Briefings in Bioinformatics, 21(6), 2020, 1886–1903

doi: 10.1093/bib/bbz109
Advance Access Publication Date: 21 November 2019
Review article

Methods and resources to access mutation-dependent effects on cancer drug treatment

Hongcheng Yao†, Qian Liang†, Xinyi Qian, Junwen Wang, Pak Chung Sham and Mulin Jun Li

Corresponding author: Mulin Jun Li, Tianjin Medical University, Tianjin, China. Tel.: +86 22 83336668; Fax: +86 22 83336668; E-mail: mulin0424.li@gmail.com
†These authors contributed equally to this work.

Abstract

In clinical cancer treatment, genomic alterations would often affect the response of patients to anticancer drugs. Studies have shown that molecular features of tumors could be biomarkers predictive of sensitivity or resistance to anticancer agents, but the identification of actionable mutations are often constrained by the incomplete understanding of cancer genomes. Recent progresses of next-generation sequencing technology greatly facilitate the extensive molecular characterization of tumors and promote precision medicine in cancers. More and more clinical studies, cancer cell lines studies, CRISPR screening studies as well as patient-derived model studies were performed to identify potential actionable mutations predictive of drug response, which provide rich resources of molecularly and pharmacologically profiled cancer samples at different levels. Such abundance of data also enables the development of various computational models and algorithms to solve the problem of drug sensitivity prediction, biomarker identification and *in silico* drug prioritization by the integration of multiomics data. Here, we review the recent development of methods and resources that identifies mutation-dependent effects for cancer treatment in clinical studies, functional genomics studies and computational studies and discuss the remaining gaps and future directions in this area.

Key words: precision medicine; actionable mutation; targeted cancer therapy; drug response prediction; bioinformatics tool

Hongcheng Yao is a PhD candidate at the School of Biomedical Sciences, the University of Hong Kong, Hong Kong. His research interests include functional variants annotation and mutation drug response association analysis.
Qian Liang is a PhD candidate at the Department of Pharmacology, Tianjin Medical University. Her research interests include exploiting noncoding functional elements conferring cancer drug resistance using CRISPR screening.
Xinyi Qian is an MS candidate at the Department of Pharmacology, Tianjin Medical University. Her research interests include validating biological functions of causal regulatory variants for particular complex disease.
Junwen Wang is a Professor of Biomedical Informatics at the Department of Health Sciences Research and Center for Individualized Medicine, Mayo Clinic Arizona. He is an affiliate professor at the Department of Biomedical Informatics, Arizona State University, Tempe Arizona, and is supervising graduate students from Arizona State University. His group has developed over a dozen tools in analyzing next-generation sequencing data, inferring regulatory networks from multidimensional OMICs data and annotating noncoding genetic variants.
Pak Chung Sham Centre of Genomics Sciences, State Key Laboratory of Brain and Cognitive Sciences, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China.
Mulin Jun Li is a Professor at the Basic Medical Research Center of Tianjin Medical University. His research interests include the development of bioinformatics tools and resources for interpreting variant effects on different types of molecular traits and human complex diseases and the dissection of biological mechanism and drug resistance mechanism for germline variants/somatic mutations using high-throughput functional genomics methods.

Submitted: 18 May 2019; **Received (in revised form):** 31 July 2019

Introduction

Dissecting mutation impacts to therapeutic response has long been a critical problem in cancer studies. Currently, mutations are suggested to play two major roles in cancer treatment. They could either confer resistance [1] or be used to identify significant responders to specific targeted drugs [2]. Both effects are continuously discovered and verified during clinical treatment of cancer, and typical examples could be found in various types of cancers. For instance, the importance of EGFR mutations to therapeutic responses was found early in the clinic, when lung adenocarcinoma patients detected with EGFR exon 19 deletion and L858R were associated with better response to gefitinib [3–5]. During clinical treatment and tumor evolution, acquired EGFR T790M mutation [6, 7] will lead to resistance to gefitinib and erlotinib. In addition, HER2 amplification [8] was considered as a clinical indication of trastuzumab in HER2-positive breast cancer. However, due to the high cost of DNA sequencing at that time, the scale of genomic profiling and the number of patients involved were both limited.

With the recent development of next-generation sequencing (NGS) [9], the expense and time required for sequencing have been greatly reduced, and it is now feasible to carry out large-scale genomic profiling of patients' tumors in cancer treatment study. Apart from the systematic characterization of tumor genome projects such as the Cancer Genome Atlas (TCGA) [10] and International Cancer Genome Consortium (ICGC) [11], a number of clinical trials with molecular characterization of patients were conducted as regional and international projects. Functional genomics studies including cancer cell line-based drug screening, CRISPR-Cas9 (the clustered regularly interspaced short palindromic repeats associated protein 9) screening and patient-derived model (PDM) studies have also employed NGS to illustrate the effects of mutations to specific therapeutic agents in particular types of cancer.

Along with the emergence of precision medicine, researchers are working to more precisely identify functional genomics alterations, which are responsive to targeted drugs. Such actionable or drugable mutations [2, 12] could guide the prescription, repositioning and development of targeted cancer therapy. In the process of identifying actionable mutations, DNA sequencing and therapeutic response data are accumulating rapidly. To make full use of such a tremendous amount of data, computational methods and bioinformatic tools are developed, and other types of data, such as transcriptomic and epigenomic profiles, are often integrated [13]. There are resources and databases collecting and formatting clinical implications of genomic alterations in cancer treatment, drug sensitivity prediction and biomarker identification methods like those submitted to the National Cancer Institute and the Dialogue on Reverse Engineering Assessment and Methods (NCI-DREAM) program [14] and *in silico* drug prioritization methods. Here, we review the progress in illustrating mutation impact on cancer treatment with an emphasis on the bioinformatics efforts involved. The development of clinical studies, functional genomics studies and computational studies, as well as the interaction of these three research fields, are highlighted.

Clinical studies

Clinical studies

Since the improvement in NGS technology has laid the foundation for large-scale genomic characterization of tumor samples, a number of clinical trials are carried out across the world to detect novel actionable mutations associated with current targeted cancer drugs and to further investigate known actionable mutations in a larger cohort [15]. For example, KRAS exon 2 mutations were confirmed to be predictive of resistance to cetuximab in colorectal cancer patients in a clinical trial [16]. Typical examples of large-scale clinical trials include the Lung Master Protocol [17] and the NCI Molecular Analysis for Therapy Choice (NCI-MATCH) [18] in the United States, which are part of the Precision Medicine Initiative of NCI, the SCRUM-Japan [19] in Japan, and the National Lung Matrix Trial [20] in the United Kingdom. Most of these clinical trials adopted either an umbrella trial design or a basket trial design. The umbrella trial [21, 22] defines that patients will first be selected by their cancer type, and those with the same type of cancer will be further assigned to different groups to receive different therapies according to their genomic alterations. The basket trial [21, 22] requires that patients are assigned into different arms based on their genomic alterations irrespective of the tumor tissue of origin. Of note, NCI-MATCH adopted the Oncomine Comprehensive Panel [23], which is an integrative NGS-based assay of Thermo and provided a classification scheme for actionable mutations, and the National Lung Matrix Trial employed the actionable mutation tiers [2] for classifying actionable mutations. The abovementioned clinical trials designed to match single mutation with monotherapy often have the problem of low matching rate and low response rate. More recently, clinical trials with novel cancer therapy-matching strategies were carried out to address this issue. For example, the I-PREDICT study [24] matched patients with drug combinations based on multiple actionable mutations. The WINTHER trial [25] compared assigning therapies to patients based on DNA sequencing data with that on gene expression data. And the TARGET study [26] demonstrated the feasibility of genomic profiling using circulating tumor DNA in the process of matching patients with cancer therapy. These clinical trials illuminate possible future directions for more efficient and rational recruitment of patients during the trial design. In addition to these regional clinical trials, The ICGC has recently launched the ICGC-ARGO project, aiming to integrate, analyze and share cancer treatment data and molecular profiling data of a million patients from all over the world [27]. There are also large-scale cancer genome sequencing projects such as the Memorial Sloan Kettering Cancer Center–Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT) project [28], which aims to facilitate cancer clinical trials by providing genomic profiling data. The MSK-IMPACT project focuses on sequencing samples of metastatic cancer patients and serves as an important complement of the TCGA project.

Apart from the classic targeted cancer therapy, recent advances in immunotherapy provide new possibilities for cancer treatment. The benefits and problems of combination of targeted therapies and immunotherapies in cancer treatment as well as the rational design of corresponding clinical trials were discussed [29, 30]. Progress in combining immunotherapies with targeted therapies dependent of specific actionable mutations were also reported in melanoma [31] and non-small cell lung cancer [32]. It has been observed that the activation of EGFR pathway induced by actionable mutations (such as EGFR L858R/T790M) positively correlates with the immunosuppression signature by upregulation of PD-1, PD-L1, CTLA-4 and proinflammatory cytokines [33]. Expression of actionable mutation (V600E) in BRAF could promote transcription of IL-1α and IL-1β and induce immunosuppression in melanocytes and melanoma cell lines [31]. Such molecular evidence illuminates the potential combinatory effect of classic targeted cancer

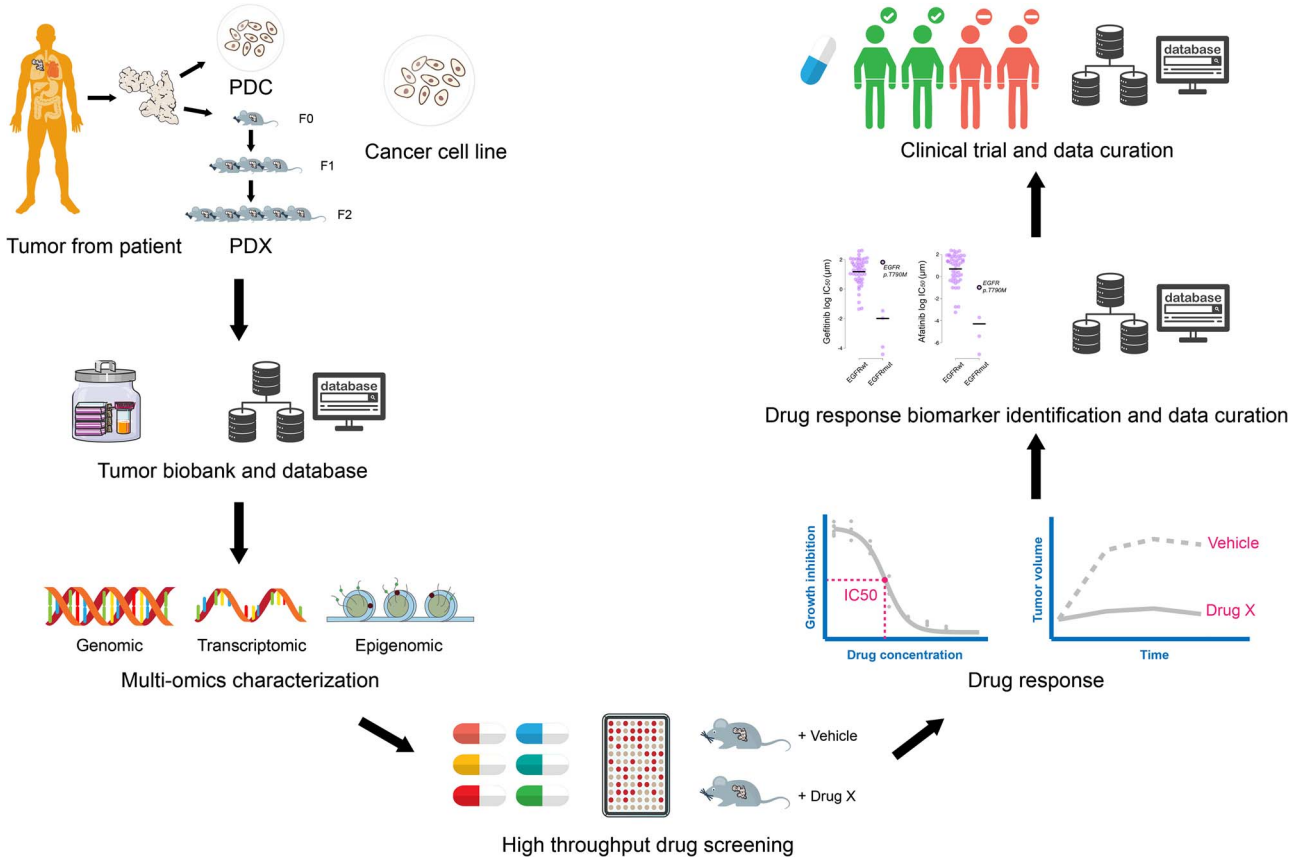

Figure 1. A flowchart showing the discovery of therapeutic biomarkers using comprehensive drug screening on cancer cell lines and PDMs. Cancer specimens are cell cultured or transplanted into immunodeficient mice. *In vitro* and *in vivo* drug screenings are carried out among these established cell or mouse models. Next, multiomics profiles together with drug response data are used to establish drug–biomarker relationship (such as specific mutations). After validation in clinical trials, patients with matched genomic features could be treated with the drug that showed ideal response.

treatment and immunotherapy. However, most current results are based on preclinical models while clinical trials are still ongoing.

Functional genomics studies

Cell line-based drug screening studies

The utilization of human cancer cell lines to study drug response started from the NCI-60 in 1980s [34]. Along with the unprecedented development in sequencing technology, cancer cell line-based drug sensitivity studies incorporating genomic profiling data emerge in succession (Figure 1). Since it would be intractable and prolix to include all relevant studies here, we will focus on the major projects and review them in sequence below.

NCI-60 was the first systematic project to employ cancer cell lines for anticancer drug screening. The project was initiated with the objective to facilitate drug discovery in human solid tumors [34] and was soon expanded to include 60 cell lines representing nine types of cancer. Tens of thousands of compounds were screened in these cell lines to determine their sensitivity or resistance patterns. Furthermore, some algorithms were developed for NCI-60 data analysis and comparison. COMPARE [35] serves to compare the cell growth inhibition patterns when treated with tested compounds with the backend NCI-60 screening database, which reflects the similarity of mechanism of actions (MOA) between compounds and help to elucidate the MOA of novel compounds. Researchers have also correlated tumor molecular profiles with drug response patterns from NCI-60. Weinstein et al. [36] integrated gene expression with drug screening data from NCI-60 to study the molecular pharmacology of cancer. In 2009, Shankavaram et al. [37] developed CellMiner, which compiles the molecular profiles of the NCI-60 cell lines and served as a relational database and query tool. To date, NCI-60 cell lines have gone through comprehensive molecular profiling, and data such as whole-exome sequencing (WES), copy number variation, single-nucleotide polymorphism array, gene expression and reverse-phase protein lysate array are available. However, in general, the number of cell lines included in NCI-60 is relatively small.

The advancement of NGS enables the implementation of large-scale cancer cell line-based drug screening with comprehensive molecular profiling of cell lines (Table 1). In 2012, Barretina et al. [38] first introduced the Cancer Cell Line Encyclopedia (CCLE). CCLE contains both genomic data of 947 cancer cell lines covering 36 tumor types and pharmacological screening data of 24 compounds in 479 of the cell lines. The genomic characterization of cancer lines in CCLE were comprehensive, including gene expression, chromosomal copy number and targeted massive sequencing of 1651 protein-coding genes. Comparison of genomic profiles between cancer cell lines and primary tumors demonstrated their similarity in terms of chromosomal copy number, gene expression patterns and point mutation frequency. With the major objective of identifying

Table 1. Resources of large-scale cell line-based drug screening studies

Name	Number of cell lines with genomic characterization	Number of cell lines with drug screening data	Number of compounds	Molecular characterization	URL	Reference
Cancer Cell Line Encyclopedia (CCLE)	947	479	24	gene expression, chromosomal copy number and targeted massive sequencing of 1,651 protein-coding genes	https://portals.broadinstitute.org/ccle	[38]
Genomic of Drug Sensitivity in Cancer (GDSC) – 1st publication	639	639	130	copy number alteration, gene expression, and targeted sequencing of 64 genes	https://www.cancerrxgene.org/	[39]
Genomic of Drug Sensitivity in Cancer (GDSC) – 3rd publication	1,001	990	265	whole exome sequencing, copy number alteration, gene expression and DNA methylation	https://www.cancerrxgene.org/	[41]
Cancer Therapeutics Response Portal (CTRP) v1	use genomic characterization from CCLE	242	354	use genomic characterization from CCLE	https://portals.broadinstitute.org/ctrp.v1/	[42]
Cancer Therapeutics Response Portal (CTRP) v2	use genomic characterization from CCLE	860	481	use genomic characterization from CCLE	https://portals.broadinstitute.org/ctrp/	[43]

correlations between genomic profiles and drug sensitivity, they further constructed predictive models using machine learning methods (naïve Bayes algorithm and elastic net algorithm), and results showed that both known and novel genomic features were identified as top predictors for drug sensitivity.

Garnett et al. [39] introduced their effort in identifying genomic biomarkers of drug sensitivity in cancer cells at the same time with CCLE. Six hundred thirty-nine cancer cell lines, which were subjected to copy number alteration analysis, gene expression profiling and targeted sequencing of 64 genes, were screened with 130 drugs. These data were then compiled in a web-based resource, the Genomic of Drug Sensitivity in Cancer (GDSC) [40] later by the same group. Recently, Iorio et al. [41] published the latest progress of GDSC, which further expanded their work to include 1001 human cancer cell lines and 265 drugs. Drug sensitivity screening and genomic profiling including WES, copy number alteration, gene expression and DNA methylation were generated for most cell lines. To compare the differences between tumor samples and cancer cell lines, they defined a catalog of clinically relevant oncogenic alterations through analysis of tumor sample data from TCGA and ICGC, and comparison demonstrated that this large panel of human cancer cell lines could well capture the oncogenic alterations identified in tumor samples. To determine genomic biomarkers predictive of drug sensitivity, analysis of variance and logic models were used for identification of single genomic alteration or combination of genomic alterations, respectively. Importantly, by using machine learning models, which employ elastic net and random forest algorithms, gene expression profiles turned out to be the best predictor of drug sensitivity in pan-cancer analysis, while in cancer-specific context, mutated genes and copy number alterations became the major predictors. In 2013, Basu et al. [42] described the Cancer Therapeutics Response Portal (CTRP). In this study, 354 small molecules were screened in 242 genetically annotated human cancer cell lines, which were a subset of the cell lines in CCLE. In 2015, Seashore-Ludlow et al. [43] introduced the updated version of CTRP (CTRP v2). CTRP v2 increased their data set to include 860 cancer cell lines (still a subset of CCLE) and 481 compounds, which was different from CCLE and GDSC; CTRP focused on increasing the number of tested drugs and developing novel analysis methods. In CTRP v1, a sorted-based enrichment analysis and elastic net algorithms were employed respectively to identify genomic biomarkers of drug response, while a new method, the annotated cluster multidimensional enrichment (ACME), was developed in CTRP v2. Compared with previous methods that focused on identifying predictive biomarkers for each compound, ACME aimed to detect correlation between clusters of cancer cell lines with similar genetic or cellular features and clusters of compounds sharing common targets.

Some debates [44] have also been raised on the consistency of drug response data in CCLE and GDSC. After a series of discussions [45–48], researchers reached the consensus that when taking the biological nature of targeted drug screening in cancer cell lines into consideration, these two series of studies reach a reasonable concordance. Several notes for the best practice of experiment and analysis were also put forward for future studies [49].

CRISPR screening studies

Over the past decades, the investigation of cancer drug response by library-based functional perturbations has also experienced rapid development. For example, RNAi screening employs small interfering RNAs or short hairpin RNAs to identify drug response genes and mutation dependencies in cancer [50, 51]. Recently, CRISPR-Cas9-based pooled libraries emerge to be a critical tool used in therapeutic targets discovery and chemical–genetic interactions dissection [52]. For example, Shi et al. [53] introduced CRISPR-Cas9 indel mutagenesis strategy to exons encoding functional protein domains and inferred their functional importance by measuring the potency of negative selection of each perturbation. This study also illustrated that indel mutation-dependent loss of function of specific protein domains allows comprehensive identification of protein units that are suitable for drug targeting. Behan et al. [54], in a recent large-scale CRISPR experiment, screened 324 human cancer cell lines from 30 cancer types and developed a computational framework to prioritize candidates for cancer therapeutics. They comprehensively inferred potential drug targets in defined tissues and genotypes and then uncovered many priority targets with mutation-linked dependencies. Pan et al. [55] combined CRISPR screening and found that the tumors harboring inactivating mutations in a specific SWI/SNF chromatin remodeling complex were more sensitive to T cell-mediated killing, which provides a promising strategy to overcome treatment-resistant tumors using immunotherapy. Although most existing studies cannot directly connect specific mutation effect with drug response, they have narrowed down the therapeutic targets to gene or protein domain level. Given the possibility of large-scale precise CRISPR editing using either homology-directed repair or deaminase base editor, CRISPR screening have been optimized to investigate the mutation-dependent effect on cancer drug resistance. Jun et al. [56] recently leveraged CRISPR RNA-guided deaminase and single-cell RNA-seq technology to dissect single-mutation consequence associated with resistance to vemurafenib in BRAF V600E-mutant melanoma. They designed 420 sgRNAs to introduce precise C to T mutations for all of the exons of three known vemurafenib-resistant genes, *MAP 2K1*, *KRAS* and *NRAS*. By CROP-seq, an experiment to analyze designed perturbation effects with each cell by integrating both the gene expression readout and CRISPR-based perturbations, they validated classical mutations in critical protein domain and identified several novel mutation-dependent transcriptome signatures for conferring resistance to vemurafenib.

The CRISPR-Cas9 system has great efficiency in genome editing, which makes it a powerful tool in studying mutation-dependent cancer drug response. However, it also has several limitations, which may affect its utilization. For example, researchers should pay extra attention to the potential off-target problem, which may lead to false interpretation of mutation effects [57]. And there could also be false positive results caused by abnormal genomic copy number in CRISPR screening studies [58]. Looking forward, CRISPR screening studies will be frequently employed to examine the mechanism of mutation-dependent anticancer drug responses, particularly to investigate the novel therapeutic methods that remit drug resistance during cancer therapy.

PDM studies

Despite the rapid progress in cancer research, only a tiny fraction of compounds can be approved for clinical research by the Food and Drug Administration (FDA). The reason is attributed to lack of clinical cancer models, as well as limitations of available panels of cancer cell lines that directly match with patients' conditions. However, recent advances in PDMs, including patient-derived tumor cell (PDC), patient-derived xenograft model (PDX) and patient-derived organoid (PDO), greatly facilitate the investigation of the association between molecular biomarkers and drug response [59, 60]. Among them, PDC is the easiest to construct and used as *in vitro* model in high-throughput drug screening and cancer research [61, 62]. However, it cannot reflect tumor microenvironment and has the problem of loss of tumor heterogeneity during the culture process, which leads to differences in drug efficacy in following clinical trials [62]. PDX, which is constructed by injecting tumor samples directly obtained from cancer patients into mice, can largely retain genomic characteristics, histological features, molecular diversity, and microenvironment. It is a great model to evaluate the efficacy of anticancer drugs and identify novel mutations associated with targeted cancer therapies (Figure 1). For example, Bertotti et al. [63] constructed PDX models with colorectal cancer samples from 85 patients to test the response of cetuximab. It turned out that the PDX models showed a similar response rate as colorectal cancer patients. HIF-2 antagonist PT2399 was also tested in 20 NOD/SCID PDX models constructed with renal cancer patient samples during preclinical trials. It suppressed tumor growth in 56% of mice models and demonstrated its efficacy [64]. Stewart et al. [65] and Childhood Solid Tumor Network used tumor samples from 168 pediatric patients to establish 67 patient-derived xenografts of 12 types of cancer. Several promising compounds were identified as potential pediatric cancer treatment based on their models. Researchers at Novartis Institute of Biomedical Research constructed about 1000 PDX models with a diverse set of driver mutations. Using these PDX models, small molecular compounds were screened *in vivo* to evaluate the population responses of 62 treatment strategies according to six indicators [66].

In order to facilitate systematic application of PDX models to drug development process, several groups in academia and industry have now attempted to develop a collaborative network for PDX biobanking (Table 2). For example, EurOPDX is a collaborative network of 16 European academic institutions. It consists of more than 1500 PDX models covering over 30 tumors types [67]. NCI launched a Patient-Derived Model Repository (PDMR) database, which consists of clinical annotations of early-passage patient-derived xenografts for quality control [68]. Besides, the NCI-supported US Pediatric Preclinical Testing Consortium (PPTC) is a program of *in vivo* testing of pediatric drug candidates using pediatric cancer PDX models [69]. Public Repository of Xenografts (PRoXe) is an open-source repository of PDXs particularly focusing on leukemia and lymphoma [70]. The children's oncology group (COG) is a clinical trials group, and they construct a cell culture and xenograft repository to provide validated cell lines and PDXs based on pediatric cancer patient samples [71]. In addition, Bruna et al. [72] established 83 PDX models of breast cancer, and they were combined with

Table 2. Major repositories of patient-derived xenograft model

Name	Institutions	Cancer type	Number of PDX models	Molecular characterization	Open-access database	URL	Reference
EUROPDX	16 European academic institutions	more than 30 different solid tumor types	1,498	WES, gene expression, CNA	Yes	http://europdx.eu/	[67]
NCI Patient-Derived Models Repository (PDMR)	US National Cancer Institute	more than 17 different solid tumor types	2,039	WES, gene expression	Yes	https://pdmr.cancer.gov/	[68]
US Pediatric Preclinical Testing Consortium (PPTC)	US National Cancer Institute	childhood cancer solid tumor and leukemia	61	RNA expression	Yes	http://www.ncipptc.org/	[69]
Public Repository of Xenografts (PRoXe)	Weinstock Laboratory	leukaemia and lymphoma	157	gene expression	Yes	https://www.proxe.org/	[70]
Children's Oncology Group (COG)	Texas Tech University Health Sciences Center	childhood cancer	NA	DNA microarray	No	https://www.cccells.org/xenografts.php	[71]
Breast Cancer PDTX Encyclopaedia (BCaPE)	Carlos Caldas' laboratory	breast cancer	104	gene expression, CNA, DNA methylation	Yes	https://caldaslab.cruk.cam.ac.uk/bcape/	[72]

PDC models of tumors to construct an integrated platform. This platform was used for both high-throughput single drug, drug–drug combination screening studies and *in vivo* drug response testing. An open-access database, the Breast Cancer PDTX Encyclopaedia, was constructed to make these data freely available. The Brain Tumor Resource Laboratory provides the first comprehensive resource for patients with pediatric brain tumors to better understand the basis of sensitivity or resistance to cancer therapies, particularly in rare brain tumor subtypes. It consists of 30 PDX models and three independent cell lines, which are representative of multiple molecular subgroups of malignant pediatric brain tumors [73]. These PDX model biobanks are valuable resources for preclinical cancer pharmacogenomic studies. Combination of the PDX biobanks with unique therapeutic strategies will promote the identification of novel predictive biomarkers and eventually facilitate the progress of precision cancer medicine.

Although PDX has many outstanding properties as a cancer model, it cannot retain intratumor heterogeneity very well. Intratumor heterogeneity, which results from intercellular genetic variation, genomic instability of tumor cells, selection pressure of microenvironment, disease progression and drug treatment [74], can accelerate the occurrence of drug resistance, and it is thus highly important to maintain this heterogeneity in cancer models to better mimic the characteristics of tumor samples [75]. PDX is also not suitable for high-throughput drug screening because of its high cost and long establishment time. Compared with PDC and PDX, PDO is able to retain the genomic, histological and morphological characteristics of tumor samples, including intratumor heterogeneity, while it allows for *in vitro* high-throughput drug screening [76]. PDOs are derived from patient samples, and it can self-organize in three-dimensional culture in a short time due to their self-renewal and differentiation capacities. More and more scientists build large biobanks of patient-derived tumor organoids that can be used to perform drug screening and assess drug response, including bladder cancer [77], colorectal cancer [78], gastric cancer [79], ovarian cancer [80], etc. Although PDO has many advantages over other PDMs, as a newly developed technology, it can only be constructed by a few laboratories. In addition, retaining tumor heterogeneity and tumor purity is a tradeoff when constructing PDOs and PDXs, which could be affected by many factors such as sample source, sampling region, tumor culture platform as well as tumor evolution.

Computational studies

Classification and resources for actionable mutations

To associate mutations with proper targeted therapies and define their actionability, multiple groups have put forward their classification schemes [2], which are similar to some extent. Most schemes [12, 22, 23, 81–83] would classify mutations into three to five groups with different levels of actionability according to an integration of existing evidences. The top level means that these mutations are actionable with high confidence, and there is usually substantial clinical evidence involved, like FDA guidelines and results from successful clinical trials. Such

mutations may be considered as good candidates for follow-up clinical trials. The medium level indicates that mutations are potentially actionable with evidence from preclinical models and experiments, and these mutations could be involved in early-phase signal-seeking trials. The bottom level usually contains all other mutations with little or no evidence about their actionability. These various schemes provide practical ways for implementation, and some classification approaches have already been applied in recent clinical trials. However, a well-accepted classification standard of actionable mutations remains to be ascertained through continuous practices as well as multicenter collaborations across the world to reach a consensus.

Following the rapid data accumulation from clinical and preclinical studies, several research groups and institutions have developed multiple knowledge base to curate and format this information. The early work started in 2011; Vanderbilt-Ingram Cancer Center launched the My Cancer Genome online resource aiming to provide up-to-date information of clinical relevance of mutations. Later in 2013, Yeh et al. [84] described the DIRECT as part of the My Cancer Genome knowledge resource. My Cancer Genome records common mutations in cancer and curates their effects on therapeutic responses from literature. Mutations are arranged by cancer type (23 types of cancer) and gene, and each webpage contains information of therapeutic response that this mutation may influence with all information fully referenced. Relevant clinical trials are also provided for each gene. In 2015, Johnson et al. [85] from MD Anderson Cancer Center introduced their Precision Oncology Decision Supports (PODS) platform. It is a precision oncology framework containing the function of determination of actionable mutations, identification of optimal available targeted therapy for genomic alteration of interest and evaluation of the confidence level for each treatment option. In this article, they also described the Personalized Cancer Therapy (PCT), which is an open-access web portal making part of the PODS available to the public. Data of PCT are sorted by actionable genes (26 genes), and each gene page includes actionable genomic alterations, relevant drugs, levels of evidence of the association between drugs and genomic alterations in tumor type-specific context and relevant clinical trials. In addition, information has undergone expert review and provided with references. Also in 2015, Dienstmann et al. [86] described the Gene-Drug Knowledge Database (GDKD). They manually curated associations between predictive genomic biomarkers and cancer drugs and translated the data into a structured format. Each association is an independent item in the database, and it consists of tumor type, gene, variant, variant effect to drug response, etc., and PubMed identifier is provided as a reference. The database is available on Synapse in the format of Excel tables and is regularly updated. It is worth mentioning that GDKD actually integrates information from other public databases, including My Cancer Genome and PCT, but My Cancer Genome and PCT are intended for data query while GDKD provides all its data in an Excel file, which makes it feasible to be incorporated into user-specific analysis pipelines.

Since 2015, there is a surge of open-access databases development, such as Cancer Driver Log (CanDL) [87], JAX Clinical Knowledgebase (JAX-CKB) [88], Precision Medicine Knowledgebase [89] (PMKB), Clinical Interpretation of Variants in Cancer (CIViC) [90], OncoKB [91], Cancer Genome Interpreter (CGI) [92] and Database of Evidence for Precision Oncology (DEPO) [93]. All databases are summarized and compared (Table 3), and some of their special features are discussed below. CanDL is a web-based database, in which functionally characterized driver mutations were manually curated by literature mining. It aims to provide mutation actionability annotation and does not contain drug information in its database. PMKB distinguishes itself from other resources by its emphasis on clinical usage. The database is developed in close collaboration with pathologists and aims to be applicable to clinical reporting. For example, it will report whether a variant is pertinent negative under specific tumor type. PMKB supports both bulk download and access through application programming interface (API), and through collaboration with hospitals and research institutions, it has been constantly used after targeted sequencing and WES, which proves its practicability in a clinical context. CIViC is a web-based knowledge base. Every clinical interpretation of variants in cancer is called an evidence record, which is the basic unit in CIViC, and each evidence record is described by a series of structured attributes, including gene, variant, disease, drug and clinical relevance of variant. CIViC differs from other databases in the way of data curation and its efforts to make the data easily accessible. It is designed to be a crowdsourced community to receive public submissions from users, and the data submitted will then be expert reviewed and transformed into a structured record if accepted. Furthermore, CIViC is open-access and open-source with public API and regular bulk download releases. OncoKB consists of similar data as CIViC and supports various types of data access methods. All data were collected and maintained by their own knowledge systems group and curators. A special feature of OncoKB is that it annotates more than 4000 genomic alterations with their mutation effect and oncogenicity. However, only around 200 items of potential treatment implications related to these annotated genomic alterations were recorded in their database. DEPO is also a database of curated drugable variants with potential clinical implications. The data type is pretty much the same as other databases, but DEPO integrates mutation functional annotation to inspect whether an input mutation is close to known drugable mutations spatially. It also enables users to visualize these potentially drugable sites through its web portal. In addition, the Catalogue of Somatic Mutations in Cancer (COSMIC) have also started adding drug resistance information as one of the mutation annotations since its v77 release in May 2016. This work is still underway with constant updates [94–96].

There are also databases focusing on integration of cancer cell line-based compound screening data and molecular profiling data. Cancer Drug Resistance Database [97] compiles pharmacological profiling data and drug targets mutation status data of cancer cell lines from CCLE and COSMIC and provides tools such as a clustering module and an NGS mapping tool to help study mutations responsible for drug resistance. While Gohlke et al. [98] integrates genomic, transcriptomic and pharmacological data from CCLE, COSMIC and CellMiner to develop CancerResource. CancerResource also includes drug–target interaction data and pathway information to make itself a comprehensive cancer cell line database.

Drug sensitivity prediction and biomarker identification

Together with the tremendous amount of molecular and pharmacological profiling data generated from the clinical and functional studies, various kinds of bioinformatics tools, methods and algorithms are developed to tackle the problem of drug sensitivity prediction and biomarker identification in cancer (Table 4). As there are numerous methods working on these two

Table 3. Databases of cancer-related clinical implications of genomic alterations

Name	Institutions	Curation method	Bulk download	URL	Year of most recent update	Reference
Cancer Driver Log (CanDL)	The Roychowdhury Lab Team at The Ohio State University	expert curation	Yes	https://candl.osu.edu/	2015	[87]
My Cancer Genome	Vanderbilt-Ingram Cancer Center	expert curation	No	https://www.mycancergenome.org/	2016	[84]
Personalized Cancer Therapy (PCT)	The University of Texas MD Anderson Cancer Center	expert curation	No	https://pct.mdanderson.org/	2019	[85]
Gene-Drug Knowledge Database (GDKD)	DIENSTMANN *et al.* at Sage Bionetworks, Fred Hutchinson Cancer Research Center	expert curation	Yes	https://www.synapse.org/#!Synapse:syn2370773/wiki/	2017	[86]
JAX Clinical Knowledgebase (JAX-CKB)	The Jackson Laboratory	expert curation	No	https://ckb.jax.org/	2019[a]	[88]
Precision Medicine Knowledgebase (PMKB)	Institute of Precision Medicine at Weill Cornell Medical College	crowdsourcing + expert review	Yes	https://pmkb.weill.cornell.edu/	2019	[89]
Clinical Interpretation of Variants in Cancer (CIViC)	The McDonnell Genome Institute at Washington University School of Medicine	crowdsourcing + expert review	Yes	https://civicdb.org/home	2019[a]	[90]
OncoKB	Knowledge Systems group at Memorial Sloan Kettering Cancer Center	expert curation	Yes	https://oncokb.org	2019	[91]
Cancer Genome Interpreter (CGI)	Barcelona Biomedical Genomics Lab at Institute for Research in Biomedicine	expert curation + community feedback	Yes	https://www.cancergenomeinterpreter.org/home	2018	[92]
Database of Evidence for Precision Oncology (DEPO)	Ding Lab at Washington University in St. Louis	expert curation	No	http://depo-dinglab.ddns.net/	2018	[93]
Catalogue Of Somatic Mutations In Cancer (COSMIC)	Wellcome Sanger Institute	expert curation	Yes	https://cancer.sanger.ac.uk/cosmic	2018	[94–96]

[a]These databases are updated daily.

related topics, here we only review some classic studies and studies with special features.

Accurate prediction of drug sensitivity using genomic features of tumors has long been the objective of researchers. Due to the limitations of availability of tumor samples, high cost and time required, it is usually unfeasible to build drug response prediction models based on patient data. Thus, cell line-based data became the common substitutions in most prediction analysis (Figure 2). In 2012, a systematic comparison of drug response prediction algorithms was carried out through the collaboration of NCI and DREAM [14]. This NCI-DREAM challenge provided participants with genomic, transcriptomic, epigenomic and proteomic profiles of human breast cancer cell lines and drug screening profiles for a part of the cell lines. The remaining drug sensitivity data were reserved for performance evaluation. Forty-four teams from all over the world attended this project and submitted diverse types of solutions, and the top-performing method used a Bayesian multitask multiple kernel learning (MKL) algorithm. Through analysis of all submitted methods, it was shown that most top-performing solutions would model the nonlinear relationship in the data set. In addition, of all profiles provided, gene expression was reported to be the most informative data type. However, when using MKL or elastic net as the testing model, combination of two different types of data would significantly improve the prediction performance, and the best-performing MKL method actually integrated all types of data and additional outside information such as biological pathways. Similar attempts were made in other studies [38, 39, 42] like CCLE, GDSC and CTRP, as mentioned above. These studies emphasize the importance of integrating various types of data, including genomic, transcriptomic, epigenomic and proteomic data, into the prediction model to different extent. As mentioned above, gene expression was often regarded as the most informative mark. As a result, it would explain most of the variation of the drug sensitivity in the final model, which sometimes makes the model hard to interpret. Aben et al. [99] developed a two-stage approach, TANDEM, which first predicts the drug sensitivity using molecular data except gene expression and then predicts the residuals according to gene expression data. TANDEM was tested using molecular and pharmacological data from GDSC, and it was shown that, when using TANDEM, a larger proportion of variation would be explained by data except gene expression while retaining similar predictive power as the conventional approach. Models

Table 4. Summary of drug response prediction and biomarker identification methods utilizing mutation information

Name	Major statistical methods	Method description	Feature	Limitation	Prior objective	Molecular data type	Drug response data type	Year of publishment	Reference
Menden et al. 2013[100]	Neural network + random forests	Network with the resilient error backpropagation implementation, then used random forests for comparison	Drug chemical properties were included as predictors	Only cell line data were used	Drug sensitivity prediction	gene mutation, CNA, microsatellite instability status, drug chemical properties	cell line	2013	[100]
Geeleher et al. [103, 104]2014, 2017	Ridge regression + linear regression	Ridge regression for drug response imputation, linear regression for associating drug response with biomarkers	Cell line data were used to generate models to impute clinical drug responses and thus improve biomarker identification	Validation of new IDWAS findings using independent clinical data is needed	Clinical drug response prediction	gene mutation, CNA, gene expression	cell line + clinical	2014, 2017	[103, 104]
TANDEM	Two step elastic net	TANDEM first regressed drug sensitivity on upstream data types, then regressed the residuals of drug sensitivity on gene expression	Two step regression method improved the interpretability of predictors	Only cell line data were used	Improvement of model interpretability	gene mutation, CNA, DNA methylation, gene expression	cell line	2016	[99]
Ding et al. 2016[102]	Elastic net + boot strap, Ensemble model of logistic regression	Elastic net and bootstrap for feature selection, the final model was built by ensemble model of logistic regression	Clinical drug response data from TCGA were collected and used	Data contained only four chemothera-peutic drugs	Comparison of predictive utility of various molecular data	gene mutation, CNA, DNA methylations, mRNA expressions, miRNA expressions, protein activities	clinical	2016	[102]
CARE	Multiple linear regression with interaction term	For each drug, generate a genome-wide gene signature score vector using multiple linear regression with interaction term based on its target gene and each objective gene. The gene signature score vector of the sample is then correlated with the gene expression vector of the sample to make predictions	Interactions between genes were considered	Only cell line data were used	Biomarker identification considering gene interactions	gene mutation, gene expression	cell line	2018	[106]

Table 4. (Continued.)

Name	Major statistical methods	Method description	Feature	Limitation	Prior objective	Molecular data type	Drug response data type	Year of publishment	Reference
Menden et al. 2018[101]	Two step elastic net	First regress drug sensitivity on somatic mutations, then regress the residuals of drug sensitivity on germline variants	Germline variants were included as predictors	Germline variants are weak predictors compared with gene expression data	Drug sensitivity prediction and biomarker identification	single variant mutation, CNA, gene fusion, germline variants	cell line	2018	[101]
MAGNETIC	Network analysis	First identify gene modules by network analysis and clustering algorithm in clinical data, then find the preserved gene modules in cell line data, finally associate gene modules with drug sensitivity	Gene modules preserved in both clinical data and cell line data were identified by network analysis	Systematic evaluation of the assumption that preserved modules could constitute biomarkers that are more likely to translate in clinical data is needed	Preserved biomarker identification	gene mutation, CNA, gene expression, DNA methylation, RPPA	cell line + clinical	2018	[105]
DeepDR	Deep neural network	Consist of a mutation encoder, expression encoder and a prediction network, synaptic parameters of the mutation encoder and the prediction encoder were initialized by results from pretraining encoders based on TCGA data	Deep learning method was used	Systematic comparison with other advanced methods besides linear regression and SVM is needed	Drug sensitivity prediction	gene mutation, gene expression	cell line + clinical	2019	[107]

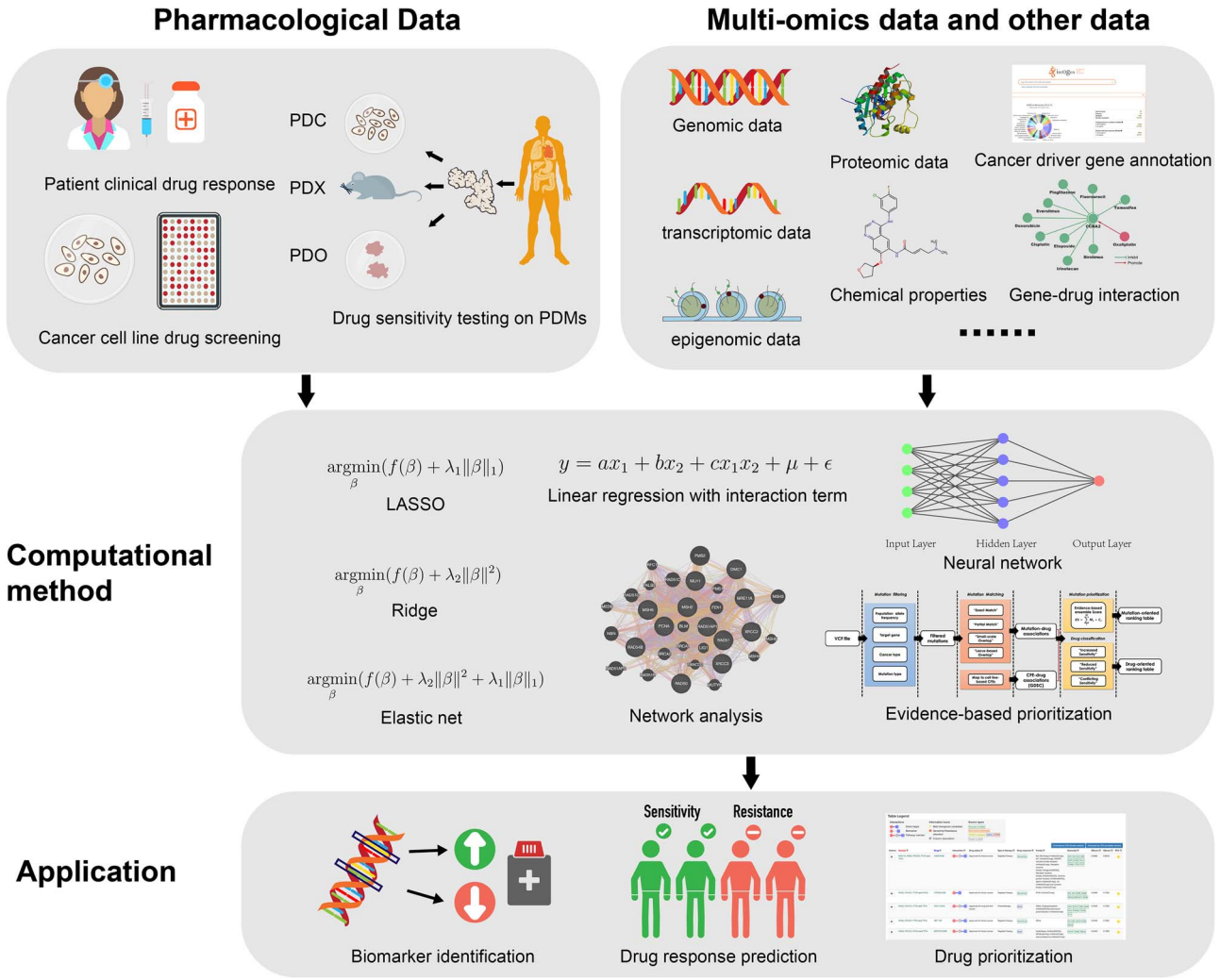

Figure 2. Summary of the workflow of computational studies. Most computational studies would take a subset of pharmacological data, multiomics data and other data as input, use certain computational methods to model the data and finally get output including biomarker identification, drug response prediction and drug prioritization. Here, we summarized all pharmacological data, multiomics data and other data as well as computational methods and applications that were related to the studies reviewed in our article.

produced by TANDEM thus became more interpretable. The integration of other types of data would provide additional predictive power. Menden et al. [100] described a machine learning method to predict drug sensitivity in preclinical cancer cell line models by using both cell line genomic features and drug chemical features. Neural network was used to handle the large amount of input features, and the model could predict drug sensitivity values with coefficient of determination R^2 of 0.72 for cross validation and 0.64 for an independent test data set. More recently, they reported the integration of germline variants data in the prediction of cancer drug response. Compared with using somatic mutation alone or a combination of somatic mutations and gene expression profiles, the combination of somatic mutations and germline variants could make better predictions in some drugs, and the prediction information from germline variants partially overlapped with that from gene expression data [101].

Another important question is how the prediction methods perform on clinical tumor sample data and how to integrate cancer cell line data with clinical data (Figure 2). Ding et al. [102] constructed predictive models with molecular data and clinical drug treatment data from TCGA. The prediction performance of genomic, epigenomic, transcriptomic and proteomic data on clinical drug response was compared. The major limitation of this study is the number of drugs involved and the diversity of their MOA. Due to the availability of clinical cancer treatment data, only four drugs were selected in this study, and all of them are chemotherapeutic drugs. Such a situation indicates the necessity for more large-scale clinical trials in the area of precision cancer medicine. Geeleher et al. [103] report a very interesting method to tackle the lack of clinical drug response data. Previously, they have developed a method, which constructed ridge regression models using gene expression and drug sensitivity data from cancer cell lines, to predict clinical drug response [104]. Based on this method, they further developed an imputed drug-wide association study to first impute drug response of 138 drugs in over 10 000 TCGA tumor samples and then associate the imputed drug response with genomic alterations of tumor samples. Both known and novel biomarkers predictive of drug response were identified, and one newly discovered biomarker was validated by experiments in cancer cell line. The comprehensive imputation improved the ability

to identify biomarkers of drug response since the imputation could increase sample size and many genomic alterations are inadequately represented in cancer cell line data. However, most findings still remain unvalidated. The combination of cancer cell line data and clinical data could also be the other way around. Webber et al. [105] proposed modular analysis of genomic networks in cancer (MAGNETIC), which conduct network analysis integrating genomic, transcriptomic, epigenomic and proteomic data to identify gene modules. Their analysis started from identifying modules in TCGA data and then whether the discovered gene modules were preserved in cancer cell lines were assessed. The preserved gene modules were shown to be more robust biomarkers of drug response than genes in preserving the interrelationship when translating from cancer cell line data to clinical data. It was also demonstrated that models based on cell line data and gene modules had better performance when applied to PDX data compared with that based on genes. This method showed a way to tackle the transferability problem between findings based on cell line data and clinical data. It also highlights the importance of rational summarization of molecular profiles by network analysis. Gene-based predictive biomarker was also seen in other studies. Computational analysis of resistance (CARE) would generate a genome-wide gene signature for each targeted cancer drug based on cancer cell line data as a biomarker to predict drug response. It differs from others in that when inferring the gene signature score for each gene, CARE would consider the interaction effect between the objective gene and the target gene of the targeted cancer drug. The genome-wide gene signature from CARE was also proved to be superior to that from experimental methods and other computational methods in cell line and clinical data [106]. Finally, deep learning, as the hottest machine learning method in recent years, was also applied in the prediction of cancer drug response. Chiu et al. [107] presented DeepDR, which uses deep neural network to predict cancer drug response according to genomic and transcriptomic data. More precisely, it consists of three neural networks, a mutation encoder and a gene expression encoder for high-dimensional data abstraction and a prediction network for final prediction. Similar to Geeleher et al.'s work and MAGNETIC, DeepDR would also combine both cell line and clinical data in the model training process during which TCGA data were used to initialize the synaptic parameters of the two encoders. When compared with linear regression, support vector machines (SVM) and two variants of DeepDR, DeepDR showed better prediction performance and stability.

Although the above studies are all about drug sensitivity prediction and biomarker identification in cancer, they have different objectives and focus on making improvements in different directions and thus have their own strengths and limitations (Table 4). Researchers should be clear about the features of these methods to make the most appropriate choice before application in their own studies.

In silico drug prioritization

Although cancer treatment has experienced great progress in recent year, clinicians often have very few options for patients resistant to current stand of care, and many cancer-specific mutations have unknown clinical implications [108]. In addition, in the era of precision medicine, it is desirable to select the most appropriate therapy for patients by making use of the rapidly accumulating multiomics patient profiling data. As a result, *in silico* drug prioritization, sometimes referred to as *in silico* drug prescription, becomes an important topic in cancer treatment studies. By making use of patient molecular profiles, gene–drug interactions, functional annotations of cancer genes and mutations as well as other types of data, *in silico* drug prescription methods could prioritize drugs for individual patient. It is notable that, in the process of *in silico* drug prioritization, sometimes the original indication of the top prioritized drug is not the same as the patient cancer type, which means the original indication could be another type of cancer or noncancer diseases. In this case, the drug prioritization process is also doing drug repositioning [109]. In the following section, some examples of *in silico* drug prioritization methods based on individual molecular profiles are introduced (Figure 2 and Table 5).

IMPACT [110] is an analysis pipeline to integrate somatic mutation calling and drug prioritization together. It served as a one-stop tool to connect patient molecular profiles with actionable therapeutics, though the drug recommendation was made mainly by drug–target interactions and only the FDA-approved mutation drug sensitivity interactions were used. Drug prioritization could be evidence based, for example, Li et al. [111] comprehensively integrated mutation–cancer drug sensitivity associations from various resources and provided evidence-based scoring scheme to score actionable mutations and drugs. Another more complex example is PanDrugs [112], which prioritizes anticancer therapies based on individual-level genomic data. PanDrugs mainly relies on genes and drugs as basic analysis unit and supports a gene list or a patient's mutation profile as input. Gscores are first calculated for all the input genes if the input is a gene list, and Vscores are calculated and then transformed to Gscores if the input is a mutation profile. The Gscore is generated by considering the gene's frequency in tumorigenesis process, potential to be a tumor driver gene, essentiality score and oncogenic score from OncoScape. Then Dscores are calculated for each drug related to the input genes based on drug indication, drug approval status, gene–drug relation type, number of genes targeted in the input list and number of curated sources that support the gene–drug relationship. The final drug prioritization list is generated relying on both Gscore and Dscore. Specifically, PanDrugsdb, the internal database of PanDrugs, contains gene annotations, drug annotations, comprehensive gene–drug interactions integrated from multiple databases and mutation drug response associations. In addition to using genomic data alone, Kalari et al. [113] described their multiomics-guided drug prioritization method PANOPLY, which was based on individual patient's multiomics data. PANOPLY identifies case-specific multiomics events by comparing the input patient data and matched controls. These case-specific multiomics events are used to generate a drug ranking score by network analysis and a random forest score by using the importance score from random forest for each drug. All drugs are then prioritized based on these two scores. Another study by Rubio-Perez et al. [109] defines the process of precision drug prescription systematically. The author presented a very comprehensive work consisting of cancer driver events identification, associating drugs with cancer driver events and prescription of drugs to patients based on their genomic profiles. The first step used three methods to identify complementary cancer drivers and constructed a Cancer Drivers Database, while the second step considered three possible situations to link anticancer therapy with cancer driver events and constructed a Cancer Drivers Actionability Database. Taken together, this work provides rich resources and serves as a classic paradigm for later studies on both mutation actionability annotation and drug prioritization. Later in 2018,

Table 5. Summary of *in silico* drug prioritization methods utilizing mutation information

Name	Prioritization method description	Feature	Limitation	Related database	Year of publishment	Reference
Rubio-Perez et al. 2015	First define the Cancer Driver Database and the Cancer Driver Actionability Database. Then driver alterations of patients are selected according to the Cancer Driver Database and therapies are assigned to patients on the basis of the driver alterations and the rules in the Cancer Driver Actionability Database	Very comprehensive study contains the whole process of *in silico* cancer drug prescription including cancer driver events identification, associating drugs with cancer driver events and prescription of drugs to patients based on their genomic profiles	No drug prioritization tools provided	Downloadable from website	2015	[109]
Cancer Genome Interpreter (CGI)	First the Cancer Biomarkers Database and the Cancer Bioactivities Database are defined. Then input genomic alterations are matched with biomarkers in the databases.	During the process of biomarker matching, information like biomarker co-occurrence, level of detail of biomarkers are considered. In addition, drug repurposing including biomarker for a different tumor type and different biomarkers with the same putative effect are considered.	The drug prioritization method is rather simple	Downloadable from website	2018	[92]
IMPACT	First tier contains drugs found by linking identified variants with FDA-approved drug indication information from NCI-Match and PCT. Second tier contains drugs prioritized by enrichment analysis of drug target genes in the variants list	IMPACT presents a whole pipeline from somatic variants calling to drug prioritization.	The drug prioritization method is rather simple	Downloadable from website	2016	[110]
mTCTScan	Evidence-based prioritization score based on variant matching types and records confidence levels	Comprehensive mutation-drug sensitivity association data collection and automatic data retrieval based on input VCF file	The drug prioritization method is rather simple	Implemented internally	2017	[111]
PANOPLY	Patient-specific multi-omics events are identified by comparison with matched controls. On the basis of patient-specific multi-omics events, the drug ranking score is generated by network analysis and the random forest score is generated by importance score from random forest analysis for each drug respectively. Drugs are prioritized by the drug ranking score and random forest score.	Integration of multi-omics data and identification of patient-specific multi-omics events by comparison with matched controls.	Appropriate matched control data are hard to find	Downloadable through R package	2018	[113]
PanDrugs	Gscores are calculated for input genes (or Vscores for input variants and then transformed to Gscores) based on gene's biological relevance to cancer. Dscores are calculated for drugs related to input genes based on drug properties and gene-drug relationship. Drugs are finally prioritized according to both Gscores and Dscores.	Pathway information is taken into consideration when constructing gene-drug relationship. Multiple genomic events are considered simultaneously when performing drug prioritization. Comprehensive drug-gene association collection.	NA	Implemented internally	2018	[112]

the same group developed CGI as a platform for automatic comprehensive annotation of genomic alterations [92]. Based on their previous work, CGI similarly defined databases including the Catalog of Cancer Genes, the Catalog of Validated Oncogenic Mutations, the Cancer Biomarkers Database and the Cancer Bioactivities Database and developed an oncogenicity prediction method, OncodriveMUT. Specifically, CGI provides *in silico* drug prescription function by matching input genomic alterations with biomarkers in the Cancer Biomarkers Database to suggest potential drugs. The *in silico* drug prescription process would take conditions like co-occurrence of multiple biomarkers and confidence level of the biomarker into consideration and support some drug repurposing strategies.

Discussions

In the past decades, great development has been seen in precision cancer medicine. Large-scale clinical trials and preclinical experiments are reported in succession, which provides extremely rich resources for developing mutation-dependent cancer treatment. However, to date, only a fraction of genomic alterations is confirmed clinically actionable, showing a deficiency in translating the abundant resource into more novel findings. One reason could be attributed to the lack of well-accepted standard to define and record the clinical relevance of mutations. Since precision medicine and companion diagnostics remain in preliminary stage, most trials are not originally designed for identifying the clinical relevance of mutations. Such a situation leads to inefficiency in extracting information from literature. Most researchers [84–86, 90] have to curate data manually from literature, and this often results in redundancy and inconsistency of the data extracted. Several groups [114] are cooperating in this direction, and hopefully more standards could be established soon.

Another major hindrance is the lack of well-characterized molecular profiles and treatment response information of patients. As mentioned in many model construction studies, predictive models based on clinical data were usually deficient in predictive power because of the limited sample size. There are multiple ongoing clinical trials specifically designed for testing molecular biomarkers of drug response, and more insights shall be revealed upon their completion. In the meantime, experiments based on PDM models may be a good substitute and supplement for clinical trials. Conducting drug screening and pharmacodynamic test prior to clinical trials can substantially improve the success rate of following studies. However, researchers should take the advantages and disadvantages of the three types of PDM models into consideration before conducting drug screening experiments. PDC models are easy to construct and culture and require short proliferation period but are inferior in retaining genomic and biological characteristics of patient tumors. PDX can well preserve genomic characteristics, histological features, tumor heterogeneity as well as microenvironment, although there are cases where the drug response results from PDX models cannot match that from clinical patients. The difference may be due to changes in clonal composition of the tumor model because of tumor evolution. Further studies are needed to assess the cancer genome evolution of PDX models and evaluate the differences of tumor microenvironment between PDX models and patient tumor samples systematically. The major limitations of PDX models are the high-cost, the long construction time and the resources needed for maintenance, which makes it unsuitable for high-throughput drug screening. For example, a PDX model suitable for preclinical study usually requires 4–8 months to develop [67]. PDO models are a recent developed alternative for PDC and PDX. They are capable of preserving tumor characteristics while allowing for high-throughput drug screening. However, as an *in vitro* model, the tumor blood vessels and immune component cannot be well reflected by PDOs. There are recent studies working on the co-culture of PDOs with blood vessels and immune cells [115], and xenotransplantation of PDOs into mice [80]. In brief, the PDO model is still in its early stage and remains to be further developed.

Algorithm development in the prediction of drug sensitivity has experienced much progress in recent years. In this process, it was shown that the integration of multiomics data can significantly improve model performance. The NCI-DREAM program drew the conclusion that the combination of multiple data types can improve prediction performance through a systematic comparison of prediction models from 44 teams [14]. Further studies revealed that the integration of chemical properties of drugs can be helpful [100] and how to better interpret the model when multiple types of data are used as predictors [99]. There are also methods to prioritize cancer drugs based on patients' multiomics data such as PANOPLY [113]. At current stage, the results from these computational methods could serve as auxiliary information for clinical trial patient selection and therapy prescription. They also provide crucial insight in mechanism explanation and experimental biomarker validation of cancer drug response. Nevertheless, most findings were made based on *in vitro* data generated by models such as cancer cell lines. The translation of *in vitro* findings to *in vivo* application is still an important yet intractable topic. Also, there is no reliable method to predict the clinical significance of novel mutations, meaning that clinical oncologists are often provided with no treatment options for those patients with novel mutations. For *in silico* drug prioritization, primary and secondary drug resistance, potential side effects and intratumor heterogeneity bring difficulties for effective drug prioritization. Following methods would benefit from integrating these information as well as multiomics data in their prioritization tools. As bioinformaticians, we believe that the integration of well-characterized multiomics data in prediction models should be the trend, and advanced machine learning algorithm such as deep learning can help in undiscovered biomarker identification and clinical relevance prediction by fully exploiting the numerous features provided by this unprecedented large amount of data.

Mutations in protein coding regions are the major focus of this review. However, there are other types of molecular biomarkers that are predictive of anticancer drug response, such as noncoding mutations and noncoding RNAs (ncRNA). Mutations in noncoding regions of CUL3 locus were associated with resistance to vemurafenib in melanoma through CRISPR-Cas9 screening [116], and increased level of miR-100 and miR-125b was associated with resistance to cetuximab in colorectal cancer [117]. However, compared with protein coding mutations, the clinical implications of noncoding mutations and ncRNAs in cancer treatment are yet to be fully described. As important components of the landscape of patient molecular profiles that affect anticancer drug responses, we expect more studies on illuminating the roles of noncoding mutations and ncRNAs in cancer treatment in the following years.

The study of mutation-dependent cancer treatment is still in its infancy, but many studies have shown promising results for its application in clinical cancer treatment. Looking forward, we expect the accumulation of large-scale patient-level multiomics data and treatment data by projects like ICGC-ARGO will pro-

mote our ability to identify promising actionable mutations and biomarkers of drug response. And the fast-developing PDMs will be able to provide personalized preclinical models, which better mimic patient tumor samples. Along with the development of molecular profiling technologies and computational models, hopefully we will see more progress in precision cancer medicine in the near future.

Key Points

- This article reviews studies to analyze associations between mutations and cancer drug response covering clinical studies, functional genomics studies and computational studies. It particularly emphasizes related bioinformatics methods and tools, including resources for actionable mutations, drug sensitivity prediction as well as *in silico* drug prioritization.
- There are a number of databases and knowledge bases that collect the associations between mutations and drug response. However, harmonization of different resources is in need both for the mutation actionability classification scheme and for the data themselves.
- Drug testing on PDMs are good substitutes of clinical trials considering the time and cost. A number of drug testing have been conducted on PDCs and PDXs, but they both have problems of retaining tumor heterogeneity of the original samples. The newly developed PDO model shows great potential but need to be further studied.
- In addition to somatic mutations, most computational methods would integrate various types of data, such as transcriptomic data and epigenomic data, to improve the accuracy of drug response prediction and drug prioritization.
- There are rich resources of preclinical drug screening data but limited clinical trial data. To connect preclinical data with clinical data, many studies have shown the consistency and transferability between them. However, we still need more systematic studies to comprehensively reveal the consistency and inconsistency between them and to establish a general pipeline consisting of preclinical testing (PDM-based), computational analysis and clinical validation for precision cancer treatment.

Acknowledgments

We appreciate all method and resource developers.

Funding

This work was supported by grants from the National Natural Science Foundation of China (31701143 and 31871327 to M.J.L.) and Natural Science Foundation of Tianjin (18JCZDJC34700 to M.J.L.).

References

1. Schmitt MW, Loeb LA, Salk JJ. The influence of subclonal resistance mutations on targeted cancer therapy. *Nat Rev Clin Oncol* 2016;**13**:335–47.
2. Carr TH, McEwen R, Dougherty B, *et al*. Defining actionable mutations for oncology therapeutic development. *Nat Rev Cancer* 2016;**16**:319–29.
3. Mok TS, Wu YL, Thongprasert S, *et al*. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* 2009;**361**:947–57.
4. Mitsudomi T, Morita S, Yatabe Y, *et al*. Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG 3405): an open label, randomised phase 3 trial. *Lancet Oncol* 2010;**11**:121–8.
5. Maemondo M, Inoue A, Kobayashi K, *et al*. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *N Engl J Med* 2010;**362**:2380–8.
6. Kobayashi S, Boggon TJ, Dayaram T, *et al*. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2005;**352**:786–92.
7. Pao W, Miller VA, Politi KA, *et al*. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med* 2005;**2**:e73.
8. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, *et al*. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* 2005;**353**:1659–72.
9. Gagan J, Van Allen EM. Next-generation sequencing to guide cancer therapy. *Genome Med* 2015;**7**:80.
10. Tomczak K, Czerwinska P, Wiznerowicz M. The cancer genome atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)* 2015;**19**:A68–77.
11. International Cancer Genome Consortium, Hudson TJ, Anderson W, *et al*. International network of cancer genome projects. *Nature* 2010;**464**:993–8.
12. Meric-Bernstam F, Johnson A, Holla V, *et al*. A decision support framework for genomically informed investigational cancer therapy. *J Natl Cancer Inst* 2015;**107**.
13. Jiang P, Sellers WR, Liu XSJARBDS. *Big data approaches for modeling response and resistance to cancer drugs*, Vol. **1**, 2018, 1–27.
14. 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.
15. Hollingsworth SJ. Precision medicine in oncology drug development: a pharma perspective. *Drug Discov Today* 2015;**20**:1455–63.
16. Karapetis CS, Khambata-Ford S, Jonker DJ, *et al*. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med* 2008;**359**:1757–65.
17. Herbst RS, Gandara DR, Hirsch FR, *et al*. Lung Master Protocol (lung-MAP)—a biomarker-driven protocol for accelerating development of therapies for squamous cell lung cancer: SWOG S 1400. *Clin Cancer Res* 2015;**21**:1514–24.
18. McNeil C. NCI-MATCH launch highlights new trial design in precision-medicine era. *J Natl Cancer Inst* 2015;**107**.
19. Shitara K, Miki I, Sudo T, *et al*. *The nationwide cancer genome screening projects for gastrointestinal cancer in Japan (SCRUM-Japan GI-SCREEN): efficient identification of actionable cancer genome alterations in advanced colorectal and non-colorectal gastrointestinal cancer (GI Screen 2013-01-CRC and 2015-01-Non CRC) 2015*. American Society of Clinical Oncology.
20. Middleton G, Crack LR, Popat S, *et al*. The National Lung Matrix Trial: translating the biology of stratification in advanced non-small-cell lung cancer. *Ann Oncol* 2015;**26**:2464–9.

21. Biankin AV, Piantadosi S, Hollingsworth SJ. Patient-centric trials for therapeutic development in precision oncology. *Nature* 2015;**526**:361–70.
22. Andre F, Mardis E, Salm M, *et al*. Prioritizing targets for precision cancer medicine. *Ann Oncol* 2014;**25**:2295–303.
23. Hovelson DH, McDaniel AS, Cani AK, *et al*. Development and validation of a scalable next-generation sequencing system for assessing relevant somatic variants in solid tumors. *Neoplasia* 2015;**17**:385–99.
24. Sicklick JK, Kato S, Okamura R, *et al*. Molecular profiling of cancer patients enables personalized combination therapy: the I-PREDICT study. *Nat Med* 2019;**1**.
25. Rodon J, Soria J-C, Berger R, *et al*. Genomic and transcriptomic profiling expands precision cancer medicine: the WINTHER trial. *Nat Med* 2019;1.
26. Rothwell DG, Ayub M, Cook N, *et al*. Utility of ct DNA to support patient selection for early phase clinical trials: the TARGET study. *Nat Med* 2019;1.
27. Johns A. *Consortium ICG*. AACR: A million years of precision oncology knowledge for the world, 2019.
28. Zehir A, Benayed R, Shah RH, *et al*. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10, 000 patients. *Nat Med* 2017;**23**:703–13.
29. Vanneman M, Dranoff G. Combining immunotherapy and targeted therapies in cancer treatment. *Nat Rev Cancer* 2012;**12**:237–51.
30. Hughes PE, Caenepeel S, Wu LC. Targeted therapy and checkpoint immunotherapy combinations for the treatment of cancer. *Trends Immunol* 2016;**37**:462–76.
31. Hu-Lieskovan S, Robert L, Homet Moreno B, *et al*. Combining targeted therapy with immunotherapy in BRAF-mutant melanoma: promise and challenges. *J Clin Oncol* 2014;**32**:2248–54.
32. Moya-Horno I, Viteri S, Karachaliou N *et al*. Combination of immunotherapy with targeted therapies in advanced non-small cell lung cancer (NSCLC), *Ther Adv Med Oncol* 2018; **10**: 1758834017745012.
33. Akbay EA, Koyama S, Carretero J, *et al*. Activation of the PD-1 pathway contributes to immune escape in EGFR-driven lung tumors. *Cancer Discov* 2013;**3**:1355–63.
34. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Cancer* 2006;**6**:813–23.
35. Paull KD, Shoemaker RH, Hodes L, *et al*. Display and analysis of patterns of differential activity of drugs against human tumor cell lines: development of mean graph and COMPARE algorithm. *J Natl Cancer Inst* 1989;**81**:1088–92.
36. Weinstein JN, Myers TG, O'Connor PM, *et al*. An information-intensive approach to the molecular pharmacology of cancer. *Science* 1997;**275**:343–9.
37. Shankavaram UT, Varma S, Kane D, *et al*. Cell miner: a relational database and query tool for the NCI-60 cancer cell lines. *BMC Genom* 2009;**10**:277.
38. 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.
39. Garnett MJ, Edelman EJ, Heidorn SJ, *et al*. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 2012;**483**:570–5.
40. Yang W, Soares J, Greninger P, *et al*. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* 2013;**41**:D955–61.
41. Iorio F, Knijnenburg TA, Vis DJ, *et al*. A landscape of Pharmacogenomic interactions in cancer. *Cell* 2016;**166**:740–54.
42. Basu A, Bodycombe NE, Cheah JH, *et al*. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* 2013;**154**:1151–61.
43. 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.
44. Haibe-Kains B, El-Hachem N, Birkbak NJ, *et al*. Inconsistency in large pharmacogenomic studies. *Nature* 2013;**504**:389–93.
45. Cancer Cell Line Encyclopedia Consortium, Genomics of Drug Sensitivity in Cancer Consortium. Pharmacogenomic agreement between two cancer cell line data sets. *Nature* 2015;**528**:84–7.
46. Bouhaddou M, DiStefano MS, Riesel EA, *et al*. Drug response consistency in CCLE and CGP. *Nature* 2016;**540**:E9–10.
47. Geeleher P, Gamazon ER, Seoighe C, *et al*. Consistency in large pharmacogenomic studies. *Nature* 2016;**540**:E1–2.
48. Mpindi JP, Yadav B, Ostling P, *et al*. Consistency in drug response profiling. *Nature* 2016;**540**:E5–6.
49. Haverty PM, Lin E, Tan J, *et al*. Reproducible pharmacogenomic profiling of cancer cell line panels. *Nature* 2016;**533**:333–7.
50. Zhu YX, Tiedemann R, Shi CX, *et al*. RNAi screen of the druggable genome identifies modulators of proteasome inhibitor sensitivity in myeloma including CDK5. *Blood* 2011;**117**:3847–57.
51. Bernards R, Brummelkamp TR, Beijersbergen RL. Sh RNA libraries and their use in cancer genetics. *Nat Methods* 2006;**3**:701–6.
52. Martinez-Lage M, Puig-Serra P, Menendez P, *et al*. CRISPR/-Cas 9 for cancer therapy: hopes and challenges. *Biomedicine* 2018;**6**.
53. Shi J, Wang E, Milazzo JP, *et al*. Discovery of cancer drug targets by CRISPR-Cas 9 screening of protein domains. *Nat Biotechnol* 2015;**33**:661–7.
54. Behan FM, Iorio F, Picco G, *et al*. Prioritization of cancer therapeutic targets using CRISPR-Cas 9 screens. *Nature* 2019.
55. Pan D, Kobayashi A, Jiang P, *et al*. A major chromatin regulator determines resistance of tumor cells to T cell-mediated killing. *Science* 2018;**359**:770–5.
56. Jun S, Lim H, Lee JH, *et al*. *Single-cell analysis of a mutant library generated using CRISPR-guided deaminase*. bio Rxiv, 2019, 610725.
57. Peng R, Lin G, Li J. Potential pitfalls of CRISPR/Cas 9-mediated genome editing. *FEBS J* 2016;**283**:1218–31.
58. Yin H, Xue W, Anderson DG. CRISPR-Cas: a tool for cancer research and therapeutics. *Nat Rev Clin Oncol* 2019;**16**:281–95.
59. Byrne AT, Alferez DG, Amant F, *et al*. Interrogating open issues in cancer precision medicine with patient-derived xenografts. *Nat Rev Cancer* 2017;**17**:254–68.
60. Kato S, Kurzrock R. An avatar for precision cancer therapy. *Nat Biotechnol* 2018;**36**:1053–5.
61. Lee JY, Kim SY, Park C, *et al*. Patient-derived cell models as preclinical tools for genome-directed targeted therapy. *Oncotarget* 2015;**6**:25619–30.
62. Lee JK, Liu Z, Sa JK, *et al*. Pharmacogenomic landscape of patient-derived tumor cells informs precision oncology therapy. *Nat Genet* 2018;**50**:1399–411.
63. Bertotti A, Migliardi G, Galimi F, *et al*. A molecularly annotated platform of patient-derived xenografts ("xenopatients") identifies HER2 as an effective therapeutic target in cetuximab-resistant colorectal cancer. *Cancer Discov* 2011;**1**:508–23.

64. Chen W, Hill H, Christie A, *et al*. Targeting renal cell carcinoma with a HIF-2 antagonist. *Nature* 2016;**539**:112–7.
65. Stewart E, Federico SM, Chen X, *et al*. Orthotopic patient-derived xenografts of paediatric solid tumours. *Nature* 2017;**549**:96–100.
66. Gao H, Korn JM, Ferretti S, *et al*. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat Med* 2015;**21**:1318–25.
67. Hidalgo M, Amant F, Biankin AV, *et al*. Patient-derived xenograft models: an emerging platform for translational cancer research. *Cancer Discov* 2014;**4**:998–1013.
68. Knoblaugh SE, Himmel LE. Keeping score: Semiquantitative and quantitative scoring approaches to genetically engineered and xenograft mouse models of cancer. *Vet Pathol* 2018;**300985818808526**.
69. Houghton PJ, Morton CL, Tucker C, *et al*. The pediatric preclinical testing program: description of models and early testing results. *Pediatr Blood Cancer* 2007;**49**:928–40.
70. Townsend EC, Murakami MA, Christodoulou A, *et al*. The public repository of Xenografts enables discovery and randomized phase II-like trials in mice. *Cancer Cell* 2016;**29**:574–86.
71. Schultz KR, Bowman WP, Aledo A, *et al*. Improved early event-free survival with imatinib in Philadelphia chromosome-positive acute lymphoblastic leukemia: a children's oncology group study. *J Clin Oncol* 2009;**27**:5175–81.
72. Bruna A, Rueda OM, Greenwood W, *et al*. A biobank of breast cancer explants with preserved intra-tumor heterogeneity to screen anticancer compounds. *Cell* 2016;**167**:260–274 e222.
73. Brabetz S, Leary SES, Grobner SN, *et al*. A biobank of patient-derived pediatric brain tumor models. *Nat Med* 2018;**24**:1752–61.
74. Burrell RA, McGranahan N, Bartek J, *et al*. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* 2013;**501**:338–45.
75. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol* 2018;**15**:81–94.
76. Yang H, Sun L, Liu M, *et al*. Patient-derived organoids: a promising model for personalized cancer treatment. *Gastroenterol Rep (Oxf)* 2018;**6**:243–5.
77. Lee SH, Hu W, Matulay JT, *et al*. Tumor evolution and drug response in patient-derived organoid models of bladder cancer. *Cell* 2018;**173**:515–528 e517.
78. van de Wetering M, Francies HE, Francis JM, *et al*. Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell* 2015;**161**:933–45.
79. Yan HHN, Siu HC, Law S, *et al*. A comprehensive human gastric cancer organoid biobank captures tumor subtype heterogeneity and enables therapeutic screening. *Cell Stem Cell* 2018;**23**:882–897 e811.
80. Kopper O, de Witte CJ, Lõhmussaar K, *et al*. An organoid platform for ovarian cancer captures intra-and interpatient heterogeneity. *Nat Med* 2019;1.
81. Van Allen EM, Wagle N, Stojanov P, *et al*. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat Med* 2014;**20**:682–8.
82. Sukhai MA, Craddock KJ, Thomas M, *et al*. A classification system for clinical relevance of somatic variants identified in molecular profiling of cancer. *Genet Med* 2016;**18**:128–36.
83. Dienstmann R, Dong F, Borger D, *et al*. Standardized decision support in next generation sequencing reports of somatic cancer variants. *Mol Oncol* 2014;**8**:859–73.
84. Yeh P, Chen H, Andrews J, *et al*. DNA-Mutation Inventory to Refine and Enhance Cancer Treatment (DIRECT): a catalog of clinically relevant cancer mutations to enable genome-directed anticancer therapy. *Clin Cancer Res* 2013;**19**:1894–901.
85. Johnson A, Zeng J, Bailey AM, *et al*. The right drugs at the right time for the right patient: the MD Anderson precision oncology decision support platform. *Drug Discov Today* 2015;**20**:1433–8.
86. Dienstmann R, Jang IS, Bot B, *et al*. Database of genomic biomarkers for cancer drugs and clinical targetability in solid tumors. *Cancer Discov* 2015;**5**:118–23.
87. Damodaran S, Miya J, Kautto E, *et al*. Cancer Driver Log (CanDL): catalog of potentially actionable cancer mutations. *J Mol Diagn* 2015;**17**:554–9.
88. Patterson SE, Liu R, Statz CM, *et al*. The clinical trial landscape in oncology and connectivity of somatic mutational profiles to targeted therapies. *Hum Genomics* 2016;**10**:4.
89. Huang L, Fernandes H, Zia H, *et al*. The cancer precision medicine knowledge base for structured clinical-grade mutations and interpretations. *J Am Med Inform Assoc* 2017;**24**:513–9.
90. Griffith M, Spies NC, Krysiak K, *et al*. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat Genet* 2017;**49**:170–4.
91. Chakravarty D, Gao J, Phillips SM, *et al*. Onco KB: a precision oncology knowledge base. *JCO Precis Oncol* 2017;**2017**.
92. Tamborero D, Rubio-Perez C, Deu-Pons J, *et al*. Cancer genome interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med* 2018;**10**:25.
93. Sun SQ, Mashl RJ, Sengupta S, *et al*. Database of evidence for precision oncology portal. *Bioinformatics* 2018;**34**:4315–7.
94. Forbes SA, Beare D, Bindal N, *et al*. COSMIC: high-resolution cancer genetics using the catalogue of somatic mutations in cancer. *Curr Protoc Hum Genet* 2016;**91**:10.11.1–10.11.37.
95. Forbes SA, Beare D, Boutselakis H, *et al*. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res* 2017;**45**:D777–83.
96. Tate JG, Bamford S, Jubb HC, *et al*. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res* 2019;**47**:D941–7.
97. Kumar R, Chaudhary K, Gupta S, *et al*. Cancer DR: cancer drug resistance database. *Sci Rep* 2013;**3**:1445.
98. Gohlke BO, Nickel J, Otto R, *et al*. Cancer resource—updated database of cancer-relevant proteins, mutations and interacting drugs. *Nucleic Acids Res* 2016;**44**:D932–7.
99. 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.
100. 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.
101. Menden MP, Casale FP, Stephan J, *et al*. The germline genetic component of drug sensitivity in cancer cell lines. *Nat Commun* 2018;**9**:3385.
102. Ding Z, Zu S, Gu J. Evaluating the molecule-based prediction of clinical drug responses in cancer. *Bioinformatics* 2016;**32**:2891–5.
103. 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.
104. 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.
105. Webber JT, Kaushik S, Bandyopadhyay S. Integration of tumor genomic data with cell lines using multi-dimensional network modules improves cancer pharmacogenomics. *Cell Syst* 2018;**7**:526–536 e526.
106. Jiang P, Lee W, Li X, *et al*. Genome-scale signatures of gene interaction from compound screens predict clinical efficacy of targeted cancer therapies. *Cell Syst* 2018;**6**:343–354 e345.
107. Chiu YC, Chen HH, Zhang T, *et al*. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Med Genomics* 2019;**12**:18.
108. Hyman DM, Taylor BS, Baselga J. Implementing genome-driven oncology. *Cell* 2017;**168**:584–99.
109. Rubio-Perez C, Tamborero D, Schroeder MP, *et al*. In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. *Cancer Cell* 2015;**27**:382–96.
110. Hintzsche J, Kim J, Yadav V, *et al*. IMPACT: a whole-exome sequencing analysis pipeline for integrating molecular profiles with actionable therapeutics in clinical samples. *J Am Med Inform Assoc* 2016.
111. Li MJ, Yao H, Huang D, *et al*. mTCTScan: a comprehensive platform for annotation and prioritization of mutations affecting drug sensitivity in cancers. *Nucleic Acids Res* 2017;**45**:W215–21.
112. Pineiro-Yanez E, Reboiro-Jato M, Gomez-Lopez G, *et al*. Pan drugs: a novel method to prioritize anticancer drug treatments according to individual genomic data. *Genome Med* 2018;**10**:41.
113. Kalari KR, Sinnwell JP, Thompson KJ, *et al*. PANOPLY: Omics-guided drug prioritization method tailored to an individual patient. *JCO Clin Cancer Inform* 2018;(2):1–11.
114. Ritter DI, Roychowdhury S, Roy A, *et al*. Somatic cancer variant curation and harmonization through consensus minimum variant level data. *Genome Med* 2016;**8**:117.
115. Nozaki K, Mochizuki W, Matsumoto Y, *et al*. Co-culture with intestinal epithelial organoids allows efficient expansion and motility analysis of intraepithelial lymphocytes. *J Gastroenterol* 2016;**51**:206–13.
116. Sanjana NE, Wright J, Zheng K, *et al*. High-resolution interrogation of functional elements in the noncoding genome. *Science* 2016;**353**:1545–9.
117. Lu Y, Zhao X, Liu Q, *et al*. Lnc RNA MIR100HG-derived mi R-100 and mi R-125b mediate cetuximab resistance via Wnt/beta-catenin signaling. *Nat Med* 2017;**23**:1331–41.