OXFORD

An omics perspective on drug target discovery platforms

Jussi Paananen and Vittorio Fortino

Corresponding author: Vittorio Fortino, Institute of Biomedicine, University of Eastern Finland, Finland. Tel.: +358 50 326 6148;
E-mail: vittorio.fortino@uef.fi

Abstract

The drug discovery process starts with identification of a disease-modifying target. This critical step traditionally begins with manual investigation of scientific literature and biomedical databases to gather evidence linking molecular target to disease, and to evaluate the efficacy, safety and commercial potential of the target. The high-throughput and affordability of current omics technologies, allowing quantitative measurements of many putative targets (e.g. DNA, RNA, protein, metabolite), has exponentially increased the volume of scientific data available for this arduous task. Therefore, computational platforms identifying and ranking disease-relevant targets from existing biomedical data sources, including omics databases, are needed. To date, more than 30 drug target discovery (DTD) platforms exist. They provide information-rich databases and graphical user interfaces to help scientists identify putative targets and pre-evaluate their therapeutic efficacy and potential side effects. Here we survey and compare a set of popular DTD platforms that utilize multiple data sources and omics-driven knowledge bases (either directly or indirectly) for identifying drug targets. We also provide a description of omics technologies and related data repositories which are important for DTD tasks.

Key words: drug target discovery; omics-informed drug discovery; drug efficacy and safety evaluations

Background

Target-based drug discovery is the most common strategy for the development of new drugs. However, when it comes to clinical trials, most of these new drugs fail due to inadequate efficacy or safety concerns [1], suggesting that a 'wrong' target has been selected during the target discovery process. Traditionally, molecular targets for drug discovery are selected on the basis of the accumulation of a series of experimental evidence supporting the hypothesis that modulating the function of the molecule will have an effect on disease [2]. This process strongly relies on the use of databases and bioinformatics tools enabling the collection and integration of multiple sources of evidence linking molecular drug targets to diseases [3]. Being an interdisciplinary branch of the life sciences, bioinformatics aims to provide methodology and computational methods needed to organize, explore and analyze large volumes of biological data, including genomic, proteomic and other 'omics' data types.

These computational tools have become essential to research progress in drug target discovery (DTD), thanks to their ability to aid in elucidation and understanding of the mechanisms of (complex) diseases. Current bioinformatics strategies for DTD use a wide range of data sources obtained from experimental, mechanistic, pharmacological and, more recently, omics-based molecular profiles. Omics technologies have brought unprecedented abilities to screen biological samples at the gene, transcript, protein, metabolite and their interaction network level in searching of novel targets [4]. In particular, genome-wide association studies (GWASs), whole genome sequencing and transcriptome analysis constitute essential tools to discover or validate new drug targets, since they can provide a systematic approach to evaluate their therapeutic efficacy and related side effects. Even though omics studies can provide information on the efficacy of drug targets, they can be exploited to better understand their mechanisms of action and, most importantly, to

detect in advance drug-induced side effects. Omics technologies generate large amount of data from single experiments, and this information is often useless in their original formats. Therefore, advanced data mining algorithms are required to identify, evaluate and rank putative drug targets from omics data. Besides, it is very difficult to systematically fuse these big volumes of data with existing scientific literature and biomedical databases. To speed up the process of collection, processing and analyzing data sources for drug discovery, many software platforms and database systems have been developed. These tools help scientists working on early drug discovery to automatically identify and extract relevant drug target–disease associations without requiring the application of sophisticated algorithms. In this review we describe and evaluate the most popular DTD platforms that directly or indirectly (through the use of external data sources) employ omics data for the identification of disease-relevant drug targets. We highlight their main functionalities and abilities in providing concise information on the therapeutic efficacy, druggability and safety of selected, putative targets. In this review, drug targets and their desired properties are discussed. Then, the application of omics information for target discovery is briefly introduced along with a summary description of databases and web platforms that can be utilized to prioritize drug targets. Finally, the selected DTD platforms are compared on the basis of drug target–disease associations and the use of omics data.

Drug targets

A drug target can be defined as a molecule in the body, usually a protein, that is intrinsically associated with a particular disease process and that could be addressed by a drug to produce a desired therapeutic [5, 6]. Drug targets should exhibit the following several, basic features: involvement in a crucial biological pathway; functionally and structurally characterized; and druggable (capable of binding to small molecules, implying the presence of a binding site). Traditionally, structure-based analysis has been used to search for good drug targets, which leads to the concept of 'druggability'. A drug target indeed is often described as proteins that possess protein folds that favor interactions with drug-like chemical compounds [7–9]. Many proteins are druggable according to their structure, but their binding will not lead to the therapeutic benefit. Over the past two decades, there have been several efforts to curate drug targets and to categorize them. Most frequent proteins for DTD include proteases, kinases, G protein-coupled receptors (GPCRs) and nuclear hormone receptors [10]. Druggability is not the only desired property for the definition of 'good' drug targets. Indeed, often scientists are first interested in selected candidates based on their participation in a biological process critical to diseases [10, 11]. Imming *et al.* [12] categorized drug targets based on 'mechanism of actions', such as enzymes, substrates, metabolites, proteins, receptors, ion channels, transport proteins, DNA, RNA, ribosomes and targets of monoclonal antibodies. Although proteins have been in the past the majority of clinically useful drugs, there are many emerging classes of drug targets, such as nucleic acid, regulatory DNA element and non-coding RNA (ncRNAs). Their importance is rapidly growing in the fields of drug development and precision medicine. Indeed, drugs which target nucleic acids, especially in the areas of antibacterial and anticancer therapy, have been already provided [13]. RNA is now being recognized as an essential component in various regulatory processes just like proteins. Indeed, RNA plays important roles in the transcription regulation, regulation of

the translation, catalysis, protein function, protein transport, peptide bond formation and RNA splicing [14]. Compared to DNA, RNA could deliver better therapeutics since RNA displays a greater structural diversity and lacks repair mechanisms. Like proteins, RNA has three-dimensional folding that gives rise to complex structures allowing the highly specific binding of effector molecules. RNA targets have been successfully employed in the antibacterial and antiviral areas [15, 16]. Moreover, with new emerging classes of RNAs and their characterization in regulatory mechanisms of mammals, their application has rapidly expanded. Among these new classes of drug targets, ncRNAs are gaining increasingly attention. ncRNA refers to a large group of endogenous RNA molecules that have no protein-coding capacity, while having specialized biological functions. While ncRNAs lack potential to encode proteins, they can affect the expression of other genes through a variety of mechanisms. In some cases, their mechanisms of action are well known, and their strategies for controlling activity are well established. The ability of ncRNAs to control gene expression makes them targets for drug development. However, the uncertainty about how ncRNAs function (and even whether they have a function) makes the drug development process even more challenging [17]. Targeting RNAs offers opportunities to therapeutically modulate numerous cellular processes, including those linked to 'undruggable' protein targets [18]. Examples include proteins with multiple key functions, which are difficult to block by using a single molecule, or proteins that are so closely related to others that it is difficult to achieve adequate selectivity.[19, 20].

Key properties of a drug target

A drug target is a key molecule interfering with biological pathways that are specific to a disease or a disease state. The difference between a drug target and other biomolecule involved in the same pathway is only in their location and role. A putative drug target should be disease-dependent, which implicates that its relevance for other disease should be minimal. However, human diseases are often complex involving many interrelated pathways, which can lead to the identification of different molecular targets. Indeed, studying complex diseases with a Mendelian perspective, or the 'one gene–one disease' theory and treated with a 'magic bullet' therapy, has been demonstrated to be ineffective [21]. It is frequently observed that the complex interplay existing between multiple molecular features leads to disease state, rather than the behavior of a single molecule. Therefore, it is often important to characterize the association between multiple drug targets and disease [22]. Drug targets can be categorized into two classes: known drug targets and novel drug targets. The former are those for which there is robust scientific evidence, supported by publications and experimental data showing how the target functions in normal physiology and how it is involved in human pathology. Furthermore, there are drugs targeting this target. Whereas, novel drug targets are those biomolecules whose functions are not fully understood and with no established drugs targeting them. These targets merit more attention since they might lead to completely new therapies. Very often the goal of a preclinical drug discovery program is to deliver a ranked list of drug target biomolecules (such as DNA, RNA, proteins and peptides), including both known and novel targets. Candidate drug targets should be then characterized by a well-balanced profile between efficacy and safety ('drug adverse reactions'). The efficacy should aim to evaluate how good a biomolecule is as a drug target. This evaluation should take into account the target 'druggability' [23], and it should be

Table 1. Key drug target properties

Property	Description	Key aspects
Efficacy	In order for a drug to have an effect, it needs to bind to its target, and then to affect the function of this target. A target can refer to a gene, a protein or other biomolecules, and it is responsible for the therapeutic efficacy of the drug [32]. Therefore, the efficacy of a target should evaluate its potential in delivering effective therapeutic treatments.	Target druggability Target disease validation Tissue-specific efficacy evaluations
Safety	Safety evaluation aims to identify potential adverse consequences of target modulation, unavoidable on-target toxicities and potential clinical adverse to support the steps of drug target identification and prioritization [33].	Drug toxicity in patients OFF/ON drug targets Unsafe biomolecules (essential genes, carcinogenic, etc.)
Novelty	It estimates the scarcity of publications and patents about a protein target [26].	Text mining of scientific and patent literature

supported by sufficient evidence showing how efficient a target is for a given disease ('target disease efficacy'). Well-balanced efficacy–safety profile will take into account the intended use of the drug (e.g. severity of the indication and treatment phase) and efficacy and safety of existing treatments. A potentially good drug target will have mild overdose effects, giving drug discovery a comparatively broad efficacy region. In practice, the acceptable efficacy–safety balance for a first-in-market drug for rapidly progressing high mortality disease will be different than for a drug targeting benign disease with existing treatments. Omics studies can be used to evaluate the following: the modulation, a target is disease modifying and/or has a proven function in the pathophysiology of a disease; the tissue specificity, a target expression is not uniformly distributed throughout the body; and the druggability, a target can be modulated by a drug [24]. However, safety-related drug attrition represents a major leap in the development of therapeutic targets. Indeed, safety issues are often the main cause of drug development failures. Key safety liabilities induced by target modulation could be evaluated by using omics studies. In particular, safety assessment could apply genomics, transcriptomics and proteomics, to pre-evaluate on-target, off-target, toxicity pathways [25] and harmful or unpleasant clinical events triggered by drug interaction with the target. Moreover, extensive studies of drug target safety by the means of omics could lead to the accumulation of enough data to develop robust *in silico* methods to predict the safety of putative drug targets or, even better, to help researchers identify safe-by-design drug targets. Omics-driven efficacy and safety evaluations of drug targets can be combined with information extracted from scientific literature in order to increase the success of identifying the right candidates. To this end, different web-based text mining tools, such as TIN-X [26], PubTator [27] and Chemotext [28], can be applied. These tools can improve the accuracy of drug target–disease associations extracted from scientific publications [29]. Moreover, network analysis aiming to integrate gene, drug and phenotype information can also be utilized to better estimate drug target efficacy [30]. Finally, in addition to efficacy and safety evaluations, drug targets can be characterized by describing their novelty [31]. Table 1 describes key drug target properties.

In summary, good drug target needs to be efficacious, druggable, safe and meets clinical and commercial needs. However, it should be noted that there are challenges in making a clean unambiguous assignment in many cases, especially regarding how to define the concept of drug target efficacy.

Multiple target strategies

Traditionally drugs have been designed by following the paradigm 'one drug, one target', which aims to find a single molecular target, usually a protein (the so-called 'on-target'), with high selectivity to avoid any unwanted effects arising from mis-targeting other biological targets ('off-targets'). While target-first strategies might prove useful to approach single gene disorders, disease is often a multifactorial condition involving a combination of constitutive and/or environmental factors. In this case scenario, single-target drugs might be inadequate to achieve a therapeutic effect [34, 35]. It is now widely accepted that complex diseases are more likely to be healed or alleviated through simultaneous modulation of multiple targets. Indeed, research on multi-target drugs has rapidly increased since 2000, and it is nowadays one of the hottest topics in drug discovery. However, there are several challenges to be addressed when designing multi-target drugs, both in terms of target selection and small molecule discovery [35, 36]. For instance, algorithms that determine multi-drug dosages are important to ensure effective treatments [36]. However, the extremely high number of possible multi-drug combinations combined with heterogeneity and resistance-related issues makes the dosage adjustment optimization an extremely challenging task [37]. Even though several online resources exist for multiple target selection, the Therapeutic Target Database [38] and DrugBank [39], there are no well-established, data-driven computational methods to identify the right combination of molecular targets for a given disease in both multi-target drugs and therapeutic combinations. The fact is that addressing multiple DTD tasks require a deeper understanding of disease mechanisms, target disease associations, pathway-target-drug-disease relationships and adverse events [40]. Besides, when selecting multiple targets additive or synergistic effects should be carefully considered [41].

Omics applications for DTD

Recent technical advances in sequencing, microarray and mass spectrometry (MS) technologies allow scientists to generate genomics, transcriptomics, proteomics, and other -omic data types at an unprecedented level of resolution. Many studies have used these technologies to better understand the molecular mechanisms underlying complex diseases and to provide information on drug treatments. The resulting rich information data can be utilized to identify drug targets, to

Table 2. Omics data types and their use for informed pharmaceutical research and development

Omics	Function	Databases
Genomics	Understanding pathogenesis Genetic association studies Identification of disease genes Discovery of putative drug targets Patient-centered efficacy and toxicity assessment of drugs/targets Patient stratification	GWAS Catalog GWAS central dbGaP PharmGKB
Transcriptomics	Disease mechanisms Mode of action of compounds Moving from disease genes to drug targets Identification/evaluation of drug target candidates Early prediction of adverse drug target effects	DrugMatrix TG-GATE LINCS 1000 Expression Atlas GEO repository ArrayExpress
Proteomics	Post-translational process Protein–protein network interaction Drug target efficacy and safety evaluation at protein level Protein toxicology	PRIDE Archive Peptide Atlas ProteomicsDB Human Proteome Map Human Proteome Atlas
Metabolomics	Novel DTD Drug target efficacy and safety evaluation at metabolomic level Metabolic toxicity	Human Metabolome Madison Metabolomics Golm Metabolome Database MassBank MetaboLights MetabolomeExpress

uncover the mechanism of action of drugs and to assess (or infer) their side effects. Omics-based studies can also provide essential information to deliver personalized medicine. For example, it has been shown that genetic variations can help clinicians assess efficacy or toxicity of some targeted agents for specific subsets of molecularly profiled patients [42]. If systematically integrated, omics-driven molecular profiles of diseases and drug treatments/exposures could significantly accelerate drug discovery and development process. Table 2 lists and briefly describes omics technologies that can be utilized in drug discovery and development and related omics data repositories.

Genomics

Genomics have provided the earliest applications for DTD. In particular, the application of DNA microarray and next-generation sequencing (NGS) technologies has enabled high-throughput analysis of genotype–phenotype relationships on human populations, opening a new era of genetics-informed drug discovery. DNA microarrays have been extensively used to conduct GWASs, which have helped scientists identify loci that harbor genetic variants (typically single nucleotide polymorphisms; SNPs) that are associated with risk for diseases and traits [43]. Studying the target at GWAS risk loci, such as genes or ncRNAs that mediate the associations observed in GWAS [44], can lead to a better understanding of the molecular mechanisms that influence disease risk and, most importantly, to new potential targets for drug development. However, the associated genes remain largely unknown for most GWAS loci. Even though the first GWASs aimed to identify as few as tens of genes contributing to genetic traits, today GWASs have helped to identify thousands of genes contributing to complex genetic traits. To date, almost 10 000 strong associations have been

reported between genetic variants and one or more complex traits. Among these findings, there are several examples of disease-associated genes that have been identified as being effective drug targets [45]. GWAS data have also been employed for the development of *in silico* methods for disease genes and drug target identification [46, 47]. GWAS can also enable the discovery of biological pathways that confer susceptibility to diseases. However, these studies alone cannot elucidate how the variants affect downstream pathways and lead to a disease. Indeed, one of the most relevant post-GWAS challenges is to combine GWAS findings with additional omics data layers in order to shed the light on the biological systems underlying complex diseases and provide more effective drug targets. Indeed, GWAS-driven analyses have been often merged with transcriptomics data, which can help identify disease-related genes by comparing gene expression profiles between disease and control groups [48]. GWASs have also rapidly changed the landscape of pharmacogenomic research. Pharmacogenomics is the study of the impact of genetic variations of individuals on their drug response or drug metabolism. It helps understand how patients with specific genomic characteristics respond to certain treatments and drugs [49]. In more practical terms, it can affect the drug development process in two primary ways: indicating how well the drug works (efficacy) and providing drug-related toxicity information [50]. Table 3 reports GWAS data repositories and pharmacogenomics databases that are currently used for discovering putative drug targets, drug repositioning, drug efficacy and safety assessments.

Transcriptomics

Genome-wide transcriptional profiling provides a global view of cellular state and how this state changes under different treatments (e.g. drugs) or conditions (e.g. healthy and diseased).

Table 3. Genomics databases for DTD

Database	Description	Application
GWAS Catalog [51]	Collects published human GWASs that are manually curated by expert scientists. GWAS Catalog provides accurate and structured metadata for publication, study design, sample and trait information and the most significant published results. https://www.ebi.ac.uk/gwas/	Mining disease genes Narrow-down/prioritize candidate loci Disease risk prediction Disease mechanisms
GWAS Central [52]	A database of summary level findings from genetic association studies, both large and small. GWAS central collects datasets from public domain projects and encourages direct data submissions from the community. http://www.gwascentral.org/index	Mining SNP-drug response associations
NCBI dbGaP [53]	The NCBI Database of Genotypes and Phenotypes archives results of studies that have investigated the interaction of genotype and phenotype and distributes these results to investigators for secondary study. It includes phenotype data, GWAS data, summary level analysis data, Short Read Archive (SRA) data, reference alignment (BAM) data, Variant Call Format (VCF) data, etc. http://www.ncbi.nlm.nih.gov/gap	Genotype studies for the identification of disease genes
PharmGKB [50]	A publicly available online knowledgebase aggregating, curating, integrating and disseminating knowledge regarding the impact of human genetic variation on drug response. http://www.pharmgkb.org	Mining drug–gene, drug–SNP, gene–disease, disease–SNP, drug–pathway, disease–pathway, drug–drug

In particular, drug-induced gene expression profiles in human cell lines and *in vivo* models can be used to elucidate the biological effects of putative targets and evaluate in advance therapeutic efficacy [25, 54]. Transcriptomics signals are important for clinical candidate selection as they provide an evaluation of potential adverse effects of drug targets at an early stage in drug development [25]. Indeed, in recent years, transcriptomics data have been intensively used in the toxicogenomics field. This has led to the development of large-scale public databases, such as DrugMatrix [55] and Open TG-GATE [56], which collect compound-induced gene expression data with *in vivo* histopathological data, and Connectivity Map [57] and the Library of Integrated Network-based Cellular Signatures L1000 dataset [58], which collect transcriptional drug perturbations for thousands of compounds tested on more than 70 cell lines. Other gene expression data following drug treatments can be retrieved from the Gene Expression Omnibus (GEO) repository [59] and the ArrayExpress Archive of Functional Genomics Data [60] which are continuously updated. The data provided by these databases can be used in combination with genomic and network features to prioritize drugs with less likelihood of causing side effects [61]. Similar computational strategies are currently proposed to support alternative methods for chemical risk assessment [25]. Moreover, gene expression and transcriptome profiling can help researchers improve the design of clinical trials in phase I and phase II studies [42]. Genome expression profiling can also be combined with the genotype of trait-associated variants using *in vivo* data, thus identifying target genes and the directionality of the effect of trait-variants. Expression quantitative trait loci (eQTL) analyses are useful in this regard as they can provide genome-wide lists of genetic variants that associate with gene expression in a particular tissue [63]. eQTL analysis can be used to identify causal and also to discover genetic networks that might play significant roles in drug resistance responses [64]. Currently, many eQTL databases exists, and they can be queried to determine if a trait-associated variant (or variants

in linkage disequilibrium) associates with the expression of a specific gene. One of the most used databases is the Genotype-Tissue Expression [65]. Over the past decade, the transcriptomics field has developed rapidly, thanks to the advent of NGS technologies. In particular, whole-transcriptome analysis with total RNA sequencing (RNA-Seq) has become an indispensable tool for gene expression profiling. Total RNA-Seq profiles provide an exceptional opportunity to study ncRNAs, including microRNAs and long non-coding RNAs [66]. Similar to the protein-coding genes, ncRNAs can play critical roles in tumor progression [67] and cancer therapy [68]. Currently available large-scale cancer genome and pharmacogenomics projects, such as The Cancer Genome Atlas (TCGA), Cancer Cell Lines Encyclopedia (CCLE) [69] and Genomics of Drug Sensitivity in Cancer (GDSC) [70, 71], can be used to systematically determine the regulatory roles of ncRNA in cancer drug response by combining RNA-seq data in conjunction with clinical and drug response data from thousands of tumor samples and cancer cell lines [66]. These databases provide information on different omics data types, including mutation and copy number variation. Table 4 introduces the main data sources of transcriptomic data for DTD applications.

Proteomics

Proteomics refers to the analysis of the entire protein content of a cell, tissue or organism under a specific condition. Several techniques have been developed to study the proteome of an organism, and among them MS has become the tool of choice. The three primary applications of MS to proteomics are cataloging protein expression, defining protein interactions and identifying sites of protein modification. These sources of information have been extensively used as a DTD tool [73]. However, while transcriptome data cover the whole range of expressed genes, a typical untargeted MS proteomics experiment can usually detect and quantify up to 5000 proteins, which is less than

Table 4. Transcriptomic databases for DTD

Database	Description	Application
DrugMatrix [55]	DM is provided by the U.S. National Toxicology Program and it gives access to large-scale gene expression data derived from standardized toxicological experiments in which rats or primary rat hepatocytes were systematically treated with therapeutic, industrial and environmental chemicals at both non-toxic and toxic doses. https://outage.niehs.nih.gov/drugmatrix/index.html	DTD Understanding drug/target toxicity
TG-GATEs [56]	TG-GATEs provides gene expression profiles and traditional toxicological data derived from *in vivo* (rat) and *in vitro* (primary rat hepatocytes, primary human hepatocytes) exposure to 170 compounds at multiple dosages and time points. http://toxico.nibio.go.jp/english/index.html	DTD Understanding drug/target toxicity
LINCS 1000 [58]	L1000 generates gene expression signatures from treatment of a variety of cell types with perturbagens that span a range of small-molecule compounds, gene overexpression and gene knockdown reagents. The gene expression profiles are generated from a method, namely L1000, which defines a reduced representation of the transcriptome. http://www.lincsproject.org	DTD Drug repositioning
Expression Atlas [72]	EA collects baseline gene expression data in different species and contexts, such as tissue, developmental stage or cell type. It also contains differential studies, reporting changes in expression between two different conditions, such as healthy and diseased tissue. https://www.ebi.ac.uk/gxa/baseline/experiments	DTD and validation Disease genes
GEO repository [59]	GEO is a database repository of high-throughput gene expression data and hybridization arrays, chips, microarrays. https://www.ncbi.nlm.nih.gov/geo/	Retrieve drug, gene and disease perturbations
ArrayExpress [60]	AE serves as an international repository for microarray data and high-throughput sequencing-based functional genomics experiments associated with scientific publications. http://www.ebi.ac.uk/arrayexpress	Retrieve drug, gene and disease perturbations
TCGA	TCGA collects and functional genomics data repository for >30 cancers across >10 K samples. Primary data types include mutation, copy number, mRNA and protein expression. https://tcga-data.nci.nih.gov/tcga	Discover novel molecular targets
GTEx [65]	GTEx provides transcriptomic profiles of normal tissues, including >7 K samples across >45 tissue types. http://www.gtexportal.org	Tissue-specific drug targets
CCLE [69]	CCLE provides genetic and pharmacologic characterization of >1000 cancer cell lines. http://www.broadinstitute.org/ccle	Identify novel drug targets and drug response biomarkers
GDSC [70]	GDSC is the largest public resource for information on drug sensitivity in cancer cells and molecular markers of drug response. https://www.cancerrxgene.org/	Identify novel drug targets and drug response biomarkers

half of the expressed human proteome. In the past few years new proteomics technologies have been proposed with the ability to identify >8000 proteins in a 5 h analysis [74]. Overall quantitative proteomic methodologies are becoming more robust and reliable with technological developments and can produce robust, reproducible and standardized data sets [75]. The differential and quantitative profiling of the dynamic protein changes in health and disease will inevitably further our understanding of the mechanistic basis of disease. Proteomics experiments can be used for different aspects of clinical and health sciences such as biomarker discovery and drug target identification. A biomarker usually refers to disease-related molecule that can be used to diagnose or monitor risk or prognosis of disease, and can also indicate opportunities for therapeutic interventions. For example, proteomic strategies have been extensively used

for discovering novel cancer biomarkers [76, 77]. Proteomics can also be used to address several steps of the drug development process, including identification and validation of drug targets, informing assay development for screening of leads and in generating *in vitro* and *in vivo* biomarkers as surrogate endpoints for efficacy, toxicology and disease stratification. Most of the MS-based proteomics research studies in drug discovery have been performed to characterize protein expression profiling, functional proteomics and phosphoproteomics. These experiments aim to measure the protein expression levels, protein–protein complexes and signal transduction relative to a control treatment. However, in drug discovery it is very important to discover protein targets from phenotypic assays and to understand on- and off-target engagement of potential therapeutic compounds. This task can be addressed by using chemoproteomics [78, 79].

Table 5. Proteomic databases for DTD

Database	Description	Application
PRIDE Archive [84]	The PRIDE is a public data repository for proteomics, including protein and peptide identifications, post-translational modifications and supporting spectral evidence. https://www.ebi.ac.uk/pride/archive/	Drug target identification
ProteomicsDB [85]	PDB is a large collection of quantitative MS-based proteomics data across various tissue types as well as protein–protein interaction information, functional annotation, target deconvolution, cell sensitivity and reference MS data. https://www.proteomicsdb.org/	Drug target identification Drug target efficacy/potency
Human Proteome Map [88]	Hosts high-resolution MS proteomic data representing 17 adult tissues, 6 primary hematopoietic cells and 7 fetal tissues resulting in >84% human proteome coverage. http://www.humanproteomemap.org/	Drug target identification Biomarkers
Human proteome atlas [89]	Collects expression and localization of majority of human protein-coding genes based on both RNA and protein data. The HPA also employs antibody-based proteomics and transcriptomics profiling methods to locate and identify proteins in tissues and cell types. http://www.proteinatlas.org/	Druggable proteome Drug target efficacy and specificity

Chemoproteomics refers to a new technology that facilitates large-scale study of proteins by combining chemical methods with MS proteomics. In particular, it provides direct binding of small molecules with protein targets, helping one to quantify the amount of drug required to bind a target and subsequently produce a therapeutic effect) and drug selectivity determination (through the assessment of off-target interactions) [78]. Proteomics is also used in drug target identification by applying protein–protein interaction networks (PPINs). PPINs are typically modeled via graphs, whose nodes represent proteins and whose edges connect pairs of interacting proteins. These connections are specific, occur between defined binding regions in the proteins and have a particular biological meaning (i.e. serve a specific function). The totality of PPIs that happen in a cell and in a given biological context is called interactome. As a result of the development of large-scale PPI screening techniques, especially high-throughput affinity purification combined with MS and the yeast two-hybrid assay, today we can access big amounts of PPI data and build very complex interactomes [80]. All this information can serve new ways for DTD. In a recent study, breast, pancreatic and ovarian cancer PPINs were employed to identify the respective sets of driver proteins [81]. In this study, the PPI was implemented as linear time-invariant dynamical systems (LTISs). Then, an efficient (low polynomial time) algorithm was provided for computing the minimal number of input nodes needed to structurally control the given LTIS representing a given cancer PPI. The identified driver proteins, called cancer survivability-essential proteins, were proved to be a key for *in vivo* cancerous cell's proliferation and survival. PPI can also be exploited to characterize topological properties of efficient drug targets and to use this information for target prediction [82]. The applications of MS in identification and quantification of proteins encoding disease genes are rapidly evolving [83]. This has led to an exponential growth of targeted quantitative proteomic analyses that aim to systematically measure the abundance of proteins in large sets of samples, generating big volume of multidimensional data. In order to disseminate these large data sets to the scientific community, researchers have recently developed central repositories to store and share MS proteomics data such as PRIDE [84] and ProteomicsDB [85]. The PRoteomics

IDEntifications, an archive database for MS proteomics data, provides protein and peptide identifications, post-translational modifications and supporting mass spectra evidence. Whereas, ProteomicsDB contains quantitative data from 78 projects, for a total of 19 k LC–MS/MS experiments. A standardized analysis pipeline enables comparisons between multiple datasets to facilitate the exploration of protein expression across hundreds of tissues, body fluids and cell lines. Table 5 reports the main proteomics databases for DTD. In addition, protein databases such as UniProt [86] and Protein Data Bank (PDB) [87] can be used to further examine individual proteins on functional and structural level.

Metabolomics

Metabolomics is the study of the metabolome, i.e. all the metabolites present in a cell, tissue or organism at a given time. It provides an overview of the metabolic status and global biochemical events associated with a cellular or biological system. Unlike genes and proteins, whose functions are influenced by intriguing regulatory mechanisms such as epigenetic regulation and protein post-translational modifications, metabolites provide direct signatures of biochemical activity and are therefore easier to correlate with phenotype [90]. Metabolomic-based clinical applications and tests are now emerging [91] and can help to understand disease mechanisms from a new perspective. Metabolomics have contributed to identifying metabolic causes and biomarkers for chronic diseases such as diabetes, Alzheimer disease, atherosclerosis and cancer. Metabolites play an important role in tumor cell proliferation. In particular, by analyzing transcriptional-metabolomic data from experiments of knock-down genes responsible for the enzyme supporting cell growth in glucose-free media, Vincent *et al.* [92] identified metabolic pathways that support glucose-independent tumor cell proliferation. Metabolomics can be used to implement effective precision medicine approaches such as personalized phenotyping and individualized drug-response monitoring. For instance, the analysis of pre-dose metabolite biofluid profiles allows clinicians to predict the effectiveness of a selected drug treatment for

Table 6. Metabolomic databases for DTD

Database	Description	Applications
The Human Metabolome Database [84]	HMDB is a freely available electronic database containing detailed information about small molecule metabolites found (and experimentally verified) in the human body. It contains experimental MS/MS data for over 5700 compounds. http://hmdb.ca	DTD
The Madison Metabolomics Consortium Database [97]	MMCD collects small molecules of biological interest gathered from electronic databases and the scientific literature. It contains approximately 10 000 metabolite entries and experimental spectral data on about 500 compounds. http://mmcd.nmrfam.wisc.edu	DTD
Golm Metabolome Database [98]	GMD represents a general MS-based repository of reference metabolite profiles for essential plant tissues and typical variations of growth conditions. http://gmd.mpimp-golm.mpg.de/	DTD
MassBank [99]	The first public repository of Electron Impact-MS data covering more than 200 000 spectra for a wide range of organic compounds. https://github.com/MassBank/	DTD
MetaboLights [96]	ML is an open-access database repository for cross-platform and cross-species metabolomics research at the European Bioinformatics Institute. It provides Metabolomics Standard Initiative-compliant metadata and raw experimental data associated with metabolomics experiments. https://www.ebi.ac.uk/metabolights/	Drug safety Drug efficacy
MetabolomeExpress [100]	MB is designated to perform three main functions: (i) store GC-MS metabolomics data, allowing for analysis without the user having to download the data, (ii) provide a GC-MS analysis pipeline and (iii) store metabolite response statistics. https://www.metabolome-express.org/	DTD Drug safety

a given individual [93]. Toxicity assessment of drug targets can be addressed by using metabolomics data. In particular, the information from metabolomic analysis can be used to determine the off-targets of a drug candidate and thus provide a mechanistic understanding of drug toxicity [94]. Moreover, metabolic profile analysis can also allow clinicians to quantify drug efficacy and safety and use this information to tailor personalized treatments. Metabolomics also has the potential for generating a new generation of biomarkers. For instance, a panel of metabolic biomarkers to monitor responses to therapeutic interventions was developed recently [95]. Metabolomics is reducing the cost of toxicological screening, enabling improved clinical trial design, allowing better patient selection and monitoring and shortening the time needed for drugs to move through the development pipeline. Considerable advances have been made in the assessment of mechanisms of action of toxicity of drugs and other substances. Comprehensive metabolomics databases include The Human Metabolome Database (HMDB) and MetaboLights [96]. Table 6 lists and briefly describes metabolic data sources that can be used for drug target efficacy and safety assessments.

Platforms for multi-omics data discovery

As discussed in the previous section, numerous repositories are available to share and disseminate omics datasets for DTD. However, these repositories do not provide tools to systematically link omics studies having similar experimental set-ups. The need to facilitate the retrieval and the integration of publicly available big omics data has recently led to web-based platforms for indexing, discovering and integrating datasets from different omics technologies and databases into common framework and web interface. Examples include Biomedical and Healthcare Data Discovery Index Ecosystem (bioCADDIE), funded by the National Institute of Health (NIH) Big Data to Knowledge (BD2K) initiative that aims to provide a platform to retrieve relevant metadata of entire datasets [101]. Another example of an omics-based search engine is the Omics Discovery Index (OmicsDI; http://www.omicsdi.org), which helps scientists integrate proteomics, genomics, metabolomics and transcriptomics datasets [102]. OmicsDI has developed a common metadata structure framework and exchange format across 11 repositories, including proteomics databases (PRIDE, MassIVE and GPMDB), metabolomics databases (MetaboLights, GNPS, MetabolomeExpress and Metabolomics Workbench) and transcriptomics databases (ArrayExpress and Expression Atlas).

Multi-omics approaches for DTD

Each type of omics data provides important information highlighting differences between normal and disease conditions. These data can be utilized to discover diagnostic and prognostic markers and to give insight as to which biological processes are different between the disease and control samples. However, single-level omics data analysis is limited, reflecting reactive processes rather than causative ones [103]. Indeed, integrating multiple omics data types could reveal important molecular mechanisms that regulate for complex diseases and help scientists understand the dynamics that lead to disease manifestations. Moreover, a more detailed understanding of disease mechanisms would be beneficial when searching for novel, more effective biomarkers. Current DTD platforms do not provide tools for multi-omics data analysis mainly because of the

Table 7. Pathway-based databases useful for DTD

Database	Description	Drug-related information
KEGG [113]	The Kyoto Encyclopedia of Genes and Genomes is a widely used database containing metabolic pathways (372 reference pathways) from a wide variety of species (>700). These pathways are hyperlinked to metabolite and protein-complex/enzyme information. https://www.genome.jp/kegg/	Drug metabolism Drug development Disease/drug information
BioCyc [114]	The BioCyc database is a set of 3000 Pathway/Genome Databases (PGDBs) for many sequenced genomes. PGDBs describe the entire genome of an organism, as well as its biochemical pathways and (when curated) its regulatory network. https://biocyc.org/	Pathway-based target selection and validation Antimicrobial drug targets
Reactome [115]	Reactome builds and maintains a peer reviewed knowledge base of biological pathways (primary species of interest is *Homo sapiens*), including metabolic pathways as well as protein complex trafficking and signaling pathways. Reactome includes several types of reactions in its pathway diagram collection including experimentally confirmed, manually inferred and electronically inferred reactions. https://reactome.org/	Simulate impact of drugs on pathway activities Drug target interaction in pathway diagrams
WikiPathways [116]	WikiPathways is an open, collaborative platform dedicated to the curation of biological pathways. It is based on the MediaWiki open source software used by Wikipedia, coupled to a custom graphical pathway editing tool and integrated databases covering major gene, protein complex and small-molecule systems. https://wikipathways.org/	Drug target search strategies
Pathway Commons [117]	PC provides a collection of publicly available pathways from multiple organisms that provide researchers with convenient access to a comprehensive collection of pathways from multiple sources represented in a common language. https://www.pathwaycommons.org/	Robust pathway analyses
Biocarta	Biocarta is an open source database of pathways highlighting molecular relationships from areas of active research as well as classical pathway maps. It also catalogs and summarizes important resources providing information for over 120 000 genes from multiple species. www.biocarta.com	Enhancing genomic information for DTD
PharmGKB [118]	PGKB is a publicly available online knowledgebase responsible for the aggregation, curation, integration and dissemination of knowledge regarding the impact of human genetic variation on drug response. It also contains manually curated pharmacokinetic and pharmacodynamics pathways. https://www.pharmgkb.org/	Drug target–side effects

absence of standardized analytical protocols. However, recent studies have proposed interesting approaches applied to the cancer drug discovery [104]. It is acknowledged that multi-omics cancer data have the potential in improvement of targeted therapy and the effectiveness of traditional therapies, in clarification of molecular mechanisms of cancer therapeutic resistance and in the discovery of novel biomarkers and targeted drugs. For example, genomic and transcriptomic data and long-term clinical outcomes were recently analyzed to detect changes of gene expression based on somatic gene copy number aberrations. This analysis revealed important targeted therapeutic response-related events [105]. An integrative analysis of genomic and proteomic data demonstrated that aberrations of the PI3K pathway are particularly common in hormone receptor-positive breast cancer, which might be important in clinical selection of targeted therapies [106]. Another multi-omics data analysis, conducted in sorafenib-treated failure HCCs aimed, combined quantitative proteomics and phosphoproteomics data to better understand the molecules targeted by this drug. This study revealed that sorafenib can indeed effectively inhibit its target kinase in Raf-Erk-Rsk pathway, but the downstream targets of Rsk-2 (eIF4B, filamin-A and so on) were not affected, which suggests that another alternative pathways might have been active and contribute to the treatment failure [107].

Pathway databases for DTD

Recent emphasis on multi-omics data analysis has helped to pave the way for more systems biology-driven approaches to DTD [108, 109]. These approaches strongly rely on the integration of omics-driven information with pathway annotations in order to more accurately identify effective drug targets. Currently, there are a large number of pathways resources that are used for system biology analysis [110, 111]. These resources have a variety of goals, ranging from identifying gene functions in model organisms to providing tools for drug discovery. Pathway-based strategies are also useful for the identification of alternative druggable targets. Sometimes targets identified with GWAS and other omics technologies may not be druggable [112]. However, these undruggable genes may occur on a pathway with a partner

Table 8. Platforms and databases for DTD and evaluation

DTD	Link	Description	Main goals	License
DrugBank [39, 119]	drugbank.ca	A bioinformatics and chemoinformatics resource that combines drug and drug target information	Drug and target information	CC BY-NC 4.0
ChEMBL [120]	ebi.ac.uk/chembl	An open large-scale bioactivity database combining molecule, target and drug data	Drug and target information	CC BY-SA 3.0
DGIdb [121]	dgidb.org	A collection of drug–gene interactions and gene druggability information	Drug–gene interactions	MIT
TTD [38]	db.idrblab.org/ttd/	A database to provide information about known and explored therapeutic protein and nucleic acid targets and related targeted disease	Drug and target information	Free access
DisGeNET [68]	disgenet.org	A collection of genes and variants associated with human diseases	Gene disease associations	CC BY-NC-SA 4.0
DTC [2]	drugtargetcommons. fimm.fi	A crowd-sourcing platform to improve the consensus and use of drug target interactions	Drug target interactions	CC BY-NC-SA 3.0
Open Targets [3]	opentargets.org	Platform for target identification and prioritization target–disease associations	Target–disease associations	APACHE LICENSE, VERSION 2.0
PHAROS [122]	pharos.nih.gov	Knowledge base for the Druggable Genome	Target–disease associations	CC BY-SA 4.0
CTD [123, 124]	http://ctdbase.org/	A literature-based, manually curated associations between chemicals, gene products, phenotypes, diseases and environmental exposures	Drug-gene interactions Drug disease associations	TM
ADReCS-Target [125]	bioinf.xmu.edu.cn	A collection of ADRs caused by drug interaction with protein, gene and genetic variation	Drug target–adverse effect associations	Non-commercial use

that is from a known druggable family. Moreover, pathway-level information can help researchers identify drug target side effect [113]. Table 7 reports a brief description of pathway databases that can be used to implement systems biology approaches to drug discovery and validation.

DTD platforms

Over the past decade many databases and computational platforms for target discovery have been created to help scientists find robust evidence linking targets to diseases. These tools aim to assess the therapeutic efficacy of targets and, more recently, also their safety aspects. Initially, DTD platforms were not built specifically for omics-driven target discovery. However, they provide access to multiple biomedical data showing relevant target diseases associations. Large-scale data, such as omics, are systematically processed into concise information about drug target and target disease associations and drug- and target-related side effects. Table 8 lists brief descriptions of the reviewed DTDs.

DrugBank

DrugBank [120] collects comprehensive molecular information about drugs, their mechanisms, interactions and targets. It is primarily focused on providing data mining tools needed to facilitate target discovery and drug development. Latest versions of the DrugBank also provide information on the effect of hundreds of drugs on metabolite levels (pharmacometabolomics data), gene expression (pharmacotranscriptomics data) and protein expression (pharmacoproteomics data). New data have also been added on the status of hundreds of new drug clinical trials and

existing drug repurposing trials. Data in DrugBank are provided in an XML format. This makes data downloads and development of data extraction routines simpler and faster for programmers and database developers.

ChEMBL

ChEMBL [121] is a well-established resource in the fields of drug discovery and medicinal chemistry research. It curates and stores standardized bioactivity, molecule, target and drug data extracted from multiple sources, including the primary medicinal chemistry literature. Moreover, the ChEMBL database includes data typically generated in the preclinical and clinical phases of drug discovery, specifically drug metabolism and disposition data. These data help researchers better understand the key aspects of successful drug discoveries. ChEMBL data can be downloaded in a number of standard formats that can be automatically imported to external applications for further analyses. Alternatively, ChEMBL provides a REST API based service, allowing the remote retrieval of ChEMBL data and its integration into other applications.

DGidb

The DGidb database [122] collects drug–gene relationships and gene druggability information from 30 distinct repositories, including papers, databases and web resources. It has been the first DTD platform providing tools to capture and prioritize genes that are known to be targeted by existing drugs, especially targeted drugs rather than broad chemotherapeutics. Drug–gene interactions have been mined from existing databases

and literature to populate DGIdb. Similarly, genes have been categorized as potentially druggable according to membership in selected pathways, molecular functions and gene families. In the latest version, druggable genes from GWASs are also included. All data from DGIdb are available as tab-delimited data downloads and also through a web services API.

Therapeutic Target Database

The TTD [38] provides information about the known therapeutic protein and nucleic acid targets described in the literature, the targeted disease conditions, the pathway information and the corresponding drugs/ligands directed at each of these targets. The database currently contains 2025 targets, 17 816 drugs and 3681 multi-target agents. TTD also provides information about drug resistance mutations, gene expression profiles in the disease-relevant drug-targeted tissue of the patients and healthy individuals and target combinations of multi-target drugs and drug combinations. The database is organized through five main panels which authorize to browse it by advanced search, patient data, targets or drugs groups or by model data. It also permits to download various datasets.

DisGeNET

The DisGeNET platform [127] aims to overcome the fragmentation and heterogeneity of available genomic data for mining gene disease associations. It integrates data from manually curated databases, GWAS catalogs, animal models and the scientific literature. DisGeNET features a score based on the supporting evidence to prioritize gene disease associations. These scores rely on data retrieved from databases of curating genetic association studies (the GWAS Catalog and the Genetic Association Database) and genomic information extracted from animal models. DisGeNET can be used for different research purposes including the analysis of properties of disease genes, the generation of hypothesis on drug therapeutic action and drug adverse effects, the validation of computationally predicted disease genes and the evaluation of text mining methods performance. DisGeNET data are available for downloading in several formats: as SQLite database, as tab-separated files and as dump files, serialized in RDF/Turtle.

Drug Target Commons

DTC [2] is a crowdsourcing web platform that aims to standardize the collection, management, curation and annotation of the notoriously heterogeneous compound–target bioactivity data to facilitate drug discovery, target identification and drug repurposing. It integrates different, publicly available bioactivity data, which are generated using various assays, on compound–target interactions. Multiple data sources are systematically used for discovery of new indications for drugs (i.e. the selection of compound affecting specific proteins or biological pathways). DTC also provides access to QSAR models that can be used to extend target spaces for drugs. Besides, the drug-related bioactivity data are combined with chemical proteomic data in order to characterize biological pathways that are affected by certain drugs, to identify makers for drug monitoring and to determine drug-cocktails [128]. These data can be used to characterize potential safety issues associated with certain drugs (e.g. *in vivo* absorption, distribution, metabolism, excretion and toxicity properties) [129].

Open Targets Platform

Open Target [3] is a platform for therapeutic target identification and validation, providing either a target-centric workflow to identify diseases that may be associated with a specific target, or a disease-centric workflow to identify targets that may be associated with a specific disease. Coverage includes genetic associations, somatic mutations, know drugs, gene expression, affected pathways, literature mining and animal models. The latest version of Open Target provides information regarding the tractability of a target, which measures the 'ligandability' of putative drug targets [130], and safety risk information associated with selected targets. The Open Targets Platform allows programmatic retrieval of data via a set of REST services or, alternatively, the access to dump files.

Pharos

Pharos [123] provides a web interface for data collected by the Illuminating the Druggable Genome initiative. It incorporates text-mined bibliometric associations and statistics from the biomedical and patent literature, mRNA and protein expression data, disease and phenotype associations, bioactivity data, drug target interactions, and omics-driven data imported from the Harmonizome. It also integrates with the functionality of the Drug Central and DTO resources. The Drug Target Ontology is a database providing tools to classify and integrate drug discovery data based on formalized and standardized classifications and annotations of druggable protein targets. DTO integrates phylogenecity, function, target development level, disease association, tissue expression, chemical ligand and substrate characteristics and target-family specific characteristics. Protein classes are linked to tissue and disease via different levels of confidence. DTO also contains drug target development classifications, a large collection of cell lines from the LINCS project and relevant cell–disease and cell–tissue relations. DTO is modeled in OWL2-DL to enable further classification by inference reasoning and SPARQL queries. DTO is implemented following a modularization approach. DTO will serve as the organizational framework for drug targets in the IDG PHAROS User Interface Portal.

Comparative Toxicogenomics Database

CTD [124] is a public resource that provides information about interactions between chemicals and gene products, and their relationships to diseases. Professional biocurators manually curate the scientific literature, transforming text, tables, figures and supplemental files into annotated data that are seamlessly integrated and available through CTD's public web application (PWA). Different data sources for toxicogenomics, phenotypes, diseases, environmental exposures and pharmaceuticals are considered in order to build drug gene and drug diseases interactions. Overall, CTD includes over 38 million toxicogenomic relationships for analysis and hypothesis development. This information is organized through community-accepted controlled vocabularies and ontologies with accession identifiers, in order to ensure that CTD's content is cohesive, manageable and computable, as well as adhering to the FAIR principle. CTD's vocabularies and content are described and made freely available for users to download in a variety of formats.

ADReCS-Target

ADReCS-Target [126] provides target profiles for aiding drug safety research and application by collecting data about adverse

Table 9. Comparison on drug target–disease associations

DTD	Main association	Drug target	Target–disease	Efficacy	Safety	Novelty
DrugBank	Drug target (RNA, DNA and other molecules)	Drug binding data Drug pharmacokinetics Drug bioavailability Drug ADMET characteristics Clinical trials (TTD, STITCH, BindingDB, ChEMBL)	External links to ChemSpider, HMDB, MMCD, SMPDB and OMIM	Yes (DrT)	Yes	Yes
ChEMBL	Molecule-target (genes/proteins)	Efficacy assays data ADME assays data Drug metabolism data Toxicity assay data	External link to ClinicalTrials.gov	Yes (DrT)	Yes	Yes
DGIdb	Drug target (genes)	Drug bioactivity data Physically binding data Modulation and indirect interaction RNA drug binding data (DrugBank, TTD, ChEMBL, TALC)	Missing	Yes (DrT)	No	No
TTD	Drug target Target disease	Drug target interaction (PubChem, DrugBank, SuperDrug and ChEBI)	Gene expression profiles	Yes (DsT)	No	No
DisGeNET	Gene diseases Variant diseases	External links to Drug activity data Drug gene interaction Drug adverse reaction (ChEMBL, CTD, Sider)	Genomic data (GWAS) Scientific literature Animal models	Yes (DsT)	No	No
DTC	Drug target interactions (proteins)	Drug activity data; clinical development information of drugs (25 databases, including ChEMBL, PubChem, DrugBank, PharmGKB and ClinicalTrials.gov)	External links to DisGeNET, Cancer Genome Interpreter)	Yes (DrT)	No	No
Open Targets	Target disease (genes/proteins)	External link to ChEMBL	Genetic associations; somatic mutations; drugs pathways & systems biology; RNA expression; text mining; animal models	Yes (DrT)	Yes	No
PHAROS	Drug target	Scientific literature mRNA and protein expression data Disease and phenotype associations Bioactivity data Drug target interactions Adverse drug reactions	External links to DisGeNET, Expression Atlas GTEx, GWAS Catalog, JensenLab data	Yes (DrT)	No	Yes
CTD	Drug target Drug disease Target disease (genes/proteins)	Curated chemical–gene interactions (bioactivity, binding, expression, mutagenesis and metabolic processing)	Curated chemical–diseases interactions Inferred gene disease associations	Yes (DsT)	Yes	No
ADReCS-Target	Drug target/side effects (genes/proteins)	A collection of ADRs caused by drug interaction with protein, gene and genetic variation	External links to CTD, DrugBank, dbSNP	No	Yes	No

Efficacy estimates can refer to drug target interaction (DrT) or disease target associations (DsT).

drug reactions (ADRs) caused by drug interactions with protein, gene and genetic variation. ADReCS-Target contains more than 66 000 association pairs with over 2200 standard ADR terms manually curated from text mining of the public scientific literatures. All the terms are standardized by using ADReCS ontology and represented as a connected network or systematic fashion. The user can download selected records via the embedded download function in six formats such as JSON, XML, CSV, TXT, SQL and MS-Excel. ADReCS-Target also allows batch data retrieval.

Comparative analysis of DTD platforms

In this section we compare the selected DTDs based on what type of associations are pursued, the information provided to inform about efficacy and safety of putative targets and the omics data employed for the discovery process.

Comparing the information presented in Table 9, we can observe that many existing DTDs help identify and prioritize drug targets solely on the basis of molecule–target interactions, without specifying whether the target is disease-modifying and/or has a proven function in the pathophysiology of a disease. On the other hand, web platforms such as Open Targets and DisGeNET provide information on drug target–disease associations. These tools provide also efficacy scores of target disease associations calculated from multiple sources of evidence, including omics and scientific literature. Both scoring systems can be utilized to summarize and rank disease-relevant targets. This is especially the case of the Open Targets

Table 10. Comparison based on the use of omics data layers and prioritization tool

DTD	Omics	Omics data types	External DB	Ranking
DrugBank	Genomic Transcriptomic Metabolomic Proteomic	SNP-drug data (*) Up/down regulation of genes due to drug metabolism Manually compile metabolomic information (*) Drug-action pathways on protein targets (*)	dbSNP, Literature, SMPDB, HMDB, T3DB, SMPDB, Uniprot, CTD	Not available
ChEMBL	Transcriptomic Genomic	Gene expression profiles induced by chemical or drug exposure (*) Drug sensitivity (*)	TG-GATE, DrugMatrix, Gene Expression Atlas, GDSC	Confidence score to rank molecule-target interactions
DGIdb	Genomic	Druggable genome/genes (*)	MyCancerGenome	Number of distinct sources of evidence and PMIDs supporting each interaction.
TTD	Transcriptomic Genomic	Tissue-specific gene expression profiles in healthy and diseased individuals (*) Drug resistance mutation (*)	Gene Expression Omnibus and ArrayExpress Literature	Not available
DisGeNET	Transcriptomic Genomic	Gene expression alteration (*) Relationships between human variants/genes and phenotypes/diseases (*) Genome association studies (*)	Gene Expression Atlas CTD, CLINVAR Orphanet, GWAS Catalog, The Genetic Association Database (GAD)	GDA score to rank the gene disease according to their level of evidence. This score compiles efficacy scores on the basis of genomic information and scientific literature.
DTC	Genomic	Gene disease associations (*) Somatic mutation information (*)	DisGeNET	Not available
Open Targets	Transcriptomic Genomic/Genetic	Expression profile of diseases (*) Genome association studies, somatic mutation (*)	Gene Expression Atlas GWAS/PheWAS Catalog, Gene2Phenotype	Multi-evidence ranking of target disease associations
PHAROS	Transcriptomic/Proteomic Genomic/Genetic	Tissue-specific RNA expression (*) Genome association studies (*)	GTEx, Expression Atlas, Harmonizome, CPR (Jensen Lab), EMBL-EBI datasets	Not available
CTD	Transcriptomic Genomic Metabolomic	Gene expression alteration (*) Genetic alteration of a gene product (*) Metabolic processing (*)	DrugBank	Not available
ADReCS-Target	Genomic/Genetic	Gene disease associations (*) Drug relevant genetic variations (*)	CTD, GWAS Catalog, DrugBank	Not available

The (*) indicates that omics-driven information is obtained from an external data source, database or literature.

platform, which allows the collection and fusion of many pieces of evidence to support target disease associations. The Open Targets platform employs data mining algorithms in order to estimate a numerical score for each type of evidence. These scores are finally merged through the harmonic sum, which gives an overall efficacy estimate indicting strength of an association between a molecular target and a disease. More recently, Open Targets platform has provided a new source of information namely 'target tractability' which can be used to collect drug target details, such as whether there is a binding site in the protein that can be used for small molecule binding, or an accessible epitope for antibody based therapy. This can assist in target prioritization, drug target inclusion in discovery pipelines and selection of therapeutic modalities that are most likely to succeed. An important aspect of drug target prioritization is target safety assessment, which should aim to identify potential unintended adverse consequences of target modulation [33]. However, to the best of our knowledge, none of the presented DTD platforms provide safety evaluation scores to be combined with efficacy estimates of target disease association. Open Target and DisGeNET do not directly provide information on drug target interactions. CTD is the only platform

aiming to bridge the information gap between drug target interaction and target disease association tasks. It integrates multiple sources of information for drug target interaction and it includes clinical development information for the compounds and target gene disease associations, as well as cancer-type indications for mutant protein targets, which are critical for precision oncology developments. However, CTD often considers the disease as an adverse outcome given by the interaction between a chemical and a gene. Another important observation is that many DTDs do not directly indicate information on potential side effects that could be caused by a drug interacting with the selected molecular target (e.g. gene or protein), if the selected targets are classified as 'essential' or play a role in oncogenic pathways. The term essential gene can refer to genes encoding proteins that are necessary to maintain a central metabolism, replicate DNA, translate genes into proteins, maintain a basic cellular structure, etc. CTD and ADReCS-Target provide information on putative side effects associated with drug target. In particular, CTD integrates chemical and biological information to elucidate toxicology relationships, while ADReCS-Target provides information on drug toxicity–target relationships. TTD and DrugBank provide information on

multi-target drugs or therapeutic combinations. TTD provides information on synergistic, additive, antagonistic, potentiative and reductive drug combinations. Whereas, the DrugBank database includes a set of 12 128 drug–drug interactions along with a brief textual description of the interaction and information about therapeutic effects [131]. Overall, the existing DTDs focus more on efficacy evaluations of drug target–disease associations and less on safety aspects of drug targets, and none of them provide ranking systems to prioritize drug targets on efficacy and safety evaluations simultaneously. Table 10 lists the omics data types that DTDs use for mining drug target or target disease associations. In particular, we can observe that genomic (genetic) and transcriptomic data are commonly used and that they are often obtained from pre-compiled omics data analysis. Moreover, this information is not always utilized to compile efficacy estimates of drug–gene or gene disease associations. Platforms such as ChEMBL, DGIdb, DisGeNET and Open Targets provide scoring methods to rank drug target or target disease interactions. However, these methods often do not fully exploit the data collected by DTDs, for instance, DGIdb simply reports the number of distinct sources and distinct PubMed IDs (PMIDs) supporting each interaction. Moreover, benchmark studies supporting the validity of these scoring methods are missing. As DTD is one of the first phases of drug development process, it seems feasible that DTD platforms would be incorporated into the early stages of the process. The platforms can be utilized to discover completely new targets, rank existing targets lists to identify the most likely candidates based on different criteria or to provide additional evidence and starting points for deep dives into specific targets. To be able to efficiently incorporate DTD platforms to an existing drug development process and pipeline, the following features are desirable: (i) possibility to batch process a set of targets, (ii) programmatic access to the platform through API and possibility to download the data, (iii) possibility to incorporate private data, (iv) standardized data formats and gene/gene product identifiers, (v) references to where and how the evidence for a specific target was gathered, (vi) rankable evidence scores and (vii) license that enables intended use. We believe these features make a platform easier to integrate with other DTD platforms and existing drug development pipelines.

Conclusion

Current DTD platforms provide alternative ways of utilizing omics data sources for improved drug target prioritization and selection. However, there could be some improvements on the data mining algorithms which are used to quantify the efficacy estimates of drug target–disease associations. In particular, genomic, transcriptomics and proteomics data could be more efficiently used to provide new ways to link targets to diseases and validate these targets. Perhaps, it would be important in the near future to develop computational tools that could assist with the integration of these complex multi-omics data sets in order to more robustly identify drug targets. Moreover, there has been a little effort to apply omics data for early identification of safety-related issues of putative drug targets. The research using clinical data by computational biologists and biostatisticians, in academia and industry, continuously work toward the development of cost-effective and sensitive diagnostic biomarkers. Overall, we identified three major technical gaps that could be bridged by the next generation of drug discovery platforms: (i) the lack of *in silico* tools for target safety assessment, (ii) comparative analysis of different efficacy and safety estimates for drug target prioritization and (iii) systematic

identification of multiple drug targets and selection of optimal therapeutic strategies.

Key Points

- Target-based drug discovery is still largely manual work that bottlenecks the whole drug discovery process.
- Many computational platforms exist to rapidly identify and prioritize genes or proteins that encode promising drug targets from hundreds of data sources, ranging from scientific publications to omics databases.
- There are a few platforms providing omics-driven efficacy estimates of target disease associations.
- No single tool, platform or database supports drug target prioritization based on efficacy and safety assessment scores.

References

1. Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. *Contemp Clin Trials Commun* 2018;**11**:156–64.
2. Tanoli Z, Alam Z, Vähä-Koskela M, *et al*. Drug Target Commons 2.0: a community platform for systematic analysis of drug-target interaction profiles. *Database (Oxford)* 2018;**2018**:1–13.
3. Koscielny G, An P, Carvalho-Silva D, *et al*. Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res* 2017;**45**:D985–94.
4. Matthews H, Hanison J, Nirmalan N. Omics-informed drug and biomarker discovery: opportunities, challenges and future perspectives. *Proteomes* 2016;**4**(3):28.
5. Amaratunga D, Göhlmann H, Peeters PJ. Microarrays. *Comprehensive Medicinal Chemistry II*. Elsevier, 2007;87–106.
6. Bemis GW, Murcko MA. The properties of known drugs. 1. Molecular frameworks. *J Med Chem* 1996;**39**:2887–93.
7. Zheng CJ, Han LY, Yap CW, *et al*. Therapeutic targets: progress of their exploration and investigation of their characteristics. *Pharmacol Rev* 2006;**58**:259–79.
8. Russ AP, Lampel S. The druggable genome: an update. *Drug Discov Today* 2005;**10**:1607–10.
9. Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov* 2002;**1**:727–30.
10. Bakheet TM, Doig AJ. Properties and identification of human protein drug targets. *Bioinformatics* 2009;**25**:451–7.
11. Kim B, Jo J, Han J, *et al*. In silico re-identification of properties of drug target proteins. *BMC Bioinformatics* 2017;**18**:248.
12. Imming P, Sinning C, Meyer A. Drugs, their targets and the nature and number of drug targets. *Nat Rev Drug Discov* 2006;**5**:821–34.
13. Diamantopoulos MA, Tsiakanikas P, Scorilas A. Non-coding RNAs: the riddle of the transcriptome and their perspectives in cancer. *Ann Transl Med* 2018;**6**:241.
14. Connelly CM, Moon MH, Schneekloth JS. The emerging role of RNA as a therapeutic target for small molecules. *Cell Chem Biol* 2016;**23**:1077–90.
15. McKnight KL. Heinz BA. RNA as a target for developing antivirals. *Antivir Chem Chemother* 2003;**14**:61–73.
16. Dersch P, Khan MA, Mühlen S, *et al*. Roles of regulatory rnas for antibiotic resistance in bacteria and their potential value as novel drug targets. *Front Microbiol* 2017;**8**:803.

17. Matsui M, Corey DR. Non-coding RNAs as drug targets. *Nat Rev Drug Discov* 2017;**16**:167–179.

18. Warner KD, Hajdin CE, Weeks KM. Principles for targeting RNA with drug-like small molecules. *Nat Rev Drug Discov* 2018;**17**:547–58.

19. Sah DWY, Aronin N. Oligonucleotide therapeutic approaches for Huntington disease. *J Clin Invest* 2011;**121**:500–7.

20. Ozcan G, Ozpolat B, Coleman RL, *et al*. Preclinical and clinical development of siRNA-based therapeutics. *Adv Drug Deliv Rev* 2015;**87**:108–19.

21. Boran ADW, Iyengar R. Systems approaches to polypharmacology and drug discovery. *Curr Opin Drug Discov Devel* 2010;**13**:297–309.

22. Sheng Z, Sun Y, Yin Z, *et al*. Advances in computational approaches in identifying synergistic drug combinations. *Brief Bioinform* 2018;**19**:1172–82.

23. Vukovic S, Huggins DJ. Quantitative metrics for drug-target ligandability. *Drug Discov Today* 2018;**23**:1258–66.

24. Gashaw I, Ellinghaus P, Sommer A, *et al*. What makes a good drug target? *Drug Discov Today* 2011;**16**:1037–43.

25. Alexander-Dann B, Pruteanu LL, Oerton E, *et al*. Developments in toxicogenomics: understanding and predicting compound-induced toxicity from gene expression data. *Mol Omics* 2018;**14**:218–36.

26. Cannon DC, Yang JJ, Mathias SL, *et al*. TIN-X: target importance and novelty explorer. *Bioinformatics* 2017;**33**:2601–3.

27. Wei C-H, Kao H-Y, Lu Z. PubTator: a web-based text mining tool for assisting biocuration. *Nucleic Acids Res* 2013;**41**:W518–22.

28. Capuzzi SJ, Thornton TE, Liu K, *et al*. Chemotext: a publicly available web server for mining drug-target–disease relationships in PubMed. *J Chem Inf Model* 2018;**58**:212–8.

29. Percha B, Altman RB. A global network of biomedical relationships derived from text. *Bioinformatics* 2018;**34**:2614–24.

30. Zhu S, Bing J, Min X, *et al*. Prediction of drug–gene interaction by using metapath2vec. *Front Genet* 2018;**9**:248.

31. Stoeger T, Gerlach M, Morimoto RI, *et al*. Large-scale investigation of the reasons why potentially important genes are ignored. *PLoS Biol* 2018;**16**:e2006643.

32. Santos R, Ursu O, Gaulton A, *et al*. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov* 2017;**16**:19–34.

33. Brennan RJ. Target safety assessment: strategies and resources. *Methods Mol Biol* 2017;**1641**:213–28.

34. Bolognesi ML, Cavalli A. Multitarget drug discovery and polypharmacology. *ChemMedChem* 2016;**11**:1190–2.

35. Talevi A. Multi-target pharmacology: possibilities and limitations of the "skeleton key approach" from a medicinal chemist perspective. *Front Pharmacol* 2015;**6**:205.

36. Vakil V, Trappe W. Drug combinations: mathematical modeling and networking methods. *Pharmaceutics* 2019;**11**.

37. Al-Lazikani B, Banerji U, Workman P. Combinatorial drug therapy for cancer in the post-genomic era. *Nat Biotechnol* 2012;**30**:679–92.

38. Chen X, Ji ZL, Chen YZ. TTD: therapeutic target database. *Nucleic Acids Res* 2002;**30**:412–5.

39. Wishart DS, Knox C, Guo AC, *et al*. DrugBank: a comprehensive resource for *in silico* drug discovery and exploration. *Nucleic Acids Res* 2006;**34**:D668–72.

40. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol* 2008;**4**:682–90.

41. Foucquier J, Guedj M. Analysis of drug combinations: current methodological landscape. *Pharmacol Res Perspect* 2015;**3**:e00149.

42. Simon R, Roychowdhury S. Implementing personalized cancer genomics in clinical trials. *Nat Rev Drug Discov* 2013;**12**:358–69.

43. Cannon ME, Mohlke KL. Deciphering the emerging complexities of molecular mechanisms at GWAS loci. *Am J Hum Genet* 2018;**103**:637–53.

44. Baxter JS, Leavy OC, Dryden NH, *et al*. Capture Hi-C identifies putative target genes at 33 breast cancer risk loci. *Nat Commun* 2018;**9**:1028.

45. Doostparast Torshizi A, Wang K. Next-generation sequencing in drug development: target identification and genetically stratified clinical trials. *Drug Discov Today* 2018;**23**:1776–83.

46. Uenaka T, Satake W, Cha P-C, *et al*. *In silico* drug screening by using genome-wide association study data repurposed dabrafenib, an anti-melanoma drug, for Parkinson's disease. *Hum Mol Genet* 2018;**27**:3974–85.

47. Yin W, Gao C, Xu Y, *et al*. Learning opportunities for drug repositioning via GWAS and phewas findings. *AMIA Jt Summits Transl Sci Proc* 2018;**2017**:237–46.

48. Ferrero E, Agarwal P. Connecting genetics and gene expression data for target prioritisation and drug repositioning. *BioData Min* 2018;**11**:7.

49. Weinshilboum RM, Wang L. Pharmacogenomics: precision medicine and drug response. *Mayo Clin Proc* 2017;**92**:1711–22.

50. Barbarino JM, Whirl-Carrillo M, Altman RB, *et al*. PharmGKB: a worldwide resource for pharmacogenomic information. *Wiley Interdiscip Rev Syst Biol Med* 2018;**10**:e1417.

51. Buniello A, MacArthur JAL, Cerezo M, *et al*. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 2019;**47**:D1005–12.

52. Beck T, Hastings RK, Gollapudi S, *et al*. GWAS Central: a comprehensive resource for the comparison and interrogation of genome-wide association studies. *Eur J Hum Genet* 2014;**22**:949–52.

53. Tryka KA, Hao L, Sturcke A, *et al*. NCBI's Database of Genotypes and Phenotypes: dbGaP. *Nucleic Acids Res* 2014;**42**:D975–9.

54. Sawada R, Iwata M, Tabei Y, *et al*. Predicting inhibitory and activatory drug targets by chemically and genetically perturbed transcriptome signatures. *Sci Rep* 2018;**8**:156.

55. Natsoulis G, El Ghaoui L, Lanckriet GRG, *et al*. Classification of a large microarray data set: algorithm comparison and analysis of drug signatures. *Genome Res* 2005;**15**:724–36.

56. Igarashi Y, Nakatsu N, Yamashita T, *et al*. Open TG-GATEs: a large-scale toxicogenomics database. *Nucleic Acids Res* 2015;**43**:D921–7.

57. Lamb J, Crawford ED, Peck D, *et al*. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006;**313**:1929–35.

58. Subramanian A, Narayan R, Corsello SM, *et al*. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 2017;**171**:1437–1452.e17.

59. Clough E, Barrett T. The gene expression omnibus database. *Methods Mol Biol* 2016;**1418**:93–110.

60. Kolesnikov N, Hastings E, Keays M, *et al*. ArrayExpress update—simplifying data submissions. *Nucleic Acids Res* 2015;**43**:D1113–6.

61. Piñero J, Gonzalez-Perez A, Guney E, *et al*. Network, transcriptomic and genomic features differentiate genes relevant for drug response. *Front Genet* 2018;**9**:412.

62. Isik Z, Baldow C, Cannistraci CV, *et al*. Drug target prioritization by perturbed gene expression and network information. *Sci Rep* 2015;**5**:17417.

63. Pritchard J-LE, O'Mara TA, Glubb DM. Enhancing the promise of drug repositioning through genetics. *Front Pharmacol* 2017;**8**:896.

64. Ho Y-Y, Cope LM, Parmigiani G. Modular network construction using eQTL data: an analysis of computational costs and benefits. *Front Genet* 2014;**5**:40.

65. Consortium GTE. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015;**348**:648–60.

66. Wang Y, Wang Z, Xu J, *et al*. Systematic identification of non-coding pharmacogenomic landscape in cancer. *Nat Commun* 2018;**9**:3192.

67. Schmitt AM, Chang HY. Long noncoding rnas in cancer pathways. *Cancer Cell* 2016;**29**:452–63.

68. Parasramka M, Yan IK, Wang X, *et al*. BAP1 dependent expression of long non-coding RNA NEAT-1 contributes to sensitivity to gemcitabine in cholangiocarcinoma. *Mol Cancer* 2017;**16**:22.

69. Barretina J, Caponigro G, Stransky N, *et al*. The Cancer Cell Line Encyclopedia enables predictive modelling of anti-cancer drug sensitivity. *Nature* 2012;**483**:603–7.

70. 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.

71. Dong Z, Zhang N, Li C, *et al*. Anticancer drug sensitivity prediction in cell lines from baseline gene expression through recursive feature selection. *BMC Cancer* 2015;**15**:489.

72. Papatheodorou I, Fonseca NA, Keays M, *et al*. Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res* 2018;**46**:D246–51.

73. Schirle M, Bantscheff M, Kuster B. Mass spectrometry-based proteomics in preclinical drug discovery. *Chem Biol* 2012;**19**:72–84.

74. Hebert AS, Prasad S, Belford MW, *et al*. Comprehensive single-shot proteomics with FAIMS on a hybrid orbitrap mass spectrometer. *Anal Chem* 2018;**90**:9529–37.

75. Deutsch EW, Mendoza L, Shteynberg D, *et al*. Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. *Proteomics Clin Appl* 2015;**9**:745–54.

76. Lee PY, Chin S-F, Low TY, *et al*. Probing the colorectal cancer proteome for biomarkers: current status and perspectives. *J Proteomics* 2018;**187**:93–105.

77. Swiatly A, Plewa S, Matysiak J, *et al*. Mass spectrometry-based proteomics techniques and their application in ovarian cancer research. *J Ovarian Res* 2018;**11**:88.

78. Jones LH, Neubert H. Clinical chemoproteomics-opportunities and obstacles. *Sci Transl Med* 2017;**9**.

79. Nguyen C, West GM, Geoghegan KF. Emerging methods in chemoproteomics with relevance to drug discovery. *Methods Mol Biol* 2017;**1513**:11–22.

80. Luck K, Sheynkman GM, Zhang I, *et al*. Proteome-scale human interactomics. *Trends Biochem Sci* 2017;**42**:342–54.

81. Kanhaiya K, Czeizler E, Gratie C, *et al*. Controlling directed protein interaction networks in cancer. *Sci Rep* 2017;**7**:10327.

82. Feng Y, Wang Q, Wang T. Drug target protein–protein interaction networks: a systematic perspective. *Biomed Res Int* 2017;**2017**:1289259.

83. Johnson ECB, Dammer EB, Duong DM, *et al*. Deep proteomic network analysis of Alzheimer's disease brain reveals alterations in RNA binding proteins and RNA splicing associated with disease. *Mol Neurodegener* 2018;**13**:52.

84. Vizcaíno JA, Csordas A, del-Toro N, *et al*. 2016 Update of the PRIDE database and its related tools. *Nucleic Acids Res* 2016;**44**:D447–56.

85. Schmidt T, Samaras P, Frejno M, *et al*. ProteomicsDB. *Nucleic Acids Res* 2018;**46**:D1271–81.

86. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019;**47**:D506–15.

87. Berman HM, Westbrook J, Feng Z, *et al*. The protein data bank. *Nucleic Acids Res* 2000;**28**:235–42.

88. Kim M-S, Pinto SM, Getnet D, *et al*. A draft map of the human proteome. *Nature* 2014;**509**:575–81.

89. Uhlén M, Fagerberg L, Hallström BM, *et al*. Proteomics. Tissue-based map of the human proteome. *Science* 2015;**347**:1260419.

90. Patti GJ, Yanes O, Innovation SG. Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* 2012;**13**:263–9.

91. Wishart DS. Emerging applications of metabolomics in drug discovery and precision medicine. *Nat Rev Drug Discov* 2016;**15**:473–84.

92. Vincent EE, Sergushichev A, Griss T, *et al*. Mitochondrial phosphoenolpyruvate carboxykinase regulates metabolic adaptation and enables glucose-independent tumor growth. *Mol Cell* 2015;**60**:195–207.

93. Balashova EE, Maslov DL, Lokhov PG. A metabolomics approach to pharmacotherapy personalization. *J Pers Med* 2018;**8**:3.

94. Wang P, Shehu AI, Ma X. The opportunities of metabolomics in drug safety evaluation. *Curr Pharmacol Rep* 2017;**3**:10–5.

95. Yang B, Wang C, Xie Y, *et al*. Monitoring tyrosine kinase inhibitor therapeutic responses with a panel of metabolic biomarkers in chronic myeloid leukemia patients. *Cancer Sci* 2018;**109**:777–84.

96. Kale NS, Haug K, Conesa P, *et al*. MetaboLights: an open-access database repository for metabolomics data. *Curr Protoc Bioinformatics* 2016;**53**:14.13:1–18.

97. Cui Q, Lewis IA, Hegeman AD, *et al*. Metabolite identification via the Madison Metabolomics Consortium Database. *Nat Biotechnol* 2008;**26**:162–4.

98. Hummel J, Strehmel N, Bölling C, *et al*. Mass spectral search and analysis using the golm metabolome database. In: *The Handbook of Plant Metabolomics*. Kahl:Plant Metabolomics O-Bk, 2013, 321–43.

99. Horai H, Arita M, Kanaya S, *et al*. MassBank: a public repository for sharing mass spectral data for life sciences. *J Mass Spectrom* 2010;**45**:703–14.

100. Carroll AJ, Badger MR, Harvey Millar A. The Metabolome-Express Project: enabling web-based processing, analysis and transparent dissemination of GC/MS metabolomics datasets. *BMC Bioinformatics* 2010;**11**:376.

101. Sansone S-A, Gonzalez-Beltran A, Rocca-Serra P, *et al*. DATS, the data tag suite to enable discoverability of datasets. *Sci. Data* 2017;**4**:170059.

102. Perez-Riverol Y, Bai M, da Veiga Leprevost F, *et al*. Discovering and linking public omics data sets using the Omics Discovery Index. *Nat Biotechnol* 2017;**35**:406–9.

103. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biol* 2017;**18**:83.

104. Lu M, Zhan X. The crucial role of multiomic approach in cancer research and clinically relevant outcomes. *EPMA J* 2018;**9**:77–102.

105. Curtis C, Shah SP, Chin S-F, *et al*. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012;**486**:346–52.

106. Stemke-Hale K, Gonzalez-Angulo AM, Lluch A, *et al*. An integrative genomic and proteomic analysis of PIK3CA, PTEN, and AKT mutations in breast cancer. *Cancer Res* 2008;**68**:6084–91.

107. Dazert E, Colombi M, Boldanova T, *et al*. Quantitative proteomics and phosphoproteomics on serial tumor biopsies from a sorafenib-treated HCC patient. *Proc Natl Acad Sci U S A* 2016;**113**:1381–6.

108. Ramos PIP, Fernández Do Porto D, Lanzarotti E, *et al*. An integrative, multi-omics approach towards the prioritization of Klebsiella pneumoniae drug targets. *Sci Rep* 2018; **8**:10755.

109. Cisek K, Krochmal M, Klein J, *et al*. The application of multi-omics and systems biology to identify therapeutic targets in chronic kidney disease. *Nephrol Dial Transplantation* 2016;**31**:2003–11.

110. Jin L, Zuo X-Y, Su W-Y, *et al*. Pathway-based analysis tools for complex diseases: a review. *Genomics Proteomics Bioinformatics* 2014;**12**:210–20.

111. Chowdhury S, Sarkar RR. Comparison of human cell signaling pathway databases–evolution, drawbacks and challenges. *Database (Oxford)* 2015;**2015**.

112. Dang CV, Reddy EP, Shokat KM, *et al*. Drugging the "undruggable" cancer targets. *Nat Rev Cancer* 2017;**17**:502–8.

113. Fukuzaki M, Seki M, Kashima H, *et al*. Side effect prediction using cooperative pathways. *2009 IEEE International Conference on Bioinformatics and Biomedicine*, Washington, DC, USA: IEEE, 2009;142–7.

114. Kanehisa M, Furumichi M, Tanabe M, *et al*. KEGG: new perspectives on genomes, pathways,diseases and drugs. *Nucleic Acids Res* 2017;**45**:D353–61.

115. Caspi R, Billington R, Ferrer L, *et al*. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 2016;**44**:D471–80.

116. Vastrik I, D'Eustachio P, Schmidt E, *et al*. Reactome: a knowledge base of biologic pathways and processes. *Genome Biol* 2007;**8**:R39.

117. Slenter DN, Kutmon M, Hanspers K, *et al*. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res* 2018;**46**: D661–7.

118. Cerami EG, Gross BE, Demir E, *et al*. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res* 2011;**39**:D685–90.

119. Thorn CF, Klein TE, Altman RB. Pharmgkb: the pharmacogenomics knowledge base. *Methods Mol Biol* 2013;**1015**:311–20.

120. Wishart DS, Feunang YD, Guo AC, *et al*. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;**46**:D1074–82.

121. Mendez D, Gaulton A, Bento AP, *et al*. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res* 2019;**47**:D930–40.

122. Cotto KC, Wagner AH, Feng Y-Y, *et al*. DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. *Nucleic Acids Res* 2018;**46**:D1068–73.

123. Nguyen D-T, Mathias S, Bologa C, *et al*. Pharos: collating protein information to shed light on the druggable genome. *Nucleic Acids Res* 2017;**45**:D995–1002.

124. Davis AP, Grondin CJ, Johnson RJ, *et al*. The Comparative Toxicogenomics Database: update 2019. *Nucleic Acids Res* 2019;**47**:D948–54.

125. Davis AP, Grondin CJ, Johnson RJ, *et al*. The Comparative Toxicogenomics Database: update 2017. *Nucleic Acids Res* 2017;**45**:D972–8.

126. Huang L-H, He Q-S, Liu K, *et al*. ADReCS-Target: target profiles for aiding drug safety research and application. *Nucleic Acids Res* 2018;**46**:D911–7.

127. Piñero J, Bravo À, Queralt-Rosinach N, *et al*. Dis GeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* 2017;**45**:D833–9.

128. Ramsay RR, Popovic-Nikolic MR, Nikolic K, *et al*. A perspective on multi-target drug discovery and design for complex diseases. *Clin Transl Med* 2018;**7**:3.

129. Ravikumar B, Aittokallio T. Improving the efficacy-safety balance of polypharmacology in multi-target drug discovery. *Expert Opin Drug Discovery* 2018;**13**:179–92.

130. Bauer U, Breeze AL. "Ligandability" of drug targets: assessment of chemical tractability via experimental and *in silico* approaches. *Lead Generation* 2016;35–62.

131. Lin H-H, Zhang L-L, Yan R, *et al*. Network analysis of drug-target interactions: a study on FDA-approved new molecular entities between 2000 to 2015. *Sci Rep* 2017;**7**: 12230.