Briefings in Bioinformatics, 2017, 1–18

doi: 10.1093/bib/bbx125
Paper

Web-based drug repurposing tools: a survey

Elizabeth Sam and Prashanth Athri

Corresponding author: Prashanth Athri, Department of Computer Science and Engineering, Amrita University, Kasavanahalli, Carmelaram PO, Bangalore – 560035, India. Tel.: +91 80 2518 3700 (Ext. 727). E-mail: prashanthathri@gmail.com

Abstract

Drug repurposing (a.k.a. drug repositioning) is the search for new indications or molecular targets distinct from a drug's putative activity, pharmacological effect or binding specificities. With the ever-increasing rates of termination of drugs in clinical trials, drug repositioning has risen as one of the effective solutions against the risk of drug failures. Repositioning finds a way to reverse the grim but real trend that Eroom's law portends for the pharmaceutical and biotech industry, and drug discovery in general. Further, the advent of high-throughput technologies to explore biological systems has enabled the generation of zeta bytes of data and a massive collection of databases that store them. Computational analytics and mining are frequently used as effective tools to explore this byzantine series of biological and biomedical data. However, advanced computational tools are often difficult to understand or use, thereby limiting their accessibility to scientists without a strong computational background. Hence it is of great importance to build user-friendly interfaces to extend the user-base beyond computational scientists, to include life scientists who may have deeper chemical and biological insights. This survey is focused on systematically presenting the available Web-based tools that aid in repositioning drugs.

Key words: drug repurposing, drug repositioning, target prediction, approved drugs, drug indications, off-target predictions

Introduction

Systematic drug repurposing is the re-evaluation of known pharmaceutically relevant compounds toward identifying new indications. It has deep impact in the area of personalized medicine, and promises rapid translation to bedside. Repurposing studies provide an alternative paradigm for therapeutic stratification that aim at individual-centric classification of disease phenotypes [1], and hence, address the hard reality of the Eroom's law [2] (it points to the halving of efficiency in research and development of new drugs every 9 years or so in the United States and indicative of the downward trend). Repurposing compounds significantly reduce R&D time and financial liabilities because they have known bioavailability and ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity) profiles [3]. It offers pharmaceutical/biotech companies and non-profits engaged in drug discovery and development an alternate route to cheaper, low-risk drug development, thus alleviating the stagnation owing to the reported steep decrease in active substance entries [4]. Finding alternate uses to old drugs has the obvious advantage of cutting down on discovery and development phases of research, and major savings on cost and time of drug development. The ideal candidates for repurposing are leads which have made it past Phase III, in terms of the American Food and Drug Administration (FDA) system, as this implies they are proven to be efficacious in larger populations and verified to be safe [5]. In effect, clinical trials can proceed at a much faster rate. Because, *in vitro* and *in vivo* screening, chemical optimization, toxicology, bulk manufacturing and clinical trials have already been completed and can be bypassed, it removes substantial risks and costs from the pathway to the market.

The recent times have seen many successes in repositioning old drugs (see RepurposeDB [6] for a list of repurposed drugs), and what initially was driven by serendipity is now driven by focused, systematic computational explorations that precede shorter experimental project cycles. The promise of teaching new tricks to old drugs [7] inspired the development of a plethora of data-driven computational prediction and analytic algorithms and services to assist. The explosion in drug-related data

Elizabeth Sam is a Research Scholar working toward her PhD at Amrita University, Department of Computer Science and Engineering. She works on developing computational tools to repurpose drugs.
Prashanth Athri is an Associate Professor at Amrita University, Department of Computer Science and Engineering. His group works in the area of big data bioinformatics and computational pharmacogenomics.
Submitted: 3 July 2017; **Received (in revised form):** 6 September 2017

like bioassays, high-throughput genomic screens, protein structures and many more has led to many promising *in silico* data integration tools that seek to find hidden relationships to find new candidates or predictive tools that present lead hypotheses in various phenotypic contexts. Many studies, and even recent ones ([8–10]), review the computational strategies and data resources used to repurpose drugs.

Repositioning involves a deep synergy of experimentalists and computational scientists to develop relevant and realistic exploration tools. Life scientists will find it hard to use many of the computational tools that require involved data preparation, installation and running packaged software. This problem is further complicated by some software being platform/operating system specific. This is similar and comparable with computer scientists not being able to carry out experimental validations to predictions. Luckily, Web-based tools serve sophisticated computational solutions and help bridge this gap between wet lab scientists and the many *in silico* tools available to repurpose drugs. In fact, many of the Web servers in this review started as stand-alone software before being offered as Internet-based solutions. We have tried to build an exhaustive list of Web-based platforms that help repurpose drugs, in addition to classifying them into specific categories. The categories are centered around the type of interaction used to explore repositioning strategies. Similar schemes to categorize computational methods were published recently [10] and can be found in older studies as well. We have divided the systematic exploration of the various Web servers into three main categories listed below:

1. Predicting drug–target interactions (see 'Predicting drug–target interactions' section)
2. Linking drugs to disease (see 'Linking drugs to disease' section)
3. Using drug-induced gene expression to predict new connections (see 'Using drug-induced gene expression to predict new connections' section)

Figure 1 helps decide which set of tools to explore, and the legend provides a guide. Further, Table 1 provides a listing based on the main categorization mentioned above. This has many useful indices to help decide which tool is the best fit for a particular use case. It provides an enumeration of all back-end databases used, a summary description of the method, the URL, a count of citations received, the date of last update, batch query and application programming interface (API) availability. All sections in each category below and the table list the tools in alphabetical order. It is important to note that some servers have other contexts for usage, but we only summarize the drug repurposing functionality. For example, if a particular tool provides repurposing hypothesis of drugs, and a parallel user interface (UI) landscape to investigate drug mechanism of action (MoA), we have only discussed the former. The following exclusions apply to our review. In some cases, examples of prominent studies left out in this survey owing to the exclusion are mentioned against the respective criterion.

1. Web services that only provide a collection of repurposed drugs: RepurposeDB [6], The Drug Repurposing Hub [11] and repoDB [12].
2. Tools whose outputs do not provide a direct way of inferring the repurposing prediction (e.g. they may generate pharmacophores, weighted or non-weighted interaction networks): PharmMapper [13], PharmaGist [14], ProSMoS [15] and VisANT [16].
3. Studies that focus on any one single family, or some functional sub-set of proteins, or a single disease: iCDI-PseFpt [17], AlzPlatform [18] and ACTP [19].
4. Resources that only aggregate databases and provide associations by connecting them, without using any predictive algorithm-based analysis: Pharos [20], SIDER [21, 22], DTome [23], ChEMBL [24] and PubChem [25].
5. Web-UI tools that predict interactions of molecules with non-protein targets: ChemiRs [26].
6. Web portals not accessible during the time of this review (authors have been contacted around the month of May 2017, if the Web site was down).
7. Studies that are not published in peer-reviewed journals.

Predicting drug–target interactions

Drug–target interactions have been extensively exploited to build tools that repurpose drugs. Methods to build repurposing hypotheses using drug–target interactions have been reviewed before [9, 72, 73]. To organize the large number of servers available within this broad paradigm, we decided to further categorize them as shown below. The categorization stems from the variations in the principal data used to make predictions and how they are parametrized. It is derived from our own analysis, and also motivated from earlier studies [74–76].

1. Ligand similarity using fingerprint encoding;
2. 3D structures of drug and targets;
3. Biological networks;
4. Binding site parametrization; and
5. Other.

Ligand-similarity using fingerprint encoding

The fundamental principle in ligand-centered predictions is that structure similarity implies comparable biological function or properties. Molecular fingerprints, be it 1D, 2D or 3D, are fundamental to encode and optimize searches in these techniques [77]. Based on the principle that similar compounds are likely to bind the same target, *a priori* knowledge of targets binding the query are used to discover previously unknown leads. This class of methods is independent of target structure or ligand–target interaction mechanisms, and reviewed in earlier studies [10, 78, 79]. The number of publicly accessible compounds are far greater than solved protein structure, hence target-agnostic ligand-similarity-based strategies effectively capitalize on the rich data available [25, 30].

ChemMapper

ChemMapper [80] uses a 3D similarity algorithm called SHAFTS (SHApe-FeaTure Similarity) [81, 82] to find polypharmacological manifestations of a query. SHAFTS uses a triplet hashing technique [83] for rapid alignment of molecular conformations and uses shape and chemotype to evaluate similarity. ChemMapper integrates pharmacology and target annotations from multiple sources, namely, ChEMBL [24], DrugBank [27], BindingDB [28], KEGG [29] and protein data bank (PDB) [30]. In other words, ChemMapper uses SHAFTS to retrieve similar molecules and the above-mentioned target annotations to identify candidate targets. Standard virtual screening data sets were used to validate SHAFTS [81], and it has been shown that the use of 3D similarity metrics improve off-target prediction accuracies [84].

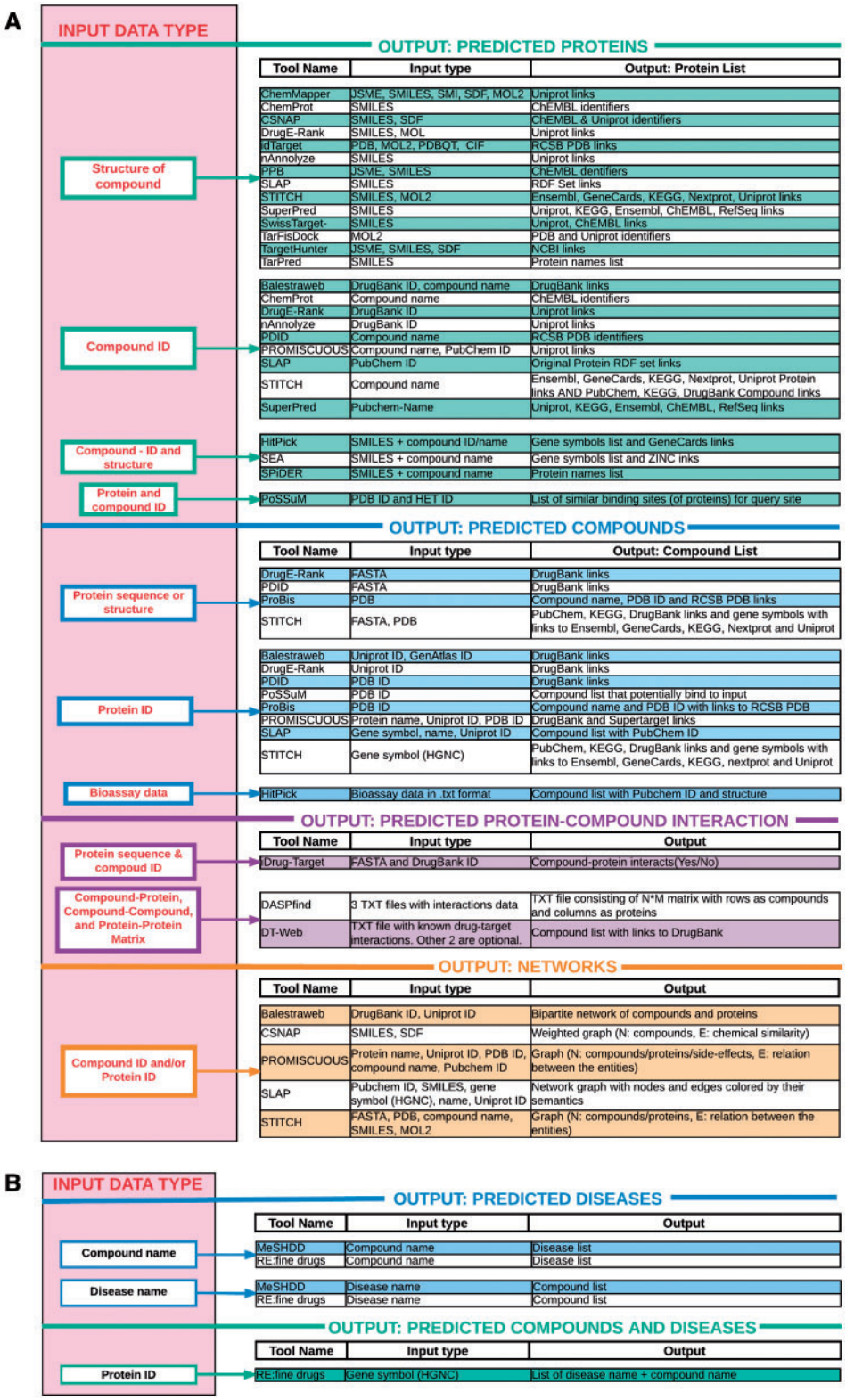
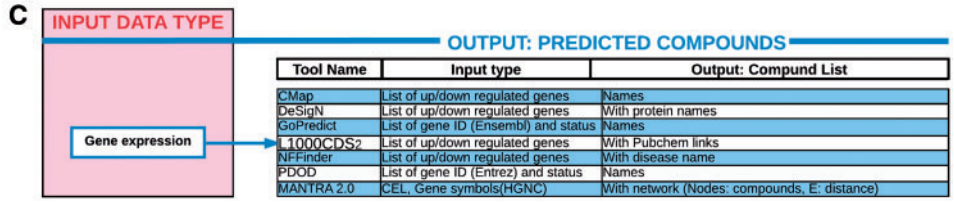


Figure 1. Illustrative guide for selecting desired Web servers based on input and output data types. The flowchart helps choose the tool or set of tools most appropriate depending on (i) the type of predictions the researcher seeks (e.g. list of plausible proteins, graphs, list of related compounds) and (ii) type of input they have at hand (e.g. Protein ID, protein sequence, compound list). We urge the readers to first browse all the 'OUTPUT', which are headings to find (i), and then scan vertically to identify (ii) under the heading 'INPUT DATA TYPE'. This would result in a set of tools that they can investigate individually. **(A)** Web servers predicting drug–target interactions. [PDB: Protein Data Bank, SMILES: Simplified Molecular-Input Line-Entry System, JSME: Molecule Editor in Javascript, Mol2: Molecule Format, SDF: Structure-Data File, PDBQT: Protein Data Bank Partial Charge (Q) and Atom Type (T), CIF: Common Intermediate Format, FASTA: Fast Approximation of Smith and Waterman Algorithm, HGNC: Hugo Gene Nomenclature Committee]. **(B)** Web servers linking drugs to disease. [CEL: Affymetrix CEL file that stores the results of intensity calculations]. **(C)** Web servers using drug-induced gene expressions to predict new connections.



Tool Name	Input type	Output: Compund List
CMap	List of up/down regulated genes	Names
DeSigN	List of up/down regulated genes	With protein names
GoPredict	List of gene ID (Ensembl) and status	Names
L1000CDS2	List of up/down regulated genes	With Pubchem links
NFFinder	List of up/down regulated genes	With disease name
PDOD	List of gene ID (Entrez) and status	Names
MANTRA 2.0	CEL, Gene symbols(HGNC)	With network (Nodes: compounds, E: distance)

Figure 1. Continued

ChemProt

ChemProt 3.0 [85–87] provides a 'global pharmacological heatmap' that connects bioactivities of compounds and proteins based on >7 million stored interactions collected from multiple databases that annotate compounds, proteins and diseases. One significant point about ChemProt is that it has one of the largest confederated sets of databases for each of the entities (drugs, protein, interactions, diseases, etc.) from various resources. In the context of drug repurposing, it provides a similarity ensemble approach (SEA)-based (see section 'Similarity ensemble approach' for SEA) reimplementation and Quantitative Structure Activity Relationships (QSAR) ensemble-based prediction of possible targets to query molecules. The QSAR prediction can be generated for two cases. The query molecule can be compared with the drug set and the map provides a method to navigate known interactions (as per integrated databases). To predict new interactions, fingerprint similarity is used to generate a set of similar drugs. Targets of these drugs with >20 ligands have QSAR models that can be used to predict the activity of the new compound. Second, a SMILES (Simplified Molecular-Input Line-Entry System) representation of the query can be input, followed by selecting proteins from the list, and then downloading predicted negative and positive bioactivities. No validations have been reported for any newly predicted interactions.

HitPick

This is the only resource [88] that can process results from chemical biology screening experiments, and provide target prediction. The user can upload explicit bioassay results, and HitPick will generate B-scores [89] to pick high-quality hits based on a statistical evaluation of many screening parameters. Following this, targets are predicted based on 2D molecular fingerprints. Target prediction can be done individually (user can upload SMILES strings). It uses an integrative approach that combines 1-nearest-neighbor (1NN) similarity metrics [90] and Laplacian-modified Naïve Bayesian target models [91]. The most similar compound from the compound–target interactions (from STITCH 3.1 [37]) is identified using pairwise Tanimoto coefficient [92]. A Laplacian-modified Naïve Bayesian method-based score is generated for targets of the most similar compound, thus providing a ranking of target predictions. Target prediction was cross-validated using 15% of ligands from the data source and found to perform better than using the two algorithms in isolation. The authors claim the performance to be comparable with SEA [93] (refer 'Similarity ensemble approach' section).

iDrug-target

iDrug-Target [94] comprises four engines to predict drug interactions with G-protein-coupled receptors (iDrug-GPCR), ion channels (iDrug-Chl), enzymes (iDrug-Ezy) and nuclear receptors (iDrug-NR) based on KEGG [38] data. The predictions seek to avoid oversampling owing to non-interactive drug–target pairs. They use the Neighborhood Cleaning Rule [95] and Synthetic Minority Over-Sampling Technique [96] to weed out redundant negative samples, and add a few hypothetical positive samples to reduce the above-mentioned biases, respectively. iDrug-Target combines protein sequence encoding, using Pseudo amino acid composition [97], with a 256-component 2D fingerprint representation of the ligand. This signature needs to be generated to construct a query as well. Support Vector Machine is used to classify the input as interactive or non-interactive. They performed jackknife validation on their optimized data set, and a targeted-jackknife on an experimentally observed interaction data set. They point out superior performance, as measured by larger receiver operating characteristic (ROC) curve areas and precision–recall curves, compared with some of their own previous studies.

Polypharmacology browser

Polypharmacology browser (PPB) [98] uses 10 different fingerprints to encode query compounds, and predicts targets based on the fingerprint's distance to the target's putative binders. Fingerprints are calculated using a diverse set of features like molecular shapes, chemical composition, substructures and pharmacophores. They also generate fused fingerprints that merge features across different fingerprints, and combination pairs. The authors use ChEMBL [24] to retrieve all targets that have at least 10 compounds with bioactivity in the ≤10 μM range. These are encoded using the 10 different fingerprints, and the same is calculated for any query molecule. Similarity is calculated using city-block distances. The number of targets of the hit compounds, to which the query matched, are also limited by a predefined value. They have validated the performance of each type of fingerprint using 670 approved drug–target pairs from ChEMBL [24] that they had left out, using target-based and 'hits'-based counts as an index. Recovery statistics they report point at better performances from fusion and combination-pair fingerprints as compared with single fingerprints. Finally, they have predicted and experimentally verified off-targets of TRPV6 inhibitor through *in vitro* profiling. The compound shows 50% inhibition at <10 μM concentrations in 12 among the 24 predicted targets.

Similarity ensemble approach

SEA [93] was the first to use ligand similarity to cluster proteins. Protein clusters thus formed represented functional themes that could be used to predict polypharmacology of the ligands. Ligands were grouped based on minimum spanning tree, Tanimoto scores was used to determine similarity and Daylight 2D fingerprints [99] were used to encode the ligands. The data set comprises a 246-receptor set and the associated ligands as annotated by the MDL Drug Data Report (MDDR) [39]. Receptors with <5 ligands and non-molecular targets were not considered. Prominent discoveries have gone through the full circle of experimental validation. For example, they experimentally

Table 1. List of Web servers

Predicting drug–target interactions							
Name	Method	Databases	URL	Date of last update	No. of citations	Batch queries	API
Ligand-similarity using fingerprint encoding							
ChemMapper[a]	3D similarity (chemotype features and molecular shape) calculation	ChEMBL [24], DrugBank [27], BindingDB [28], KEGG [29] and PDB [30]	http://lilab.ecust.edu.cn/chemmapper/	Dec 2016	52	No	No
ChemProt 3.0	2D similarity-based approach	ChEMBL [24], DrugBank [27], BindingDB [28], STITCH [31], PharmGKB [32], IUPHAR [33], Ki Database [34], CTD [35], WOMBAT [36]	http://potentia.cbs.dtu.dk/ChemProt/	Jan 2015	105[b]	No	No
HitPick	1NN similarity search and Laplacian-modified Naïve Bayesian target models	STITCH 3.1 [37]	http://mips.helmholtz-muenchen.de/proj/hitpick	May 2013[c]	28	Yes	No
iDrug-Target	Fingerprint-based approach with machine learning	KEGG [38]	http://www.jci-bioinfo.cn/iDrug-Target/	Jan 2015[c]	72	Yes	No
PPB	Multi fingerprint-based approach. Ten different fingerprints	ChEMBL [24]	http://gdbtools.unibe.ch:8080/PPB/	Nov 2016	0	No	No
SEA	2D similarity-based approach	MDDR [39], ChEMBL [24], WOMBAT [36]	http://sea.bkslab.org/	Feb 2007	826	No	No
SuperPred	2D similarity-based approach	SuperTarget [40], ChEMBL [24], BindingDB [28]	http://prediction.charite.de	Apr 2014	107[b]	No	No
SwissTarget-Prediction	Combination of 2D and 3D similarity approach	ChEMBL [24]	http://www.swisstargetprediction.ch	Jul 2014[c]	52	No	No
TarPred	KNN-based data fusion with 2D fingerprint-based similarity	DrugBank [27], BindingDB [28], TTD [41]	http://www.dddc.ac.cn/tarpred	Jun 2015[c]	8	No	No
TargetHunter	2D similarity-based approach	ChEMBL [24]	http://www.cbligand.org/TargetHunter/	Oct 2016	46	No	No
3D structures of drug and targets							
idTarget	Divide-and-conquer-based docking approach	PDB [42]	http://idtarget.rcas.sinica.edu.tw/	Aug 2015	59	No	No
PDID	Predictions generated using ILbind, SMAP and eFindSite	PDB [42]	http://biomine.ece.ualberta.ca/PDID/	May 2015	3	No	No
TarFisDock	Reverse ligand-protein docking approach	PDB [42]	http://www.dddc.ac.cn/tarfisdock/	Aug 2014	234	No	No
Biological networks							
Balestraweb	PMF method	DrugBank [27]	http://balestra.csb.pitt.edu/	Jun 2015	8	No	No
CSNAP	CSN-based approach	ChEMBL [24]	https://services.mbi.ucla.edu/CSNAP/index.html	Aug 2015	12	No	No
DASPfind	Network-based approach	BRENDA [43], SuperTarget [40], DrugBank [27] and KEGG [38]	http://www.cbrc.kaust.edu.sa/daspfind/	Mar 2016[c]	6	Yes	No
DT-Web	Recommendation-based approach	DrugBank [27]	http://alpha.dmi.unict.it/dtweb/	Jun 2015[c]	11	Yes	No
PROMISCUOUS	Network-based approach evolving around target–target and drug–target interactions along with side effects	SuperDrug [44], UniProt [45], PDB [46] and SIDER [21]	http://bioinformatics.charite.de/promiscuous/	May 2011	134	Yes	No
SLAP	Semantic Link Association-based Prediction	Chem2Bio2RDF [47], Chem2Bio2OWL ontology [48], DrugBank [27]	http://chem2bio2rdf.org/slap	Dec 2010	66	Yes	Yes
STITCH	Search Tool for Interacting Chemicals	DrugBank [27], GLIDA [49], MATADOR [50], TTD [41], CTD [35], KEGG [38], PID [51],	http://stitch.embl.de/	Dec 2016	890 [b]	Yes	Yes

Continued

Table 1. (continued)

Predicting drug–target interactions							
Name	Method	Databases	URL	Date of last update	No. of citations	Batch queries	API
		Reactome [52], BioCyc [53], ChEMBL [24], PDSP Ki Database [34], PDB [45]					
Binding site parameterization							
ProBis	Fast maximum clique algorithm	ProBiS-Database [54]	http://probis.cmm.ki.si/	Jul 2015	60	No	No
PoSSuM	All-pairs similarity	PDB [42], ChEMBL [24], UniProt [45], GO [55], PDBSprotEC [56], CATH [57], SCOP [58], SCOPe [59]	http://possum.cbrc.jp/PoSSuM/	Oct 2014	31 [b]	No	No
Other approaches							
DrugE-Rank	Feature-based and similarity-based approach	DrugBank [27]	http://datamining-iip.fudan.edu.cn/service/DrugE-Rank	Jun 2016	6	No	No
DR.PRODIS[a]	FINDSITEcomb algorithm	DrugBank [27], ChEMBL [24], PDB [46], SIDER [21], COSMIC [60], OMIM [61]	http://cssb.biology.gatech.edu/dr.prodis/	Jun 2015[c]	20	No	No
SPiDER	Self-organizing map-based prediction	COBRA [62]	http://modlabcadd.ethz.ch/software/spider/	Jun 2017	56	No	No
LINKING DRUGS TO DISEASE							
MeSHDD	Repositioning based on drug–drug similarity	MEDLINE, DrugBank [27]	http://apps.chiragjpgroup.org/MeSHDD/	Dec 2016[c]	3	No	No
RE:fine drugs[a]	Repositioning based on drug–protein and Gene–Disease interaction	DrugBank [27], GWAS [63], PheWAS [64]	http://drug-repurposing.nationwidechildrens.org	Dec 2016[c]	8	No	No
Using drug-induced gene expression profiles to predict new connections							
CMap	Discovering patterns of association between drug sensitivity and gene expression signatures	GEO [65]	http://www.broad.mit.edu/cmap	Jun 2017	2801[b]	Yes	Yes
DeSigN	Global baseline DEGs to drug response	GDSC [66]	http://design.cancerresearch.my/	May 2016	0	Yes	No
GoPredict	Integration of genomic, transcriptomic and signaling pathway data to allow drug repurposing	TCGA [67], Tumorscape [68], COSMIC [60], DrugBank [27], KEGGDrug, GO [55]	http://csblcanges.fimm.fi/GOPredict/	Aug 2014	2	Yes	No
L1000CDS[2]	Predict compounds that either reverse or mimic an input gene expression signature	GEO [65], CCLE [69], CMap [70]	http://amp.pharm.mssm.edu/L1000CDS2/	Mar 2017	9	Yes	Yes
MANTRA 2.0	Network-based and nonparametric statistics on gene expression data	CMap [70]	http://mantra.tigem.it/	Feb 2014 [c]	431[b]	Yes	No
NFFinder	Matching similar gene expression signatures and discovering new connections between drugs, genes and diseases related to the same biological process	GEO [65], CMap [70], DrugMatrix [71]	http://nffinder.cnb.csic.es/	Jul 2015[c]	12	Yes	No
PDOD	Predict drugs having opposite effects on altered states of disease genes	KEGG [38], DrugBank [27], CTD [35], GEO [65]	http://gto.kaist.ac.kr/pdod/index.php/main	Jan 2016	1	Yes	No

[a]Need to contact authors for commercial usage.
[b]Sum total of citations of references of all versions of the tool.
[c]Date of publication is provided because the authors were unreachable to confirm the date of last update.

confirm the predicted affinity of methadone to M3 Muscarine receptors, emetine to adrenergic α2 and loperamide to neurokinin NK2. All of the above were tested by direct binding and a cell-based functional assay. It is our view that any resource created to propose repositioning hypothesis must, as in the case of SEA [93], at least using a few cases, validate using experimental techniques.

SuperPred

SuperPred [100] predicts targets based on similarity of query to similarity distribution among the target's ligand sets as one of its two functionalities. Ligand–target interactions are aggregated from SuperTarget [40], ChEMBL [24] and BindingDB [28]. They normalize/clean the ligand set from the above sources using JChem [101] to get a unique set of ligands. Only molecular targets are extracted, filtered based on strict binding affinity thresholds, and finally 'successful targets' as classified by Therapeutic Target Database [41] are used for target prediction. To predict targets, the similarity of query molecule is calculated using 2D Tanimoto-based strategy against the target's ligand sets. The specificity of a prediction is evaluated using E-value and Z-scores. For diverse target sets a weighting factor is introduced to indicate the intra-target-set similarity between ligands, to make the scores more usable. The authors report a target prediction success rate of 94.1% within the data set that they have used.

SwissTargetPrediction

SwissTargetPrediction [102] combines 2D and 3D similarity metrics to predict targets of bioactive molecules, using ChEMBL [24] as a data source. It has been shown in previous studies [103], from the same group, that if the query molecule is novel or an outlier compared with the training compound series, the target prediction accuracy is better when 2D and 3D similarity measures are combined. FP2 fingerprints (as implemented by OpenBabel, v 2.2.0 [104]) are used with Tanimoto coefficient [105] as a similarity measure. 3D similarity is calculated using Electroshape vectors [106] and calculated for 20 conformations of a molecule, and Manhattan distances are used to compare the conformation. The prioritization of a target is based on a logistic regression of the above-mentioned 2D-3D similarity values [103]. Target predictions using this algorithm have been rigorously validated by the authors using significantly higher percentage of negative interactions, as compared with positive interactions. They report an average area under ROC curve (AUC) value of 87% on both negative and positive interactions. Further, in their validation it was found that 70% of the known targets were ranked under 15 in their results, and in 31% of the test cases the predicted protein was a true positive. SwissTargetPrediction allows users to map predictions between and within organisms based on the target homology, and choosing an organism when performing target prediction.

TarPred

TarPred [107] also uses a combination of ECFP4 [108] with Tanimoto co-efficient [105] to calculate similarity of query with what is referred to as set of Drug-Related Targets. Prediction of targets is based on their previously published algorithm [109]. The prioritized list of targets produced are closely associated with FDA-approved drugs. First, sequences of proteins interacting with FDA-approved drugs are retrieved from DrugBank [27] (FDA-approved drug targets). These sequences are BLASTed [110] against proteins from BindingDB [28], and ligands associated with these targets are retrieved. After filtering them based on multiple criteria, targets are compiled with their respective ligand sets. Given a compound as query, TarPred calculates the ECFP4 similarity scores between the query compound and the ligand sets, and produces a ranked list of targets are results. Similarity fusion scores, which comprises K nearest neighbors (KNN in ligand set of a target), Max (KNN, with $K = 1$) and C-score (average similarity of ligands of a target) determine the priority of targets. The authors use DrugBank [27] and Therapeutics Target Database (TTD) [41] drug sets to validate results of the algorithm used by the Web server. They mostly compare values with one other Web server's prediction program [93], justifying that it uses an algorithm that is similar to theirs. They use Precision, Recall and F measure to evaluate 10-fold cross-validation results and compare with the other Web server/algorithm. Based on the validation results, the authors set the value of K (in KNN) to 3, and perform all other validations based on 3NN implementation. Extensive performance results for four other use cases are also provided, namely, performance with large reference sets, predicting targets specifically for approved drug list from TTD [41] and DrugBank [27], predicting new target–drug interactions and toxicity prediction based on hERG interactions targets.

TargetHunter

TargetHunter [111] uses an algorithm based on Tanimoto similarity index [105], which they call TAMOSIC (Targets Associated with its MOst SImilar Counterparts). Targets associated to the top 'N' most similar compounds (to the query compound) are produced as possible targets. Similarities may be calculated using three different 2D fingerprints, and the choice is made by the user. ChEMBL [24] is used to retrieve hit compounds and their respective targets. Proteins that score higher than their assigned threshold are suggested as possible targets. The Tanimoto threshold is calculated by examining the relationship between Tanimoto similarity and prediction accuracy, while partitioning the ChEMBL data set into test and training data. If multiple targets are reported, the conditional probability of the compound being active against each target is calculated using logistic regression to rank them. TargetHunter claims to have a better capability to sieve out false positives using the Tanimoto threshold they implement. The predictions can only be made for query compounds that have structural features represented in the training set that the authors have used. They assess and report the prediction accuracy of TAMOSIC by using a set of high-potency compounds from ChEMBL (test set). The average accuracy, using a 7-fold cross-validation technique was 90.9%, and overall 91.1% of the compounds were correctly assigned to their targets within three guesses. They also report relative prediction improvements in identifying popular drug targets that include hydrolases, kinases and GPCRs.

3D structures of drug and targets

Structure-based design is essentially centered around binding site similarity explorations, and has been a cornerstone in the area of rational drug design efforts (see [10, 112, 113] for a review of methods). The pre-requisite is the availability of PDB crystal structures, so the search may be driven by the 3D structure of the active site or binding site of the protein. They are more specific because ligand promiscuity is fundamentally driven by binding site similarity, and can be directly evaluated owing to the availability of the 3D structure. In general, they mostly utilize molecular modeling techniques such as docking and pharmacophore models calculate binding affinities of leads. Nonetheless, they tend to be computationally more expensive, and thus, most computational research in this area is to help

build predictive software (and make it faster), rather than build real-time Web-based applications. In addition, some of the modeling tasks are best done using actual visually driven human interaction (molecular modeling).

idTarget

idTarget [114] seeks to reverse dock the query ligand to all PDB [42] structures. They optimize the time required to explore the vast search space by using a divide-and-conquer approach adopted from the BDT method [115], which parallelizes the search space through the use of small overlapping grids. The group has previously helped develop scoring functions based on Autodock4 [116] that they have rigorously benchmarked for binding affinity prediction [117]. In the Web-based tool they provide the option of choosing from multiple charge models that are supported by the above scoring functions, and pre-structured controls on few search parameters. The search engine they use is MEDock [118]. Search space is contracted using CD-hit-based [119] sequence similarity-based clustered sets, and then expanded using homology such that only the most relevant hits are retrieved. They finally report three examples in which the algorithm reproduced known (as shown by published literature) off-targets.

PDID

Binding sites of drug–protein complexes of the PDB [42] are queried by PDID (protein–drug interaction database) [120] against all other proteins found in PDB, based on few strict filters, to find probable off-targets of the original drugs. They provide a ranked list of targets based on aggregated scores from three different methods, namely, ILbind [121], SMAP [122] and eFindSite [123, 124]. The resource contains millions of predictions, and thus serves as start point of repurposing studies.

TarFisDock

TarFisDock (TARget FIShing DOCKing) [125] is one of the earlier reverse docking target identification resources. For a given input drug, it performs docking based on DOCK 4.0 [126] algorithm using protein structures they have stored in PDTD (Potential Drug–Target Database) [127]. During docking, the molecule is treated flexibly, and the protein is held rigid. The targets can be provided by user or retrieved from PDTD. TarFisDock provides the top 2, 5 or 10% of the results from the reverse docking. As a validation, the authors report the results for two searches—Vitamin E and 4 H-tamoxifen. For vitamin E, the top 10% targets predicted by TarFisDock covers 50% of known targets, and the top 5% candidates of 4 H-tamoxifen binding proteins identified by TarFisDock covers 50% of experimentally confirmed targets.

Biological networks

Integrating system-wide biological networks helps opportunistically targeting pharmacological properties of drugs to repurpose them. Data on system-wide network topologies, as well as deep annotations on various drug associated entities, like targets, are stored in many interaction databases. In these network-based approaches, the analysis of ligand–target interactions are multidimensional. Data are significantly more diverse, and as seen in many prominent tools below, database integration that support these tools need to take into account semantic integration of data points and big data-related optimization. They are also referred to as network-based polypharmacology studies, and the types of algorithms/computational methods developed are reviewed in earlier studies [128–130].

BalestraWeb

BalestraWeb [131] predicts the interactions of leads (with off-targets), targets (with approved drugs) or all associations centered on a drug–target interaction of interest, based on an active learning (AL), collaborative filtering technique. User can also explore repositioning opportunities using drug–drug and target–target similarity in separate tabs. Approved drugs and their targets from DrugBank [27] form the nodes of a bipartite graph. The known interactions (edges of the graph) are used to learn the latent variable (LV) vectors that represent each drug and protein. The dot product of the LVs of the drug–target pair is the strength of the interaction as predicted by this method. They adopted an AL method that is based on the probabilistic matrix factorization (PMF) [132] method to calculate the statistical weight of each approved drug to all targets associated with the entire set of approved drugs [133]. In effect, BalestraWeb relies on these interaction profiles to make predictions, and hence not dependent on any structural or chemical similarities. As validated by the authors, the PMF-based method they use compares competitively with SEA [93] and STITCH [37]. The PMF method is (also) extensively validated using a 5-fold cross-validation across four different target classes. Similar validations with two other drug–target network topology-based learning algorithms [134, 135] demonstrate that the strengths of the three algorithms vary across the four target classes. Nonetheless, the algorithm in BalestraWeb may be used for extended data sets as the PMF model scales linearly with the number of interactions.

CSNAP

Chemical similarity network analysis pull-down (CSNAP) [136] uses a combination of chemical similarity networks (CSN) and chemical consensus to form chemotype-based sub-networks to predict targets over a range of drug classes. Bioactivity databases, ChEMBL [24] and PubChem [25], are used to retrieve compounds, and hence implicitly include bioactivity information to match profiles of hit compounds to the query. Query and target-annotated compounds from the databases are clustered into CSNs, and the target prediction is prioritized based on a consensus statistic determined by the frequency of the target shared by the first neighbors centered on the query compound. CSNAP clusters compounds into distinct sub-networks that represent a particular 'chemotype'. The cluster's nodes represent compounds, and the edges represent similarity. Ligands, represented by FP2 fingerprints, are compared using both the Tanimoto coefficient [105] and Z-score similarity measures. Structurally diverse ligands can be part of the same subnetwork because the similarity metric is based on chemotypes. Targets of first neighbor compounds are ranked using the S-score [137], and significance of each compound protein pair is calculated using a H-score [138]. The method is benchmarked against SEA approach [93], and has a higher predictive capability in the case of pre-annotated protein targets across six drug classes, and especially pronounced for promiscuous drugs. The diversity data set was taken from the Directory of Useful Decoys (DUD LIB VS 1.0) [139]. CSNAP can elucidate targets for compounds from high-throughput chemical screens, as they typically have nonspecific binding patterns. CSNAP was also applied to 212 mitotic compounds as a test case in the primary publication. They were able to identify known compounds against new mitotic targets, and formulate novel chemotypes that target microtubules. In general, CSNAP is most effective for large-scale profiling of compounds from multivariate chemical screens.

DASPfind

DASPfind [140] uses three sub-graphs, namely, drug–drug, target–target and drug–target similarities, to identify new drug–target interactions. The above-mentioned classes of interaction are retrieved from BRENDA [43], SuperTarget [40], DrugBank [27] and KEGG [38] to form an inter-connected heterogeneous network that rank new relations. A weighted graph is constructed with nodes as drugs and proteins. The weight of an edge between two drugs is the similarity calculated using SIMCOMP [141], and that between two proteins is calculated using Smith–Waterman algorithm [142]. A drug–protein edge represents a known interaction with a weight of 1. DASPfind uses only 'simple paths' to find new interactions, and longer paths are penalized when generating the score. Results were validated using established data sets [134, 143–145], HGBI data set [146] and approved drugs from DrugBank [27]. While the AUC has been compared with few other methods of prediction, DASPfind performs best when a subjective test that uses only the 'top 1' candidate is used. The predictive capability on the previously mentioned Ion Channel data set is confirmed by validating all (210) but three predictions, either by literature or database searches. The authors hypothesize that there may be new targets, but do not validate.

DT-Web

DT-Web (domain tuned) [147] extends a recommendation based on bipartite network projection by integrating previously known drug–target interactions, drug–drug similarity and target–target similarity into a heterogeneous network. It is a Web-based interface to the DT-Hybrid algorithm [148]. DT-Web takes as input three matrices, namely, drug–drug similarity matrix, target–target similarity matrix and known drug–target matrix. Drug–drug similarities are scored using SIMCOMP [141], and hence construct a drug similarity matrix. Target–target similarity score is the sequence similarity between the two proteins. The target similarity matrix can be obtained either by running BLAST [149] or Smith–Waterman local alignment technique [142]. Known drug–target interactions can be obtained from DrugBank [27], and an adjacency matrix can be constructed from the same. A drug–target interaction network is constructed using the three matrices. Each target in the drug–target interaction network is mapped to Entrez identifier [150] and annotated with Gene Ontology (GO) terms [55]. For each pair of GO terms, based on the node distance in ontology-directed acyclic graph, the similarity measure is computed. For every drug, *P*-value is used to score the association between predicted and validated targets. Another feature of DT-Web is that, given a set of candidate disease genes as input, DT-Web predicts combinations of drugs whose targets are at an optimal distance from those genes. DT-Web is evaluated using the four standard drug–target interaction benchmarks [145] and complete DrugBank [27]. Results of DT-Web are evaluated by applying a 10-fold cross-validation and repeating the experiments 30 times. Performance is evaluated using precision and recall enhancement, and average AUC for the top 20 predictions. DT-Web shows improved results compared with NBI [145, 151] and Hybrid [152], both of which are network-based interaction prediction algorithms.

nAnnolyze

nAnnolyze [153] provides a Web-based interface to the network-based comparative docking method called Annolyze [130]. Only proteins that have solved 3D structure are used for prediction. They use a bi-partite network of structural similarities and interactions. There are four main components that are merged into this network, namely, compounds in PDB [46] that exert a pharmacological effect on their co-crystalized protein, protein binding sites from LigBase [154], the human structural proteome from ModBase [155] and DrugBank [27] compounds. First, a ligand sub-network is formed using the PDB ligands that have drug-likeness above a certain threshold. The subnetwork is minimized using similarities derived from Random Forest Classifier (RFC), into a k-core network to avoid redundancy. Next, the protein sub-network is constructed using targets that bind ligands above a threshold of drug-likeness. The network is linked using structural similarity of binding sites as calculated by ProBis [156]. A similar filtering, as the ligand sub-network, is applied. The two sub-networks are merged if there is a known ligand–target interaction as validated by a solved PDB structure. The human structural proteome is derived by filtering structures from ModBase, and using ProBis to compare with the above protein network to derive the human network. Finally, a large part of the DrugBank compounds are integrated by using RFC to calculate similarity, and thus adding an edge to the most similar compound in the ligand sub-network. To benchmark, the authors used a 'positive' benchmark set containing all drug–protein annotated pairs between structures in PDB and FDA-approved drugs. The Receiver Operating Characteristic curve was 0.96 when drug names were used, and 0.7 for compounds not in the training set. The authors also provide several examples of predictions by nAnnolyze that can be validated through known experimental studies.

Promiscuous

PROMISCUOUS [157] is one of the first public network-based Web servers for drug repurposing. The network consists of drugs, proteins and side-effects as the nodes, with drug–side effect, drug–target, protein–protein and drug–drug interactions serving as edges. They integrate data from publicly available databases such as SuperDrug [44], UniProt [45], PDB [46] and SIDER [21]. As the targets are predicted by transitive mapping of the network, PROMISCUOUS does not rank or prioritize the predicted targets in any manner, but provides an explorative network as output. They point out the capability of the tool to identify new indications using examples like Memantine-originally used for dementia, but has been repositioned for Parkinson's treatment. In PROMISCUOUS, Memantine is similar to Amantadine which is an anti-Parkinson drug that shares the NMDA glutamate receptor as a target with Memantine.

SLAP

SLAP (semantic link association prediction) [158] predicts associations between drugs and targets, primarily through semantic database integration and statistical modeling. SLAP predicts associations using 'path patterns', which are various pre-defined association paradigms comprising nodes and edges. These, in turn, are part of a semantic network constructed using drug–drug and protein–protein similarity, and drug–target interactions from Chem2Bio2RDF [47] and semantic annotations from Chem2Bio2OWL ontology [48]. The actual drug–target pairs used to build the above association network were retrieved from DrugBank [27]. Heap-based Dijkstra algorithm is used to find the shortest paths of length <3 between two nodes [159, 160]. The predicted targets are ranked in terms of *P*-value and association score, which is the summation of the z-scores of all valid paths between two nodes. The interface can accept as input a (i) drug–pair, and predict association, or (ii) drug-to-output predicted targets and drugs with similar biological function, or (iii) protein alone, and get associated ligands. Validation

of predicted drug–target interactions through SLAP was performed using MATADOR [50] interactions. They have shown improved performance over similar link prediction methods by measuring AUROCs. Their results are comparable with SEA [93] (refer 'Similarity ensemble approach' section) for drug–target predictions, and that of CMap [70] (refer 'CMap' section) for predictions of drug associations.

STITCH

STITCH (search tool for interacting chemicals) [31, 37, 161, 162, 163] has gone through many versions of development. It focuses on providing a substantially comprehensive map of drug–target interactions with sophisticated filters and visualization. Over the many years of development, a large number of databases are connected and we refer the reader to the original publication or the Web server for a full list (Table 1) (Examples include [27, 35, 38, 41, 49–53]). In essence, they provide a common interface that integrates data resources of drug–target interactions originating from high-throughput experiments, manually curated databases, as well as many predictive algorithms. Further, they have also implemented automated text mining algorithms that predict interactions based on co-occurrence in PubMed, MEDLINE and NIH Re-PORTER [164]. Each version has brought in degrees of selectivity and addition of resources—for example, in version 5 users can now filter out interactions based on tissue specificity. Each imported source of information is scored individually and then combined with data from text-mining [161]. The confidence scores reveal the level of significance and certainty of an interaction. STITCH takes as input a chemical name, gene name, chemical structure or protein sequence, to generate a network of chemicals and proteins interacting with the query. A binding affinity view displays all known Ki values of drug–target interactions. In this view the edge width of drug–target interactions is scaled according to binding affinity. STITCH is a well-established resource, and in the latest update [163] the authors provide several published studies from other groups [165–167] that directly use results from STITCH.

Binding site parametrization

Binding sites are regions in the protein structure that bind ligands via non-bonded interactions. These regions often comprise conserved regions, and hence can be used to search for other proteins, even spanning other fold families, based on binding region structural similarity [76]. Target hunting based on binding site similarity mapping algorithms are reviewed in earlier studies [76, 112].

ProBis

ProBis (protein binding sites) [156] uses local (spatially constrained) binding site similarity as the primary index to find targets similar to the query. It uses a fast maximum clique algorithm [168] with the same name that uses both structure and physicochemical properties of the constituent amino acids and their backbones to compare two protein binding sites. For a protein as input, ProBiS gives a list of proteins with similar binding sites, predicted ligand and small molecule binding partners and nucleic acid residues, if found. ProBiS-Database [54] is a repository of non-redundant binding sites and associated PDB structures, which is updated weekly. Users have the option of choosing pre-calculated data in ProBiS-Database to get instant results. ProBis accepts only proteins as input and not drugs.

PoSSuM

PoSSuM (pocket similarity search using multiple-sketches) [169, 170] is based on an algorithm that searches the entire PDB [30] for all-pairs binding similarity. A potential ligand-binding region is considered if a probe-cluster with $\leq$200 probes is generated using Ghecom [171]. A set of amino acids close to a non-polymer molecule is called the known ligand-binding site and a set of amino acids close to a probe-cluster is called a putative binding site. Three types of inputs are accepted by PoSSuM: (i) a protein structure, (ii) a ligand-binding site, (iii) a ligand. For a protein structure in the PDB as input, PoSSuM will search similar known ligand binding sites for the query structure. For a ligand-binding site in the PDB as input, PoSSuM will search similar sites for the query site. Users can search the query against any of the three types of databases, namely, known ligand-binding sites, putative binding sites or both known and putative sites. PoSSuM uses an ultrafast method [74, 172] that detects similar binding sites spanning over 1 million binding sites. Based on the physicochemical and geometric properties, the ligand-binding sites are encoded as feature vectors and similar sites are computed using a fast neighbor search algorithm called SketchSort [173]. PoSSuM also takes a ligand as input and gives as output a set of binding sites similar to the pockets to which the query ligand is known to bind. Measure of similarity is given by cosine similarity and P-value [169]. Dissimilarity values are given by root mean square deviation. This all-pair similarity search was applied to 3.4 million known and potential ligand-binding sites, and over 24 million similar binding sites aligned with six or more residues constitutes the PoSSuM relational database. Validation of the results obtained in PoSSuM is yet to be done.

Others

A few other servers that do not fall in the above categories, nonetheless, are drugs and/or target predictors based on drug–target association. DR. PRODIS (DRugome, PROteome and DISeasome) [174] uses an algorithm called FINDSITEcomb [175], which in turn is based on the assumption that evolutionarily related proteins have similar functions and thus bind similar ligands. DrugE-Rank [176] uses hybrid, feature- and similarity-based descriptors with an ensemble learning approach, while SPiDER (self-organizing map-based prediction of drug equivalence relationships) [177] predicts targets using SOM (Self Organizing Maps) projections of pharmacophore descriptors and physico-chemical properties.

Linking drugs to disease

In this section we review tools that use annotations dependent on disease associations. Disease-based approaches are utilized when pharmacology of drugs is not present, or not being considered. Computational strategies that use drug–disease relations have been reviewed earlier [178], and we present two of them that provide access through the Web.

MeSHDD

MeSH-based drug–drug similarity and repositioning (MeSHDD) [179] clusters drugs based on drug–drug similarity derived based on disease-centered Medical Subject Heading (MeSH) terms found in the MEDLINE Baseline Repository® to predict shared indications. The basic drug-–MeSH term overlap database they derived uses drug names if they match with approved drugs from DrugBank [27]. The co-occurrence of drug-MeSH term is calculated

using a hypergeometric *P*-value, followed by a Bonferroni correction [180]. Drug–drug similarity is measured by calculating the bitwise distance that is derived from converting the above *P*-values to a binary representation. They have clustered the drugs based on pair-wise distances and bootstrap means clustering techniques [181] (implemented in R), and the Jaccard index [182] was used to compare clustering of various k values. Enrichment for disease indications in the clusters is evaluated by comparing against data from TTD [41]. As a form of validation, the authors point out the ability of MeSHDD to discover a cystic fibrosis indication for the putative antidiabetic drug Metformin.

RE:fine drugs

RE:fine drugs [183] integrates drug–gene–disease data in a transitive manner to yield drug–disease relationships, predicting new indications for existing drugs. Given a disease as input to the Web server, we receive a list of drugs that can be used for the treatment of that disease. RE:fine drugs classifies the predicted drug–disease pairs as Known/Rediscovered if the interaction is present in DrugBank, Strongly Supported if it is present in NIH clinical trial registry and biomedical literature, Likely if evidence is in either NIH clinical trial registry or biomedical literature and Novel if not present in either. Daclizumab was originally indicated for the prevention of renal transplant rejection, but was predicted to be active against asthma by RE:fine as seen in other studies [184–186].

Using drug-induced gene expression to predict new connections

Drug-induced gene expression (see [187, 188] for a review) refers to comparing mRNA expression profiles in a cell line, before and after drug treatment. They act as a signature to characterize a drug's effect on a biological system. Among other important applications like reveal insights into drug mechanism and hypothesize novel disease indications for a given drug, these signatures identify similarities to other drugs based on similarities of their respective expression signatures or identify negative expression profiles as compared with that of a particular disease and thus discover repurposing candidates. This repurposing approach is accomplished by comparing disease-associated expression signatures to these drug-induced expression signatures, seeking drugs that have opposing effects to the disease, and therefore may be efficacious.

CMap

CMap (Connectivity map) [70] is supported by a database of cellular responses to various chemical perturbagens, as well as normal controls. CMap provides mRNA expression data from DNA microarrays for researchers who want to check for differential expression so as to identify drugs that produce reverse signatures to the query expression signatures. Connectivities are measured using Kolmogorov–Smirnov statistical test [189]. Hence, in the context of repurposing, CMap can identify both agonists and antagonists. They provide experimental studies in various classes like estrogens, HDAC inhibitors and phenothiazines that CMap can reproduce only using their Differential Gene Expression (DGE) data, as a way of validation. Similarly, they have identified reverse drug signatures to disease states, namely, diet-induced obesity and Alzheimer's. Further, they have provided instances of discovering molecules that can revert drug resistance in the case of acute lymphoblastic leukemia.

Since its creation, CMap has had a deep impact on therapeutic research and has opened up completely new lines of scientific enquiry in the areas of drug repurposing, lead discovery, MoA elucidation, biological understanding and systems biology [190]. CMap provides one of the most valuable, direct methods to investigate alternate therapeutic potential of drugs, and CMap approaches have been extensively exploited by various groups across many therapeutic areas to repurpose drugs (e.g. muscle atrophy [191], lung cancer [192], hair growth [193] and others).

DeSigN

DeSigN (differentially expressed gene signatures—inhibitors) [194] is similar to CMap, and associates disease signatures with drug response associating gene signatures with drugs based on IC50 data. DeSigN is constructed using GDSC [195]. CMap uses pre- and post-gene expression profiles, whereas DeSigN uses only baseline gene expression profiles. DeSigN was tested on four GEO studies [65], and for each of these studies, the connectivity scores correlated with the drug response and it was consistent with the published GEO studies.

GoPredict

GoPredict [196] integrates data from heterogeneous public sources, signaling pathway databases and drug target information with genomic data in cancer to suggest drugs for gene expressions. It takes gene expression as input and gives drugs as output. The reference database used in GoPredict are TCGA [67], KEGGDrug [197], DrugBank [27] and Gene Ontology [198]. They calculate a gene rank correlated to its impact on a pathway's regulation. Any gene–drug pair is prioritized based on the respective GO processes [55], the drug's rank is an average of all genes it regulates. As a validation, authors have reproduced novel drug–DGE associations that have been reported in literature and clinical trials.

L1000CDS

L1000CDS^2 [199] is a Web interface to access and use the Characteristic Direction (CD) [200] signatures of the LINCS-L1000 data to predict new indications. CMap (refer 'CMap; section) and earlier works used moderated Z-scores, whereas the multivariate method, CD, developed here is shown to be more sensitive to identifying DGE. It calculates the angle between the input signature vector and the LINCS-1000 data to produce a list of probable candidate molecules that reverse or mimic (endogenous ligands) the input gene expression. As a use case, they have predicted candidates for diseases signatures from GEO [65]. Further, they have predicted a drug called Kenpaullone to be active against Ebola virus, and provide many relevant studies and experiments that support their claim.

MANTRA 2.0

Molecular targets for drugs can be explored using MANTRA 2.0 (mode of action by NeTwoRk analysis) [201] by uploading gene expression profiles before and after drug perturbation, which gets embedded into a collaborative, additive learning environment. A visual network with this new 'node' lets users explore near neighbors to find new indications. They compute a prototype ranked list (PRL) [202] for each drug, followed by a method to compare two PRLs using a Gene Set Ensemble Approach (GSEA)-based method [203]. They provide a collaborative, exploratory network visualization environment, and an opportunity for users to share their data.

NFFinder

While having a comparable work flow with others in this category, NFFinder [204] uses MARQ [205] method to compare the signatures. Two sets of up- and down-regulated genes are needed that are subsequently compared with GEO [65], CMap [70] and DrugMatrix [206] data. They carried out a two-step validation: first, found that Trichostatin A (TCA) was effective in killing malignant peripheral nerve sheath tumors (MPNST) cells. And next, retrieved TCA as a hit when the gene expression profile of the known tumor suppressor cocktail PD901/JQ1 against MPNST cells was used as a query.

PDOD

In addition to gene expression signatures, PDOD (prediction of drugs having opposite effects on disease genes) [207] considers 'effect-type' and 'effect-direction' using pathway informations (KEGG), drug–drug target information from DrugBank [27] and GEO [65] for expression data to check if there is a drug that can compensate for differentially regulated disease genes. Differentially expressed genes were identified using Limma [208] and a function they developed to evaluate drug–disease score based on parameterizing the above relationships. The study cannot predict drugs for a few diseases, and is based on availability of relevant data in the databases they use. As concluded by the authors, the study has selective successes in predicting new indications in a few disease classes.

Discussion

The review provides a classification of Web servers to help wet-lab researchers choose the right tool to assist studies in repurposing drugs. Among the three main classes discussed in this survey, studies focused on predicting drug–target interactions (see 'Predicting drug–target interactions' section) are the most well-studied. These methods are largely derived from algorithms and techniques from Computer-Aided Drug Design (CADD)/computational chemistry. This is a well-explored field of study with deep applications in drug discovery developed over tens of years, and hence we see many tools in this category. Disease-based studies originate from a clinical perspective of disease [178]. While newer algorithmic/computational development seeking to assist repurposing efforts may still tend to borrow substantially from these canonical areas, it would be prescient to include data from modern molecular measurements (x-omics data like transcriptomics, proteomics, etc.) as data resources. Owing to the bioinformatics data deluge [209], it is our view that algorithms from the realm of big data science can be adopted more extensively, so system-level approaches to discover new hypothesis can emerge.

As algorithms start using various types of data resources, Web-based tools must also be developed parallelly to serve UIs that help explore the massive sets of bioinformatics, biomedical and x-omics data. We suggest computational tools have both a Web-based UI, as well as stand-alone versions. While the latter confers the advantages of open-source development and crowd sourcing, Web UIs are a critical bridge to lab scientists who can move repurposing predictions from computer screens to lab validations through animal models and clinical trials. As noted in previous reviews [178], these are critical steps to fully realize the utility of computational predictions and pre-requisites for predictions to have any translational outcomes.

It is equally important for algorithms to validate their predictions experimentally, as it is to build Web-based UIs. An observation common to all of the highly cited tools (see Table 1) is that the respective authors have experimentally verified a few representative, high-confidence predictions. This encourages scientists to adopt the predictive techniques with a degree of confidence. If developers of tools still wish (or able) to only provide computational validation, we urge them to not validate just against their own, previous algorithms, or choose an arbitrary set of algorithms that do not represent any universal gold standard for benchmarking. The results of such validation may be viewed as subjective and hence self-defeating if they are trying to convince a broader scientific community to use their tools. Finally, update cycles that the authors wish to commit to must be published on the URL and associated journal article, especially if they are using third-party databases.

The review has listed and summarized, what we believe to be, most of the currently maintained Web servers focused on drug repurposing. The overview can assist computer scientists to evaluate the current state of art and explore ways to develop cutting-edge repositioning tools based on advancements in various fields of computer science like data analytics, big data technologies and machine learning techniques. They can get formative ideas on how various areas in computer science can contribute to discovering novel indications to known drugs/leads. During the course of our review, we have found many advanced computational algorithms, and work flows that aid in repositioning. Ideally, all biologists are part computer scientists [210] or have fairly advanced knowledge about developing computational biology software to mine the vast bio-oriented databases. Until that happens, Web servers provide a critical bridge between the two disciplines. Life scientists like biologists, medicinal chemists and pharmacogenomic researchers will find a good use for it toward choosing the right exploration tool. We hope the review will open up more discussions between the life scientists and the developers of the above tools, both to improve the resources and make feature requests that will help improve usage experience.

Key Points

- We enumerate available and active Web servers that assist drug repurposing studies.
- The tools have been classified based on the fundamental biological/chemical relations used in the respective servers to make the predictions.
- The categorization comprises three main associations: predicting drug–target interactions, linking drugs and diseases and using drug-induced gene expression to predict new connections.
- Experimentalists can explore the category (or subsections therewith) based on what type of data their labs work with, to help them develop (or validate) drug repurposing studies.
- Computer scientists who wish to develop Web-based interfaces to explore drug repurposing paradigms can use the compendium to understand the current state of art.

Acknowledgements

The authors wish to thank the anonymous reviewers for their valuable and extensive suggestions to improve the manuscript.

Funding

ES was funded by the Amrita University Full Time PhD Scholarship. The research was supported by funding provided by the Department of Biotechnology, Government of India (Grant Reference Number: BT/PR16476/BID/7/631/2016).

References

1. Power A, Berger AC, Ginsburg GS. Genomics-enabled drug repositioning and repurposing: insights from an IOM roundtable activity. *JAMA* 2014;**311**:2063–4.
2. Scannell JW, Blanckley A, Boldon H, *et al*. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat Rev Drug Discov* 2012;**11**(3):191–200.
3. Padhy B, Gupta Y. Drug repositioning: re-investigating existing drugs for new therapeutic indications. *J Postgrad Med* 2011;**57**(2):153.
4. Novac N. Challenges and opportunities of drug repositioning. *Trends Pharmacol Sci* 2013;**34**(5):267–72.
5. DiMasi JA, Feldman L, Seckler A, *et al*. Trends in risks associated with new drug development: success rates for investigational drugs. *Clin Pharmacol Ther* 2010;**87**(3):272–7.
6. Shameer K, Glicksberg BS, Hodos R, *et al*. Systematic analyses of drugs and disease indications in RepurposeDB reveal pharmacological, biological and epidemiological factors influencing drug repositioning. *Brief Bioinform* 2017. doi: 10.1093/bib/bbw136.
7. Shaughnessy AF. Old drugs, new tricks. *BMJ* 2011;**342**:d741.
8. Li J, Zheng S, Chen B, *et al*. A survey of current trends in computational drug repositioning. *Brief Bioinform* 2016;**17**(1):2–12.
9. Dai YF, Zhao XM. A survey on the computational approaches to identify drug targets in the postgenomic era. *Biomed Res Int* 2015;**2015**:239654.
10. March Vila E, Pinzi L, Sturm N, *et al*. On the integration of *in silico* drug design methods for drug repurposing. *Front Pharmacol* 2017;**8**:298.
11. Corsello SM, Bittker JA, Liu Z, *et al*. The drug repurposing hub: a next-generation drug library and information resource. *Nat Med* 2017;**23**(4):405–8.
12. Brown AS, Patel CJ. A standard database for drug repositioning. *Sci Data* 2017;**4**:170029.
13. Wang X, Shen Y, Wang S, *et al*. PharmMapper 2017 update: a web server for potential drug target identification with a comprehensive target pharmacophore database. *Nucleic Acids Res* 2017; **45**(W1):W356–60.
14. Schneidman Duhovny D, Dror O, Inbar Y, *et al*. PharmaGist: a webserver for ligand-based pharmacophore detection. *Nucleic Acids Res* 2008;**36(Suppl 2)**:W223–8.
15. Shi S, Chitturi B, Grishin NV. ProSMoS server: a pattern-based search using interaction matrix representation of protein structures. *Nucleic Acids Res* 2009;**37(Suppl 2)**:W526–31.
16. Hu Z, Mellor J, Wu J, *et al*. VisANT: data-integrating visual framework for biological networks and modules. *Nucleic Acids Res* 2005;**33(Suppl 2)**:W352–7.
17. Xiao X, Min JL, Wang P, *et al*. iCDI-PseFpt: identify the channel–drug interaction in cellular networking with PseAAC and molecular fingerprints. *J Theor Biol* 2013;**337**:71–9.
18. Liu H, Wang L, Lv M, *et al*. AlzPlatform: an Alzheimers disease domain-specific chemogenomics knowledgebase for polypharmacology and target identification research. *J Chem Inf Model* 2014;**54**(4):1050–60.
19. Xie T, Zhang L, Zhang S, *et al*. ACTP: a webserver for predicting potential targets and relevant pathways of autophagy-modulating compounds. *Oncotarget* 2016;**7**(9):10015.
20. Nguyen DT, Mathias S, Bologa C, *et al*. Pharos: collating protein information to shed light on the druggable genome. *Nucleic Acids Res* 2016;**45**(D1):D995–1002.
21. Kuhn M, Letunic I, Jensen LJ, *et al*. The SIDER database of drugs and side effects. *Nucleic Acids Res* 2015;**44**:D1075–9.
22. Kuhn M, Campillos M, Letunic I, *et al*. A side effect resource to capture phenotypic effects of drugs. *Mol Syst Biol* 2010;**6**(1):343.
23. Sun J, Wu Y, Xu H, *et al*. DTome: a web-based tool for drug-target interactome construction. *BMC Bioinformatics* 2012;**13**(9):S7.
24. Gaulton A, Bellis LJ, Bento AP, *et al*. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res* 2012;**40**(D1):D1100–7.
25. Wang Y, Bryant SH, Cheng T, *et al*. PubChem BioAssay: 2017 update. *Nucleic Acids Res* 2017;**45**(D1):D955–63.
26. Su ECY, Chen YS, Tien YC, *et al*. ChemiRs: a web application for microRNAs and chemicals. *BMC Bioinformatics* 2016;**17**(1):167.
27. Law V, Knox C, Djoumbou Y, *et al*. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res* 2014;**42**(D1):D1091–7.
28. Liu T, Lin Y, Wen X, *et al*. BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Res* 2007;**35(Suppl 1)**:D198–201.
29. Kanehisa M, Goto S, Furumichi M, *et al*. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 2010;**38(Suppl 1)**:D355–60.
30. Berman HM, Westbrook J, Feng Z, *et al*. The protein data bank. *Nucleic Acids Res* 2000;**28**(1):235–42.
31. Kuhn M, Szklarczyk D, Pletscher Frankild S, *et al*. STITCH 4: integration of protein–chemical interactions with user data. *Nucleic Acids Res* 2014;**42**(D1):D401–7.
32. Hewett M, Oliver DE, Rubin DL, *et al*. PharmGKB: the pharmacogenetics knowledge base. *Nucleic Acids Res* 2002;**30**(1):163–5.
33. Harmar AJ, Hills RA, Rosser EM, *et al*. IUPHAR-DB: the IUPHAR database of G protein-coupled receptors and ion channels. *Nucleic Acids Res* 2008;**37(Suppl 1)**:D680–5.
34. Roth BL, Lopez E, Patel S, *et al*. The multiplicity of serotonin receptors: uselessly diverse molecules or an embarrassment of riches?. *Neuroscientist* 2000;**6**(4):252–62.
35. Mattingly C, Rosenstein M, Colby G, *et al*. The comparative toxicogenomics database (CTD): a resource for comparative toxicological studies. *J Exp Zool A Ecol Genet Physiol* 2006;**305**(9):689–92.
36. Olah M, Mracec M, Ostopovici L, *et al*. WOMBAT: world of molecular bioactivity. *Chemoinform Drug Discov* 2004;**23**:223–39.
37. Kuhn M, Szklarczyk D, Franceschini A, *et al*. STITCH 3: zooming in on protein–chemical interactions. *Nucleic Acids Res* 2012;**40**(D1):D876–80.
38. Kotera M, Moriya Y, Tokimatsu T, *et al*. KEGG and GenomeNet, new developments, metagenomic analysis. In: *Encyclopedia of Metagenomics*. Boston, MA: Springer, 2015, 329–39.
39. Systems MI. MDL drug data report, 2006.1. 2006. San Leandro, CA: MDL Information Systems Inc.
40. Hecker N, Ahmed J, von Eichborn J, *et al*. SuperTarget goes quantitative: update on drug–target interactions. *Nucleic Acids Res* 2011;**40**:D1113–17.

41. Chen X, Ji ZL, Chen YZ. TTD: therapeutic target database. *Nucleic Acids Res* 2002;**30**(1):412–5.
42. Kouranov A, Xie L, de la Cruz J, *et al*. The RCSB PDB information portal for structural genomics. *Nucleic Acids Res* 2006; **34(Suppl 1)**:D302–5.
43. Schomburg I, Jeske L, Ulbrich M, *et al*. The BRENDA enzyme information system–from a database to an expert system. *J Biotechnol* 2017. doi: 10.1016/j.jbiotec.2017.04.020.
44. Goede A, Dunkel M, Mester N, *et al*. SuperDrug: a conformational drug database. *Bioinformatics* 2005;**21**(9):1751–3.
45. Boutet E, Lieberherr D, Tognolli M, *et al*. UniProtKB/Swiss-Prot: the manually annotated section of the uniprot knowledgebase. *Methods Mol Biol* 2007;**406**:89–112.
46. Ito JI, Tabei Y, Shimizu K, *et al*. PDB-scale analysis of known and putative ligand-binding sites with structural sketches. *Proteins* 2012;**80**(3):747–63.
47. Chen B, Dong X, Jiao D, *et al*. Chem2Bio2RDF: a semantic framework for linking and data mining chemogenomic and systems chemical biology data. *BMC Bioinformatics* 2010; **11**(1):255.
48. Chen B, Ding Y, Wild DJ. Improving integrative searching of systems chemical biology data using semantic annotation. *J Cheminform* 2012;**4**(1):6.
49. Okuno Y, Tamon A, Yabuuchi H, *et al*. GLIDA: GPCR-ligand database for chemical genomics drug discovery-database and tools update. *Nucleic Acids Res* 2008;**36(Suppl 1)**:D907–12.
50. Günther S, Kuhn M, Dunkel M, *et al*. SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Res* 2008;**36(Suppl 1)**:D919–22.
51. Schaefer CF, Anthony K, Krupa S, *et al*. PID: the pathway interaction database. *Nucleic Acids Res* 2009;**37(Suppl 1)**:D674–9.
52. Joshi Tope G, Gillespie M, Vastrik I, *et al*. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res* 2005; **33(Suppl 1)**:D428–32.
53. Caspi R, Altman T, Billington R, *et al*. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* 2014; **42**(D1):D459–71.
54. Konc J, Česnik T, Konc JT, *et al*. ProBiS-Database: precalculated binding site similarities and local pairwise alignments of pdb structures. *J Chem Inf Model* 2012;**52**:604–12.
55. Consortium GO, *et al*. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 2004;**32(Suppl 1)**:D258–61.
56. Martin AC. PDBSprotEC: a web-accessible database linking PDB chains to EC numbers via SwissProt. *Bioinformatics* 2004; **20**(6):986–8.
57. Sillitoe I, Cuff AL, Dessailly BH, *et al*. New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. *Nucleic Acids Res* 2012; **41**(D1):D490–8.
58. Andreeva A, Howorth D, Chandonia JM, *et al*. Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res* 2007;**36(Suppl 1)**:D419–25.
59. Fox NK, Brenner SE, Chandonia JM. SCOPe: structural classification of proteinsextended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Res* 2013;**42**(D1):D304–9.
60. Bamford S, Dawson E, Forbes S, *et al*. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer* 2004;**91**(2):355.
61. McKusick V. *Online Mendelian Inheritance in Man (OMIM) database [Internet]*. Bethesda, MD: National Center for Biotechnology Information for the National Institute of Health, 2004.
62. Schneider P, Schneider G. Collection of bioactive reference compounds for focused library design. *Mol Inform* 2003;**22**(7): 713–8.
63. Burdett T, Hall P, Hastings E, *et al*. The NHGRI-EBI Catalog of published genome-wide association studies, 2016. www.ebiacuk/gwas.
64. Denny JC, Bastarache L, Ritchie MD, *et al*. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol* 2013;**31**(12):1102–11.
65. Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 2002;**30**(1):207–10.
66. Garnett MJ, Edelman EJ, Heidorn SJ, *et al*. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 2012;**483**(7391):570.
67. Weinstein JN, Collisson EA, Mills GB, *et al*. The cancer genome atlas pan-cancer analysis project. *Nat Genet* 2013;**45**(10): 1113–20.
68. Beroukhim R, Mermel C, Porter D, *et al*. The landscape of somatic copy-number alteration across human cancers. *Nature* 2010;**463**:899e905.
69. Barretina J, Caponigro G, Stransky N, *et al*. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;**483**(7391):603–7.
70. 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**(5795): 1929–35.
71. Ganter B, Tugendreich S, Pearson CI, *et al*. Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J Biotechnol* 2005;**119**(3):219–44.
72. Cheng T, Hao M, Takeda T, *et al*. Large-scale prediction of drug-target interaction: a data-centric review. *AAPS J* 2017; **19**(5):1264–75.
73. Chen X, Yan CC, Zhang X, *et al*. Drug–target interaction prediction: databases, web servers and computational models. *Brief Bioinform* 2015;**17**(4):696–712.
74. Reddy AS, Pati SP, Kumar PP, *et al*. Virtual screening in drug discovery-a computational perspective. *Curr Protein Pept Sci* 2007;**8**(4):329–51.
75. Li P, Fu Y, Wang Y. Network based approach to drug discovery: a mini review. *Mini Rev Med Chem* 2015;**15**(8):687–95.
76. Haupt J. *Protein Binding Site Similarities as Driver for Drug Repositioning*. 2014. http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-144517
77. Fara DC, Oprea TI. *Cheminformatics-Basics: Molecular Descriptors and Fingerprints*. 2011. http://biocomp.health.unm.edu/biomed505/Course/Cheminformatics/basic/descs_fingers/molec_descs_fingerprints.htm (27 September 2016, date last accessed).
78. Ding H, Takigawa I, Mamitsuka H, *et al*. Similarity-based machine learning methods for predicting drug–target interactions: a brief review. *Brief Bioinform* 2013;**15**(5):734–47.
79. Koutsoukas A, Simms B, Kirchmair J, *et al*. From *in silico* target prediction to multi-target drug design: current databases, methods and applications. *J Proteomics* 2011;**74**(12): 2554–74.
80. Gong J, Cai C, Liu X, *et al*. ChemMapper: a versatile web server for exploring pharmacology and chemical structure association based on molecular 3d similarity method. *Bioinformatics* 2013;**29**(14):1827–9.

81. Liu X, Jiang H, Li H. SHAFTS: a hybrid approach for 3d molecular similarity calculation. 1. method and assessment of virtual screening. *J Chem Inf Model* 2011;**51**(9):2372–85.
82. Lu W, Liu X, Cao X, *et al*. SHAFTS: a hybrid approach for 3d molecular similarity calculation. 2. prospective case study in the discovery of diverse p90 ribosomal s6 protein kinase 2 inhibitors to suppress cell migration. *J Med Chem* 2011;**54**(10):3564–74.
83. Liu X, Ouyang S, Yu B, *et al*. PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Res* 2010;**38(Suppl 2)**:W609–14.
84. Yera ER, Cleves AE, Jain AN. Chemical structural novelty: on-targets and off-targets. *J Med Chem* 2011;**54**(19):6771–85.
85. Taboureau O, Nielsen SK, Audouze K, *et al*. ChemProt: a disease chemical biology database. *Nucleic Acids Res* 2010;**39(Suppl 1)**:D367–72.
86. Kim Kjærulff S, Wich L, Kringelum J, *et al*. ChemProt-2.0: visual navigation in a disease chemical biology database. *Nucleic Acids Res* 2012;**41**(D1):D464–9.
87. Kringelum J, Kjaerulff SK, Brunak S, *et al*. ChemProt-3.0: a global chemical biology diseases mapping. *Database* 2016. doi: 10.1093/database/bav123.
88. Liu X, Vogt I, Haque T, *et al*. HitPick: a web server for hit identification and target prediction of chemical screenings. *Bioinformatics* 2013;**29**:1910–12.
89. Malo N, Hanley JA, Cerquozzi S, *et al*. Statistical practice in high-throughput screening data analysis. *Nat Biotechnol* 2006;**24**(2):167–75.
90. Schuffenhauer A, Floersheim P, Acklin P, *et al*. Similarity metrics for ligands reflecting the similarity of the target proteins. *J Chem Inf Comput Sci* 2003;**43**(2):391–405.
91. Nidhi GM, Davies JW, *et al*. Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. *J Chem Inf Model* 2006;**46**(3):1124–33.
92. Willett P, Barnard JM, Downs GM. Chemical similarity searching. *J Chem Inf Comput Sci* 1998;**38**(6):983–96.
93. Keiser MJ, Roth BL, Armbruster BN, *et al*. Relating protein pharmacology by ligand chemistry. *Nat Biotechnol* 2007;**25**(2):197–206.
94. Xiao X, Min JL, Lin WZ, *et al*. iDrug-Target: predicting the interactions between drug compounds and target proteins in cellular networking via benchmark dataset optimization approach. *J Biomol Struct Dyn* 2015;**33**(10):2221–33.
95. Hu J, He X, Yu DJ, *et al*. A new supervised over-sampling algorithm with application to protein-nucleotide binding residue prediction. *PLoS One* 2014;**9**(9):e107676.
96. Pears R, Finlay J, Connor AM. Synthetic Minority oversampling technique (SMOTE) for predicting software build outcomes. *arXiv preprint* arXiv:1407.2330, 2014.
97. Chou KC. Prediction of protein cellular attributes using pseudo-amino acid composition. *Proteins* 2001;**43**(3):246–55.
98. Awale M, Reymond JL. The polypharmacology browser: a web-based multi-fingerprint target prediction tool using chembl bioactivity data. *J Cheminform* 2017;**9**(1):11.
99. James C, Weininger D, Delany J. *Daylight Theory Manual 4.71*. Irvine, CA: Daylight Chemical Information Systems. Inc., 2000.
100. Nickel J, Gohlke BO, Erehman J, *et al*. SuperPred: update on drug classification and target prediction. *Nucleic Acids Res* 2014;**42**(W1):W26–31.
101. Weber L. JChem base—chemAxon. *Chemistry World* 2008;**5**(10):65–6.
102. Gfeller D, Grosdidier A, Wirth M, *et al*. SwissTargetPrediction: a web server for target prediction of bioactive small molecules. *Nucleic Acids Res* 2014;**42**(W1):W32–8.
103. Gfeller D, Michielin O, Zoete V. Shaping the interaction landscape of bioactive molecules. *Bioinformatics* 2013;**29**(23):3073–9.
104. O'Boyle NM, Banck M, James CA, *et al*. Open Babel: an open chemical toolbox. *J Cheminform* 2011;**3**(1):33.
105. Willett P. Similarity-based virtual screening using 2D fingerprints. *Drug Discov Today* 2006;**11**(23):1046–53.
106. Armstrong MS, Finn PW, Morris GM, *et al*. Improving the accuracy of ultrafast ligand-based screening: incorporating lipophilicity into ElectroShape as an extra dimension. *J Comput Aided Mol Des* 2011;**25**(8):785.
107. Liu X, Gao Y, Peng J, *et al*. TarPred: a web application for predicting therapeutic and side effect targets of chemical compounds. *Bioinformatics* 2015;**31**:2049–51.
108. Rogers D, Hahn M. Extended-connectivity fingerprints. *J Chem Inf Model* 2010;**50**(5):742–54.
109. Liu X, Xu Y, Li S, *et al*. In Silicotarget fishing: addressing a" Big Data" problem by ligand-based similarity rankings with data fusion. *J Cheminform* 2014;**6**(1):33.
110. Altschul SF, Madden TL, Schäffer AA, *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;**25**(17):3389–402.
111. Wang L, Ma C, Wipf P, *et al*. TargetHunter: an *in silico* target identification tool for predicting therapeutic potential of small organic molecules based on chemogenomic database. *AAPS J* 2013;**15**(2):395–406.
112. Ehrt C, Brinkjost T, Koch O. Impact of binding site comparisons on medicinal chemistry and rational molecular design. *J Med Chem* 2016;**59**(9):4121–51.
113. Katsila T, Spyroulias GA, Patrinos GP, *et al*. Computational approaches in target identification and drug discovery. *Comput Struct Biotechnol J* 2016;**14**:177–84.
114. Wang JC, Chu PY, Chen CM, *et al*. idTarget: a web server for identifying protein targets of small chemical molecules with robust scoring functions and a divide-and-conquer docking approach. *Nucleic Acids Res* 2012;**40**(W1):W393–9.
115. Vaqué M, Arola A, Aliagas C, *et al*. BDT: an easy-to-use front-end application for automation of massive docking tasks and complex docking strategies with AutoDock. *Bioinformatics* 2006;**22**(14):1803–4.
116. Huey R, Morris GM, Olson AJ, *et al*. A semiempirical free energy force field with charge-based desolvation. *J Comput Chem* 2007;**28**(6):1145–52.
117. Wang JC, Lin JH, Chen CM, *et al*. Robust scoring functions for protein–ligand interactions with quantum chemical charge models. *J Chem Inf Model* 2011;**51**(10):2528–37.
118. Chang DTH, Oyang YJ, Lin JH. MEDock: a web server for efficient prediction of ligand binding sites based on a novel optimization algorithm. *Nucleic Acids Res* 2005;**33(Suppl 2)**:W233–8.
119. Huang Y, Niu B, Gao Y, *et al*. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 2010;**26**(5):680–2.
120. Wang C, Hu G, Wang K, *et al*. PDID: database of molecular-level putative protein–drug interactions in the structural human proteome. *Bioinformatics* 2015;**32**(4):579–86.
121. Hu G, Gao J, Wang K, *et al*. Finding protein targets for small biologically relevant ligands across fold space using inverse ligand binding predictions. *Structure* 2012;**20**(11):1815–22.
122. Xie L, Bourne PE. Detecting evolutionary relationships across existing fold space, using sequence order-independent

profile–profile alignments. *Proc Natl Acad Sci USA* 2008; **105**(14):5441–6.
123. Brylinski M, Feinstein WP. eFindSite: improved prediction of ligand binding sites in protein models using meta-threading, machine learning and auxiliary ligands. *J Comput Aided Mol Des* 2013;**27**(6):551–67.
124. Feinstein WP, Brylinski M. eFindSite: enhanced fingerprint-based virtual screening against predicted ligand binding sites in protein models. *Mol Inform* 2014;**33**(2):135–50.
125. Li H, Gao Z, Kang L, *et al*. TarFisDock: a web server for identifying drug targets with docking approach. *Nucleic Acids Res* 2006;**34(Suppl 2)**:W219–24.
126. Ewing TJ, Makino S, Skillman AG, *et al*. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des* 2001;**15**(5):411–28.
127. Gao Z, Li H, Zhang H, *et al*. PDTD: a web-accessible protein database for drug target identification. *BMC Bioinformatics* 2008;**9**(1):104.
128. Lotfi Shahreza M, Ghadiri N, Mousavi SR, *et al*. A review of network-based approaches to drug repositioning. *Brief Bioinform* 2017. doi: 10.1093/bib/bbx017.
129. Lecca P, Priami C. Biological network inference for drug discovery. *Drug Discov Today* 2013;**18**(5):256–64.
130. Harrold J, Ramanathan M, Mager D. Network-based approaches in drug discovery and early development. *Clini Pharmacol Ther* 2013;**94**(6):651–8.
131. Cobanoglu MC, Oltvai ZN, Taylor DL, *et al*. BalestraWeb: efficient online evaluation of drug–target interactions. *Bioinformatics* 2014;**31**:131–3.
132. Salakhutdinov R, Mnih A. Probabilistic matrix factorization. In *NIPS'07 Proceedings of the 20th International Conference on Neural Information Processing Systems*, vol. 1, 2007, 1257–64. Curran Associates, New York, USA.
133. Cobanoglu MC, Liu C, Hu F, *et al*. Predicting drug–target interactions using probabilistic matrix factorization. *J Chem Inf Model* 2013;**53**(12):3399–409.
134. Yamanishi Y, Araki M, Gutteridge A, *et al*. Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 2008;**24**(13):i232–40.
135. Yamanishi Y, Kotera M, Kanehisa M, *et al*. Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics* 2010; **26**(12):i246–54.
136. Lo YC, Senese S, Li CM, *et al*. Large-scale chemical similarity networks for target profiling of compounds identified in cell-based chemical screens. *PLoS Comput Biol* 2015;**11**(3): e1004153.
137. Schwikowski B, Uetz P, Fields S. A network of protein–protein interactions in yeast. *Nat Biotechnol* 2000;**18**(12):1257–61.
138. Hishigaki H, Nakai K, Ono T, *et al*. Assessment of prediction accuracy of protein function from protein–protein interaction data. *Yeast* 2001;**18**(6):523–31.
139. Jahn A, Hinselmann G, Fechner N, *et al*. Optimal assignment methods for ligand-based virtual screening. *J Cheminform* 2009;**1**(1):14.
140. Ba Alawi W, Soufan O, Essack M, *et al*. DASPfind: new efficient method to predict drug–target interactions. *J Cheminform* 2016;**8**(1):15.
141. Hattori M, Okuno Y, Goto S, *et al*. Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. *J Am Chem Soc* 2003;**125**(39):11853–65.
142. Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol* 1981;**147**(1):195–7.
143. van Laarhoven T, Nabuurs SB, Marchiori E. Gaussian interaction profile kernels for predicting drug–target interaction. *Bioinformatics* 2011;**27**(21):3036–43.
144. Bleakley K, Yamanishi Y. Supervised prediction of drug–target interactions using bipartite local models. *Bioinformatics* 2009;**25**(18):2397–403.
145. Cheng F, Liu C, Jiang J, *et al*. Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput Biol* 2012;**8**(5):e1002503.
146. Wang W, Yang S, Li J. Drug target predictions based on heterogeneous graph inference. *Pac Symp Biocomput* 2013; 53–64.
147. Alaimo S, Bonnici V, Cancemi D, *et al*. DT-Web: a web-based application for drug-target interaction and drug combination prediction through domain-tuned network-based inference. *BMC Syst Biol* 2015;**9**(3):S4.
148. Alaimo S, Pulvirenti A, Giugno R, *et al*. Drug–target interaction prediction through domain-tuned network-based inference. *Bioinformatics* 2013;**29**(16):2004–8.
149. Altschul SF, Gish W, Miller W, *et al*. Basic local alignment search tool. *J Mol Biol* 1990;**215**(3):403–10.
150. Maglott D, Ostell J, Pruitt KD, *et al*. Entrez gene: gene-centered information at NCBI. *Nucleic Acids Res* 2005; **33(Suppl 1)**:D54–8.
151. Zhou T, Ren J, Medo M, *et al*. Bipartite network projection and personal recommendation. *Phys Rev E* 2007;**76**:046115.
152. Zhou T, Kuscsik Z, Liu JG, *et al*. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proc Natl Acad Sci USA* 2010;**107**(10):4511–5.
153. Martínez Jiménez F, Marti Renom MA. Ligand-target prediction by structural network biology using nAnnoLyze. *PLoS Comput Biol* 2015;**11**(3):e1004157.
154. Stuart AC, Ilyin VA, Sali A. LigBase: a database of families of aligned ligand binding sites in known protein sequences and structures. *Bioinformatics* 2002;**18**(1):200–1.
155. Pieper U, Eswar N, Davis FP, *et al*. MODBASE: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res* 2006;**34(Suppl 1)**:D291–5.
156. Konc J, Janežič D. ProBiS-2012: web server and web services for detection of structurally similar binding sites in proteins. *Nucleic Acids Res* 2012;**40**:W214–21.
157. Von Eichborn J, Murgueitio MS, Dunkel M, *et al*. PROMISCUOUS: a database for network-based drug-repositioning. *Nucleic Acids Res* 2011;**39(Suppl 1)**:D1060–6.
158. Chen B, Ding Y, Wild DJ. Assessing drug target association using semantic linked data. *PLoS Comput Biol* 2012;**8**(7):e1002574.
159. Wang H, Ding Y, Tang J, *et al*. Finding complex biological relationships in recent PubMed articles using Bio-LDA. *PLoS One* 2011;**6**(3):e17243.
160. He B, Tang J, Ding Y, *et al*. Mining relational paths in integrated biomedical data. *PLoS One* 2011;**6**(12):e27506.
161. Kuhn M, von Mering C, Campillos M, *et al*. STITCH: interaction networks of chemicals and proteins. *Nucleic Acids Res* 2008;**36(Suppl 1)**:D684–8.
162. Kuhn M, Szklarczyk D, Franceschini A, *et al*. STITCH 2: an interaction network database for small molecules and proteins. *Nucleic Acids Res* 2010;**38(Suppl 1)**:D552–6.
163. Szklarczyk D, Santos A, von Mering C, *et al*. STITCH 5: augmenting protein–chemical interaction networks with tissue and affinity data. *Nucleic Acids Res* 2016;**44**(D1):D380–4.
164. NIH. NIH research portfolio online reporting tools, 2014.
165. Kalinina OV, Wichmann O, Apic G, *et al*. Combinations of protein-chemical complex structures reveal new targets for established drugs. *PLoS Comput Biol* 2011;**7**(5):e1002043.

166. Meslamani J, Rognan D. Enhancing the accuracy of chemogenomic models with a three-dimensional binding site kernel. *J Chem Inf Model* 2011;**51**(7):1593–603.
167. Heiskanen MA, Aittokallio T. Predicting drug–target interactions through integrative analysis of chemogenetic assays in yeast. *Mol Biosyst* 2013;**9**(4):768–79.
168. Konc J, Janezic D. An improved branch and bound algorithm for the maximum clique problem. *Proteins* 2007;**4**(5).
169. Ito JI, Tabei Y, Shimizu K, *et al*. PoSSuM: a database of similar protein–ligand binding and putative pockets. *Nucleic Acids Res* 2012;**40**(D1):D541–8.
170. Ito JI, Ikeda K, Yamada K, *et al*. PoSSuM v. 2.0: data update and a new function for investigating ligand analogs and target proteins of small-molecule drugs. *Nucleic Acids Res* 2015; **43**(D1):D392–8.
171. Kawabata T. Detection of multiscale pockets on protein surfaces using mathematical morphology. *Proteins* 2010;**78**(5): 1195–211.
172. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 2005; **33**(7):2302–9.
173. Tabei Y, Uno T, Sugiyama M, *et al*. Single versus multiple sorting in all pairs similarity search. In: *The 2nd Asian Conference on Machine Learning*, 2010, 145–60. Tokyo, Japan. JMLR, Inc. and Microtome Publishing, USA.
174. Zhou H, Gao M, Skolnick J. Comprehensive prediction of drug-protein interactions and side effects for the human proteome. *Sci Rep* 2015;**5**:11090.
175. Zhou H, Skolnick J. FINDSITEcomb: a threading/structure-based, proteomic-scale virtual ligand screening approach. *J Chem Inf Model* 2012;**53**(1):230–40.
176. Yuan Q, Gao J, Wu D, *et al*. DrugE-Rank: improving drug–target interaction prediction of new candidate drugs or targets by ensemble learning to rank. *Bioinformatics* 2016; **32**(12):i18–27.
177. Reker D, Rodrigues T, Schneider P, *et al*. Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus. *Proc Natl Acad Sci USA* 2014;**111**(11):4067–72.
178. Dudley JT, Deshpande T, Butte AJ. Exploiting drug–disease relationships for computational drug repositioning. *Brief Bioinform* 2011;**12**(4):303–11.
179. Brown AS, Patel CJ. MeSHDD: literature-based drug-drug similarity for drug repositioning. *J Am Med Inform Assoc* 2016; **24**:614–18.
180. Abdi H. Bonferroni and Šidák corrections for multiple comparisons. *Encyclopedia Measur Stat* 2007;**3**:103–7.
181. Jain AK, Moreau J. Bootstrap technique in cluster analysis. *Pattern Recogn* 1987;**20**(5):547–68.
182. Kodinariya TM, Makwana PR. Review on determining number of Cluster in K-means clustering. *Int J* 2013;**1**(6):90–5.
183. Moosavinasab S, Patterson J, Strouse R, *et al*. RE: fine drugs: an interactive dashboard to access drug repurposing opportunities. *Database* 2016. doi: 10.1093/database/baw083.
184. Menzella F, Lusuardi M, Galeone C, *et al*. Tailored therapy for severe asthma. *Multidiscip Respir Med* 2015;**10**(1):1.
185. Colice GL. Emerging therapeutic options for asthma. *Am J Manag Care* 2011;**17**(3):S82.
186. Antoniu SA. Daclizumab a novel corticosteroid-sparing therapy for asthma? Evaluation of: dusse WW, Israel E, Nelson HS, et al. Daclizumab improves asthma control in patients with moderate to severe persistent asthma: a randomized, controlled trial. Am J Respir Crit Care Med 2008; 178: 1002-8. *Exp Opin Investig Drugs* 2009;**18**(3):369–71.
187. Iorio F, Rittman T, Ge H, *et al*. Transcriptional data: a new gateway to drug repositioning? *Drug Discovery Today* 2013; **18**(7):350–7.
188. Hurle M, Yang L, Xie Q, *et al*. Computational drug repositioning: from data to therapeutics. *Clin Pharmacol Ther* 2013;**93**(4): 335–41.
189. Massey FJ. Jr The Kolmogorov-Smirnov test for goodness of fit. *J Am Stat Assoc* 1951;**46**(253):68–78.
190. Qu XA, Rajpal DK. Applications of connectivity map in drug discovery and development. *Drug Discovery Today* 2012; **17**(23):1289–98.
191. Kunkel SD, Suneja M, Ebert SM, *et al*. mRNA expression signatures of human skeletal muscle atrophy identify a natural compound that increases muscle mass. *Cell Metabolism* 2011; **13**(6):627–38.
192. Sirota M, Dudley JT, Kim J, *et al*. Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci Transl Med* 2011;**3**(96): 96ra77–96ra77.
193. Ishimatsu Tsuji Y, Soma T, Kishimoto J. Identification of novel hair-growth inducers by means of connectivity mapping. *FASEB J* 2010;**24**(5):1489–96.
194. Lee BKB, Tiong KH, Chang JK, *et al*. DeSigN: connecting gene expression with therapeutics for drug repurposing and development. *BMC Genomics* 2017;**18**(1):934.
195. 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**(D1): D955–61.
196. Louhimo R, Laakso M, Belitskin D, *et al*. Data integration to prioritize drugs using genomics and curated data. *BioData Min* 2016;**9**(1):21.
197. Kanehisa M, Goto S, Sato Y, *et al*. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 2011;**40**:D109–14.
198. Ashburner M, Ball CA, Blake JA, *et al*. Gene Ontology: tool for the unification of biology. *Nat Genet* 2000;**25**(1):25–9.
199. Duan Q, Reid SP, Clark NR, *et al*. L1000CDS2: LINCS L1000 characteristic direction signatures search engine. *NPJ Syst Biol Appl* 2016;**2**:16015.
200. Clark NR, Hu KS, Feldmann AS, *et al*. The characteristic direction: a geometrical approach to identify differentially expressed genes. *BMC Bioinformatics* 2014;**15**(1):79.
201. Carrella D, Napolitano F, Rispoli R, *et al*. Mantra 2.0: an online collaborative resource for drug mode of action and repurposing by network analysis. *Bioinformatics* 2014;**30**(12): 1787–8.
202. Iorio F, Bosotti R, Scacheri E, *et al*. Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc Natl Acad Sci USA* 2010;**107**(33):14621–6.
203. Subramanian A, Tamayo P, Mootha VK, *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005;**102**(43):15545–50.
204. Setoain J, Franch M, Martínez M, *et al*. NFFinder: an online bioinformatics tool for searching similar transcriptomics experiments in the context of drug repositioning. *Nucleic Acids Res* 2015;**43**:W193–9.
205. Vazquez M, Nogales Cadenas R, Arroyo J, *et al*. MARQ: an online tool to mine geo for experiments with similar or

opposite gene expression signatures. *Nucleic Acids Res* 2010; **38(Suppl 2)**:W228–32.
206. Ganter B, Snyder RD, Halbert DN, *et al*. Toxicogenomics in drug discovery and development: mechanistic analysis of compound/class-dependent effects using the DrugMatrix® database. *Pharmacogenomics* 2006;**7**:1025–44.
207. Yu H, Choo S, Park J, *et al*. Prediction of drugs having opposite effects on disease genes in a directed network. *BMC Syst Biol* 2016;**10**:2.
208. Smyth GK, Ritchie M, Thorne N, *et al*. LIMMA: linear models for microarray data. In: *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. 2005. Springer Science+Bussiness Media Inc., New York, USA.
209. Stephens ZD, Lee SY, Faghri F, *et al*. Big data: astronomical or genomical?. *PLoS Biol* 2015;**13**(7):e1002195.
210. Loman N, Watson M. So you want to be a computational biologist? *Nat Biotechnol* 2013;**31**(11):996–8.