Journal of Genetics and Genomics 48 (2021) 540–551

Contents lists available at ScienceDirect

Journal of Genetics and Genomics

Journal homepage: www.journals.elsevier.com/journal-of-genetics-and-genomics/

Review

Large-scale pharmacogenomic studies and drug response prediction for personalized cancer medicine

Fangyoumin Feng [a], Bihan Shen [a], Xiaoqin Mou [a], Yixue Li [a, b], Hong Li [a, *]

[a] CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China
[b] Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 330106, China

ARTICLE INFO

Article history:
Received 22 January 2021
Received in revised form
26 March 2021
Accepted 28 March 2021
Available online 14 April 2021

Keywords:
Pharmacogenomics
Personalized medicine
Drug response
Biomarkers
Deep learning

ABSTRACT

The response rate of most anti-cancer drugs is limited because of the high heterogeneity of cancer and the complex mechanism of drug action. Personalized treatment that stratifies patients into subgroups using molecular biomarkers is promising to improve clinical benefit. With the accumulation of preclinical models and advances in computational approaches of drug response prediction, pharmacogenomics has made great success over the last 20 years and is increasingly used in the clinical practice of personalized cancer medicine. In this article, we first summarize FDA-approved pharmacogenomic biomarkers and large-scale pharmacogenomic studies of preclinical cancer models such as patient-derived cell lines, organoids, and xenografts. Furthermore, we comprehensively review the recent developments of computational methods in drug response prediction, covering network, machine learning, and deep learning technologies and strategies to evaluate immunotherapy response. In the end, we discuss challenges and propose possible solutions for further improvement.

Introduction

Cancer is a highly complex disease, including not only intra-tumor heterogeneity but also inter-tumor heterogeneity. Although cancer treatments such as chemotherapy, targeted therapy, and immunotherapy significantly improve clinical efficacy, a part of patients suffers partial response or no response due to the heterogeneity and complexity of tumor and microenvironment (Sharma et al., 2011). The average response rate for FDA-approved cancer drugs from 2006 to 2018 indication is only around 40%, which varies from 20% to 60% for different patient subpopulations (Chen et al., 2019). Therefore, it is important to identify the associations between molecular features and drug response, discover novel predictive biomarkers, and estimate drug response to guide personalized medicine. Large-scale pharmacogenomic screening of preclinical models has been performed to represent the heterogeneous responses of cancer patients. Accumulation of drug sensitivity and multi-omics data makes it possible to apply computational methods to identify gene-drug associations and make the prediction. Many researchers have put effort into developing computational methods. However, finding predictive biomarkers from numerous genes is still complicated due to the relatively small sample size and the high dimension of data. Predicting drug response is an even more challenging task. Here, we review the clinical applications of pharmacogenomic biomarkers, the data resources for pharmacogenomic screening and molecular omics, the utility of computational approaches to build drug response prediction models to aid in patient stratification, and discuss the challenges in pharmacogenomics development.

FDA-approved pharmacogenomic biomarkers for cancer therapy

The core of personalized cancer treatment is patient stratification based on predictive pharmacogenomic biomarkers. The efficacy of biomarker candidates needs to be evaluated by clinical trials. Basket trials investigate the effect of common biomarkers in a variety of cancers, while umbrella trials enroll patients with one cancer type and test different exploratory biomarkers (Park et al., 2020). We systematically collected and manually curated biomarkers from the list of FDA-approved drugs whose labels include pharmacogenomic information (U.S. Food and Drug Administration, 2020). Up till June

* Corresponding author.
E-mail address: lihong01@sibs.ac.cn (H. Li).

https://doi.org/10.1016/j.jgg.2021.03.007

2020, there are 47 biomarkers associated with the effective responses of 75 anti-cancer drugs or combinations involving 25 cancer types or subtypes (Table 1). These molecular biomarkers can be used as guidelines on clinical medicine decisions with more suitable treatments for patients. Most of the biomarkers are germline or somatic variants, while a few of them are chromosomal abnormalities or protein expression differences. Some biomarkers are unique to one type of cancer, while others are applicable to multiple cancer types. Drugs with the same target genes or mechanism of action tend to share the same biomarkers. Many biomarkers belong to target genes or genes in the action pathways of drugs.

Take gefitinib, a targeted therapy drug for lung cancer, as an example. The effective rate of gefitinib was only 20%–30% by 2002. Till 2005, researchers found that EGFR mutation is a sensitive biomarker for gefitinib. Patients with EGFR mutation are more likely to benefit from gefitinib, with a response rate of around 65% and an overall survival time above 30 months (Han et al., 2005). Microsatellite instability (MSI)/mismatch repair deficiency (dMMR) and tumor mutational burden (TMB) have been approved as response indicators to PD-1 inhibitors such as pembrolizumab (U.S. Food and Drug Administration, 2020). Patients with MSI-high/dMMR or high TMB show significantly higher overall response rates and longer progression-free survival in various solid tumors treated with PD-1 inhibitors (Boyiadzis et al., 2018). In summary, patient stratification based on predictive molecular biomarkers can greatly improve the response rates of drugs for a certain patient subpopulation and promote the development of precision medicine. However, only 75 out of 292 FDA-approved anti-neoplastic agents (derived from DrugBank [Wishart et al., 2018]) have predictive biomarkers in the drug labeling. The finding of new biomarkers for personalized medicine have become an important research direction.

Large-scale pharmacogenomic resources

In addition to FDA-approved pharmacogenomic biomarkers, gene-drug associations as biomarker candidates can be identified from large-scale preclinical pharmacogenomic datasets. Patient-derived cancer cell lines, organoids, and xenografts (PDXs) are commonly used as preclinical cancer models for drug screening, biomarker identification, and drug response prediction (Fig. 1A). Cancer cell lines are a population of cells propagated in two-dimension *in vitro* culture; Organoids are three-dimension cultured models composed of multiple cells to mimic the architectures and functions of original tumors; PDXs are generated by transplanting patient-derived tumor tissues into immunodeficient mice (Ibarrola-Villava et al., 2018). The histology, genomics, transcriptome, and other characteristics of the patient's tumors can be retained in the above cancer models to some extent.

With the development of experimental technologies, the success rates of establishing preclinical cancer models have been improved, and a growing number of large-scale platforms of cancer models have been established (Shoemaker, 2006; Conte et al., 2019; Ghandi et al., 2019; Bock et al., 2021) (Fig. 1B and 1C; Table S2). Each model may be characterized by multiple molecular profiles such as gene expression, somatic mutations, copy number variations (CNV), alternative splicing, miRNA expression, chromatin profile, DNA methylation, and proteome (Fig. 1B). High-throughput drug screening (HTS) enables simultaneously testing the efficacy of dozens of drugs against different cancer models. Drug responses can be measured by quantitative indexes: IC50 (half maximal inhibitory concentration), EC50 (concentration for 50% of maximal effect), and AUC (Area under the dose-response curve) calculated from response-curves of cell lines or organoids; tumor growth inhibition rate calculated by comparing the tumor volume of the treatment group with the control group of PDXs. Combining both drug responses and molecular characteristics makes it possible to identify drug-gene associations and construct computational models to estimate drug response. We will depict the progress of representative pharmacogenomic researches using three kinds of preclinical cancer models, respectively (Fig. 1B and 1C).

Cancer cell lines

Pan-cancer high-throughput drug screens are majorly performed on cancer cell lines. The NCI-60 Human Tumor Cell Lines Screen (NCI60) has been established since 1990, which evaluated growth inhibition for more than 52,000 small molecules (synthetic or purified natural products) on 60 cancer cell lines from nine different tissue types (Shoemaker, 2006). The relatively small sample size may affect the accuracy and reliability when applying NCI60 in pharmacogenomic studies. The Cancer Cell Line Encyclopedia (CCLE) (Ghandi et al., 2019), the Cancer Therapeutics Response Portal (CTRP) (Basu et al., 2013), and the Genomics of Drug Sensitivity in Cancer (GDSC) (Yang et al., 2013) were set up to contain more cell lines and cover diverse tumor types. CCLE provides molecular profiling data of 1457 cancer cell lines covering over 40 cancer types, which contains gene expression profiles from microarray and RNA sequencing methods, gene mutations, copy number variants, DNA methylations, gene fusion calls, structural variation, chromatin profiles and reverse-phase protein array (RPPA) data, as well as short hairpin RNA knockdown and CRISPR-Cas9 knockout data. Moreover, CCLE generated pharmacological profiles for 24 anticancer drugs across 479 cell lines. CTRP V1 quantitatively measured the sensitivity of 242 cell lines to 185 small molecules; CTRP V2 measured the sensitivity of 860 cancer cell lines to 481 small molecules. The molecular data of cell lines in CTRP can be obtained from CCLE. The GDSC collection comprises over 1000 tumor cell lines, ~60% of which are overlapped with CCLE. GDSC V1 dataset contains dose-response results for 367 compounds on 987 diverse cancer cell lines. GDSC V2 provides 198 compounds on 809 cancer cell lines where some experiments from GDSC V1 have been repeated with advanced equipment and procedures.

In addition to the prominent resources discussed above, other research institutes and companies also undertook independent drug screen projects. Institute for Molecular Medicine Finland generated drug response and molecular profiling data covering 106 drugs and 308 cancer cell lines (Mpindi et al., 2016). GlaxoSmithKline performed high-throughput drug screening for 19 drugs on 311 cancer cell lines and analyze the relationship between drug response and oncogenic patterns (Greshock et al., 2010). The Genentech Cell Line Screening Initiative produced pharmacogenomic profiling with 16 drugs and 410 cancer cell lines (Haverty et al., 2016). XenoBase provided drug response data of 11 compounds on 707 cancer cell lines and linked with genetic profiles (Crown Bioscience Inc., 2021).

Pan-cancer pharmacogenomics datasets are helpful to obtain a large sample size and increase statistical power, but they ignore the distinctive molecular features among different cancer types. In order to address this problem, some studies perform drug screening for cell lines from a specific cancer type. Heiser et al. (2012) screened 77 therapeutic compounds on more than fifty breast cancer lines to discover responses mechanism in subtype, pathway, and genomic aberration levels, respectively. Daemen et al. (2013) tested for 138 compounds on 70 breast cancer cells and constructed a machine learning model to identify predictive genetic signatures of therapeutic response. The National Cancer Institute (NCI) also provided cancer-specific pharmacogenomic screening datasets to help identify drug response-related genetic patterns. For example, the small cell lung cancer (SCLC) project generated pharmacogenomic data of 70 SCLC cell lines to 526 compounds (Polley et al., 2016); similarly, the sarcoma project targeted 445 compounds and 64 cell lines (Teicher

Table 1
FDA-approved pharmacogenomic biomarkers in the labeling of anti-cancer drugs (up to 2020).

Cancer Types	Biomarkers	Drugs
ALL	BCR-ABL1	Dasatinib; Imatinib; Ponatinib
ALL (r)	BCR-ABL1 (−)	Vincristine*
AML	CD33 exp	Gemtuzumab Ozogamicin
	FLT3 mut	Midostaurin and chemotherapy
	IDH1 mut	Ivosidenib
	IDH2 mut	Enasidenib
AML (r/r)	FLT3 mut	Gilteritinib
APL	t(15; 17) translocation; PML/RARA exp	Arsenic trioxide; Tretinoin
BRCA	HER (+)	Trastuzumab
BRCA (a/m)	HER (+)	Lapatinib; Trastuzumab; Paclitaxel and Trastuzumab; Tucatinib and Trastuzumab and Capecitabine; Pertuzumab and Trastuzumab and Docetaxel; Neratinib and Capecitabine*
	HR (+)	Fulvestrant; Anastrozole*, Letrozole*
	HR (+), HER2 (−)	Fulvestrant; Fulvestrant and Palbociclib; Fulvestrant and Abemaciclib; Fulvestrant and Ribociclib; Abemaciclib*; Everolimus*; Everolimus and Exemestane*; Verolimus and Exemestane*
	HR (+), HER2 (−), PIK3CA mut	Alpelisib and Fulvestrant
BRCA (e)	HER (+)	Pertuzumab and Trastuzumab and Chemotherapy; Neratinib; Exemestane*
	HR (+)	Letrozole*
BRCA (m)	ER-positive	Tamoxifen
	Germline BRCA mut, HER2 (−)	Olaparib*, Talazoparib*
CEL	FIP1L1-PDGFRα fusion	Imatinib
CHOL (a/m)	FGFR2 fusion; FGFR2 rearrangement	Pemigatinib
CLL	CD20 (+)	Rituximab and fludarabine and cyclophosphamide
CLL; SLL	17p deletion	Ibrutinib*
CML	BCR-ABL1	Bosutinib; Dasatinib; Imatinib; Nilotinib; Ponatinib
CRC (m)	EGFR exp, wt-RAS	Cetuximab; Panitumumab
CRC (m)	MSI-H or dMMR	Ipilimumab and Nivolumab*
CRPC	HRR gene muts	Olaparib*
CTCL (p/r)	CD25 exp	Denileukin diftitox
GEP-NETs	SSTR-positive	Lutetium Dotatate Lu-177
GISTs	KIT exp	Imatinib
GISTs (u/m)	PDGFRA Exon 18 mut	Avapritinib
MTC (m)	RET mut	Selpercatinib
NHL	CD20 (+)	Rituximab
NSCLC (m)	ALK rearrangement	Alectinib; Brigatinib; Ceritinib; Crizotinib; Lorlatinib
	EGFR exon 19 deletions; L858R substitution	Dacomitinib; Erlotinib; Gefitinib; Osimertinib; Ramucirumab*
	EGFR mut	Afatinib
	EGFR T790M mut	Osimertinib
	MET exon 14 skipping	Capmatinib
	PD-L1 exp, no EGFR or ALK aberrations	Pembrolizumab*
	PD-L1 exp, no EGFR or ALK aberrations	Atezolizumab; Ipilimumab and Nivolumab
	ROS1 rearrangement	Entrectinib; Crizotinib*, Lorlatinib*
NSCLC (m); ATC (m)	BRAF V600E mut	Dabrafenib and Trametinib
	RET fusion	Selpercatinib
OV (a)	HRD	Niraparib*; Bevacizumab and Olaparib*
OV; CRPC (m)	BRCA mut	Rucaparib*
OV; PAAD (m)	BRCA mut	Olaparib*
sALCL	CD30 exp	Brentuximab Vedotin
SKCM, (u/m)	BRAF V600E mut	Dabrafenib; Vemurafenib; Encorafenib and Cetuximab
	BRAF V600E or V600K mut	Binimetinib and Encorafenib; Cobimetinib and Vemurafenib; Dabrafenib and Trametinib
Solid Tumors	MSI-H or dMMR; TMB-H	Pembrolizumab*
	NTRK fusion	Larotrectinib; Entrectinib
UCC	FGFR3 or FGFR2 genetic alterations	Erdafitinib
UCC; TNBC (u/m)	PD-L1 exp	Atezolizumab
UCEC (a)	not MSI-H/dMMR	Lenvatinib and Pembrolizumab*

1) Information was manually collected from the US FDA Table of Pharmacogenomic Biomarkers in Drug Labeling (Last Updated: 06/2020). 2) Cancer types full names and abbreviations display in Table S1; cancer status or stages: r/r, relapsed/refractory; m, metastatic; a, advanced; e, early; p/r, persistent/recurrent; u/m, unresectable/metastatic. 3) Biomarkers: semicolon and comma represent logic representations "or" and "and,", representatively; (+) and (−) denote "positive" and "negative"; mut, mutation; exp, expression; MSI-H, microsatellite instability-high; dMMR, mismatch repair deficiency; HRR, homologous recombination repair; HRD, homologous recombination deficiency; TMB-H, tumor mutational burden-high; HER2-positive, HER2 protein overexpression or HER2 gene amplification; wild-type RAS, the status of "absence of a RAS mutation in exon 2 (codons 12 and 13), exon 3 (codons 59 and 61), and exon 4 (codons 117 and 146) of both KRAS and NRAS". 4) Drugs: Targeted genes were collected from Drugbank database; asterisk indicates biomarkers of drugs are not included in their targeted genes; drugs separated by a semicolon; "and" were used to present combinational therapy.

et al., 2015). The Liver Cancer Model Repository provides drug sensitivity data of 90 compounds on 81 liver cancer cell lines to screen drug-gene associations and discover predictive biomarkers (Qiu et al., 2019). Similar large-scale drug response screening has also been performed on cancer such as lung cancer (LC), colon cancer (CRC), head and neck squamous cell carcinoma (HNSCC), glioblastoma (GBM), acute myeloid leukemia (AML), ovarian cancer (OV) and so on (Fig. 1C; Table S2). Moreover, databases such as the connectivity map (CMAP) (Lamb et al., 2006) and the Library of Integrated Network-Based Cellular Signatures (LINCS) (Koleti et al., 2018) have also been set up, which provides gene expression profiles before and after drug treatment, as well as other molecular data.

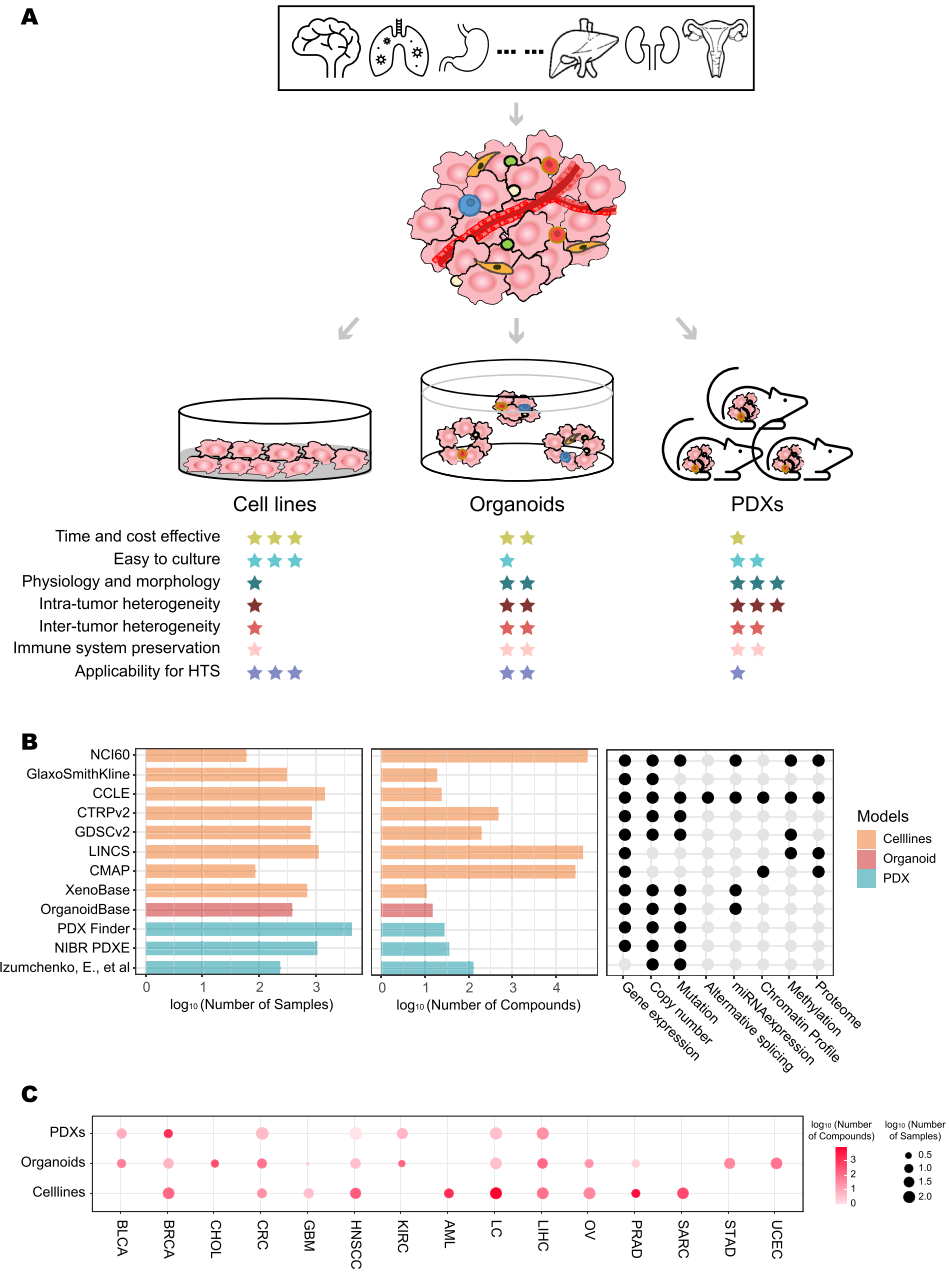

Fig. 1. Preclinical cancer models and the representative pharmacogenomic resources. **A**: Introduction and comparison of three kinds of cancer models (including patient-derived cancer cell lines, organoids, and xenografts). The number of stars indicates strengths (more stars) and limitations (less stars). **B**: Sample size, number of screened compounds, and molecular characteristics of the predominant pan-cancer pharmacogenomic resources that include multiple tumor types and the number of drugs should be above ten. **C**: Large-scale cancer-specific pharmacogenomic datasets are generated from drug sensitivity researches, where one single and common type of cancer is studied, and at least one kind of omics data is available. For the cancer types with more than one data source, the dataset with the highest number of sample-compound pairs is kept. The summary of datasets in (**B**) and (**C**) is provided in Table S2.

Large scaled drug perturbated protein response has also been profiled using RPPA in cancer cell lines and provided a new insight to identify protein-drug association for drug sensitivity prediction and understanding drug mechanism of action (Zhao et al., 2020).

Tumor organoids

The success rate of establishing patient-derived tumor organoids varies from different tumor types, ranging from 15% in prostate cancer to 90% in colorectal cancer (CRC) (Weeber et al., 2017). Drug screening experiments have been executed on tumor organoids to support personalized therapy design. Pauli et al. (2017) established 56 organoids from more than 17 metastatic or primary tumor types as living biobanks, and four of them had been used to prioritize therapeutics in *ex-vivo* drug screening. Although there are few new projects like OrganoidBase (Crown Bioscience Inc., 2021) and the newly launched Organoid Cell Atlas pilot project by Human cell atlas (Bock et al., 2021), unlike large-scaled pharmacogenomic resources of cancer cell lines that covers diverse tumor types, pharmacogenomic studies of organoids are usually more focused on a single tumor type. Wetering et al. (2015) successfully generated 20 matched tumor and healthy organoids from colorectal cancer patients and developed an automatic drug screening assay to test 83 compounds. They well characterized the genomic and transcriptomic profile of organoid

heterogeneity and identified genetic correlations between oncogenic mutations and drug response. Subsequent studies of CRC organoids have revealed more genetic variations that are associated with drug response (Kondo et al., 2019). Similar to CRC, organoid biobanks have been established for gastric cancer (STAD), advanced prostate cancers (PRAD), glioblastoma (GBM), liver cancer (LIHC), bladder cancer (BLCA), biliary tract carcinoma (CHOL), and kidney cancers (KIRC) et al. (Fig. 1C; Table S2). Breast cancer, known to be difficult to construct organoid model owing to possible newly occurred clonal drift, has also been successfully established (Sachs et al., 2018). Meanwhile, these researches have conducted genomic, transcriptomic, epigenetic, as well as phenotypic analyses to prove the ability of these organoids to capture inter-tumor and intra-tumor heterogeneity and restore the genetic landscape of their tumor origins. These biobanks also provide drug HTS data and have been used on pharmacogenomic biomarker identification and drug sensitivity prediction (Liu et al., 2020a).

Patient-derived tumor xenografts

Similar to cancer cell lines, there are several pan-cancer PDX repositories providing accessibility of experimental models, genetic profiles, and drug response data. Mouse Models of Human Cancer Database (MMHCdb) is an expertly curated library of cancer mouse models constructed by the Jackson Laboratory (Woo et al., 2019). There are more than 580,000 mouse models in MMHCdb, 4372 of which are PDX models, and 301 of these PDXs have screened with 23 single drug treatments and nine combinations. MMHCdb also provides mutational status, gene expression, the status of microsatellite instability, as well as histopathology images of these models. The European cancer research institutions have also built a data portal (EurOPDX) for sharing more than 1500 PDX models from various cancer types (Cerami et al., 2012). MMHCdb have collaborated with EurOPDX and other PDX resources to form a searchable platform, PDX Finder, to support the studies of tumor mechanism and drug response (Conte et al., 2019). So far, PDX Finder has provided 4372 PDX models, 942 of which went through treatments or drug dosing experiments. The patient-derived model repository (PDMR) from NCI provides access to more than 3600 PDX models with RNA sequencing and whole-exome sequencing data available (Evrard et al., 2018). Interestingly, some of the PDX models in PDMR have treatment histories of patients, which would be valuable to study the mechanism of drug resistance.

Next, we reviewed the applications of PDX models in individual pharmacogenomic researches. The Novartis Institutes for BioMedical Research PDX encyclopedia (NIBR PDXE) provided 1075 PDX models from more than 16 organ origins (Gao et al., 2015). They also designed PDX clinical trial to measure drug response (involving 36 monotherapies and 26 combination therapies). Together with genetic analysis, researchers have helped to identify predictive biomarkers and reveal clinically related resistance mechanisms. Izumchenko et al. (2017) established 237 PDX models from 1163 patients with various advanced solid tumors and proved the reproductivity of both positive and negative clinical outcomes with the corresponding patients against the same treatments. PDXliver specified on liver cancer (116 PDX models) stored not only gene expression profiles, germline and somatic mutations, copy number variants but also clinical characteristics of patients and sorafenib response data of PDX mice (He et al., 2018). PDX models also have been successfully established for monotherapy or drug combination sensitivity screening on lung squamous cell carcinoma (LUSC), bladder cancer (BLCA), breast cancer (BRCA), HNSCC, KIRC, and CRC (Fig. 1C; Table S2). Although most PDX platforms only have pharmacogenomic data for a small number of drugs, PDX models have shown their potential to select drugs, identify biomarkers and guide patient stratification. For example, the study of HNSCC PDXs identified CCND1 and CDKN2A alterations as predictive biomarkers towards the response of CDK4/6 inhibitors (Karamboulas et al., 2018); another study of liver cancer PDXs found the association between MAP3K1 expression and sorafenib sensitivity (Hu et al., 2020).

Model comparison and improvement

The following describes the comparison of preclinical cancer model properties along with their advantages and limitations. Compared to organoids and PDXs, cell lines have been more widely used in various cancer types because of simpler culture protocols, less cost, and more effective high-throughput screening. The main bottleneck of cancer cell lines is the lack of phenotypic and genetic heterogeneity presented in original tumors. Thus, it is difficult for cell lines to imitate the mixed context of tumor tissue and their response towards drugs.

Compared with cancer cell lines, organoids can better reveal drug sensitivity or resistance-associated genetic patterns closer to what happened in patients (Liu et al., 2020a). With relatively less time and expense than animal models, organoids are more suitable to enlarge sample size to improve reproductivity and reliability of drug discovery (Liu et al., 2020a). Meanwhile, organoids can avoid possible interspecies reactions, which may occur in animal models regarding the measurement of drug response and toxicity, and help further reduce the failure rate on preclinical experiments (Shanks et al., 2009). With special organoid culture systems or co-culture techniques, organoids also show the potential to imitate tumor microenvironment, especially immune systems, which offers applicable models for tumor immunotherapy response prediction and biomarker discovery (Yuki et al., 2020). On the other hand, organoid models also have some shortcomings. Tumor organoids require a complex cultural environment, and the success rate is relatively low for some cancer types (Liu et al., 2020a). Research teams have made efforts to modify traditional HTS assay to fit the organoid culture system, but it is still harder to perform high-throughput drug screening on organoids (Liu et al., 2020a). Furthermore, the therapeutic response of organoids may be influenced by cultural methods. Therefore, the accumulation of drug response data on organoids is much slower.

Engrafted in tumor tissue form, PDX models can retain more morphological characteristics, three-dimensional spatial structure, as well as partial tumor microenvironments components of their patient origins (Goto, 2020). Moreover, PDX models show more consistent genetic and histological patterns with patient cancer tissues, and these patterns are stable while passaging (Goto, 2020). These give PDX models the ability to maintain human tumor biological characteristics while stably enlarge sample size to preserve valuable clinical resources. PDXs have shown closer consistence with clinical responses of patients on the evaluation of drug efficacy, as well as resistance and toxicity, which may effectively promote translational medicine from basic to clinical science (Izumchenko et al., 2017; Woo et al., 2019). Furthermore, PDXs can reflect genetic characteristics and heterogeneity of patients more accurately so that it serves a more solid experimental model to identify genetic biomarkers to guide patient stratification clinically. The limitations of the PDX model cannot be ignored. The establishment of PDX and drug screening are quite expensive and time-consuming. The stromal components have been reported to be replaced by mouse-originate stroma while passaging (Goto, 2020). Therefore, the drug response may get askew as the times of passaging increase.

Another shortcoming of PDX mice is the lack of an immune system. Technologies to "humanize" the mouse immune system have been developed to study the interactions between tumor cells, immune cells, and stromal cells together with their impacts on anticancer treatments (Olson et al., 2018). A genetically engineered

mouse (GEM) is another model used in cancer research. GEMs are established by editing the genome or introducing specific genetic alterations into cells of interest (Sharpless and Depinho, 2006). GEMs develop tumors in a natural microenvironment and provide the most complete representation of cancer progression from initiation to progression. However, high expense, long establishing time and experimental difficulties hinder the use of these new models in pharmacogenomic screening.

Computational methods for drug response prediction

With the accumulation of drug screening and multi-omics data, drug-gene associations and biomarker identification can be addressed by feature selection, and drug response prediction can be abstracted as classification or regression problems (Fig. 2). These computational models are usually evaluated by cross-validation and independent datasets, some of which are also validated by animal experiments or clinical trials. In this session, we review some recent computational methods addressing these tasks.

Feature selection

Different feature selection strategies have been used to eliminate redundant and irrelevant features to identify gene-drug associations and improve the accuracy of drug prediction. Statistical tests such as *t*-test, chi-square test, analysis of variance, and linear discriminant analysis have been used to filter variables based on correlations with outcome variables (Vidyasagar, 2015). More sophisticated methods combine feature selection with prediction models; variables with the best performance are selected as features. Elastic net regression with bootstrap and random forest model has been applied to select features associated with drug responses (Ding et al., 2016; Fang et al., 2018) (Table 2). Another strategy of feature selection is utilizing networks of known biological relationships. ProGENI (Prioritization of Genes Enhanced with Network Information) support vector regression (SVR) model firstly performed network information on gene expression profiles and then used Random Walk with Restart to rank genes (Emad et al., 2017); NETPHIX (NETwork-to-PHenotype association with eXclusivity) identified subnetworks of genes whose

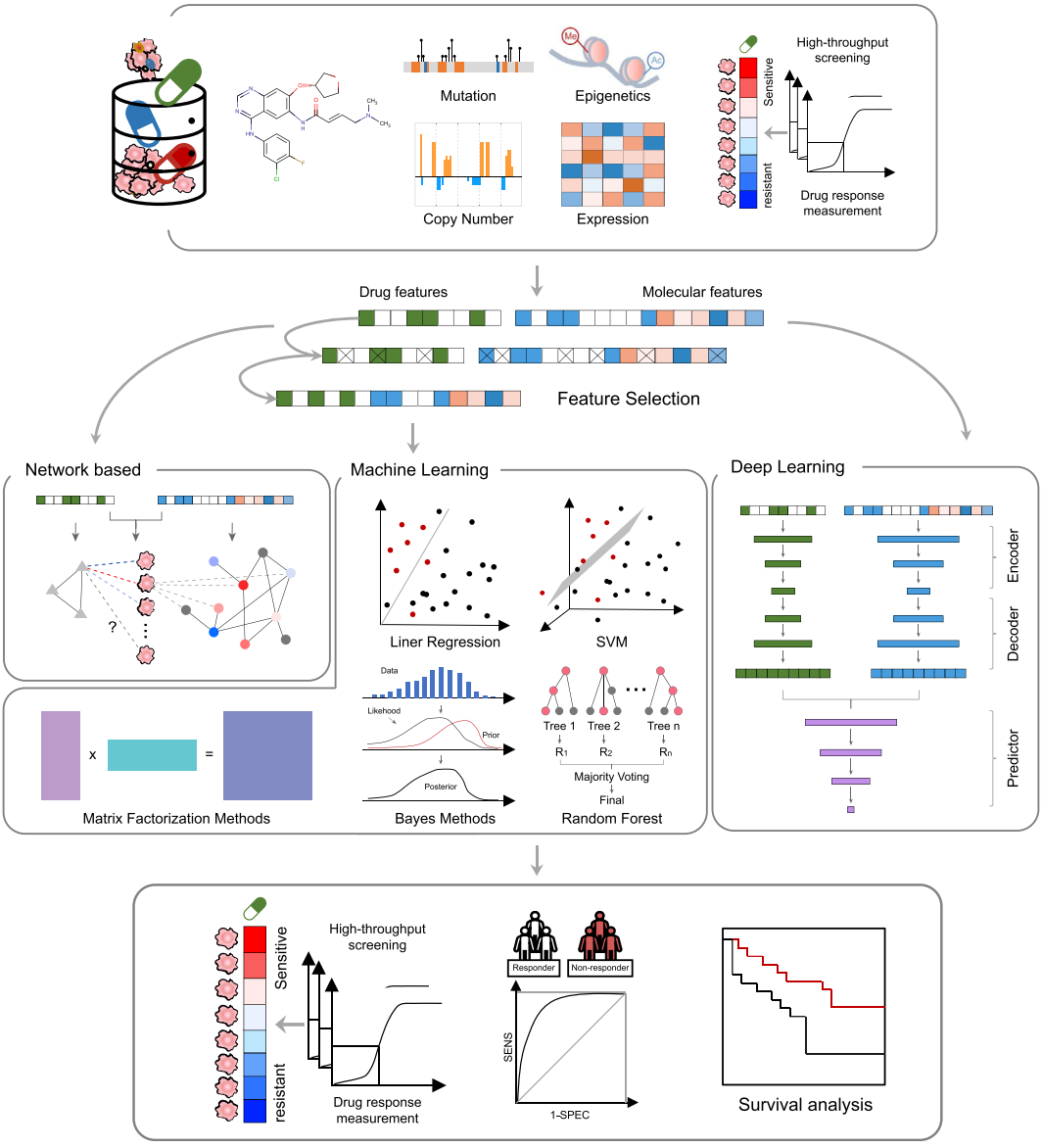

Fig. 2. Workflow of drug response prediction models.

Table 2
Representative computational models for drug response prediction.

Algorithms	Features	Model description	Validation schemes	Data source	Performance	References
Network-based prediction models						
Dual-Layer Network model	T+E	Drug-cell line dual-Layer Network; Linear weighted models	LOOCV	CCLE GDSC	$PCC_{CCLE} \in [0.51,0.88]$ $PCC_{GDSC} > 0.5$	Zhang et al., 2015
GloNetDRP	T+EM	Drug-cell line dual-Layer Network; Random walk with restart	5F-CV	CCLE GDSC	PCC = 0.8 PCC = 0.45	Le and Pham, 2018
HNMDRP	T+E+PPI	Drug-cell-targeted gene networks; Informative flow-based algorithm	LOOCV	GDSC	AUC = 0.87	Zhang et al., 2018
Zachary et al.	M+PPI	Drug-cell-gene sensitive and resistant networks; Difference between drug sensitive profile and resistant profile	LOOCV	GDSC CCLE	AUC = 0.8813 AUC = 0.8474	Stanfield et al., 2017
Machine earning models						
Geeleher et al.	E	Most differentially expressed genes (t-tests) as features; Ridge regression	LOOCV	CGP	$AUC_{docetaxel} = 0.81$ $AUC_{bortezomib} \in (0.63, 0.71)$	Geeleher et al., 2014
Gu et al.	E/C/Me/Mi/R	Univariate logistic regression or elastic net regression for different input feature selection; Ensemble multivariate logistic regression with bootstrap	5F-CV	TCGA	AUC ∈ (0.33, 0.84)	Ding et al., 2016
RFE -SVM	E	Recursive feature elimination (RFE); SVM with linear kernel	LOOCV	NCI60	ACC = 0.841	Huang et al., 2017
ProGENI-SVR	E+PPI	Top 100 highest correlated genes as input features via Prioritization of genes enhanced with network information (ProGENI); Support vector regression with nonlinear kernel	5F-CV	LCL	PCI ∈ [0.45,0.75]	Emad et al., 2017
pairwiseMKL	T+EMC	Multiple pairwise nonlinear kernel combined with ridge regression	10F-CV	GDSC	PCC = 0.858	Cichonska et al., 2018
BMTMKL	EMCRMe	Pathway-based multiple data views; Bayesian multiview multitask multi-kernel (nonlinear) learning	Individual test	GDSC DREAM	$PCI_{DREAM} = 0.583$	Costello et al., 2014
NBSBM	EMCMe+PPI	Disease-specific network; Sparse Bayesian classifier	5F-CV	Cancer specific datasets	$AUC_{PRAD} = 0.942$ $AUC_{BRCA} = 0.737$	Liu et al., 2019b
SRMF	Y+E	Similarity-regularized matrix factorization	10F-CV	GDSC CCLE	$PCC_{GDSC} = 0.71$ (±0.15) $PCC_{CCLE} = 0.78$ (±0.07)	Wang et al., 2017
DSPLMF	T+EMC	Similarity-regularized logistic matrix factorization	10F-CV	GDSC CCLE	$AUC_{GDSC} = 0.76$ $AUC_{CCLE} = 0.776$	Emdadi and Eslahchi, 2020
MC-RR	E	Ensemble of matrix completion and ridge regression	10F-CV	GDSC CCLE	$PCC_{CCLE\text{-}MEAN} = 0.7$	Liu et al., 2020b
QRF	EMC	Random forest; Quantile regression forest model	Out of bag	CCLE	PCC = 0.6	Fang et al., 2018
Deep-Resp-Forest	EC	Multi-info multi-grained scanning (MIMGS), cascade forest with feature optimization (CFFO)	5F-CV	GDSC CCLE	$ACC_{GDSC} = 0.863$ $ACC_{CCLE} = 0.833$	Su et al., 2019
Deep learning models						
DNN model	E	DNN; Interpreted with GSEA analysis on the weights of the first hidden layer	5F-CV	GDSC	AUC = 0.75	Sakellaropoulos et al., 2019
CDRscan	T+M+G	Ensembled five CNN	5F-CV	GDSC	RMSE = 1.069, $R^2 > 0.84$, AUC > 0.98	Chang et al., 2018
MOLI	EMC	3 DNN encoders + 1-layer classifier	5F-CV Individual test	GDSC PDX TCGA	$AUC_{PDX\_TCGA} \in [0.53,0.74]$	Sharifi-Noghabi et al., 2019
tCNNS	Y+MC	2 CNN encoders + DNN	0.8/0.1/0.1	GDSC	PCC = 0.909	Liu et al., 2019a
PaccMann web server	T+E+PPI	Attention-based encoders + DNN; Interpreted with Gene attention weights	25F-CV	CCLE GDSC	RMSE = 0.89, $R^2 = 0.86$	Cadow et al., 2020
DrugCell	T+M+GO	VNN encoder + ANN encoder + 1-layer classifier; Interpreted with Relative Local Improvement in Predictive Power (RLIPP) score	5F-CV	GDSC CTRP	SCC = 0.8	Kuenzi et al., 2020

Abbreviations: (1) "Features" column: T, chemical structure; E, expression profile; M, mutation status; C, CNV; R: RPPA; Me, DNA methylation; Mi, miRNA expression; V, variant genes; DT, drug targeted genes; PPI, protein-protein interaction network; P, biological pathways and gene sets; G, cancer-associated genes; GO, gene ontology. (2) Performance measurements used in the original publications: PCC, Pearson correlation coefficient; AUC, area under ROC curves; SCC, Spearman correlation coefficient; PCI, weighted concordance index developed by NCI-DREAM drug sensitivity challenge team; MSE, mean squared error; RMSE, Root-mean-squared error; R^2, coefficient of determination.

genetic alterations are associated with drug response (Kim et al., 2020); Kong et al. (2020) identified proximal pathways of drug targets by computing average shortest-path lengths between drug targets and pathways in protein-protein interaction network as inputs of machine learning models trained on colorectal and bladder organoid models to obtain predictive biomarkers and tested on patient data.

Network-based prediction models

Network-based models have the capacity to imitate complex systems of cancer at the sample level by integrating high-dimension multimodal omics data and identifying gene-drug associations, which may help reveal the underlying mechanism of action of drugs on specific cancer samples. The commonly used network structure for drug response prediction is the dual-layer integrated cell line-drug network connecting a drug similarity network and a cell line similarity network by known drug-cell line responses (Zhang et al., 2015) (Table 2). Drug similarity can be derived from chemical structure fingerprints; cell line similarity can be defined via gene expression and genetic variants. Based on the assumption that similar cell lines might show similar responses to similar drugs, a linear weighted model based on the known responses of neighboring nodes (Zhang et al., 2015) or a random walk with restart (RWR) algorithm was used to make predictions (Le and Pham, 2018). HNMDRP (Heterogeneous network-based method for drug response prediction) expanded network architecture to three layers by adding protein-protein interaction (PPI) network and drug-target links and used informative flow-based algorithm to estimate drug response (Zhang et al., 2018). Stanfield et al. (2017) also integrated the PPI network, links between drug and cell lines, cell lines, and mutated genes. Their novelty is to use RWR to compute the network profile, a vector representing proximity to genes mutated in cell lines. Each cell line is represented by a cell line profile; each drug has two network profiles representing the cell lines that are sensitive or resistant towards the corresponding drug. The difference between the sensitivity score (the correlation between sensitive profile and cell line profile) and the resistance score (the correlation between resistant profile and cell line profile) were used to predict drug sensitivity.

Machine learning models

Machine learning (ML) is an application of artificial intelligence that can automatically learn patterns from complex observational data by a statistics method. The underlying linear or nonlinear associations between genes and drug response are complex. Therefore, the combination of multiple strategies such as linear regression models with a penalty, kernel-based models like support vector machine (SVM), Bayesian methods, matrix factorization-based methods, as well as ensemble models are commonly integrated for drug response prediction.

Linear regression is one of the basic models, which is fast and easy-to-interpret. In practice, the regularization penalty is often used together to avoid possible overfitting (Friedman et al., 2010). Classic penalized linear regression models such as ridge regression, least absolute shrinkage and selection operator (LASSO) regression, and elastic net regression have been applied to drug response prediction (Geeleher et al., 2014; Huang et al., 2020) (Table 2). In order to combine more biological information, various molecular profiles, as well as prior knowledge such as protein-protein interactions and tumor tissue origins, were incorporated in linear regression models (Ding et al., 2016; Huang et al., 2020). For example, Huang et al. (2020) constructed a tissue-guide linear regression model with LASSO penalty; Ding et al. (2016) applied elastic net regression with bootstrapping to select features from multi-types of molecular data and designed an ensembled multivariate logistic regression to predict the patient clinical response.

Kernel trick is an important strategy in ML models, which project input features to higher dimension space to learn the separating hyperplane. Kernel support vector machine (SVM) is a typical approach that handles both classification and regression on linear and nonlinear relationships (Emad et al., 2017; Cichonska et al., 2018) (Table 2). SVM is also commonly combined with feature selection methods. For example, Huang et al. (2017) trained support vector machine (RFE-SVM) model with a recursive feature elimination (RFE) feature selection approach; Emad et al. (2017) trained nonlinear support vector regression with a network-based gene prioritization method (ProGENI) as feature filter. Multiple-kernel learning methods offer promising benefits since they can incorporate different information sources simultaneously. The multiple pairwise kernel regression model (pairwiseMKL) achieved good performance (Pearson's correlation coefficient, PCC $=$ 0.858) in anticancer drug response prediction (Cichonska et al., 2018). PairwiseMKL firstly designed multiple drug kernels and multiple cell line kernels from the chemical structure and genomic data, respectively, and then pairwise kernels were calculated as Kronecker products of drug kernels and cell line kernels; finally, a weighted combination of pairwise kernels was used for prediction.

Bayesian inference is an effective method utilizing prior information to compute the posterior probability. The problem of drug response prediction needs to integrate multiple data sources and output multiple variables (responses of multiple drugs). Therefore, Bayesian inference coupled with multiview and multitask learning have been increasingly used (Costello et al., 2014; Liu et al., 2019b) (Table 2). Multiview learning represents heterogeneous input data such as gene expression, mutation, copy number variation, drug target genes, and biological pathways as multiple data views. Multitask learning trains all drugs simultaneously, which can obtain better performance than single-task algorithms. A typical example is the Bayesian multitask multiple kernel learning method (BMTMKL), the top-performing model in the NCI-DREAM drug sensitivity prediction challenge (Costello et al., 2014). In addition to multiview learning, multitask learning, and Bayesian inference, BMTMKL utilized the kernel method to reduce dimensionality and captures nonlinear relationships. Moreover, Liu et al. built a cancer-specific network-based spare Bayesian model from multi-omics data and put it as prior information for drug sensitivity prediction (Liu et al., 2019b).

Matrix factorization (MF) learns latent representations by decomposing a target matrix into two low-dimensional matrices and estimating the predicted matrix from the dot product of the two latent matrices. Thus, it is possible to project drug and cell line features into a latent feature space and reconstruct the predicted drug sensitivity matrix (Wang et al., 2017; Emdadi and Eslahchi, 2020) (Table 2). Another advantage of MF-based models is that latent representations may contain informative biological associations between drug and pathway or drug and gene. For example, a similarity-regularized matrix factorization method (SRMF) not only showed promising prediction power (average PCC $= 0.71 \pm 0.09$) but also found novel drug-cancer gene associations related to drug sensitivity (Wang et al., 2017).

Ensemble learning is a popular ML technique that combines multiple learners to improve performance. Random forest is the most commonly used ensemble method to predict drug response. Boosting algorithm such as Extreme Gradient Boosting (xgBoost) was also used to train prediction model on cell line data and further validated on individual datasets, including cell lines and xenografts (Kurilov et al., 2020). Liu et al. combine matrix completion and ridge regression to develop an ensemble drug sensitivity prediction model (Liu et al., 2020b). Deep-Resp-Forest is a more complicated model

that combines multi-info multi-grained scanning and cascade forest with feature optimization to predict drug response and achieves high prediction accuracy (Su et al., 2019).

Deep learning models

Deep learning (DL) is a special subset of machine-learning methods. It consists of multiple layers of the neural network, which is inspired by how biological neurons connect and pass messages. Applying nonlinear activation functions to transmit signals between different layers, DL can abstract nonlinear relationships from a complex biological system. DL has the ability to deal with large-scale, high-dimensional data, so it is suitable for the scenario of estimating drug response from pharmacological and multi-omics data (Chiu et al., 2020) (Table 2).

The simplest architecture of DL models is the deep neural network (DNN), which comprises a multi-layer neural network with interconnected neuron units. Sakellaropoulos trained a three-layer DNN model with gene expression profiles to predict drug sensitivity (Sakellaropoulos et al., 2019). With batch effect adjusted, authors trained the DNN model on cancer cell line data and applied it on patient data to predict drug response (AUC = 0.75). Convolutional neural network (CNN) is another classic architecture of the DL model. Instead of fully connected neurons in DNN, convolutional layers can extract local features with the convolutional kernel so that they can efficiently reduce dimension while preserving important information. Incorporated mutation position information of curated cancer-associated genes and molecular descriptors of drugs, Cancer Drug Response Profile scan (CDScan) ensembled five CNN models with different structures to predict drug IC50 values and achieved high performance (Chang et al., 2018). Autoencoder (AE) is an unsupervised learning technique using several hidden layers to compress the dimension of inputs (encoder) and reconstruct the inputs from the compressed outputs (decoder) (Baptista et al., 2021). Some large cancer repositories like TCGA have molecular data but lack drug response data. AEs can utilize these unlabeled data to obtain the low dimension representation of gene expression, mutation status, CNV, and drug structure, and then DNN and CNN were built to train prediction models from samples with drug responses (Liu et al., 2019a; Sharifi-Noghabi et al., 2019).

One of the obstacles to using DL models is interpretability. Attention mechanism refers to assigning different weights to each input feature according to importance. PaccMann applied a multi-scale convolutional attention mechanism to encode chemical structure and gene expression (Cadow et al., 2020). Gene attention weights could reflect the importance of genes to drug sensitivity. Other interpretability methods like Shapley additive explanations and Layer-wise Relevance Propagation can be used to obtain feature importance, which facilities model interpretation (Baptista et al., 2021). For gaining better interpretability, a visible neural network (VNN) scheme was designed based on biological knowledge to regularize DL model architecture (Chiu et al., 2020). DrugCell, an improved VNN model, used the hierarchy of Gene Ontology (GO) to model functional subsystems of a cell (Kuenzi et al., 2020). It achieved a high correlation between the predicted value and ground truth (Spearman correlation coefficient = 0.8). More importantly, the contribution of each GO biological process term can be quantified so that it can learn underlying mechanisms via functional analysis.

Important factors for improving drug response prediction

In the drug sensitivity prediction challenge cohosted by DREAM and NCI, 44 algorithms covering a range of methodologies were collected and benchmarked with the same accuracy measurements for predicting the response of 28 compounds in breast cancer cell lines (Costello et al., 2014). Another recent review article assessed 17 representative methods on four large public datasets (GDSC, CCLE, NCI60, and CTRP) in nine evaluation metrics (Chen and Zhang, 2021). Both studies highlight that the common important factors for drug sensitivity prediction models are the integration of multi-omics data, the contribution of prior knowledge, the utility of nonlinear algorithms, and multitask learning. The DREAM challenge organizers noticed that the performance of models with same methodology varies greatly, even when using similar data types (Costello et al., 2014). It illustrates the importance of feature selections and algorithm-specific data processing ability. The performance of some models dramatically changed across different datasets and evaluation metrics, indicating the importance of robustness and generalization in method development (Chen and Zhang, 2021). Moreover, the contributions of different omics data to drug response prediction were discussed. Gene expression profile shows best predictive power, followed by protein expression, methylation data, and other molecular data.

Biomarkers and predictions of immunotherapy response

Although preclinical models such as humanized mice have been developed to examine responses to immunotherapy, the techniques, cost and time of model establishment limit the accumulation of large-scale pharmacogenomics data for immunotherapy; therefore, biomarkers of immunotherapy are dominantly found by clinical trials; prediction approaches of immunotherapy response are developed based on the understanding of immune escape mechanism instead of learning from large-scale pharmacogenomic data of cancer models.

In addition to FDA-approved immune checkpoint blockade (ICB) biomarkers (microsatellite status, mutation burden, and PD-L1 expression), the expression level of CD8A and CD8B also have shown a potential association with ICB response (Chen et al., 2016). However, a single biomarker is not powerful enough to precisely stratify patients because of tumor heterogeneity and complex interaction between tumor cells and immune cells (Chen et al., 2016; Jiang et al., 2018). Gene expression signatures consisting of multiple genes have been identified to associate with immunotherapy response, such as the 10-gene and 28-gene expanded signatures linking to IFN-γ expression (Ayers et al., 2017), the 18-gene tumor inflammation signature (Damotte et al., 2019), and the immunopredictive score encompassing 15 pairwise transcriptomics relations (Auslander et al., 2018). Tumor-infiltrating lymphocytes are also in association with immunotherapy response and patient survival, such as $CD8^+$ cytotoxic T cells, B cells and dendritic cells are positively associated with immunotherapy response and prognostics, while myeloid-derived suppressor cells, macrophage type 2, and regulatory T cells are negatively associated immunotherapy resistance and prognostics in some cancer types (Bruni et al., 2020; Petitprez et al., 2020). More complex prediction approaches for ICB have also been built with large-scale patient omics data such as from The Cancer Genome Atlas (TCGA). Immunophenoscore is the weighted sum of four Z-scores that represent four types of immune-related signatures involving in MHC-related antigen presentation, immune checkpoints and immune modulators, immune effector cells, and immune inhibitory cells derived from TCGA patient data (Charoentong et al., 2017). TIDE algorithm firstly classifies patients according to the level of cytotoxic T lymphocyte and then calculates the correlation of gene expression profiles with the signatures of T-cell dysfunction and immune exclusion to predict response (Jiang et al., 2018). Fifteen-gene signature related to hot and cold tumor microenvironment were mined from large biological literatures and verified the prediction performance on melanoma and breast cancer data cohorts (Wang et al., 2021). ICB resistance-related cancer cell program was obtained from single-cell sequencing data of

melanoma patients before and after receiving immune checkpoint inhibitor therapy and TCGA melanoma bulk-RNASeq dataset (Jerby-Arnon et al., 2018). With the development of computation methods and the deeper understanding of cancer immunity, microbiome and pathology images have also be considered to construct prediction models. However, due to the limitation of immunotherapy response data, current predictive models were validated on several publicly available data sets such as melanoma and lung cancer (Charoentong et al., 2017; Jiang et al., 2018; Wang et al., 2021). The generation of a more powerful model to predict ICB response remains a future challenge.

Another newly emerged immunotherapy, such as chimeric antigen receptor (CAR) T cell therapy, has received widespread attention due to its high response rate in the treatment of diffuse large B-cell lymphoma and B-cell precursor acute lymphocytic leukemia (Jafarzadeh et al., 2021). However, the response rates of CAR-T cell therapy varied widely among different cancers and patients. Although identifying therapeutic biomarkers of immune cell therapy response research is still in its early stage, some potential predictive biomarkers of CAR-T cell therapy have been identified by the single-cell analysis, such as polyfunctionality of CAR T cells, TNF-α^{+}IFN-γ^{+} polyfunctional MART-1-specific T cells, as well as $CD8^{+}$ T cell exhaustion signatures (e.g. PD-1, LAG-3 and TIM-3) and memory-related-genes (e.g. STAT3-related cytokines and serum IL-6) (Feins et al., 2019; Jafarzadeh et al., 2021). With limited data sets, it is tough to build a prediction model to estimate patient responses.

Challenge in personalized cancer medicine

Large-scale pharmacogenomic screening and drug response prediction have made dramatic improvements in the past decades, but the consistency between datasets, robustness, generalization capability, and interpretability of biomarkers and prediction models remain extremely hard tasks. We will discuss these problems from the perspective of data and algorithms. On the one hand, inconsistency between multiple pharmacogenomic datasets is common. Drug response data may be inconsistent due to the differences in cancer model establishment methods, drug doses and screening protocols, and measurement indices of drug sensitivity (Bouhaddou et al., 2016; Mpindi et al., 2016). Even for the same drug on same cancer cell lines, drug response profiles are not ideally consistent (Rahman et al., 2019). Similarly, molecular characteristics of cancer models or patients may not be comparable because of different experiment platforms and analysis pipelines. The problem is more serious for expression profiles due to the batch effect of different datasets. Even if there are methods to reduce batch effects, the impact cannot be completely removed (Espin-Perez et al., 2018). Therefore, before using the accumulated pharmacological and omics data, suitable standardization and batch-effect correction strategies should be thoroughly designed to reduce the effects. On the other hand, the accuracy of prediction algorithms usually decreases when applying to independent test datasets (Liu et al., 2019a). Most of the drug response prediction models were tested using k-fold or leave-one-out cross-validation (Table 2); these models are likely to be overfitting. The prediction accuracies among different drugs or cancer types may vary greatly (Zhang et al., 2015; Emad et al., 2017; Sharifi-Noghabi et al., 2019; Huang et al., 2020). The imbalanced number of drugs and the distribution of samples by cancer types in multiple datasets will also affect the generalization capability of models (Rahman et al., 2019). There are some DL models that tried to balance this problem with the transfer learning concept (Kim et al., 2021) which suggests that it is possible to improve the model as the method is improved.

Translation of biomarker candidates and prediction models obtained from cancer models into clinical practice is another major challenge in personalized cancer medicine. First, the pharmacogenomic studies and computational prediction methods based on cancer models have not well considered different pharmacokinetics (absorption, distribution, metabolism, and excretion) in the human body. Second, there exists intra-tumor heterogeneity. Molecular profiles obtained from bulk tissues may not reflect the features of some tumor cells and microenvironment. The single-cell technique has been emerged to dissect tumor microenvironment and has been used to identify aggressive cell population and their association with drug response (Farkkila et al., 2020; Roider et al., 2020). Therefore, incorporating a single cell technique with drug screening may help analyze the effects of intra-tumor heterogeneity. Third, patient tumors constantly evolve. Sensitive clones are killed during treatment, and resistant clones grow (Brady et al., 2017). Prediction based on the molecular characteristics of a tumor before treatment may not be suitable for the tumor after evolution. The single-cell technique also has been used to elucidate the genetic profile changes at multiple time points under drug perturbation at the cell level (Su et al., 2017; Maynard et al., 2020). It may help sort out a deeper mechanism of acquired resistance of drugs and optimize medical regime (Anchang et al., 2018) to achieve better response for patients. Although some studies have reported the consistency of drug response between preclinical tumor models and patient tumors (Gao et al., 2015; Ooft et al., 2019; Qiu et al., 2019), and a few biomarkers and prediction models achieved satisfactory results in patients (Sakellaropoulos et al., 2019; Sharifi-Noghabi et al., 2019), these studies generally only have a small number of patients and individual drugs. It is hard to estimate the effectiveness and accuracy of biomarker candidates or prediction models in the large number of heterogeneous patients. Therefore, it will be essential to build a comprehensive biomedicine knowledge base to integrate representative preclinical and patients' pharmacogenomic data (Caroli et al., 2020). New computation methods, such as transfer learning, can be embedded in the database's backend and provides drug response prediction and therapeutic biomarker candidates identification for further verification by clinical trials.

Conflict of interest

The authors declare that they have no conflict of interests.

Acknowledgements

This work was supported by National Key Research and Development Project (2019YFC1315804), National Natural Science Foundation of China (31771472), Chinese Academy of Sciences (ZDBS-SSW-DQC-02), SA-SIBS Scholarship Program, Shanghai Municipal Science and Technology Major Project (No.2018SHZDZX01), CAS Youth Innovation Promotion Association (2018307), Chinese Academy of Sciences (KFJ-STS-QYZD-126).

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jgg.2021.03.007.

References

Anchang, B., Davis, K.L., Fienberg, H.G., Williamson, B.D., Bendall, S.C., Karacosta, L.G., Tibshirani, R., Nolan, G.P., Plevritis, S.K., 2018. DRUG-NEM: optimizing drug combinations using single-cell perturbation response to account for intratumoral heterogeneity. Proc. Natl. Acad. Sci. U. S. A. 115, E4294–E4303.

Auslander, N., Zhang, G., Lee, J.S., Frederick, D.T., Miao, B., Moll, T., Tian, T., Wei, Z., Madan, S., Sullivan, R.J., et al., 2018. Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma. Nat. Med. 24, 1545–1549.

Ayers, M., Lunceford, J., Nebozhyn, M., Murphy, E., Loboda, A., Kaufman, D.R., Albright, A., Cheng, J.D., Kang, S.P., Shankaran, V., et al., 2017. IFN-gamma-

related mRNA profile predicts clinical response to PD-1 blockade. J. Clin. Invest. 127, 2930–2940.
Baptista, D., Ferreira, P.G., Rocha, M., 2021. Deep learning for drug response prediction in cancer. Briefings Bioinf. 22, 360–379.
Basu, A., Bodycombe, N.E., Cheah, J.H., Price, E.V., Liu, K., Schaefer, G.I., Ebright, R.Y., Stewart, M.L., Ito, D., Wang, S., et al., 2013. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. Cell 154, 1151–1161.
Bock, C., Boutros, M., Camp, J.G., Clarke, L., Clevers, H., Knoblich, J.A., Liberali, P., Regev, A., Rios, A.C., Stegle, O., et al., 2021. The organoid cell atlas. Nat. Biotechnol. 39, 13–17.
Bouhaddou, M., DiStefano, M.S., Riesel, E.A., Carrasco, E., Holzapfel, H.Y., Jones, D.C., Smith, G.R., Stern, A.D., Somani, S.S., Thompson, T.V., et al., 2016. Drug response consistency in CCLE and CGP. Nature 540, E9–E10.
Boyiadzis, M.M., Kirkwood, J.M., Marshall, J.L., Pritchard, C.C., Azad, N.S., Gulley, J.L., 2018. Significance and implications of S approval of pembrolizumab for biomarker-defined disease. J. Immunother. Cancer 6, 35.
Brady, S.W., McQuerry, J.A., Qiao, Y., Piccolo, S.R., Shrestha, G., Jenkins, D.F., Layer, R.M., Pedersen, B.S., Miller, R.H., Esch, A., et al., 2017. Combating subclonal evolution of resistant cancer phenotypes. Nat. Commun. 8, 1231.
Bruni, D., Angell, H.K., Galon, J., 2020. The immune contexture and immunoscore in cancer prognosis and therapeutic efficacy. Nat. Rev. Cancer 20, 662–680.
Cadow, J., Born, J., Manica, M., Oskooei, A., Rodriguez Martinez, M., 2020. PaccMann: a web service for interpretable anticancer compound sensitivity prediction. Nucleic Acids Res. 48, W502–W508.
Caroli, J., Dori, M., Bicciato, S., 2020. Computational methods for the integrative analysis of genomics and pharmacological data. Front. Oncol. 10, 185.
Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al., 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Canc. Discov. 2, 401–404.
Chang, Y., Park, H., Yang, H.J., Lee, S., Lee, K.Y., Kim, T.S., Jung, J., Shin, J.M., 2018. Cancer drug response profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature. Sci. Rep. 8, 8857.
Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., Hackl, H., Trajanoski, Z., 2017. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. Cell Rep. 18, 248–262.
Chen, J., Zhang, L., 2021. A survey and systematic assessment of computational methods for drug response prediction. Briefings Bioinf. 22, 232–246.
Chen, P.L., Roh, W., Reuben, A., Cooper, Z.A., Spencer, C.N., Prieto, P.A., Miller, J.P., Bassett, R.L., Gopalakrishnan, V., Wani, K., et al., 2016. Analysis of immune signatures in longitudinal tumor samples yields insight into biomarkers of response and mechanisms of resistance to immune checkpoint blockade. Canc. Discov. 6, 827–837.
Chen, E.Y., Raghunathan, V., Prasad, V., 2019. An overview of cancer drugs approved by the US Food and Drug Administration based on the surrogate end point of response rate. JAMA Intern. Med. 179, 915–921.
Chiu, Y.C., Chen, H.H., Gorthi, A., Mostavi, M., Zheng, S., Huang, Y., Chen, Y., 2020. Deep learning of pharmacogenomics resources: moving towards precision oncology. Briefings Bioinf. 21, 2066–2083.
Cichonska, A., Pahikkala, T., Szedmak, S., Julkunen, H., Airola, A., Heinonen, M., Aittokallio, T., Rousu, J., 2018. Learning with multiple pairwise kernels for drug bioactivity prediction. Bioinformatics 34, i509–i518.
Conte, N., Mason, J.C., Halmagyi, C., Neuhauser, S., Mosaku, A., Yordanova, G., Chatzipli, A., Begley, D.A., Krupke, D.M., Parkinson, H., et al., 2019. PDX Finder: a portal for patient-derived tumor xenograft model discovery. Nucleic Acids Res. 47, D1073–D1079.
Costello, J.C., Heiser, L.M., Georgii, E., Gonen, M., Menden, M.P., Wang, N.J., Bansal, M., Ammad-ud-din, M., Hintsanen, P., Khan, S.A., et al., 2014. A community effort to assess and improve drug sensitivity prediction algorithms. Nat. Biotechnol. 32, 1202–1212.
Crown Bioscience Inc, 2021. Oncology Databases. https://www.crownbio.com/oncology/oncology-databases.
Daemen, A., Griffith, O.L., Heiser, L.M., Wang, N.J., Enache, O.M., Sanborn, Z., Pepin, F., Durinck, S., Korkola, J.E., Griffith, M., et al., 2013. Modeling precision treatment of breast cancer. Genome Biol. 14, R110.
Damotte, D., Warren, S., Arrondeau, J., Boudou-Rouquette, P., Mansuet-Lupo, A., Biton, J., Ouakrim, H., Alifano, M., Gervais, C., Bellesoeur, A., et al., 2019. The tumor inflammation signature (TIS) is associated with anti-PD-1 treatment benefit in the CERTIM pan-cancer cohort. J. Transl. Med. 17, 357.
Ding, Z., Zu, S., Gu, J., 2016. Evaluating the molecule-based prediction of clinical drug responses in cancer. Bioinformatics 32, 2891–2895.
Emad, A., Cairns, J., Kalari, K.R., Wang, L., Sinha, S., 2017. Knowledge-guided gene prioritization reveals new insights into the mechanisms of chemoresistance. Genome Biol. 18, 153.
Emdadi, A., Eslahchi, C., 2020. DSPLMF: a method for cancer drug sensitivity prediction using a novel regularization approach in logistic matrix factorization. Front. Genet. 11, 75.
Espin-Perez, A., Portier, C., Chadeau-Hyam, M., van Veldhoven, K., Kleinjans, J.C.S., de Kok, T., 2018. Comparison of statistical methods and the use of quality control samples for batch effect correction in human transcriptome data. PLoS One 13, e0202947.
Evrard, Y.A., Ahalt, M.M.G., Alcoser, S.Y., Arthur, K., Baldwin, M., Blumenauer, L.L., Bonomi, C., Borgel, S., Bradtke, E., Camalier, C., et al., 2018. Abstract 986: the national cancer institute's patient-derived models repository (PDMR). Cancer Res. 78, 986 - 986.
Fang, Y., Xu, P., Yang, J., Qin, Y., 2018. A quantile regression forest based method to predict drug response and assess prediction reliability. PLoS One 13, e0205155.
Farkkila, A., Gulhan, D.C., Casado, J., Jacobson, C.A., Nguyen, H., Kochupurakkal, B., Maliga, Z., Yapp, C., Chen, Y.A., Schapiro, D., et al., 2020. Immunogenomic profiling determines responses to combined PARP and PD-1 inhibition in ovarian cancer. Nat. Commun. 11, 1459.
Feins, S., Kong, W., Williams, E.F., Milone, M.C., Fraietta, J.A., 2019. An introduction to chimeric antigen receptor (CAR) T-cell immunotherapy for human cancer. Am. J. Hematol. 94, S3–S9.
Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Software 33, 1–22.
Gao, H., Korn, J.M., Ferretti, S., Monahan, J.E., Wang, Y., Singh, M., Zhang, C., Schnell, C., Yang, G., Zhang, Y., et al., 2015. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. Nat. Med. 21, 1318–1325.
Geeleher, P., Cox, N.J., Huang, R.S., 2014. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. Genome Biol. 15, R47.
Ghandi, M., Huang, F.W., Jane-Valbuena, J., Kryukov, G.V., Lo, C.C., McDonald 3rd, E.R., Barretina, J., Gelfand, E.T., Bielski, C.M., Li, H., et al., 2019. Next-generation characterization of the cancer cell line encyclopedia. Nature 569, 503–508.
Goto, T., 2020. Patient-derived tumor xenograft models: toward the establishment of precision cancer medicine. J. Personalized Med. 10, 64.
Greshock, J., Bachman, K.E., Degenhardt, Y.Y., Jing, J., Wen, Y.H., Eastman, S., McNeil, E., Moy, C., Wegrzyn, R., Auger, K., et al., 2010. Molecular target class is predictive of *in vitro* response profile. Cancer Res. 70, 3677–3686.
Han, S.W., Kim, T.Y., Hwang, P.G., Jeong, S., Kim, J., Choi, I.S., Oh, D.Y., Kim, J.H., Kim, D.W., Chung, D.H., et al., 2005. Predictive and prognostic impact of epidermal growth factor receptor mutation in non-small-cell lung cancer patients treated with gefitinib. J. Clin. Oncol. 23, 2493–2501.
Haverty, P.M., Lin, E., Tan, J., Yu, Y., Lam, B., Lianoglou, S., Neve, R.M., Martin, S., Settleman, J., Yauch, R.L., et al., 2016. Reproducible pharmacogenomic profiling of cancer cell line panels. Nature 533, 333–337.
He, S., Hu, B., Li, C., Lin, P., Tang, W.G., Sun, Y.F., Feng, F.Y., Guo, W., Li, J., Xu, Y., et al., 2018. Pdxliver: a database of liver cancer patient derived xenograft mouse models. BMC Canc. 18, 550.
Heiser, L.M., Sadanandam, A., Kuo, W.L., Benz, S.C., Goldstein, T.C., Ng, S., Gibb, W.J., Wang, N.J., Ziyad, S., Tong, F., et al., 2012. Subtype and pathway specific responses to anticancer compounds in breast cancer. Proc. Natl. Acad. Sci. U. S. A. 109, 2724–2729.
Hu, B., Li, H., Guo, W., Sun, Y.F., Zhang, X., Tang, W.G., Yang, L.X., Xu, Y., Tang, X.Y., Ding, G.H., et al., 2020. Establishment of a hepatocellular carcinoma patient-derived xenograft platform and its application in biomarker identification. Int. J. Canc. 146, 1606–1617.
Huang, C., Mezencev, R., McDonald, J.F., Vannberg, F., 2017. Open source machine-learning algorithms for the prediction of optimal cancer drug therapies. PLoS One 12, e0186906.
Huang, E.W., Bhope, A., Lim, J., Sinha, S., Emad, A., 2020. Tissue-guided LASSO for prediction of clinical drug response using preclinical samples. PLoS Comput. Biol. 16, e1007607.
Ibarrola-Villava, M., Cervantes, A., Bardelli, A., 2018. Preclinical models for precision oncology. Biochim. Biophys. Acta Rev. Canc 1870, 239–246.
Izumchenko, E., Paz, K., Ciznadija, D., Sloma, I., Katz, A., Vasquez-Dunddel, D., Ben-Zvi, I., Stebbing, J., McGuire, W., Harris, W., et al., 2017. Patient-derived xenografts effectively capture responses to oncology therapy in a heterogeneous cohort of patients with solid tumors. Ann. Oncol. 28, 2595–2605.
Jafarzadeh, L., Khakpoor-Koosheh, M., Mirzaei, H., Mirzaei, H.R., 2021. Biomarkers for predicting the outcome of various cancer immunotherapies. Crit. Rev. Oncol. Hematol. 157, 103161.
Jerby-Arnon, L., Shah, P., Cuoco, M.S., Rodman, C., Su, M.J., Melms, J.C., Leeson, R., Kanodia, A., Mei, S., Lin, J.R., et al., 2018. A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. Cell 175, 984–997. e24.
Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., Li, Z., Traugh, N., Bu, X., Li, B., et al., 2018. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat. Med. 24, 1550–1558.
Karamboulas, C., Bruce, J.P., Hope, A.J., Meens, J., Huang, S.H., Erdmann, N., Hyatt, E., Pereira, K., Goldstein, D.P., Weinreb, I., et al., 2018. Patient-derived xenografts for prognostication and personalized treatment for head and neck squamous cell carcinoma. Cell Rep. 25, 1318–1331 e1314.
Kim, Y.A., Sarto Basso, R., Wojtowicz, D., Liu, A.S., Hochbaum, D.S., Vandin, F., Przytycka, T.M., 2020. Identifying drug sensitivity subnetworks with NETPHIX. iScience 23, 101619.
Kim, Y., Zheng, S., Tang, J., Jim Zheng, W., Li, Z., Jiang, X., 2021. Anticancer drug synergy prediction in understudied tissues using transfer learning. J. Am. Med. Inf. Assoc. 28, 42–51.
Koleti, A., Terryn, R., Stathias, V., Chung, C., Cooper, D.J., Turner, J.P., Vidovic, D., Forlin, M., Kelley, T.T., D'Urso, A., et al., 2018. Data portal for the library of integrated network-based cellular signatures (LINCS) program: integrated access to diverse large-scale cellular perturbation response data. Nucleic Acids Res. 46, D558–D566.
Kondo, J., Ekawa, T., Endo, H., Yamazaki, K., Tanaka, N., Kukita, Y., Okuyama, H., Okami, J., Imamura, F., Ohue, M., et al., 2019. High-throughput screening in colorectal cancer tissue-originated spheroids. Canc. Sci. 110, 345–355.

Kong, J., Lee, H., Kim, D., Han, S.K., Ha, D., Shin, K., Kim, S., 2020. Network-based machine learning in colorectal and bladder organoid models predicts anti-cancer drug efficacy in patients. Nat. Commun. 11, 5485.
Kuenzi, B.M., Park, J., Fong, S.H., Sanchez, K.S., Lee, J., Kreisberg, J.F., Ma, J., Ideker, T., 2020. Predicting drug response and synergy using a deep learning model of human cancer cells. Canc. Cell 38, 672–684. e6.
Kurilov, R., Haibe-Kains, B., Brors, B., 2020. Assessment of modelling strategies for drug response prediction in cell lines and xenografts. Sci. Rep. 10, 2849.
Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N., et al., 2006. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313, 1929–1935.
Le, D.H., Pham, V.H., 2018. Drug response prediction by globally capturing drug and cell line information in a heterogeneous network. J. Mol. Biol. 430, 2993–3004.
Liu, P., Li, H., Li, S., Leung, K.S., 2019a. Improving prediction of phenotypic drug response on cancer cell lines using deep convolutional network. BMC Bioinf. 20, 408.
Liu, Q., Muglia, L.J., Huang, L.F., 2019b. Network as a biomarker: a novel network-based sparse bayesian machine for pathway-driven drug response prediction. Genes (Basel) 10, 602.
Liu, C., Qin, T., Huang, Y., Li, Y., Chen, G., Sun, C., 2020a. Drug screening model meets cancer organoid technology. Transl. Oncol. 13, 100840.
Liu, C., Wei, D., Xiang, J., Ren, F., Huang, L., Lang, J., Tian, G., Li, Y., Yang, J., 2020b. An improved anticancer drug-response prediction based on an ensemble method integrating matrix completion and ridge regression. Mol. Ther. Nucleic Acids 21, 676–686.
Maynard, A., McCoach, C.E., Rotow, J.K., Harris, L., Haderk, F., Kerr, D.L., Yu, E.A., Schenk, E.L., Tan, W., Zee, A., et al., 2020. Therapy-induced evolution of human lung cancer revealed by Single-Cell RNA sequencing. Cell 182, 1232–1251. e22.
Mpindi, J.P., Yadav, B., Ostling, P., Gautam, P., Malani, D., Murumagi, A., Hirasawa, A., Kangaspeska, S., Wennerberg, K., Kallioniemi, O., et al., 2016. Consistency in drug response profiling. Nature 540, E5–E6.
Olson, B., Li, Y., Lin, Y., Liu, E.T., Patnaik, A., 2018. Mouse models for cancer immunotherapy research. Canc. Discov. 8, 1358–1365.
Ooft, S.N., Weeber, F., Dijkstra, K.K., McLean, C.M., Kaing, S., van Werkhoven, E., Schipper, L., Hoes, L., Vis, D.J., van de Haar, J., et al., 2019. Patient-derived organoids can predict response to chemotherapy in metastatic colorectal cancer patients. Sci. Transl. Med. 11, eaay2574.
Park, J.J.H., Hsu, G., Siden, E.G., Thorlund, K., Mills, E.J., 2020. An overview of precision oncology basket and umbrella trials for clinicians. CA. Cancer J. Clin. 70, 125–137.
Pauli, C., Hopkins, B.D., Prandi, D., Shaw, R., Fedrizzi, T., Sboner, A., Sailer, V., Augello, M., Puca, L., Rosati, R., et al., 2017. Personalized *in vitro* and *in vivo* cancer models to guide precision medicine. Canc. Discov. 7, 462–477.
Petitprez, F., Meylan, M., de Reynies, A., Sautes-Fridman, C., Fridman, W.H., 2020. The tumor microenvironment in the response to immune checkpoint blockade therapies. Front. Immunol. 11, 784.
Polley, E., Kunkel, M., Evans, D., Silvers, T., Delosh, R., Laudeman, J., Ogle, C., Reinhart, R., Selby, M., Connelly, J., et al., 2016. Small cell lung cancer screen of oncology drugs, investigational agents, and gene and microrna expression. J. Natl. Cancer Inst. 108, djw122.
Qiu, Z., Li, H., Zhang, Z., Zhu, Z., He, S., Wang, X., Wang, P., Qin, J., Zhuang, L., Wang, W., et al., 2019. A pharmacogenomic landscape in human liver cancers. Canc. Cell 36, 179–193 e111.
Rahman, R., Dhruba, S.R., Matlock, K., De-Niz, C., Ghosh, S., Pal, R., 2019. Evaluating the consistency of large-scale pharmacogenomic studies. Briefings Bioinf. 20, 1734–1753.
Roider, T., Seufert, J., Uvarovskii, A., Frauhammer, F., Bordas, M., Abedpour, N., Stolarczyk, M., Mallm, J.P., Herbst, S.A., Bruch, P.M., et al., 2020. Dissecting intratumour heterogeneity of nodal B-cell lymphomas at the transcriptional, genetic and drug-response levels. Nat. Cell Biol. 22, 896–906.
Sachs, N., de Ligt, J., Kopper, O., Gogola, E., Bounova, G., Weeber, F., Balgobind, A.V., Wind, K., Gracanin, A., Begthel, H., et al., 2018. A living biobank of breast cancer organoids captures disease heterogeneity. Cell 172, 373–386. e10.
Sakellaropoulos, T., Vougas, K., Narang, S., Koinis, F., Kotsinas, A., Polyzos, A., Moss, T.J., Piha-Paul, S., Zhou, H., Kardala, E., et al., 2019. A deep learning framework for predicting response to therapy in cancer. Cell Rep. 29, 3367–3373. e64.
Shanks, N., Greek, R., Greek, J., 2009. Are animal models predictive for humans? Philos. Ethics Humanit. Med. 4, 2.
Sharifi-Noghabi, H., Zolotareva, O., Collins, C.C., Ester, M., 2019. MOLI: multi-omics late integration with deep neural networks for drug response prediction. Bioinformatics 35, i501–i509.
Sharma, P., Wagner, K., Wolchok, J.D., Allison, J.P., 2011. Novel cancer immunotherapy agents with survival benefit: recent successes and next steps. Nat. Rev. Cancer 11, 805–812.
Sharpless, N.E., Depinho, R.A., 2006. The mighty mouse: genetically engineered mouse models in cancer drug development. Nat. Rev. Drug Discov. 5, 741–754.
Shoemaker, R.H., 2006. The NCI60 human tumour cell line anticancer drug screen. Nat. Rev. Cancer 6, 813–823.
Stanfield, Z., Coskun, M., Koyuturk, M., 2017. Drug response prediction as a link prediction problem. Sci. Rep. 7, 40321.
Su, R., Liu, X., Wei, L., Zou, Q., 2019. Deep-Resp-Forest: a deep forest model to predict anti-cancer drug response. Methods 166, 91–102.
Su, Y., Wei, W., Robert, L., Xue, M., Tsoi, J., Garcia-Diaz, A., Homet Moreno, B., Kim, J., Ng, R.H., Lee, J.W., et al., 2017. Single-cell analysis resolves the cell state transition and signaling dynamics associated with melanoma drug-induced resistance. Proc. Natl. Acad. Sci. U. S. A. 114, 13679–13684.
Teicher, B.A., Polley, E., Kunkel, M., Evans, D., Silvers, T., Delosh, R., Laudeman, J., Ogle, C., Reinhart, R., Selby, M., et al., 2015. Sarcoma cell line screen of oncology drugs and investigational agents identifies patterns associated with gene and microRNA expression. Mol. Canc. Therapeut. 14, 2452–2462.
U.S. Food and Drug Administration, 2020. Table of pharmacogenomic biomarkers in drug labeling with labeling text. https://www.fda.gov/drugs/science-and-research-drugs/table-pharmacogenomic-biomarkers-drug-labeling.
van de Wetering, M., Francies, H.E., Francis, J.M., Bounova, G., Iorio, F., Pronk, A., van Houdt, W., van Gorp, J., Taylor-Weiner, A., Kester, L., et al., 2015. Prospective derivation of a living organoid biobank of colorectal cancer patients. Cell 161, 933–945.
Vidyasagar, M., 2015. Identifying predictive features in drug response using machine learning: opportunities and challenges. Annu. Rev. Pharmacol. Toxicol. 55, 15–34.
Wang, H., Li, S., Wang, Q., Jin, Z., Shao, W., Gao, Y., Li, L., Lin, K., Zhu, L., Wang, H., et al., 2021. Tumor immunological phenotype signature-based high-throughput screening for the discovery of combination immunotherapy compounds. Sci. Adv. 7, eabd7851.
Wang, L., Li, X., Zhang, L., Gao, Q., 2017. Improved anticancer drug response prediction in cell lines using matrix factorization with similarity regularization. BMC Canc. 17, 513.
Weeber, F., Ooft, S.N., Dijkstra, K.K., Voest, E.E., 2017. Tumor organoids as a preclinical cancer model for drug discovery. Cell Chem. Biol. 24, 1092–1100.
Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al., 2018. Drugbank 5.0: a major update to the drugbank database for 2018. Nucleic Acids Res. 46, D1074–D1082.
Woo, X.Y., Srivastava, A., Graber, J.H., Yadav, V., Sarsani, V.K., Simons, A., Beane, G., Grubb, S., Ananda, G., Liu, R., et al., 2019. Genomic data analysis workflows for tumors from patient-derived xenografts (PDXs): challenges and guidelines. BMC Med. Genom. 12, 92.
Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al., 2013. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 41, D955–D961.
Yuki, K., Cheng, N., Nakano, M., Kuo, C.J., 2020. Organoid models of tumor immunology. Trends Immunol. 41, 652–664.
Zhang, N., Wang, H., Fang, Y., Wang, J., Zheng, X., Liu, X.S., 2015. Predicting anticancer drug responses using a dual-layer integrated cell line-drug network model. PLoS Comput. Biol. 11, e1004498.
Zhang, F., Wang, M., Xi, J., Yang, J., Li, A., 2018. A novel heterogeneous network-based method for drug response prediction in cancer cell lines. Sci. Rep. 8, 3355.
Zhao, W., Li, J., Chen, M.M., Luo, Y., Ju, Z., Nesser, N.K., Johnson-Camacho, K., Boniface, C.T., Lawrence, Y., Pande, N.T., et al., 2020. Large-scale characterization of drug responses of clinically relevant proteins in cancer cell lines. Canc. Cell 38, 829–843. e4.