] and the Genomics of Drug Sensitivity in Cancer (GDSC) [15]. CCLE is a collaboration of the Broad Institute and Novartis, and GDSC is a collaboration between The Cancer Genome Project at the Wellcome Trust Sanger Institute (UK) and the Center for Molecular Therapeutics at the Massachusetts General Hospital Cancer Center (USA). Both datasets contain GE, CNV, DM, and mutational status (MS) of around 1000 human cancer cell lines from various tumour types, together with pharmacological profiles across a large fraction of these cell lines, for 24 and 518 drugs, respectively, in CCLE and GDSC. Although several studies pointed at inconsistencies in the drug response measurements between CCLE and GDSC [28, 29], others argued that the inconsistencies disappear under a specific, reasonable, interpretation of the data [18, 30–33]. To shed more light on this issue, the Genentech Cell Line Screening Initiative [18] generated profiled 410 cell lines common to GDSC and CCLE using various approaches, including RNA-seq GE, MS and CNV, along with their responses to 16 drugs tested in both studies. The outcomes of this study supported the previously proposed mode of consistent interpretability across CCLE and GDSC data. PharmacoDB [34] provides large publicly available preclinical datasets that have been prepared for integrated use, with special emphasis on consistency.

The Cancer Therapeutics Response Portal (CTRP) [17] provided another pharmacogenomic dataset with a variety of different molecular features of cell lines, including MS, GE, CNV and tissue type. The last published version of this dataset includes 860 characterized cancer cell lines and their responses to 481 compounds. DepMap is a data repository that aggregates a number of different preclinical datasets (including GDSC, CCLE and CTRP).

DRP models

The high complexity of the biological response to drugs and the increasing feasibility of large-scale molecular profiling experiments make ML approaches increasingly attractive for modelling and prediction of drug responses. DRP models are most often established using a set of known phenotypic drug responses (supervised learning). Generally, the objective of DRP models is the prediction of the response of drugs for a given sample from its molecular profile (Figure 1). Regression estimates a quantitative drug response (Box I) for each sample, while classification predicts discrete labels that stratify samples into sensitive or resistant. To obtain ground truth labels in preclinical samples, the continuous measure of drug response is then thresholded by methods such as median-based [35], maximum concentration-based [14], waterfall-based [16] and

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbab408/6383559 by BIUS Jussieu user on 13 October 2021

Table 1. Publicly available drug screening datasets on cell lines

Datasets	Drug response measurement	Sample size	Drug number	Tissue number	Reference	URL
NCI60	GI50, IC50, LC50, TGI	60	<3000	9	Shoemaker *et al.* (2006)	https://dtp.cancer.gov/discovery_development/nci-60/
Genomics of Drug Sensitivity in Cancer (GDSC)	IC50, AUC	988	518	36	Garnett *et al.* (2012) Yang *et al.* (2012) Iorio *et al.* (2016)	https://www.cancerrxgene.org
Cancer Cell Line Encyclopedia (CCLE)	IC50, EC50, AA	504	24	36	Barretina *et al.* (2012)	https://portals.broadinstitute.org/ccle
GRAY	GI50	70	90	1	Heiser *et al.* (2012) Daemen *et al.* (2013)	https://link.springer.com/article/10.1186/gb-2013-14-10-r110#additional-information
Institute for Molecular Medicine Finland (FIMM)	EC50, DSS	106	308	1	Pemovska *et al.* (2013)	https://cancerdiscovery.aacrjournals.org/content/3/12/1416.figures-only
Genentech (GNE)	IC50	675	>350	17	Klijn *et al.* (2015)	https://www.nature.com/articles/nbt.3080#MOESM21
Human B-cell Cancer Cell Lines (HBCCL)	AUC, Sen/Res	26	3	2	Falgreen *et al.*, (2015)	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53798
Cancer Therapeutics Response Portal (CTRPv2)	AUC	860	481	25	Seashore-Ludlow *et al.* (2015)	http://portals.broadinstitute.org/ctrp.v2.1/
Genentech Cell Line Screening Initiative (gCSI)	AUC, IC50	410	16	23	Haverty *et al.* (2016)	https://pharmacodb.pmgenomics.ca/datasets/4
Head and neck squamous cell carcinomas cell lines	IC50	8	276	1	Chia *et al.* (2017)	https://www.chip-phenomics.org/DOWNLOAD/Chia2017/
Texas Southwestern Medical Center non-small cell lung cancer cell line	AUC, ED50	100	222	2	McMillan *et al.* (2018)	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104757

predefined [36–38]. From a practical perspective, discretization enables the use of classification methods, but it also loses information, and, if predictions are to be interpreted in a clinical context, the clinical relevance of the resulting classes requires special attention.

Feature selection, i.e. the maximal removal of data features with a minimal loss of problem-relevant information, is an important part of DRP, as it is a major antidote to the statistical and computational problems that the high-dimensional omics input data typically entail. For focused discussions of feature selection, we refer the reader to Webb and Copsey [39], and, for their use in DRP, to the study by Koras *et al.* [40].

Network-based machine learning models are relatively recent, and increasingly popular for DRP. Here, networks (often equivalent to mathematical graphs) can refer to sets of interactions between cell lines, drugs, genes, proteins or their combination. Networks carry different types of information, and their use and interpretation vary greatly across the different uses for DRP.

DRP models can be classified as single-drug learning (SDL) and multi-drug learning (MDL). To predict the response to a given drug, MDL leverages data from other drugs; SDL exclusively uses data about the given drug. SDL and MDL differ in their capabilities and validation process. Table 2 provides a systematic overview of all DRP articles covered here.

SDL models

SDL models are trained to predict drug responses for a given drug, independent of other drugs. In this section, we survey the SDL approaches in the literature across seven subclasses.

Linear regression

Regression analysis is a set of statistical methods for estimating the relationship between a dependent variable and one or more independent variables. As its simplest form, linear regression uses a linear function to model this relationship, and enjoys remarkable popularity due to its simplicity and interpretability.

In DRP, the number of available observations often does not suffice to determine a unique model. Regularization techniques address this issue by limiting the possible solutions through additional assumptions [41, 42], and are therefore a popular choice for DRP. Shrinkage is a form of regularization in which parameter values close to zero are preferred when they do not clearly contribute to the predictive performance of the model. Ridge regression [43] includes a form of shrinkage, and was used to assess whether models trained on preclinical samples can predict clinical drug responses [30, 44]. Ridge regression typically does not yield sparse models (i.e. models with few independent variables that are easier to interpret). Lasso regularization, on the other hand, forces coefficients to zero, their number being controlled by a hyperparameter [45]. In a recent study, Huang *et al.* proposed Tissue-Guided Lasso to derive tissue-specific models through tissue-specific choices for their hyperparameter. They defined tissue-specific hyperparameters by training each lasso model on all cell lines, excluding those from the tissue of interest, and then calibrated the hyperparameter using data from that tissue [46]. One principled limitation of Lasso is that the number of non-zero coefficients cannot be higher than the number of data samples [42]. To derive sparse models even in cases when the number of features is larger than

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbab408/6383559 by BIUS Jussieu user on 13 October 2021

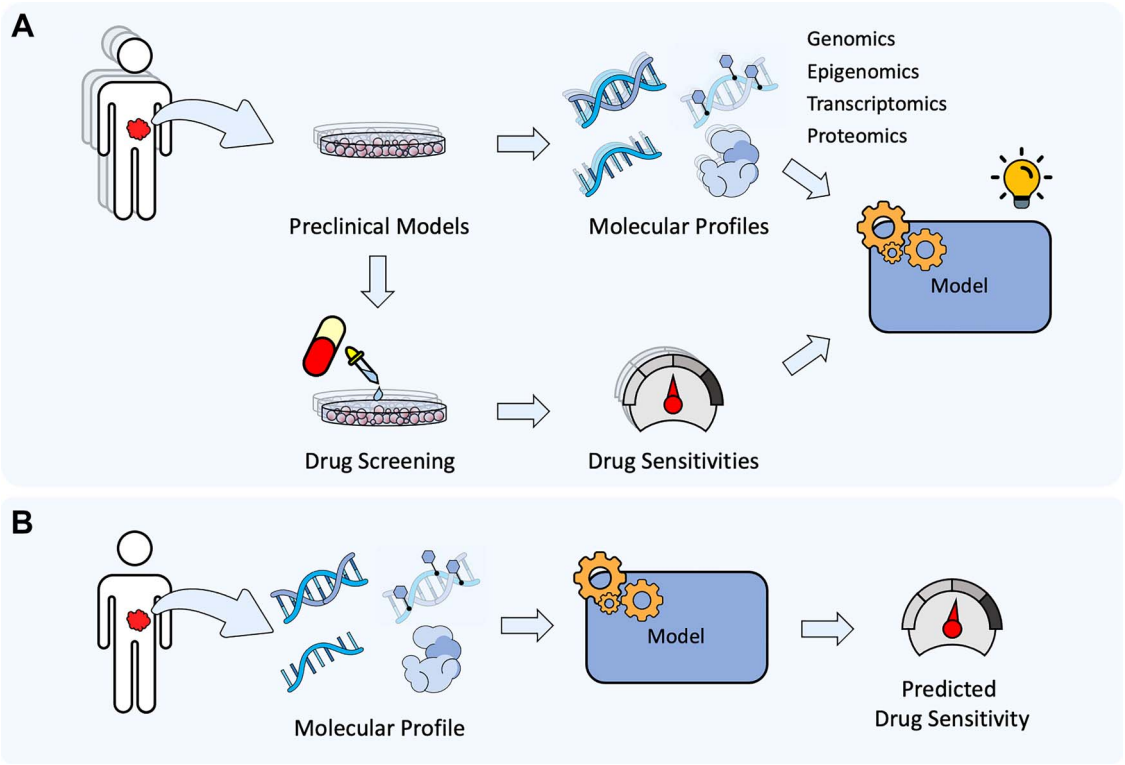


Figure 1. (A) Preclinical samples (e.g. cell lines), on one hand, are profiled to generate omics data, and on the other hand, they are screened over different drugs for obtaining their sensitivities. This information together is used to train machine learning models. (b) The trained model predicts the sensitivity for a new cell line.

the sample size, the elastic net [41] combines ridge regression and lasso regularization, and has been also used for DRP [16, 47–50].

Biomolecular networks represent interactions of biomolecules in the same functional context that can be leveraged for DRP. These interactions can be derived from prior knowledge or independent experimental data. A common use of networks for DRP is to use them for regularization, by assuming that model coefficients for functionally related molecules are similar. Sokolov *et al.* [51] introduced the generalized elastic net model that incorporates the Pathway Commons [52] network in an elastic net approach through a network-based regularization term.

Partial least squares regression (PLSR) is another type of linear method, based on latent variable projection [53]. In PLSR, input and output data are each projected into a (usually low-dimensional) new space, and then a relationship between them is established using linear regression. This makes PLSR suitable for analysing high-dimensional data and DRP [35, 54].

Logistic regression

Instead of predicting drug response as a continuous variable, samples can also be classified as sensitive or resistant. Logistic regression is a common classification approach that can be derived, in a Bayesian framework, from a simple probabilistic model that expresses drug response as a function of molecular profiles [30, 44, 55, 56]. Classifiers derived from these models have linear decision boundaries, i.e. lines or hyperplanes that separate different classes in the input space [42, 57]. The assumption of linearity can be a strength in the case of small datasets, but a limitation when larger datasets would be able to resolve more complex decision boundaries. Huang *et al.* [58] proposed a logistic regression model that is additionally constrained based on biomolecular networks.

Support vector machines

Support vector machines (SVMs) [42, 57, 59] are ML models that can be used for classification and regression. SVMs are efficiently learnable, and can employ nonlinear kernels (nonlinear transformation functions). Dong *et al.* used a SVM to classify samples as sensitive and resistant. SVM regression (SVR) [60] has also been used for drug prioritization [61] and response prediction [62, 63]. The selection of a kernel function and the interpretation of SVMs are typically not straightforward.

Nearest neighbour method

Nearest neighbour methods are simple and intuitively understandable classification and regression methods. In the *K*-Nearest Neighbour (KNN) method, the output for each new sample is on the basis of the outputs of a smaller number (commonly denoted by *K*) of molecular profiles that are the most similar to the profile of the new sample [42]. Li *et al.* [64] employed KNN regression for DRP, in which the sensitivity of a new sample is modelled as the average of the observed sensitivity values of its KNNs. Nearest neighbour methods tend to be problematic in the inadvertent, and typically unknown, presence of many irrelevant data features, and they are computationally intensive.

Artificial neural networks and deep learning

Artificial neural networks (ANNs) are powerful ML models that can approximate arbitrary input–output relationships, given a

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbab408/6383559 by BIUS Jussieu user on 13 October 2021

Table 2. Systematic overview of drug response prediction methods

Model	Paper	Method name	Dataset	Input	Output	Evaluation	Code/software availability
Single-drug learning (SDL) models							
Linear Regression							
	Barretina *et al.* (2012)	-	CCLE	MAGE, CNV, Mu, TT	IC50, AA	KFCV	-
	Papillon-Cavanagh *et al.* (2013)	-	GDSC, CCLE	MAGE	IC50	KFCV, CDV	-
	Niepel *et al.* (2013)	-	NCI	Mu	GI50	LOOCV	-
	Geeleher *et al.* (2014)	-	GDSC, three clinical datasets	MAGE	IC50	LOOCV, CDV	http://geeleherlab.org/cgpPrediction/
	Falgreen *et al.* (2015)	-	HBCCL, GDSC, IDRC, LLMPP, MDFCI, UAMS	MAGE	AUC	KFCV, LOOCV	-
	Covell *et al.* (2015)	-	GDSC, CCLE	MAGE	GI50	CDV	-
	Li *et al.* (2015)	-	OncoPanel, BATTLE	MAGE	IC50	RRSSCV	-
	Sokolov *et al.* (2016)	GELnet	LBNLBC	MAGE	GI50	LPOCV	https://cran.r-project.org/web/packages/gelnet/index.html
	Aben *et al.* (2016)	TANDEM	GDSC	MAGE, Mu, CNV, DM, TT	IC50	KFCV	https://cran.r-project.org/web/packages/TANDEM/index.html
	Huang *et al.* (2020)	TG-LASSO	GDSC, TCGA	MAGE	IC50	RRSSCV, CDV	https://github.com/emad2/TG-LASSO
Logistic Regression							
	Geeleher *et al.* (2014)	-	GDSC, three clinical datasets	MAGE	Res/Non-Res	LOOCV, CDV	http://geeleherlab.org/cgpPrediction/
	Falgreen *et al.* (2015)	-	HBCCL, GDSC, IDRC, LLMPP, MDFCI, UAMS	MAGE	Sen/Res	KFCV, LOOCV	-
	Ding *et al.* (2016)	-	TCGA	MAGE, CNV, DM, miRNA	Sen/Res		-
	Geeleher *et al.* (2017)	-	GDSC, TCGA	MAGE	IC50	KFCV, CDV	https://osf.io/pwm4z/
	Ding *et al.* (2018)	-	GDSC, CCLE	MAGE, CNV, Mu	Sen/Res	RRSSCV	-
	Huang *et al.* (2018)	Lq-NLR	GDSC, BATTLE	MAGE	IC50	KFCV	-
Maximum Margin Models							
	Dong *et al.* (2015)	-	GDSC, CCLE	MAGE	Sen/Res	KFCV, CDV	-
	Gupta *et al.* (2016)	-	CCLE	MAGE, CNV, Mu	IC50	-	http://crdd.osdd.net/raghava/cancerdp/
	Parca *et al.* (2019)	-	GDSC	MAGE	IC50	KFCV	https://github.com/lucaparca/dre
Ensemble Learning Methods							
	Riddick *et al.* (2010)	-	NCI-60	MAGE	IC50	OOBP	-
	Daemen *et al.* (2013)	-	LBNLBC, TCGA	MAGE, DM, ES, RSGE, PA, CNV	Sen/Res	RRSSCV, CDV	-
	Stetson *et al.* (2014)	-	NCI60, CCLE, GDSC	MAGE, SNP, CNV	IC50	KFCV, CDV	-
	Wan & Pal (2014)	-	CCLE, NCI-DREAM	MAGE, CNV, PA, DM, RSGE	GI50, IC50	KFCV, LOOCV	-
	Fang *et al.* (2018)	QRF	CCLE	MAGE, Mu, CNV	AA	OOBP	-
	Xu *et al.* (2019)	AutoBorutaRF	GDSC, CCLE	MAGE, SNP, CNV	Sen/Res	KFCV	https://github.com/bioinformatics-xu/AutoBorutaRF
	Oskooei *et al.* (2019)	NetBiTE	GDSC	MAGE	IC50	KFCV	-
	Su *et al.* (2019)	Deep-Resp-Forest	GDSC, CCLE	MAGE, CNV	Sen/Res	KFCV	https://github.com/RanSuLab/Deep-Resp-Forest
	Kurilov *et al.* (2020)	-	GDSC, CCLE, CTRP, gCSI, NIBR PDXE	MAGE, CNV, Mu, RSGE, TT	IC50, AUC, V1, TVC	RRSSCV	https://github.com/RomaHD/DrugRespPrediction
Nearest Neighbour Method							
	Li *et al.* (2021)	GA/KNN	GDSC, CCLE, TCGA, GTEx	MAGE, RSGE	IC50	RRSSCV, CDV	-

(Continued)

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbab408/6383559 by BIUS Jussieu user on 13 October 2021

Table 2. Continued

Model	Paper	Method name	Dataset	Input	Output	Evaluation	Code/software availability
Artificial Neural Networks and Deep Learning							
	Sakellaropoulos *et al.* (2019)	-	GDSC, TCGA, MD Anderson, OCCAMS, multiple myeloma	MAGE	IC50	KFCV	https://github.com/TeoSakel/deep-drug-response
	Sharifi-Noghabi *et al.* (2019)	MOLI	GDSC, PDX, TCGA	MAGE, SNP, CNV	Res/Non-Res	CDV	https://github.com/hosseinshn/MOLI
	Ahmed *et al.* (2020)	-	NSCLCCLP	RSGE	AUC, ED50	RRSSCV	https://github.com/compbiolabucf/drug-sensitivity-prediction
	Ma *et al.* (2021)	TCRP	GDSC, CCLE, DepMap, PDTC BioBank, PDX Encyclopedia	MAGE, Mu	AUC, TVC	CDV	https://github.com/idekerlab/TCRP/
	Malik *et al.* (2021)	-	GDSC, TCGA	MAGE, Mu, CNV, DM	IC50	KFCV, CDV	https://github.com/TeamSundar/BRCA_multiomics
Molecular Network Similarity-Based Methods							
	Kim *et al.* (2016)	NBC	GDSC, CCLE	MAGE	Sen/Res	KFCV	-
	Stanfield *et al.* (2017)	-	GDSC, CCLE	SNP	CDSS	LOOCV, CDV	-
Multi-drug learning (MDL) models							
Linear Regression							
	Costello *et al.* (2014)	Bayesian multitask MKL	NCI-DREAM, GDSC	MAGE, RSGE, CNV, Mu, DM, PA	GI50	challenge test cell lines	https://www.nature.com/articles/nbt.2877#MOESM3
	Yuan *et al.* (2016)	-	CCLE, CTRPv2, NCI60	MAGE, CNV, Mu	AA, AUC, GI50	KFCV	-
	Ammad-ud-din *et al.* (2017)	MVLR	GDSC, FIMM	MAGE	IC50, CDSS	LOOCV	https://github.com/suleimank/mvlr
Ensemble Learning Methods							
	Matlock *et al.* (2018)	-	GDSC, CCLE	MAGE, DT, DD	AUC	RRSSCV	-
	Liu *et al.* (2020)	-	GDSC, CCLE	MAGE	AA, AUC	KFCV	https://zenodo.org/record/1325121#.YNHB_i0RozV
	Sharma & Rani (2019)	-	GDSC, CCLE, NCI-Dream	MAGE	IC50, GI50	KFCV	-
	Su *et al.* (2020)	Meta-GDBP	GDSC, CCLE	MAGE, DD	IC50, AA	KFCV, CDV	https://github.com/RanSuLab/Meta-GDBP
Artificial Neural Networks and Deep Learning							
	Menden *et al.* (2013)	-	GDSC	MS, MU, CNV, DD	IC50	KFCV	-
	Chang *et al.* (2018)	CDRscan	GDSC, CCLP	SNP, DD	IC50	KFCV	-
	Li *et al.* (2019)	Deep DSC	GDSC, CCLE	MAGE, DD	IC50	KFCV, LOTO, LOCO	-
	Chiu *et al.* (2019)	DeepDR	GDSC, TCGA	MAGE, Mu	IC50	RRSSCV	-
	Joo *et al.* (2019)	DeepIC50	GDSC, CCLE, TCGA	Mu, DD	3-class sensitivity	RRSSCV, CDV	https://github.com/labnams/DeepIC50
	Liu *et al.* (2019)	tCNNS	GDSC	Mu, CNV, DD	IC50	RRSSCV, LOTO	https://github.com/Lowpassfilter/tCNNS-Project
	Manica *et al.* (2019)	MCA	GDSC	MAGE, DD	IC50	KFCV, RRSSCV	https://github.com/drugilsberg/paccmann https://ibm.biz/paccmann-aas
	Choi *et al.* (2020)	RefDNN	GDSC, CCLE	MAGE, DD	Sen/Res	KFCV	https://github.com/mathcom/RefDNN
	Zhu *et al.* (2020)	-	GDSC, CCLE, CTRP, gCSI	MAGE, DD	AUC	KFCC	-
	Bazgir *et al.* (2020)	REFINED	GDSC, NCI-60	MAGE, DD	IC50, GI50, Sen/Res	KFCV, RRSSCV	https://github.com/omidbazgirTTU/REFINED
	Liu *et al.* (2020)	DeepCDR	GDSC, CCLE, TCGA	MAGE, Mu, DM, DD	IC50, Sen/Res	KFCV, CDV	https://github.com/kimmo1019/DeepCDR

(Continued)

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbab408/6383559 by BIUS Jussieu user on 13 October 2021

Table 2. Continued

Model	Paper	Method name	Dataset	Input	Output	Evaluation	Code/software availability
	Tang & Gottlieb (2021)	PathDSP	GDSC, CCLE	MAGE, CNV, Mu, DD	IC50	KFCV, CDV	https://github.com/TangYiChing/PathDSP
	Nguyen *et al.* (2021)	GraphDRP	GDSC	Mu, CNV, DD	IC50	RRSSCV	https://github.com/hauldhut/GraphDRP
Recommender Systems (neighbourhood-based)							
	Zhang *et al.* (2015)	Dual-layer integrated cell line-drug network	GDSC, CCLE	MAGE, DD	AA, IC50	LOOCV	-
	Sheng *et al.* (2015)	-	GDSC, CCLE	MAGE, DD	IC50	KFCV, CDV	-
	Liu *et al.* (2018)	NCFGER	GDSC, CCLE	MAGE, DD	IC50, AA	KFCV	-
	Zhang *et al.* (2018)	HIWCF	GDSC, CCLE	MAGE, DD, DT	IC50, AA	KFCV	https://github.com/laureniezhang/HIWCF
	Le & Pham (2018)	GloNetDRP	GDSC, CCLE	MAGE, Mu	AUC, IC50	KFCV	-
	Wei *et al.* (2019)	CDCN	GDSC, CCLE	MAGE, DD	AA, Sen/Res	LOOCV	https://zenodo.org/record/1403638#.YNHVyi0RozV
Recommender Systems (model-based)							
	Ammad-ud-din *et al.* (2014)	KBMF	GDSC	MAGE, CNV, Mu, DD, DT	IC50	KFCV	https://research.cs.aalto.fi/pml/software/kbmf/
	Ammad-ud-din *et al.* (2016)	cwKBMF	GDSC, CTRPv1, AMLCLP	MAGE, DD	IC50, AUC	KFCV, CDV	https://research.cs.aalto.fi/pml/software/cwkbmf/
	Wang *et al.* (2017)	SRMF	GDSC, CCLE	MAGE, DD	AA, IC50	KFCV	https://github.com/linwang1982/SRMF
	Suphavilai *et al.* (2018)	CaDRReS	GDSC, CCLE, HNCCLP	MAGE	IC50	KFCV	https://github.com/CSB5/CaDRReS
	Yang *et al.* (2018)	Macau	GDSC	MAGE, DT	IC50	KFCV	https://github.com/saezlab/Macau_project_1
	Guan *et al.* (2019)	WGRMF	GDSC, CCLE	MAGE, DD	IC50, AA	KFCV	-
	Moughari & Eslahchi (2020)	ADRML	GDSC, CCLE	MAGE, Mu, CNV, DD, DT	IC50	KFCV	https://github.com/fahmadimoughari/ADRML
	Emdadi & Eslahchi (2020)	DSPLMF	GDSC, CCLE	MAGE, Mu, CNV, DD	Sen/Res	KFCV	https://github.com/emdadi/DSPLMF
	Emdadi & Eslahchi (2021)	Auto-HMM-LMF	GDSC, CCLE	MAGE, CNV, Mu, TT, DD	Sen/Res	KFCV	https://github.com/emdadi/Auto-HMM-LMF
Network Representation Learning-Based Models							
	Yang *et al.* (2019)	NRL2DRP	GDSC	Mu, CNV, DM	Sen/Res	KFCV	https://github.com/USTC-HIlab/NRL2DRP
	Yu *et al.* (2020)	DREMO	GDSC, CCLE	MAGE, Mu, CNV, DD, DT	Sen/Res	KFCV	https://github.com/LiangYu-Xidian/MNDRP
Network Propagation-Based Method	Zhang *et al.* (2018)	HNMDRP	GDSC	MAGE, DD	Sen/Res	LOOCV	https://github.com/USTC-HIlab/HNMDRP

AA = activity area; AUC = area under curve; CDSS = custom drug sensitivity score; CDV = cross-dataset validation; CNV = copy number variation; DD = drug descriptors; DM = DNA methylation; ES = exome sequencing; MAGE = microarray gene expression; KFCV = K-fold cross validation; LOCO = leave-one-compound-out; LOOCV = leave-one-out cross validation; LOTO = leave-one-tissue-out; LPOCV = leave-pair-out cross validation; miRNA = micro-RNA; MS = microsatellite; Mu = mutation; OOBP = out-of-bag prediction; PA = protein abundance; Res/Non-Res = responder/non-responder; RSGE = RNA-seq gene expression; RRSSCV = repeated random sub-sampling cross validation; Sen/Res = sensitive/resistance; SNP = single nucleotide polymorphism; TT = tissue type; TVC = tumor volume changes; V1 = viability at 1 μm.

large and diverse enough set of training samples. Multilayer feedforward neural networks are one of the most commonly used ANNs that consist of multiple layers [57]; each layer consists of a number of computational units, called neurons (Figure 2A). These ANNs are usually fully connected, i.e. the first layer of neurons receives the input data, and the outputs of all neurons in any given layer are used as inputs for all neurons in the next layer.

In their generic form, ANNs require more input data than simpler models. Ma *et al.* [65] used a transfer learning approach to address the problem of relatively little available DRP data for advanced preclinical models. In a first pre-training phase

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbab408/6383559 by BIUS Jussieu user on 13 October 2021

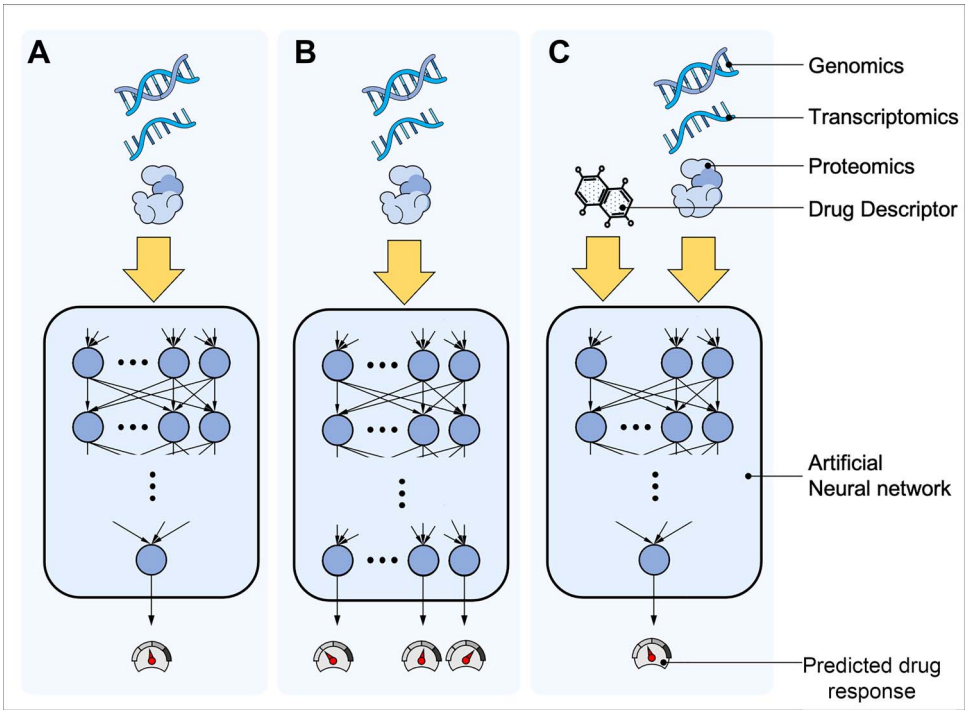


Figure 2. Artificial neural networks in drug response prediction can be configured as (A) single drug learning (SDL) or (B, C) multi-drug learning (MDL). SDL ANNs learns the response of a single drug for a sample using omics data, whereas MDL ANN learns the response to multiple drugs, which can have (B) multiple outputs, such that each output corresponds to the prediction for a specific drug, or (C) single output with the additional input of drug features.

of their TCRP method, a multilayer ANN is learned using a larger number of basic preclinical training samples (cell lines). In the following learning phase, a small dataset from advanced preclinical models is used to recalibrate the model to the new context.

ANNs with many layers are called DNNs. Deep learning has recently attracted considerable attention across many fields and has become popular for DRP as well [66–68]. Encoders are particular types of ANNs, typically used in DNNs for feature reduction, i.e. the encoding of higher dimensional information using fewer dimensions with no or little loss of information. Stacked encoders–decoder networks can be trained simultaneously as autoencoders without the need for output data [69, 70]. Ding *et al.* trained a four-layer autoencoder, and the output of each layer was separately used to train a logistic elastic net to predict drug responses [56]. Sharifi-Noghabi *et al.* [71] introduced a DNN that learns encoders of different molecular data types separately, and whose outputs are fed to a multilayer neural network for DRP. Determining the best structure of ANNs and interpreting the resulting models are non-trivial, and learning ANNs tends to be computationally expensive.

Graph neural networks (GNNs) are ANNs that operate on graphs (i.e. sets of interactions) as inputs. Recently, deep GNNs have been used to learn low-dimensional representations of biomolecular networks [72, 73]. Ahmed *et al.* [74] employed two different GNN methods to use GE and a gene co-expression network, i.e. a network representing the correlation between the expression of gene pairs, to construct a GNN. In contrast to the use of GNNs for MDL (discussed in Section ANNs and Deep Learning below), GNNs have not been shown to improve significantly over the SDL performance of classical ML models.

Network similarity-based methods

Network similarity-based methods are based on the assumption that cell lines with similar network-level information respond similarly to the same drug. Class-specific networks, typically a sensitive network and a resistant network are then constructed to capture features characteristic for the samples from both classes. Although the above assumption underlying this relatively recent class of methods is simple, its validity needs to be established for each particular interpretation of network similarity.

Stanfield *et al.* [75] construct a 'base network' of genes and cell lines, in which cell lines are linked to their mutated genes, which, in turn, are connected by BioGRID protein–protein interactions [76]. To inject information about drug responses, the base network is then augmented with links between cell lines and the drug they are sensitive to, to obtain a sensitive network, and, analogously, a resistant network. The augmentation is designed to lead to tighter network connectivity between the drug, cell lines and genes involved in sensitivity or resistance, respectively. A new cell line is then classified based on the similarity of the connectivity of its base network with each of the class-specific networks. Kim *et al.* [77] proposed another biomolecular network-based classifier that exploits the presence of class-specific gene co-expression patterns. First, sensitive and resistant gene co-expression networks are constructed from the GE values of the corresponding cell lines. Next, for each gene in each of these networks, class-specific SVR models are trained to predict GE values on the basis of the GE of neighbouring genes in the class-specific network. For a new sample, the accuracy of the class-specific SVR prediction for any given gene is then evaluated, and interpreted as evidence for, or against, the membership of the new sample in the corresponding class.

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbab408/6383559 by BIUS Jussieu user on 13 October 2021

Ensemble learning methods

By aggregating the predictions of multiple models, ensemble learning methods are often able to improve over the performance of any individual model [78]. Random forests (RFs) are established ensemble learning models that aggregate several regression or classification trees [79]. Trees are nonlinear ML models that predict the output by partitioning the input space into a hierarchical set of rectangles in the input space. The predicted output for the classification problem is the class label associated with the rectangle that corresponds to a new input. For the regression problem, the output is commonly the average value of the training outputs corresponding to each rectangle. RF trees are constructed from a random selection of the samples, and the remaining samples, called out-of-bag *samples*, can be used for validation. Riddick *et al.* [80] and Daemen *et al.* [81] initially used RF to predict drug sensitivity and identify molecular features associated with drug responses. Stetson *et al.* [82] selected differentially expressed genes between the most sensitive and resistant cell lines, and then used to train three classifiers using RF, SVM and logistic elastic net. Wan & Pal [83] proposed a RF regression-based approach using multiple omics data. To this end, they trained multiple RF regressions for different data types, and combined the individual predictions using linear regression. Moreover, Xu *et al.* [84] proposed a method called *AutoBorutaRF*, which employs an autoencoder-based feature selection algorithm and the Boruta algorithm [85] to improve the performance of RF.

To incorporate biological prior knowledge into the RF, Oskooei *et al.* [86] presented *NetBiTE*, in which a RF is biased towards using the known drug targets and their STRING PPI [87] interactors based on a network diffusion algorithm, i.e. a mathematical transformation of omics data that allows the detection of hypothetically relevant network regions [88]. Zhou & Feng [89] proposed an ensemble method called gcForest, whose performance was found to be competitive with DNNs. A modified version of gcForest, was used for drug response classification [38].

Fang *et al.* [90] argue that, due to the uncertainty inherent in DRP prediction, clinical applications benefit from richer output than a single predicted drug response. Their RF-based predictor is adapted to non-standard output distributions, and yields prediction intervals that, as they argue, may provide a better basis for the choice of appropriate treatments.

MDL models

MDL models use data from multiple drugs to predict the response of a given drug. This may lead to improved predictions for drugs with training data, and, potentially, the prediction of responses to drugs for which no training data are available.

Multitask linear regression

Linear regression is also employed for MDL. In a DRP challenge competition performed by NCI-DREAM [22] the best-performing method was Bayesian multi-task multiple kernel learning (MKL), in which multiple data types (including original omics datasets and computationally derived data) are first kernelized (i.e. converted into nonlinear transfer functions), and then a global similarity matrix is constructed as their weighted sum. Next, a set of drug-specific weights, determined through multitask linear regression, are used to provide the final drug response values. Ammad-ud-din *et al.* [91] pointed out that, although kernels may enhance the model predictions, the resulting models are typically hard to interpret, in terms of the relationship of their inputs to their outputs. To address this issue, they created the *MVLR model*, which aggregates groups of molecular features that are known to be biologically connected, and use them as inputs to a multitask linear regression. In another study, Yuan *et al.* [92] jointly trained multilinear regression models based on trace norm regularization, which can increase the dependencies between model parameters to yield a more manageable number of biologically meaningful features.

ANNs and deep learning

ANNs are widely used in MDL. Through shared model parameters, MDL models can leverage information between different drugs and reduce the problem of overfitting, i.e. the learning of parameters driven by characteristics of the training data that are not necessarily present in other data. MDL ANNs have either single or multiple outputs. Each output of a multi-output ANN predicts the response for a specific drug (Figure 2B). Chiu *et al.* [93] proposed a multitask DNN model for DRP called *DeepDR*. Their method is composed of mutation and GE encoders that reduce data dimensionality, followed by a multilayer ANN with multiple outputs.

Single-output ANNs, on the other hand, accept omics data along with drug features as input (e.g. chemical descriptors and biological targets; Figure 2C). Unlike multi-output ANNs, they share all model parameters among multiple drugs. The potential advantage of single-output ANNs is their ability to provide predictions for new drugs whose data have not been used during the training phase. Menden *et al.* [94] employed a fully connected ANN to predict drug response. *DeepDSC* [95] first reduces the dimension of GEs using an encoder whose output is combined with the chemical drug features to be fed to a multilayer neural network. *RefDNN* [96] trains, for each drug, an SDL logistic elastic net regression model that predicts the probability of drug resistance. To predict the resistance of a given cell line to a new query drug, the vector of predicted resistance probabilities is then weighted by a structure similarity profile of the query drug with all the drugs as an input to a neural network to classify the response as sensitive or resistant. In an effort to leverage the observed activity of signalling pathways, Tang & Gottlieb [97] proposed a pathway-based method named *PathDSP*. In their approach, drug structure data along with the pathways enriched for drug targets, GE, somatic mutation and CNV are fed as inputs to a DNN. Zhu *et al.* [98] used transfer learning to overcome the challenge of learning with limited amounts of data by pre-training their model on a dataset from a related context, and then finalizing the training process in the context in which the model is designed to predict.

Convolutional neural networks (CNNs) are specific types of non-fully connected feedforward ANNs that can exploit natural adjacency structures in the data (such as physical or temporal adjacency) to more efficiently detect global patterns in the input data. CNNs work well for the inputs that contain spatial or temporal patterns (such as images) [70, 99, 100]. Joo *et al.* [101] proposed a CNN method, called *DeepIC50*, to predict three classes of high, intermediate and low responsiveness based on mutation status and drug descriptors. Moreover, in *CDRscan* [102], the mutation data of cell lines and the molecular fingerprints of drugs are fed to five different versions of a CNN, whose average predicted output value is reported as the final output. Liu *et al.* [103] introduced a model named *tCNNS*, in which two CNN branches are respectively used for cell line and drug features to predict drug sensitivities. Bazgir *et al.* [104] proposed a method called *REFINED* to convert GE profiles to an image that encodes

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbab408/6383559 by BIUS Jussieu user on 13 October 2021

relevant patterns of the molecular features in specific regions. The resulting images are then fed to a CNN for DRP.

GNNs can be used to encode drug features as points in a low-dimensional space, in which similar drugs are close to each other, by representing molecular structure of compounds as networks formed by sets of chemical bonds between atoms [105, 106]. In *DeepCDR* [107] and *GraphDRP* [108], deep drug features learnt by GNNs are concatenated with deep multi-omics features, to be fed to feedforward ANNs for DRP. Manica *et al.* [109] proposed the *MCA* approach, in which genes are selected based on a network diffusion algorithm [86]. Attention-based convolutional encoders are a recent new development from the field of natural language processing that has been useful to reduce the dimensionality of the GE and drug chemical data, to be used as input to a fully connected neural network for DRP [110].

Recommender systems

Recommender systems (RS) are algorithms that yield a set of personalized item suggestions to a user. RS were developed in the 1990s to automatically recommend movies and other products to users, based on their similarity of previously observed choices by the same user, or to the choices of other users with similar preferences [111, 112]. When transposed to DRP, this principle implies that a cell line is highly sensitive to a drug when similar cell lines are highly sensitive to similar drugs. Generally, RSs can be neighbourhood based or model based [111]. The main input used for the RS-based DRP models is a cell line-drug sensitivity matrix of sensitivity values of cell lines to the screened drugs. Complementary information, such as cell line molecular features and drug descriptor features, can also be used since they can enhance the similarity between different cell lines or different drugs.

Neighbourhood-based RS. Neighbourhood-based methods (also called memory-based methods) are simple types of RS, in which the cell line-drug sensitivity prediction is based on the sensitivities of cell lines with similar characteristics to similar drugs, on the basis of cell line similarity network and drug similarity networks (Figure 3A) [113].

Zhang *et al.* [114] constructed cell line similarity and drug similarity networks using pairwise correlation of GEs and pairwise correlation of quantitative drug features, respectively. For a given cell line-drug pair, the predicted response is based on a weighted average of the known responses of cell lines to the given drug and all the drugs to the given cell line. These weights are normalized versions of cell line similarities and drug similarities. Le and Pham [115] tried to improve on this by basing their prediction upon all input cell line-drug responses.

One important drawback of the above approaches is their inability to predict in the case that drugs and cell lines are new (i.e. not part of the training data). To address this issue, Wei *et al.* [116] proposed an extension of Zhang's method called *CDCN*. Conversely, Sheng *et al.* [117] raised the possibility that distinct drug responses between two cell lines might only manifest as GE differences in a few specific genes directly related to the drug resistance. To take this into account, they defined cell line similarity scores based on drug-specific sets of genes. To identify this set of genes for any given drug, their method first identifies drugs that are similar in terms of their chemical structures. Second, for each of those similar drugs, those genes whose GEs are highly associated with drug response are determined. Third, the similarity of all other cell lines to the given cell line is defined, based on the expressions of those genes. Finally, a similarity-weighted average of the IC50 values observed in other cell lines, based on the drug and cell line similarities, is used to predict the most sensitive drugs for the given cell line.

The cell line similarity and drug similarity measures in the above studies are only based on the molecular features of cell lines and drug descriptors, respectively. Conversely, the observed cell line-drug responses can also be used to compute meaningful similarity measures [118, 119].

Overall, neighbourhood-based RSs yield results based on a simple and intuitive principle. However, these methods tend to be heuristic in nature, due to their, typically ad hoc, selection of similarity measures and their subsequent heuristic downstream use. Further drawbacks of these methods are their typically high memory consumption and computational cost for testing samples, as well their difficult application on sparse input response matrices [113].

Model-based RS. In contrast to neighbourhood-based methods, model-based RSs require an explicit model and a training process that determines the model parameters. Once learned, models are fast to apply to new data, and relatively little memory is needed to store model parameters. In addition, model-based RSs typically perform better in cases where the cell line-drug sensitivity matrix is sparse [113].

Different learning algorithms can be used to train parameters of model-based methods. Typical RS-based methods for DRP are matrix factorization methods that construct latent factor models [113]. Such methods commonly represent the cell line-drug sensitivity matrix as a product of two matrices, allowing an interpretation of cell lines and drugs as vectors in a latent space (Figure 3B).

Ammad-ud-din *et al.* [120] extended a kernelized Bayesian matrix factorization (KBMF) method that was originally developed for drug-protein interaction analysis [121], to combine different omics data and drug properties for DRP. Here, kernels allow capturing nonlinear relationships between drug response and both omics and drug features. To improve model interpretability, they further introduced component-wise MKL. In this method, biological pathway-based groups of features are used as model inputs, thus incorporating prior biological knowledge in the model [122].

Similarity-regularized matrix factorization by Wang *et al.* [123] employs regularization by GE-based cell line similarity and chemical feature-based drug similarity measures. More recently, a modified regularization term for an analogous model was proposed by Moughari and Eslahchi [124]. They showed that cell line similarities based on GE and drug similarity-based drug targets provide the best performance among a variety of cell line and drug features. To account for redundant information in similarities, the *WGRMF* method [125] approximates the cell line similarity and drug similarity matrix with a sparse matrix.

As the three above methods do not define any transformation matrix that projects cell lines and drug features into a latent space, they cannot predict drug sensitivity for new cell lines or drugs. Using a cell line projection matrix, Suphavilai *et al.* [126] developed a matrix factorization-based RS, which enables the prediction for new cell lines. In addition, Yang *et al.* [127] used a drug projection matrix that additionally enables the prediction for a new drug. They used a method called *MACAU* [128] to incorporate GE and drug target information to train parameters. Moreover, to make predictions for new cell lines, Emdadi & Eslahchi [129] employ a neighbourhood-based approach, in which the latent vector for the new cell line is calculated as the average latent vector of its neighbours. Their *DSPLMF* method trains a classification model based on logistic matrix factorization to predict the probability of a cell line being sensitive to a drug. Such

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbab408/6383559 by BIUS Jussieu user on 13 October 2021

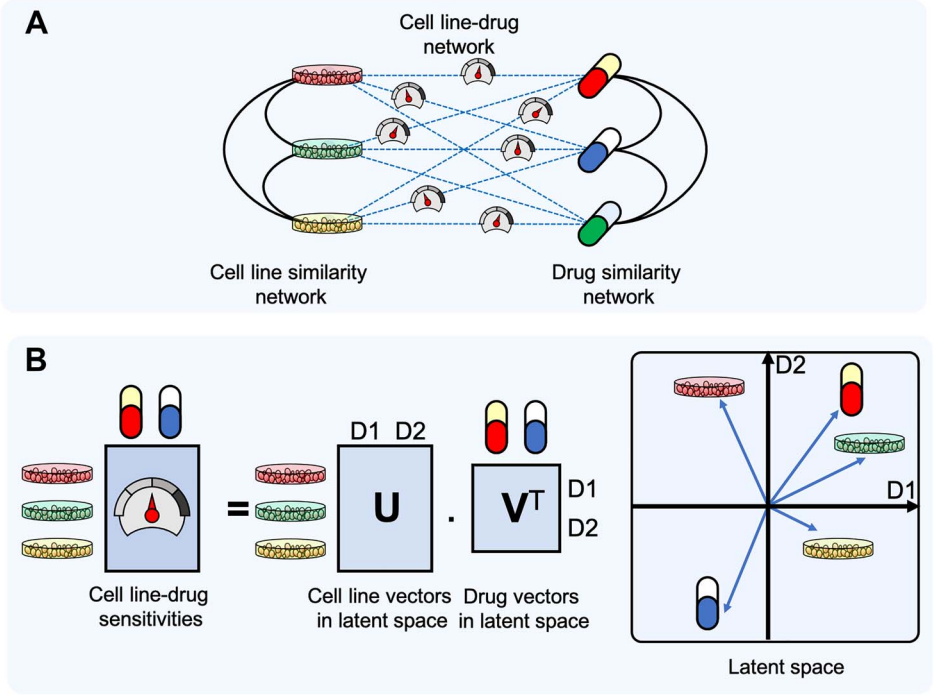


Figure 3. Two principal types of recommender systems (RS) for drug response prediction. (A) Neighbourhood-based RSs, which use cell line similarity and drug similarity measures to predict the response of a cell line to a drug. (B) Model-based RSs for DRP, which typically use matrix factorization (*left*), in which the cell line and drugs are represented as vectors in a *latent space* (*right*). The response of a cell line to a drug is then modelled to be proportional to the length of each of the vectors, and a decreasing function of the angle between cell line vector and the drug vector.

models can be categorized as hybrid model-/neighbourhood-based RS. To further improve the performance of their method, they developed a feature selection method called *Auto-HMM-LMF* [130].

Network representation learning-based models

Network representation learning methods project networks into a low-dimensional space [131] which can be used as the input features for ML. Yang *et al.* [132] proposed a method, in which a response network is first constructed by combining cell line genetic aberrations, drug responses and a PPI network. In the response network, cell lines with similar drug responses are close to each other. Next, a graph representation learning algorithm is used to learn a projection of the cell lines from the response network to a space that preserves their neighbourhood proximity. Finally, an SVM is trained on the representation vector of cell lines for each drug to predict the drug response of cell lines. The idea behind this method is that the cell lines with similar neighbourhood topology profiles have similar drug responses. The DREMO method by Yu *et al.* [133] constructs two multilayer similarity networks for cell lines and drugs based on different molecular data types, and then diffusion component analysis [134] is used to represent the cell lines and drugs into low-dimensional feature vector spaces. Finally, the sensitivity of the cell lines to each drug is predicted using logistic regression.

Network diffusion-based methods

Zhang *et al.* [135] proposed a DRP method on the basis of network diffusion on a heterogeneous network, consisting of cell line similarities, drug similarities, target similarities, cell line-drug sensitivities and drug–target interactions. Their target similarity network is obtained by merging PPIs and co-expression networks. In this method, a diffusion algorithm, called information flow-based method [136], is used to infer unknown cell line-drug response to be sensitive on the basis of the known responses by considering the density of the connectivities associated with the cell lines sensitive to drugs.

Ensemble learning methods

Ensemble learning (introduced for SDL models in Section Ensemble Learning Methods) can also be applied for MDL models. Matlock *et al.* [137] used an ensemble of RF, KNN and ANN methods for DRP. Liu *et al.* integrated a low-rank matrix completion and a ridge regression model for DRP [138]. Sharma & Rani [139] proposed another ensemble learning approach incorporating KBMTL [140], STREAM [141], L21 [142] and multitask regression. They first trained four different regression methods for predicting drug responses and then combined the output of each model using weighted averaging. Su *et al.* [143] proposed the two-level *Meta-GDBP* method. The first level contains four independent predictive models, based on GE, chemical properties of drugs, biological processes (based on GO terms [144]) and biological pathways (based on KEGG pathways [145]). For each of these information sources, they evaluated the performances of elastic net and SVR predictors. The second-level model is a weighted linear combination of the outputs of the best-performing models in the first level.

Evaluation of DRP models

Model evaluation, i.e. the assessment of the expected performance of a trained model on future data, needs to occur on test

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbab408/6383559 by BIUS Jussieu user on 13 October 2021

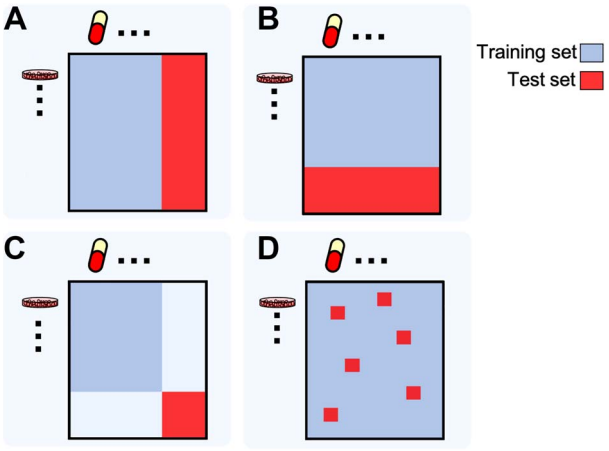


Figure 4. Different MDL model validation schemes. (A) Known cell line/new drug; (B) new cell line/known drug; (C) new cell line/new drug; (D) known cell line/known drug.

data that are independent from the training data. Otherwise, during training, the learned model parameters may be better adapted to the test data than could be expected for future data, thus leaving the test data evaluation optimistically biased. Data leakage, e.g. the—sometimes subtle—presence of information about the test data in the training data, may, for instance, occur when very similar data, such as drugs with similar structures, are present between training and test sets, or when the entire dataset is used for data preprocessing, e.g. for normalization across cell lines. There are different common ways of splitting a single dataset into training and test data [57]. In repeated random subsampling, the dataset is randomly split into training and test datasets repeatedly, typically, of given sizes. In each iteration, the model is trained on the training set and an evaluation metric (e.g. mean square error) is calculated on the test set. Finally, the model performance is estimated as the average evaluation metric over all iterations. In k-fold cross validation, the dataset is split into k subsets (folds, typically of identical size). In k iterations, each fold is used as a test set, and the remaining data are used for training. K-fold cross validation ensures that each data sample will appear in both train and test sets with the same frequency, and avoids stochastic fluctuations, particularly for small datasets. Moreover, in cases when k-fold cross validation would lead to training sets that are too small, typically, leave-one-out cross validation (LOOCV) is used, in which test sets of only a single sample are used in each iteration. It is worth noting that LOOCV is not recommended for large datasets since it is computationally expensive and leads to estimates of the model performance with high statistical variance [146]. Leave-pair-out cross validation (LPOCV) is a special case where only two samples are held out. Finally, there is a validation method specifically for RF called out-of-bag validation (refer to section Ensemble Learning Methods). In models with hyperparameters, an additional subset of the training data (validation set) is typically split-off from the training set before training to evaluate the model under different hyperparameter values before its version with the best hyperparameter is evaluated on the test set.

As the performance of MDL models for new data may depend on whether both drugs and cell lines are present or not in the training data (i.e. known or new), four different evaluation modes exist (Figure 4):

1. Known cell line/new drug. The prediction on new drugs typically uses known drug features. Leave-one-compound-out is a special case where only one drug is left out.
2. New cell line/known drug: This case is typical for personalized medicine applications. Leaving all cell lines derived from one tissue out as the test set is called leave-one-tissue-out.
3. New cell line/new drug: This evaluation corresponds to the most ambitious prediction task.
4. Known cell line/known drug: This evaluation corresponds to the least ambitious prediction task, which can be used to impute missing values.

Training and test sets can also be selected from different datasets. Cross-dataset evaluation thus also incorporates potential batch effects, i.e. systematically different characteristics between datasets, such as scale. In many real-world cases, heterogeneous output data present an additional challenge. For instance, regression models that predict IC50 values observed on cell lines may be difficult to evaluate against the drug effects as they are commonly characterized in clinical datasets.

In biomedical classification scenarios, the amount of available data may strongly vary across the different classes, which presents special challenges for learning and evaluation. To balance classes before learning, one can undersample the majority class, over-sample the minority class with replacement, or use synthetic data (e.g. SMOTE) [145, 146]. For a discussion of subtleties and pitfalls of these approaches, we refer the reader to the review by Chawla [147].

Discussion and outlook

Machine learning is a subfield of artificial intelligence whose classification and prediction algorithms learn patterns underlying the data from examples. Omics data usable for DRP consists of different types of molecular profiles, typically from genomics, epigenomics, transcriptomics and proteomics. The different available data types capture different (potentially overlapping) facets of the biological drug response. Comparative studies found that, among different individual data types, GE enables the best predictive performance, which can then be improved only slightly by adding other data types [22, 24, 50, 148]. Iorio *et al.* [14] reported that GE is an individual data type that leads to the best prediction performance across a large set of cancers. Interestingly, they also show that the prediction performance of GE is highly correlated with tissue types [14]. However, employing tissue types along with GE as model input so far did not lead to performance improvements [40, 46, 50, 148]. In a study on myeloma, GE did not provide satisfactory results [149]. When considering a broader set of cancer types by themselves, genomic features appear to be the most informative single data type [14].

Simple linear regression methods are easily interpretable and have been used for drug-specific biomarker detection [16]; however, they cannot capture more complex associations in the data. On the other hand, the complexity of the output of many more advanced DRP ML methods—for instance, those based on many different transcripts—limits their use for the development of drug resistance biomarkers, and the extraction of new knowledge about the molecular mechanisms behind drug responses. We expect that ongoing improvements in the interpretation of ANNs [150, 151] will continue to be reflected in new approaches to DRP [65, 67].

DRP ML models can be categorized as single-drug and multi-drug models. MDLs can leverage data from multiple drugs, thus conferring robustness and the ability to predict responses to new

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbab408/6383559 by BIUS Jussieu user on 13 October 2021

drugs. MDL models have also been seen empirically to show better prediction performance than SDLs [94, 116, 126, 129], and to lead to more robust biomarkers [126]. Overall, RS and deep learning have been the most widely MDL methods used for DRP. The results reported in the literature suggest the superiority of RSs [24, 114, 116, 123, 124, 126, 129, 130] and deep learning methods [93, 96, 101, 104, 107] over more traditional approaches, such as linear regression, RF and SVM.

At this stage, statements about the current performance of ML-based DRP methods overall are difficult to make. In the clinic, molecular patient data are already an established input for treatment decisions [152], and the potential of ML-based DRP models to predict tumoral responses to anticancer drugs has been established [30, 44, 46, 58, 67, 71, 101]. However, ML-based DRP seems to be rarely used in clinical practice.

One major reason for the currently limited utility of DRP in the clinic may be the difficulty of addressing random and systematic variations between datasets, which is an issue in many applications of ML. Another reason may be the lack of interpretability of many DRP models in the context of other rich, and often unstructured, relevant additional clinical information that is currently not used by DRP models.

The future of drug prioritization in the clinic will undoubtedly be influenced by ML [153]. We believe that future improvements—especially, towards the use of DRP models in the clinic—will be catalyzed in a few distinct ways. Firstly, it can be expected that, as in other ML applications, larger systematic datasets will be able to drive performance increases, especially if new datasets are aimed at the more comprehensive characterization of the relevant biology. This is particularly true if one considers the current paucity of systematic data outside of the cell line context. Future datasets with the largest positive impact on DRP performance may therefore well come from advanced preclinical models. Although more, and more diverse, data can be thought to help in any ML approach, it is in particular deep learning, as a maximally flexible class of models, that is fundamentally capable of capturing the more subtle characteristics of larger datasets. The increasing availability of computing power—and computing as a part of the biomedical research enterprise—will likely enable the continued application of new, and computationally ambitious, ML technology to DRP.

A second, complementary, route towards higher performance of DRP models is the incorporation of prior knowledge. Until now, the incorporation of prior knowledge through biomolecular networks has led to only slight overall improvements at best [44, 49, 60]. However, Oskooei *et al.* [84] has found clear improvements for specific drugs, when additional information about drug targets was provided.

Various missing aspects of drug responses in the clinic may be addressable by either large datasets, or the incorporation of knowledge (or a combination of both), as long as various current systematic limitations of data and models are addressed. A first, major, systematic limitation of most DRP models is their focus on the short-term drug response in homogeneous cell populations. Neither the available cell line input data nor the learning approaches themselves currently address the complex evolutionary process that often leads to drug resistance in heterogeneous tumours [154]. Further development of clinically relevant machine learning-based DRP models may require the use of input data, such as genomic and transcriptomic single-cell data, that resolve the subclonal tumour structure [155], and explicit modelling of tumour evolution [156].

Another aspect of biological complexity that is typically not addressed by current data-driven approaches to DRP, or data, is the immune system, which plays a major part in many diseases, including cancer [157]. The state of the immune system in relevant tissue can be highly associated with chemotherapy resistance [158], and immunomodulators have, in the last decade, become major weapons themselves against cancer and other diseases [159]. The dbGap database [160] covers immune responses to drugs with RECIST parameters, and thus provides future opportunities for the incorporation of the immune system into DRP.

The microbiome is yet another complex part of human physiology known to regulate many aspects of human health, and, increasingly, it is also recognized as a mediator of drug effects [161, 162]. Evolving insight into the microbiome and its role in drug action may stimulate the adoption of the microbiome as part of DRP models [163], or lead to models in which the microbiome can then either act as an enhancer, or even become itself an active part of therapy through the molecules it secretes [164]. Related datasets, such as TCGA [25] microbiome data have, so far, not been included in DRP approaches.

Another area of potential future significance for machine learning-based DRP is the prediction of unwanted side effects or adverse drug reactions. Although methods for the prediction of more general adverse patient events currently appear to focus on electronic health records [165], increasing insight into the molecular basis of adverse drug reactions [166] and associated data may also permit the future integration with principles used today for DRP.

Eventually, DRP methods and their underlying principles may also be applicable to the study of environmental- or food-related molecules on human health. Perhaps spurred by the human genome as the exhaustive parts list of human biomolecules, environmental science has started to adopt comprehensive approaches to profile exogenous chemicals, and to study their impact on human health [167].

As the effects of environmental chemicals on human biology can, at some level, be considered as similar to drug effects, DRP principles and methods to predict the impact of drug combinations on human health may be transposable to the prediction of the effects of chemical mixtures, where, similarly to drug combinations, exhaustive testing of chemical mixtures is typically not possible. Analogously, the study of interactions between drugs and environmental and food-related chemicals [168] will have to rely on mechanistic insight and data-driven models increasingly obtained with the help of machine learning methods, such as those developed for DRP.

Box I. Common Preclinical drug response metrics

Several common drug sensitivity metrics are used in DRP input data. All are obtained from sigmoidal dose–response curves that record cell viability (the activity of a specific biological process, or the number of cells) across a range of drug concentrations. Drug sensitivity is commonly determined relative to two reference values: *control* (i.e. cell viability in the absence of the drug) and maximal *response* (the maximally observed difference of cell viability, relative to control). Based on these two concepts, pharmacogenomic databases use a variety of metrics to quantify growth inhibition [169, 170], the most important being:

- IC50 (inhibition concentration of 50%): the drug concentration at which cell viability is reduced to the half of control.

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbab408/6383559 by BIUS Jussieu user on 13 October 2021

- GI50 (growth inhibition of 50%): the drug concentration that reduces growth by half.
- EC50 (effective concentration of 50%): the drug concentration that causes the half of the maximal response. Also called 'relative IC50'.
- AUC (area under the dose–response curve): the area under the dose–response curve.
- AA (Activity *area*): the area over the dose–response curve.

It is worth noting that the cell lines that are sensitive to a drug will get lower IC50, GI50, EC50 and AUC values, and higher AA value.

Key Points

- Biological knowledge can be leveraged for machine learning of drug response prediction, e.g. through biomolecular networks.
- Recent advances in machine learning, e.g. in the areas of deep learning and multitask learning, have translated into increased performance of drug response prediction.
- Large-scale pharmacogenomic datasets for advanced preclinical models are urgently needed for the future development of data-driven drug response prediction.

Acknowledgements

We thank the anonymous reviewers and Krister Wennerberg for helpful comments on a previous version of this article.

Funding

The European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie (grant agreement No. 860895 for TranSYS B.Y., B.S.), the European Union's Horizon 2020 research and innovation programme (grant agreement No. 667403 for HERCULES F.F., B.Y., B.S. and grant agreement No. 965193 for DECIDER B.S.), and ANR (PARIS), under the frame of ERA PerMed (B.S.).

References

1. Troyanskaya O, Trajanoski Z, Carpenter A, *et al*. Artificial intelligence and cancer. *Nature Cancer* 2020;**1**:149–52.
2. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. *Nat Rev Genet* 2015;**16**:321–32.
3. Ely S. Personalized medicine: individualized care of cancer patients. *Transl Res* 2009;**154**:303–8.
4. van't Veer LJ, Bernards R. Enabling personalised cancer medicine through analysis of gene-expression patterns. *Nature* 2008;**452**:564–70.
5. Bode AM, Dong Z. Recent advances in precision oncology research. *NPJ Precis Oncol* 2018;**2**:11.
6. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Cancer* 2006;**6**:813–23.
7. Weeber F, Ooft SN, Dijkstra KK, *et al*. Tumor organoids as a pre-clinical cancer model for drug discovery. *Cell Chem Biol* 2017;**24**:1092–100.
8. Tentler JJ, Tan AC, Weekes CD, *et al*. Patient-derived tumour xenografts as models for oncology drug development. *Nat Rev Clin Oncol* 2012;**9**:338–50.
9. Griffith LG, Swartz MA. Capturing complex 3D tissue physiology in vitro. *Nat Rev Mol Cell Biol* 2006;**7**:211–24.
10. Aboulkheyr Es H, Montazeri L, Aref AR, *et al*. Personalized cancer medicine: an organoid approach. *Trends Biotechnol* 2018;**36**:358–71.
11. Rosfjord E, Lucas J, Li G, *et al*. Advances in patient-derived tumor xenografts: from target identification to predicting clinical response rates in oncology. *Biochem Pharmacol* 2014;**91**:135–43.
12. Weinstein JN, Myers TG, O'Connor PM, *et al*. An information-intensive approach to the molecular pharmacology of cancer. *Science* 1997;**275**:343–9.
13. Weinstein JN. Cell lines battle cancer. *Nature* 2012;**483**:544–5.
14. Iorio F, Knijnenburg TA, Vis DJ, *et al*. A landscape of pharmacogenomic interactions in cancer. *Cell* 2016;**166**:740–54.
15. Garnett MJ, Edelman EJ, Heidorn SJ, *et al*. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 2012;**483**:570–5.
16. Barretina J, Caponigro G, Stransky N, *et al*. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;**483**:603–7.
17. Seashore-Ludlow B, Rees MG, Cheah JH, *et al*. Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discov* 2015;**5**:1210–23.
18. Haverty PM, Lin E, Tan J, *et al*. Reproducible pharmacogenomic profiling of cancer cell line panels. *Nature* 2016;**533**:333–7.
19. De Niz C, Rahman R, Zhao X, *et al*. Algorithms for drug sensitivity prediction. *Algorithms* 2016;**9**:77.
20. Adam G, Rampášek L, Safikhani Z, *et al*. Machine learning approaches to drug response prediction: challenges and recent progress. *NPJ Precis Oncol* 2020;**4**:19.
21. Ali M, Aittokallio T. Machine learning and feature selection for drug response prediction in precision oncology applications. *Biophys Rev* 2019;**11**:31–9.
22. Costello JC, Heiser LM, Georgii E, *et al*. A community effort to assess and improve drug sensitivity prediction algorithms. *Nat Biotechnol* 2014;**32**:1202–12.
23. GüvençPaltun B, Mamitsuka H, Kaski S. Improving drug response prediction by integrating multiple data sources: matrix factorisation, kernel and network-based approaches. *Brief Bioinform* 2021;**22**:346–59.
24. Chen J, Zhang L. A survey and systematic assessment of computational methods for drug response prediction. *Brief Bioinform* 2021;**22**:232–46.
25. Weinstein JN, Collisson EA, Mills GB, *et al*. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 2013;**45**:1113–20.
26. Eisenhauer EA, Therasse P, Bogaerts J, *et al*. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;**45**:228–47.
27. Solit DB, Garraway LA, Pratilas CA, *et al*. BRAF mutation predicts sensitivity to MEK inhibition. *Nature* 2006;**439**:358–62.
28. Haibe-Kains B, El-Hachem N, Birkbak NJ, *et al*. Inconsistency in large pharmacogenomic studies. *Nature* 2013;**504**:389–93.
29. Safikhani Z, Smirnov P, Freeman M, *et al*. Revisiting inconsistency in large pharmacogenomic studies. *F1000Res* 2016;**5**:2333.

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbab408/6383559 by BIUS Jussieu user on 13 October 2021

30. Geeleher P, Cox NJ, Huang RS. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biol* 2014;**15**:R47.
31. Stransky N, Ghandi M, Kryukov GV. Pharmacogenomic agreement between two cancer cell line datasets. *Nature* 2015;**528**:84–7.
32. Bouhaddou M, DiStefano MS, Riesel EA, *et al*. Drug response consistency in CCLE and CGP. *Nature* 2016;**540**:E9–10.
33. Mpindi JP, Yadav B, Östling P, *et al*. Consistency in drug response profiling. *Nature* 2016;**540**:E5–6.
34. Smirnov P, Kofia V, Maru A, *et al*. PharmacoDB: an integrative database for mining in vitro anticancer drug screening studies. *Nucleic Acids Res* 2018;**46**:D994–1002.
35. Li B, Shin H, Gulbekyan G, *et al*. Development of a drug-response modeling framework to identify cell line derived translational biomarkers that can predict treatment outcome to Erlotinib or Sorafenib. *PLoS One* 2015;**10**:e0130700.
36. Brubaker D, Difeo A, Chen Y, *et al*. Drug intervention response predictions with PARADIGM (DIRPP) identifies drug resistant cancer cell lines and pathway mechanisms of resistance. *Pac Symp Biocomput* 2014;125–35.
37. Dong Z, Zhang N, Li C, *et al*. Anticancer drug sensitivity prediction in cell lines from baseline gene expression through recursive feature selection. *BMC Cancer* 2015;**15**:489.
38. Su R, Liu X, Wei L, *et al*. Deep-Resp-Forest: a deep forest model to predict anti-cancer drug response. *Methods* 2019;**166**:91–102.
39. Webb A, Copsey K. *Statistical Pattern Recognition*. John Wiley & Sons 2011; 3rd ed.
40. Koras K, Juraeva D, Kreis J, *et al*. Feature selection strategies for drug sensitivity prediction. *Sci Rep* 2020;**10**:9377.
41. Zou H, Hastie T. Regularisation and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol* 2005;**67**:301–20.
42. Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning data mining, inference, and prediction*, New York, NY: Springer Science & Business Media, 2009.
43. Hoerl A, Kennerd R. *Ridge regression: biased estimation for nonorthogonal problems Technometrics* , 1970;**12**:55–67.
44. Geeleher P, Zhang Z, Wang F, *et al*. Discovering novel pharmacogenomic biomarkers by imputing drug response in cancer patients from large genomics studies. *Genome Res* 2017;**27**:1743–51.
45. Tibshirani R. *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society: Series B (Methodological) 1996;**58**:267–88.
46. Huang E, Bhope A, Lim J, *et al*. Tissue-guided LASSO for prediction of clinical drug response using preclinical samples. *PLoS Comput Biol* 2020;**16**:e1007607.
47. Papillon-Cavanagh S, De Jay N, Hachem N, *et al*. Comparison and validation of genomic predictors for anticancer drug sensitivity. *J Am Med Inform Assoc* 2013;**20**:597–602.
48. Falgreen S, Dybkær K, Young KH, *et al*. Predicting response to multidrug regimens in cancer patients using cell line experiments and regularised regression models. *BMC Cancer* 2015;**15**:235.
49. Covell DG. Data mining approaches for genomic biomarker development: applications using drug screening data from the Cancer Genome Project and the Cancer Cell Line Encyclopedia. *PLoS One* 2015;**10**:e0127433.
50. 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**:i413–20.
51. Sokolov A, Carlin DE, Paull EO, *et al*. Pathway-based genomics prediction using generalized elastic net. *PLoS Comput Biol* 2016;**12**:e1004790.
52. Cerami EG, Gross BE, Demir E, *et al*. Pathway commons, a web resource for biological pathway data. *Nucleic Acids Res* 2011;**39**:D685–90.
53. Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemometrics Intellig Lab Syst* 2001;**58**:109–30.
54. Niepel M, Hafner M, Pace EA, *et al*. Profiles of Basal and stimulated receptor signaling networks predict drug response in breast cancer lines. *Sci Signal* 2013;**6**:ra84.
55. Ding Z, Zu S, Gu J. Evaluating the molecule-based prediction of clinical drug responses in cancer. *Bioinformatics* 2016;**32**:2891–5.
56. Ding MQ, Chen L, Cooper GF, *et al*. Precision oncology beyond targeted therapy: combining omics data with machine learning matches the majority of cancer cells to effective therapeutics. *Mol Cancer Res* 2018;**16**:269–78.
57. Haykin SO. *Neural Networks and Learning Machines* 3rd edn. New Jersey: Pearson Education, 2009.
58. Huang H-H, Dai J-G, Liang Y. Clinical drug response prediction by using a Lq penalized network-constrained logistic regression method. *Cell Physiol Biochem* 2018;**51**: 2073–84.
59. Bishop CM. *Pattern Recognition and Machine Learning*, 2nd edn. New York, NY: Springer Science, 2006.
60. Drucker H, Burges C, Kaufman L, *et al*. Support vector regression machines. *Advances in neural information processing systems* 1997;**9**:155–61.
61. Gupta S, Chaudhary K, Kumar R, *et al*. Prioritization of anticancer drugs against a cancer using genomic features of cancer cells: a step towards personalised medicine. *Sci Rep* 2016;**6**:23857.
62. Parca L, Pepe G, Pietrosanto M, *et al*. Modeling cancer drug response through drug-specific informative genes. *Sci Rep* 2019;**9**:15222.
63. Kurilov R, Haibe-Kains B, Brors B. Assessment of modelling strategies for drug response prediction in cell lines and xenografts. *Sci Rep* 2020;**10**:2849.
64. Li Y, Umbach DM, Krahn JM, *et al*. Predicting tumor response to drugs based on gene-expression biomarkers of sensitivity learned from cancer cell lines. *BMC Genomics* 2021;**22**:272.
65. Ma J, Fong SH, Luo Y, *et al*. Few-shot learning creates predictive models of drug response that translate from high-throughput screens to individual patients. *Nature Cancer* 2021;**2**:233–44.
66. Baptista D, Ferreira PG, Rocha M. Deep learning for drug response prediction in cancer. *Brief Bioinform* 2021;**22**:360–79.
67. Sakellaropoulos T, Vougas K, Narang S, *et al*. A deep learning framework for predicting response to therapy in cancer. *Cell Rep* 2019;**29**:3367–73.
68. Malik V, Kalakoti Y, Sundar D. Deep learning assisted multi-omics integration for survival and drug-response prediction in breast cancer. *BMC Genomics* 2021;**22**:214.
69. Patterson J, Gibson A. *Deep Learning: A Practitioner's Approach*. California: O'Reilly, 2017.
70. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge, MA: MIT Press, 2016.
71. Sharifi-Noghabi H, Zolotareva O, Collins CC, *et al*. MOLI: multi-omics late integration with deep neural networks for drug response prediction. *Bioinformatics* 2019;**35**:i501–9.

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbab408/6383559 by BIUS Jussieu user on 13 October 2021

72. Hamilton WL. *Graph Representation Learning*. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 2020;**14**:1–159.
73. Wu Z, Pan S, Chen F, *et al*. A comprehensive survey on graph neural networks. *IEEE Trans Neural Netw Learn Syst* 2020;**32**:4–24.
74. Ahmed KT, Park S, Jiang Q, *et al*. Network-based drug sensitivity prediction. *BMC Med Genomics* 2020;**13**:193.
75. Stanfield Z, Coşkun M, Koyutürk M. Drug response prediction as a link prediction problem. *Sci Rep* 2017;**7**:40321.
76. Oughtred R, Stark C, Breitkreutz B-J, *et al*. The BioGRID interaction database: 2019 update. *Nucleic Acids Res* 2019;**47**:D529–41.
77. Kim S, Sundaresan V, Zhou L, *et al*. Integrating domain specific knowledge and network analysis to predict drug sensitivity of cancer cell lines. *PLoS One* 2016;**11**:e0162173.
78. Sagi O, Rokach L. Ensemble learning: a survey. *Wiley Interdiscip Rev Data Min Knowl Discov* 2018;**8**:e1249.
79. Breiman L. Random Forests. *Mach Learn* 2001;**45**:5–32.
80. Riddick G, Song H, Ahn S, *et al*. Predicting in vitro drug sensitivity using Random Forests. *Bioinformatics* 2011;**27**: 220–4.
81. Daemen A, Griffith OL, Heiser LM, *et al*. Modeling precision treatment of breast cancer. *Genome Biol* 2013;**14**:R110.
82. Stetson LC, Pearl T, Chen Y, *et al*. Erratum: computational identification of multi-omic correlates of anticancer therapeutic response. *BMC Genomics* 2015;**16**:481.
83. Wan Q, Pal R. An ensemble based top performing approach for NCI-DREAM drug sensitivity prediction challenge. *PLoS One* 2014;**9**:e101183.
84. Xu X, Gu H, Wang Y, *et al*. Autoencoder based feature selection method for classification of anticancer drug response. *Front Genet* 2019;**10**:233.
85. Kursa MB, Rudnicki WR, Others. Feature selection with the Boruta package. *J Stat Softw* 2010;**36**:1–13.
86. Oskooei A, Manica M, Mathis R, *et al*. Network-based biased tree ensembles (NetBiTE) for drug sensitivity prediction and drug sensitivity biomarker identification in cancer. *Sci Rep* 2019;**9**:15918.
87. Szklarczyk D, Gable AL, Lyon D, *et al*. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019;**47**:D607–13.
88. Cowen L, Ideker T, Raphael BJ, *et al*. Network propagation: a universal amplifier of genetic associations. *Nat Rev Genet* 2017;**18**:551–62.
89. Zhou Z-H, Feng J. *Deep Forest: Towards an Alternative to Deep Neural Networks*, Conference on Artificial Intelligence, 2017.
90. Fang Y, Xu P, Yang J, *et al*. A quantile regression forest based method to predict drug response and assess prediction reliability. *PLoS One* 2018;**13**:e0205155.
91. 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**:i359–68.
92. Yuan H, Paskov I, Paskov H, *et al*. Multitask learning improves prediction of cancer drug sensitivity. *Sci Rep* 2016;**6**:31619.
93. Chiu Y-C, Chen H-IH, Zhang T, *et al*. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Med Genomics* 2019;**12**:18.
94. Menden MP, Iorio F, Garnett M, *et al*. Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS One* 2013;**8**:e61318.
95. Li M, Wang Y, Zheng R, *et al*. DeepDSC: a deep learning method to predict drug sensitivity of cancer cell lines. *IEEE/ACM Trans Comput Biol Bioinform* 2019;**18**:575–82.
96. Choi J, Park S, Ahn J. RefDNN: a reference drug based neural network for more accurate prediction of anticancer drug resistance. *Sci Rep* 2020;**10**:1861.
97. Tang Y-C, Gottlieb A. Explainable drug sensitivity prediction through cancer pathway enrichment. *Sci Rep* 2021;**11**:3128.
98. Zhu Y, Brettin T, Evrard YA, *et al*. Ensemble transfer learning for the prediction of anti-cancer drug response. *Sci Rep* 2020;**10**:18040.
99. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;**521**:436–44.
100. Ching T, Himmelstein DS, Beaulieu-Jones BK, *et al*. Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface* 2018;**15**:20170387.
101. Joo M, Park A, Kim K, *et al*. A deep learning model for cell growth inhibition IC50 prediction and its application for gastric cancer patients. *Int J Mol Sci* 2019;**20**:6276.
102. Chang Y, Park H, Yang H-J, *et al*. Cancer drug response profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature. *Sci Rep* 2018;**8**:8857.
103. Liu P, Li H, Li S, *et al*. Improving prediction of phenotypic drug response on cancer cell lines using deep convolutional network. *BMC Bioinformatics* 2019;**20**:408.
104. Bazgir O, Zhang R, Dhruba SR, *et al*. Representation of features as images with neighborhood dependencies for compatibility with convolutional neural networks. *Nat Commun* 2020;**11**:4391.
105. Duvenaud D, Maclaurin D, Aguilera-Iparraguirre J, *et al*. Convolutional networks on graphs for learning molecular fingerprints. 2015. arXiv:1509.09292
106. Kearnes S, McCloskey K, Berndl M, *et al*. Molecular graph convolutions: moving beyond fingerprints. *J Comput Aided Mol Des* 2016;**30**:595–608.
107. Liu Q, Hu Z, Jiang R, *et al*. DeepCDR: a hybrid graph convolutional network for predicting cancer drug response. *Bioinformatics* 2020;**36**:i911–8.
108. Nguyen T-T, Nguyen GTT, Nguyen T, *et al*. Graph convolutional networks for drug response prediction. *IEEE/ACM Trans Comput Biol Bioinform* 2021.
109. Manica M, Oskooei A, Born J, *et al*. Toward explainable anticancer compound sensitivity prediction via multimodal attention-based convolutional encoders. *Mol Pharm* 2019;**16**:4797–806.
110. Oskooei A, Born J, Manica M, *et al*. PaccMann: prediction of anticancer compound sensitivity with multi-modal attention-based neural networks. 2018. *arXiv:1811.06802*.
111. Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Trans Knowl Data Eng* 2005;**17**:734–49.
112. Koren Y, Volinsky C. Matrix factorization techniques for recommender systems. *Computer* 2009;**42**:30–7.
113. Aggarwal CC. *Recommender Systems*. New Jersey: Pearson Education, 2016.
114. 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**: e1004498.
115. Le D-H, Pham V-H. Drug response prediction by globally capturing drug and cell line information in a heterogeneous network. *J Mol Biol* 2018;**430**:2993–3004.

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbab408/6383559 by BIUS Jussieu user on 13 October 2021

116. Wei D, Liu C, Zheng X, *et al*. Comprehensive anticancer drug response prediction based on a simple cell line-drug complex network model. *BMC Bioinformatics* 2019;**20**:44.
117. Sheng J, Li F, Wong STC. Optimal drug prediction from personal genomics profiles. *IEEE J Biomed Health Inform* 2015;**19**:1264–70.
118. Liu H, Zhao Y, Zhang L, *et al*. Anti-cancer drug response prediction using neighbor-based collaborative filtering with global effect removal. *Mol Ther Nucleic Acids* 2018;**13**: 303–11.
119. Zhang L, Chen X, Guan N-N, *et al*. A hybrid interpolation weighted collaborative filtering method for anti-cancer drug response prediction. *Front Pharmacol* 2018;**9**:1017.
120. Ammad-ud-din M, Georgii E, Gönen M, *et al*. Integrative and personalised QSAR analysis in cancer by kernelized Bayesian matrix factorisation. *J Chem Inf Model* 2014;**54**:2347–59.
121. Gönen M. Predicting drug–target interactions from chemical and genomic kernels using Bayesian matrix factorisation. *Bioinformatics* 2012;**28**:2304–10.
122. Ammad-Ud-Din M, Khan SA, Malani D, *et al*. Drug response prediction by inferring pathway-response associations with kernelized Bayesian matrix factorisation. *Bioinformatics* 2016;**32**:i455–63.
123. Wang L, Li X, Zhang L, *et al*. Improved anticancer drug response prediction in cell lines using matrix factorisation with similarity regularisation. *BMC Cancer* 2017;**17**:513.
124. Moughari FA, Eslahchi C. Author correction: ADRML: anticancer drug response prediction using manifold learning. *Sci Rep* 2020;**10**:22360.
125. Guan N-N, Zhao Y, Wang C-C, *et al*. Anticancer drug response prediction in cell lines using weighted graph regularized matrix factorization. *Mol Ther Nucleic Acids* 2019;**17**:164–74.
126. Suphavilai C, Bertrand D, Nagarajan N. Predicting cancer drug response using a recommender system. *Bioinformatics* 2018;**34**:3907–14.
127. Yang M, Simm J, Lam CC, *et al*. Linking drug target and pathway activation for effective therapy using multi-task learning. *Sci Rep* 2018;**8**:8322.
128. Simm J, Arany A, Zakeri P, *et al*. Macau: scalable Bayesian factorisation with high-dimensional side information using MCMC. In: *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*. 2017, 1–6.
129. Emdadi A, Eslahchi C. DSPLMF: a method for cancer drug sensitivity prediction using a novel regularisation approach in logistic matrix factorization. *Front Genet* 2020;**11**:75.
130. Emdadi A, Eslahchi C. Auto-HMM-LMF: feature selection based method for prediction of drug response via autoencoder and hidden Markov model. *BMC Bioinformatics* 2021;**22**:33.
131. Hamilton WL, Ying R, Leskovec J. Inductive representation learning on large graphs. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017.
132. Yang J, Li A, Li Y, *et al*. A novel approach for drug response prediction in cancer cell lines via network representation learning. *Bioinformatics* 2019;**35**:1527–35.
133. Yu L, Zhou D, Gao L, *et al*. Prediction of drug response in multilayer networks based on fusion of multiomics data. *Methods* 2020;**192**:85–92.
134. Cho H, Berger B, Peng J. Diffusion component analysis: unraveling functional topology in biological networks. *Res Comput Mol Biol* 2015;**9029**:62–4.
135. 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**:3355.
136. Wang W, Yang S, Zhang X, *et al*. Drug repositioning by integrating target information through a heterogeneous network model. *Bioinformatics* 2014;**30**:2923–30.
137. Matlock K, De Niz C, Rahman R, *et al*. Investigation of model stacking for drug sensitivity prediction. *BMC Bioinformatics* 2018;**19**:71.
138. Liu C, Wei D, Xiang J, *et al*. An improved anticancer drug-response prediction based on an ensemble method integrating matrix completion and ridge regression. *Mol Ther Nucleic Acids* 2020;**21**:676–86.
139. Sharma A, Rani R. Drug sensitivity prediction framework using ensemble and multi-task learning. *Int J Mach Learn Cybern* 2019;**11**:1231–40.
140. Gönen M, Margolin AA. Drug susceptibility prediction against a panel of drugs using kernelized Bayesian multitask learning. *Bioinformatics* 2014;**30**:i556–63.
141. Neto EC, Jang INS, Friend SH, *et al*. The Stream algorithm: computationally efficient ridge-regression via Bayesian model averaging, and applications to pharmacogenomic prediction of cancer cell line sensitivity. *Biocomputing* 2014;27–38.
142. Argyriou A, Evgeniou T, Pontil M. Convex multi-task feature learning. *Machine Learning* 2008;**73**:243–72.
143. Su R, Liu X, Xiao G, *et al*. Meta-GDBP: a high-level stacked regression model to improve anticancer drug response prediction. *Brief Bioinform* 2020;**21**:996–1005.
144. Ashburner M, Ball CA, Blake JA, *et al*. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;**25**:25–9.
145. Kanehisa M, Furumichi M, Tanabe M, *et al*. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017;**45**:D353–61.
146. Olson DL, Delen D. *Advanced Data Mining Techniques*. Berlin, Heidelberg: Springer, 2008.
147. Chawla NV. Data mining for imbalanced datasets: an overview. Boston, MA: Springer, 2005.
148. Jang IS, Neto EC, Guinney J, *et al*. Systematic assessment of analytical methods for drug sensitivity prediction from cancer cell line data. *Pac Symp Biocomput* 2014; 63–74.
149. Amin SB, Yip W-K, Minvielle S, *et al*. Gene expression profile alone is inadequate in predicting complete response in multiple myeloma. *Leukemia* 2014;**28**:2229–34.
150. Montavon G, Lapuschkin S, Binder A, *et al*. Explaining nonlinear classification decisions with deep Taylor decomposition. *Pattern Recognition* 2017;**65**:211–22.
151. Ribeiro MT, Singh S, Guestrin C. "why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 2016.
152. Luchini C, Lawlor RT, Milella M, *et al*. Molecular tumor boards in clinical practice. *Trends Cancer* 2020;**6**: 738–44.
153. Seyhan AA. Lost in translation: the valley of death across preclinical and clinical divide – identification of problems and overcoming obstacles. *Translational Medicine Communications* 2019;**4**:18.
154. McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell* 2017;**168**:613–28.

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbab408/6383559 by BIUS Jussieu user on 13 October 2021

155. Lee SH, Hu W, Matulay JT, *et al*. Tumor evolution and drug response in patient-derived organoid models of bladder cancer. *Cell* 2018;**173**:515–28.
156. Yin A, Moes DJAR, van Hasselt JGC, *et al*. A review of mathematical models for tumor dynamics and treatment resistance evolution of solid tumors. *CPT Pharmacometrics Syst Pharmacol* 2019;**8**:720–37.
157. Gonzalez H, Hagerling C, Werb Z. Roles of the immune system in cancer: from tumor initiation to metastatic progression. *Genes Dev* 2018;**32**:1267–84.
158. Vadakekolathu J, Minden MD, Hood T, *et al*. Immune landscapes predict chemotherapy resistance and immunotherapy response in acute myeloid leukemia. *Sci Transl Med* 2020;**12**:1–17.
159. Allard B, Aspeslagh S, Garaud S, *et al*. Immuno-oncology-101: overview of major concepts and translational perspectives. *Semin Cancer Biol* 2018;**52**:1–11.
160. Mailman MD, Feolo M, Jin Y, *et al*. The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet* 2007;**39**:1181–6.
161. Savage N. The complex relationship between drugs and the microbiome. *Nature* 2020;**577**:S10–1.
162. Pryor R, Martinez-Martinez D, Quintaneiro L, *et al*. The role of the microbiome in drug response. *Annu Rev Pharmacol Toxicol* 2020;**60**:417–35.
163. Cammarota G, Ianiro G, Ahern A, *et al*. Gut microbiome, big data and machine learning to promote precision medicine for cancer. *Nat Rev Gastroenterol Hepatol* 2020;**17**:635–48.
164. Wong AC, Levy M. New approaches to microbiome-based therapies. *mSystems* 2019;**4**:e00122–19.
165. Tomašev N, Harris N, Baur S, *et al*. Use of deep learning to develop continuous-risk models for adverse event prediction from electronic health records. *Nat Protoc* 2021;**16**:2765–87.
166. Micaglio E, Locati ET, Monasky MM, *et al*. Role of pharmacogenetics in adverse drug reactions: an update towards personalized medicine. *Front Pharmacol* 2021;**12**:651720.
167. Escher BI, Hackermüller J, Polte T, *et al*. From the exposome to mechanistic understanding of chemical-induced adverse effects. *Environ Int* 2017;**99**:97–106.
168. Pristner M, Warth B. Drug-exposome interactions: the next frontier in precision medicine. *Trends Pharmacol Sci* 2020;**41**:994–1005.
169. Brooks EA, Galarza S, Gencoglu MF, *et al*. Applicability of drug response metrics for cancer studies using biomaterials. *Philosophical Transactions of the Royal Society B,* 2019;**374**:20180226.
170. Hafner M, Niepel M, Chung M, *et al*. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. *Nat Methods* 2016;**13**:521–7.

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbab408/6383559 by BIUS Jussieu user on 13 October 2021