



Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines

Guangxu Jin^{1,2} and Stephen T.C. Wong^{1,2}

¹ Department of Systems Medicine and Bioengineering, Houston Methodist Research Institute, Weill Cornell Medical College, Houston, TX 77030, USA

² NCI Center for Modeling Cancer Development, Houston Methodist Research Institute, Houston, TX 77030, USA

Recycling old drugs, rescuing shelved drugs and extending patents' lives make drug repositioning an attractive form of drug discovery. Drug repositioning accounts for approximately 30% of the newly US Food and Drug Administration (FDA)-approved drugs and vaccines in recent years. The prevalence of drug-repositioning studies has resulted in a variety of innovative computational methods for the identification of new opportunities for the use of old drugs. Questions often arise from customizing or optimizing these methods into efficient drug-repositioning pipelines for alternative applications. It requires a comprehensive understanding of the available methods gained by evaluating both biological and pharmaceutical knowledge and the elucidated mechanism-of-action of drugs. Here, we provide guidance for prioritizing and integrating drug-repositioning methods for specific drug-repositioning pipelines.

A repositioned drug does not need the initial six to nine years typically required for the development of new drugs, but instead goes directly to preclinical testing and clinical trials, thus reducing risk and costs [1]. Repositioning or repurposing drugs has been implemented in several ways. One of the well-known examples is sildenafil citrate (brand name: Viagra), which was repositioned from a common hypertension drug to a therapy for erectile dysfunction [2]. Similarly, off-label use of Food and Drug Administration (FDA)-approved drugs for cancer medical practice is also popular. The National Comprehensive Cancer Network (NCCN) estimates off-label use accounts for 50–75% of drugs or biologic therapies for cancer in the USA [3]. It has been reported that 78% and 75% of patients with breast or lung cancer, respectively received FDA-approved drugs, although 68% and 95% of these drugs, respectively, were used for off-label indications not approved by the FDA [4]. Obviously, these examples were serendipitously identified and these repositioning strategies lack guidance and information to support clinical decision.

Pharmaceutical companies rely on traditional drug discovery methods to seek repositioning opportunities. Among the 75 agents (50 small molecules and 25 biologics) approved between 1999 and

2008, 28 first-in-class small molecules were discovered by phenotypic drug screening and 17 were identified by target-based methods [5,6], accounting for more than 50% of the FDA-approved small molecules and biologics. Phenotypic drug-screening approaches discover drug candidates from libraries serendipitously. Alternatively, target-based methods improve the repositioning process by including known target information into drug-repositioning studies.

Nevertheless, the low knowledge content of elucidated mechanisms for traditional drug-repositioning methods makes it hard to satisfy unmet medical needs by successfully repositioning a large number of existing or shelved drugs. Computational methods are able to alleviate this problem by high-level integration of available knowledge and elucidation of unknown mechanisms. These computational methods significantly improve the discovery process in which new indications for a drug or new drugs for a disease can be identified. They take advantage of the methods and tools available in chemoinformatics [7–9], bioinformatics [10–14], network biology [15–17] and systems biology [18–20] to make full use of known targets, drugs, and disease biomarkers or pathways, thus leading to the development of proof-of-concept methods and the design of clinical studies with accelerated timelines. Accordingly, computational drug-repositioning methods can be classified into

Corresponding author: Wong, Stephen T.C. (stwong@tmhs.org), (EWenzel@tmhs.org)

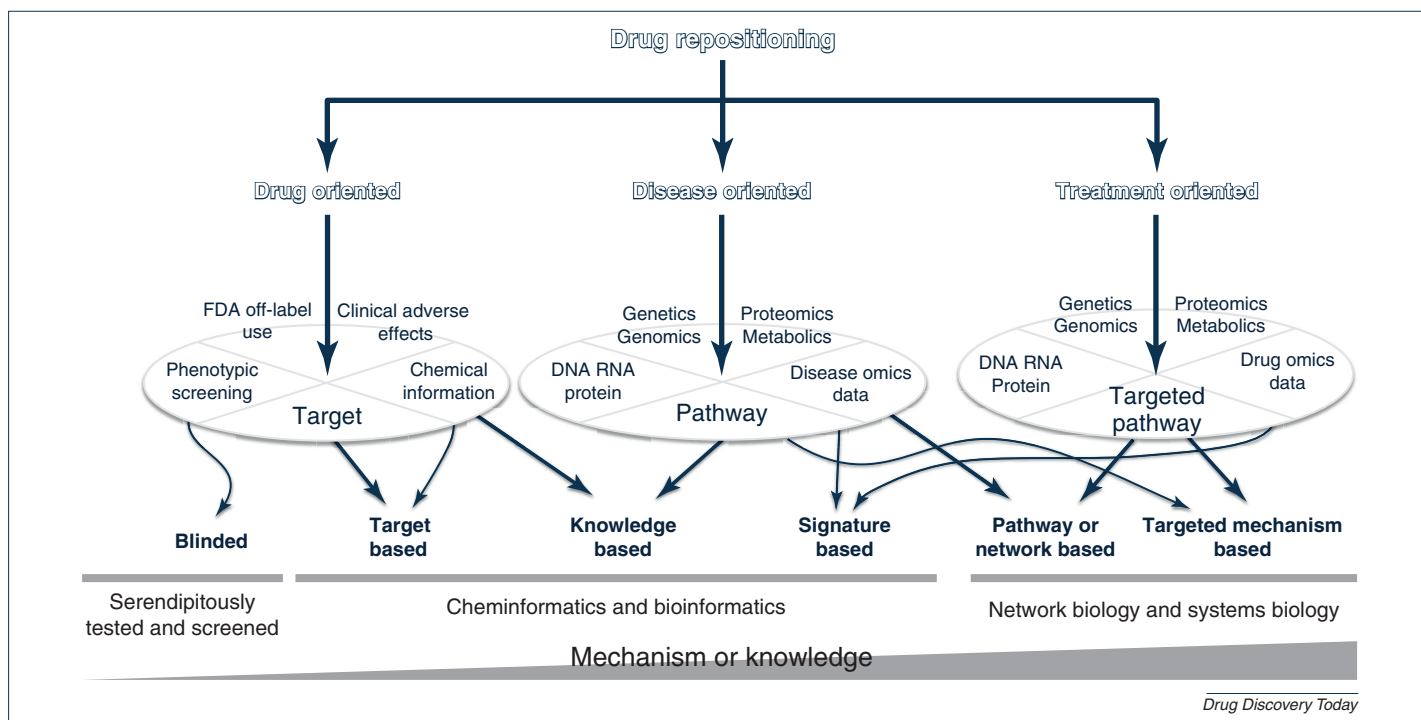


FIGURE 1

Lotus leaves flowchart (LLF) for categorization of existing drug-repositioning methods. Drug repositioning takes advantage of different potential avenues to repurpose drugs for new indications, including drug, disease and treatment oriented. These avenues were developed according to the availability of biological and pharmaceutical knowledge and requirement of understanding the mechanisms of action of drugs. Traditional phenotype-based screening methods do not need prior knowledge, and the repositioned drugs are just serendipitously tested. Targeted-based methods need specific knowledge about the targets, such as 3D protein structures, whereas knowledge-based methods require the knowledge about the drugs or diseases, such as adverse effects, FDA approval labels, records of clinical trials and published disease biomarkers (potential targets) or disease pathways. Signature-based methods mainly make use of gene signatures defined by ‘omics’ data (for diseases, drug treatments, or both). Pathway- or network-based methods generally use pathway analysis or network biology methods to discover essential pathways from genetic, genomic, proteomic and metabolic data of diseases to find new targets for repositioned drugs. More advanced drug-repositioning methods, such as targeted mechanism-based methods, aim to discover mechanisms of action of drugs by identification of off-targets or targeted pathways of treated drugs using drug omics data (before and after drug treatments). Details of these methods are in Table 1 (main text). Integrated knowledge and elucidated mechanisms of drug actions increases with the complexity of modeling methods.

target-based, knowledge-based, signature-based, pathway- or network-based, and targeted-mechanism-based methods, as shown in Fig. 1. These methods focus on different orientations defined by available information and elucidated mechanisms, such as drug oriented, disease oriented and treatment oriented. These computational drug-repositioning methods enable researchers to examine nearly all drug candidates and test on a relatively large number of diseases within significantly shortened time lines.

In recent years, the number of drug-repositioning methods has dramatically increased. It is essential to better understand these existing methods and prioritize them based on specific studies. Application of an efficient drug-repositioning pipeline to a specific study needs identification of feasible methods based on available information of the drugs or diseases of interest. In this review, we link existing drug-repositioning methods with their integrated biological and pharmaceutical knowledge and discuss how to customize a new drug-repositioning pipeline for specific studies.

Prioritize available drug-repositioning methods

Figure 1 is a top-down flowchart that we developed to better understand orientations, integrated information types, categories and complexities of existing drug-repositioning methods. We call this flowchart a lotus leaves flowchart (LLF). It enables better understanding of repositioning methods from the top down while

customizing new repositioning pipelines from the bottom up. As an example, if one wants to reposition drugs for an orphan disease, one needs to identify how much pharmaceutical or biological knowledge is available for this disease and whether understanding the mechanisms of action of repositioned drugs is necessary. There are several options to do such drug repositioning. Option 1: when little information is available for the disease, phenotypic screening or FDA off-label use would be the best option. Option 2: if there exists one protein biomarker for the disease, target-based or knowledge-based methods should be prioritized for the study. Option 3: if there is more disease information available, either knowledge-based or signature-based methods can be deployed to integrate available disease pathways or disease omics data (i.e. omics data generated from diseases) into the drug-repositioning process. Lastly, option 4: if treatment omics data (i.e. omics data generated from drug treatment) are available, it is possible to use signature-based or targeted-mechanism-based methods to elucidate unknown targeted mechanisms, such as off-targets and targeted signaling pathways.

It is easy to see that the development of an efficient drug-repositioning pipeline is a process of tradeoff among purposes, methods and available information. Here, we introduce the repositioning methods shown in the LLF to facilitate understanding the purpose, the integrated information and the complexities of

these methods. The LLF flowchart will be helpful for scientists and researchers to better understand existing computational methods and customize these methods into their own pipelines for drug-repositioning studies.

Blinded search or screening methods

Blinded drug-repositioning methods do not include pharmaceutical or biological information and are less likely to help elucidate any mechanisms of action of drugs. Most of them depend on serendipitous identification from tests aimed at specific diseases and drugs [3,4,21]. The advantage of these methods, which include FDA off-label use and phenotypic screening, is that they have high flexibility for application to a large number of drugs or diseases. This explains why the phenotypic screening method was used in the discovery of 28 of 75 small molecules and biologics approved by the FDA between 1999 and 2008.

Target-based methods

Target-based drug-repositioning methods comprise *in vitro* and *in vivo* high-throughput and/or high-content screening (HTS/HCS) of drugs for a protein or a biomarker of interest [7–9] and *in silico* screening of drugs or compounds from drug libraries [7,22], such as ligand-based screening or docking [23,24]. Compared with blinded methods, targeted-based methods significantly improve the likelihood of drug discovery because most targets link directly with the disease mechanisms. Integration of target information into the drug repositioning process ensures a higher possibility of finding useful drugs compared with traditional blinded methods. The advantage of targeted-based methods, such as docking, is that these methods enable researchers to screen nearly all drugs or compounds with known chemical structure information (e.g. Simplified Molecular-Input Line-Entry System SMILES) within a few days. This is why so many pharmaceutical companies, including Genentech and Melior, have been using these methods to find new indications.

Knowledge-based methods

Knowledge-based drug-repositioning methods are those applying bioinformatics or cheminformatics approaches to include the available information of drugs, drug–target networks [10–14], chemical structures of targets and drugs [14], clinical trial information (adverse effects) [25,26], FDA approval labels [27], signaling or metabolic pathways [28], and so on, into drug-repositioning studies. The information content of blinded and target-based methods are poor and they cannot be used to identify new mechanisms beyond the known targets. By contrast, knowledge-based methods incorporate known information into predicting unknown mechanisms, such as unknown targets for drugs, unknown drug–drug similarities, and new biomarkers for diseases. The advantage of knowledge-based methods is that they include a large amount of known information into the drug-repositioning process to improve its prediction accuracy. For example, THOMSON REUTERS™ has used this strategy to do drug repositioning based on its rich volumes of accumulated prior knowledge. Moreover, these methods have been applied to repurpose known drugs to pediatric hematology oncology. Blatt and Corey describe how the knowledge in the *Harriet Lane Handbook* (HLH) of the Johns Hopkins School of Medicine (compiled based on perceived interest to the general pediatric practitioners) and information acquired by

searching PubMed and Google.com might also be helpful to repurpose drugs for children [29].

Signature-based methods

Signature-based drug-repositioning methods make use of gene signatures derived from disease omics data with or without treatments [30–37] to discover unknown off-targets or unknown disease mechanisms. As the advancement of microarray and next generation sequencing techniques speed up the generation of vast volumes of genomics data pertinent for drug-repositioning studies, gene signatures can be used to discover unknown mechanisms. One can easily access such genomics data in publicly available databases, such as NCBI-GEO (<http://www.ncbi.nlm.nih.gov/geo/>), SRA Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/Traces/sra/>), CMAP Connectivity Map [32] and CCLE Cancer Cell Line Encyclopedia [38]. More details on these databases are shown in Table 2. The advantage of signature-based methods is that they are useful to uncover unknown mechanisms of action of molecules and drugs. Compared with knowledge-based methods, signature-based methods involve more molecular-level mechanisms, such as the significantly changed genes, by using computational approaches.

Pathway- or network-based methods

Pathway- or network-based drug-repositioning methods utilize disease omics data, available signaling or metabolic pathways, and protein interaction networks to reconstruct disease-specific pathways that provide the key targets for repositioned drugs [15–17]. The advantage of these methods is that they are helpful in narrowing general signaling networks from a large number of proteins down to a specific network with a few proteins (or targets). A recent study of drug repositioning addressed distinct signaling mechanisms of metastatic subtypes of breast cancer [17]. Neither knowledge-based nor signature-based methods can address these repositioning results because the subtype signaling mechanisms are hard to elucidate from existing breast cancer pathways or the gene signatures.

Targeted mechanism-based methods

Targeted mechanism-based drug-repositioning methods integrate treatment omics data, available signaling pathway information and protein interaction networks to delineate the unknown mechanisms of action of drugs [18–20]. The era of precision medicine motivates such drug-repositioning studies. For instance, drug resistance remains an unresolved issue in cancer therapy. Although patients respond well to a drug initially, they often acquire resistance to that drug after a few months of treatment. This indicates that deriving a successful drug treatment needs additional information about the mechanisms of action of drugs to find better drug targets. Systems biology approaches are promising in addressing this challenge. The advantage of these methods is that their goals are not only to discover the mechanisms related to diseases or drugs, but also to identify those directly related to treatments of drugs to specific diseases. Owing to the difficulties in deriving effective computational models, there are only a few studies on these targeted mechanism-based methods [18–20] that developed elegant computational models to predict the drug effects and related targeted pathways. Comprehensive overviews of these drug-repositioning methods are given in Tables 1 and 2.

TABLE 1

Details of available drug-repositioning methods

Fields	Methods	Categories	Mechanism or knowledge complexity	Approach summaries with references	Examples
Drug-oriented					
FDA off-label use	Blinded	Serendipitously tested	Extremely low	Clinical decisions [3,4,21]	Sildenafil citrate (erectile dysfunction); rituximab (breast cancer); fluorouracil (lung cancer) and etoposide (bladder cancer)
Phenotypic screening	Blinded	Screening	Extremely low	<i>In vivo</i> and <i>in vitro</i> HTS/HCS screening [44–47]	HDECC inhibitors for stem-like lung cancer cells; widely used in GSK and Novartis
Phenotypic screening	Target based	Screening	Low	<i>In vivo</i> and <i>in vitro</i> HTS/HCS screening [7–9]	Widely used in Genentech
Target 3D structure, chemical structure information of drugs and ligands	Target based	Cheminformatics	Low	<i>In silico</i> screening [7,22], ligand based and docking [23,24]	MLR-1023 for diabetes (Melior)
Drug–target information, chemical structure information of targets and drugs	Knowledge based	Bioinformatics, Cheminformatics	Moderate	Drug–target prediction [10–14]	Simvastatin and Ketoconazole (breast cancer)
Adverse effects (clinical trial information)	Knowledge based	Bioinformatics	Moderate	Using correlation to define disease adverse effect associations [25], and adverse effects to define drug similarity [26]	
FDA approval labels and adverse effects	Knowledge based	Bioinformatics	Moderate	Using principal component analysis to define drug similarity measurement [27]	
Disease-oriented					
Available Pathway information	Knowledge based	Bioinformatics	Moderate	Discovery of disease mechanism and address of key targets [28]	Vismodegib for skin cancer
Disease omics data	Signature based	Bioinformatics	Moderate	From gene signature (most changed genes) in disease of interest to identify key targets [30]	
Genetics data	Signature based	Bioinformatics	Moderate	Genome-wide association study analysis to identify key targets [31]	
Disease omics data, available pathway information, and protein interaction network	Pathway or network based	Network biology	High	Reconstruction of disease-specific pathways and networks to identify key targets [15–17]	Sunitinib and dasatinib for breast cancer brain metastases
Treatment-oriented					
Drug omics data	Signature based	Bioinformatics	Moderate	Connectivity Map, linking diseases with treated drugs by using gene signatures [32–34]	Sirolimus for patients with acute lymphoblastic leukemia with dexamethasone resistance
	Signature and network based	Bioinformatics, network biology	Moderate	Using gene signatures to define the distances among drugs and then use community structure to classify drug [48]	Fasudil (a Rho-kinase inhibitor) for neurodegenerative disorders
Disease omics and drug omics data	Signature based	Bioinformatics	High	Using both drug and disease gene signatures to define similarities between drugs and diseases [35–37]	Cimetidine for lung cancer and topiramate for Inflammatory bowel disease
Drug omics data, disease pathway and protein interaction network	Targeted-mechanism based	Network biology and systems biology	Extremely high	Elucidating targeted pathways for each treated drugs to define potential repositioning score [18–20]	Daunorubicin and clomifene for breast cancer

Customize and generate new drug-repositioning pipelines

The low-risk and low-cost drug-repositioning strategies have been widely used to identify new clinical opportunities for old drugs. Accordingly, numerous strategies have been developed from drug-repositioning studies. On the basis of the LLF, we use a fishbone flowchart to present the existing methods with preclinical and

clinical validations (Fig. 2). The fishbone flowchart helps readers to understand the existing drug-repositioning pipelines and shows how to generate new drug-repositioning pipelines.

A general fishbone flowchart of existing methods

The fishbone flowchart includes general methods used for drug-repositioning studies. Starting with a disease or a drug, one can

TABLE 2

Databases used for drug-repositioning studies

Fields	Databases	Website	Refs
Chemical structure	PubChem	http://pubchem.ncbi.nlm.nih.gov	
	Collaborative Drug Discovery Vault	https://www.collaborativedrug.com	
	Drugbank	http://www.drugbank.ca/	[49]
	Therapeutic Target Database (TTD)	http://bidd.nus.edu.sg/group/cjttd/	[50]
	Pharmacogenetics Knowledge Base (PharmGKB)	http://www.pharmgkb.org/	[51]
	DrugMap Central (DMC)	http://r2d2drug.org/index.html	[42]
	ChemSpider	http://www.chemspider.com	
	ChemFrog	http://www.chemfrog.com	
	ChemDB	http://www.chemdb.com	
	iScienceSearch	http://cwmglobalsearch.com/gs/Default.aspx	
	Chemicalize (ChemAxon)	http://www.chemicalize.org	
	DistilBio	http://distilbio.com	
Target 3D structure	RCSB Protein Data Bank (PDB)	http://www.rcsb.org	
	OCA	http://oca.weizmann.ac.il/oca-bin/ocamain	
	OPM (membrane proteins)	http://opm.phar.umich.edu	
	Proteopedia	http://proteopedia.org	
Drug–target information	TOPSAN	http://www.topsan.org	
	Drugbank	http://www.drugbank.ca/	[49]
	Therapeutic Target Database (TTD)	http://bidd.nus.edu.sg/group/cjttd/	[50]
	Pharmacogenetics Knowledge Base (PharmGKB)	http://www.pharmgkb.org/	[51]
	DrugMap Central (DMC)	http://r2d2drug.org/index.html	[42]
	MATADOR (manually annotated)	http://matador.embl.de	
	SuperTarget	http://bioinf-apache.charite.de/supertarget_v2/	[52]
	STITCH (Chemical-Protein Interactions)	http://stitch.embl.de/	[53]
	GPCR-Ligand Database (GLIDA)	http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/	[54]
	Psychoactive Drug Screening Program Ki (PDSP Ki)	http://pdsp.med.unc.edu/pdsp.php	[55]
Adverse effects and clinical trial information	BindingDB	http://www.bindingdb.org/bind/index.jsp	[56]
	SIDER	http://sideeffects.embl.de/	[57]
	FAERS (US FDA)	http://www.fda.gov/Drugs/	
	Adverse Reaction Database (Canada)	http://www.hc-sc.gc.ca	
	Iowa Drug Information Service (IDIS)	http://itsnt14.its.uiowa.edu	
	Clinicaltrials.gov	http://clinicaltrials.gov	
FDA label information	DrugMap Central (DMC)	http://r2d2drug.org/index.html	[42]
	FDALABEL (US FDA)	http://www.fda.gov/ScienceResearch/BioinformaticsTools/ucm289739.htm	
	DailyMed (US FDA)	http://dailymed.nlm.nih.gov/dailymed/about.cfm	
Pathway information	Structured Product Labeling (SPL)	http://www.fda.gov/ForIndustry/DataStandards/StructuredProductLabeling/default.htm	
	NCI Pathway Interaction Database (NCI-PID)	http://pid.nci.nih.gov/	[58]
	Kyoto Encyclopedia of Genes and Genomes (KEGG)	http://www.genome.jp/kegg/	[59]
	BioCarta	http://www.biocarta.com	
	Reactome	http://www.reactome.org	
	PathwayCommons	http://www.pathwaycommons.org/about/	[60]
Protein interaction information	DrugMap Central (DMC)	http://r2d2drug.org/index.html	[42]
	Human Protein Reference Database (HPRD)	http://www.hprd.org/	[61]
	Biological General Repository for Interaction Datasets (BioGRID)	http://thebiogrid.org/	[62]
	STRING	http://string-db.org/	[63]
	PathwayCommons	http://www.pathwaycommons.org/about/	[60]
	MIPS (mammalian protein-protein interaction database)	http://mips.helmholtz-muenchen.de/proj/ppi/	[64]
	IntAct	http://www.ebi.ac.uk/intact/	[65]
	Database of Interacting Proteins (DIP)	http://dip.doe-mbi.ucla.edu/dip/Main.cgi	[66]
Molecular omics data	NCBI-GEO	http://www.ncbi.nlm.nih.gov/geo/	
	Sequence Read Archive (SRA)	http://www.ncbi.nlm.nih.gov/Traces/sra/	
	Stanford Microarray Database	http://smd.princeton.edu	
	ArrayExpress	http://www.ebi.ac.uk/arrayexpress/	
	Princeton University MicroArray database (PUMAdb)	http://puma.princeton.edu	
	CellMiner (for NCI-60)	http://discover.nci.nih.gov/cellminer/	
	Oncomine	http://www.oncomine.org	
Genetic data or information	Cancer Cell Line Encyclopedia (CCLE)	http://www.broadinstitute.org/ccle/home	[38]
	dbSNP	http://www.ncbi.nlm.nih.gov/projects/SNP/	
	Sequence Read Archive (SRA)	http://www.ncbi.nlm.nih.gov/Traces/sra/	
Drug omics data	Online Mendelian Inheritance in Man (OMIM)	http://www.omim.org/	[67]
	Connectivity Map (CMAP)	http://www.broadinstitute.org/cmap/	[32]
	Cancer Cell Line Encyclopedia (CCLE)	http://www.broadinstitute.org/ccle/home	[38]
	NCBI-GEO	http://www.ncbi.nlm.nih.gov/geo/	
Drug omics data	Sequence Read Archive (SRA)	http://www.ncbi.nlm.nih.gov/Traces/sra/	

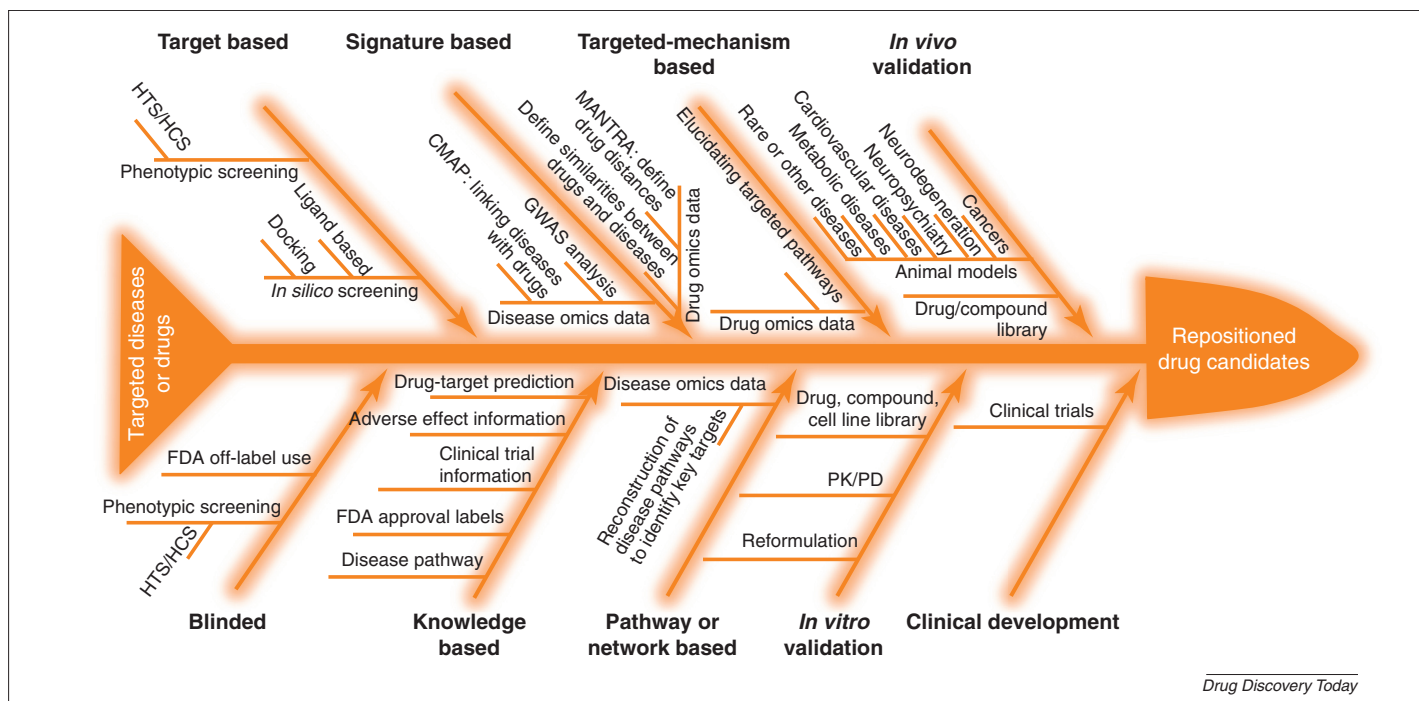


FIGURE 2

Fishbone flowchart of drug-repositioning pipelines. Developed drug-repositioning pipelines comprise at least one of these methods (i.e. blinded, target based, knowledge based, signature based, pathway or network based, and targeted-mechanism based), preclinical studies (*in vitro* and/or *in vivo* validations) and clinical development (testing the new indications identified from the preclinical development). Development of a new drug-repositioning pipeline for a drug or a disease of interest should evaluate the priorities of these repositioning methods based on the available information of the drug or the disease. For example, an infectious disease that only has limited available signaling information related to cell wall and cytoplasmic membrane proteins would lead to high priority of target-based drug-repositioning studies focusing on these cell wall and cytoplasmic membrane proteins. Many drug-repositioning pipelines can reverse the order of the listed methods in the fishbone flowchart and make use of them flexibly. As an example, several existing drug-repositioning pipelines first consider pathway- or network-based drug-repositioning methods to reconstruct disease pathways and then use knowledge-based or targeted-based methods to identify candidate drugs. The fishbone provides all relevant components of general drug-repositioning pipelines, and one can customize specific drug-repositioning pipelines according to the available knowledge and information of targeted drugs or diseases. *Abbreviations:* FDA, Food and Drug Administration; GWAS, genome-wide association study; HTS/HCS, high-throughput and/or high-content screening; PK/PD, pharmacokinetics/pharmacodynamics.

choose the feasible methods determined by the available information or prior knowledge about the disease or drug. Then, during the validation stages, one can decide which *in vitro* and *in vivo* validations are needed for further testing of the newly identified indication(s). Eventually, the repositioned drugs will enter into clinical trials to evaluate their efficacy and performance in patients. Here, we take advantage of already published pipelines to discuss how to select the best drug-repositioning methods for specific studies.

Example pipeline 1: neglected tropical diseases (malaria) + phenotypic screening methods + drug libraries + *in vivo* validations ⇒ astemizole

Orphan or rare diseases are those diseases that affect small numbers of people compared with the general population (<200,000 patient population per year in the USA) [39,40]. The orphan and/or rare diseases in developing regions of Africa, Asia and the Americas are also known as neglected tropical diseases. Most of these diseases receive less treatment and research funding, which results in little information about the disease mechanisms. According to the fishbone flowchart in Fig. 2, the methods involving little information would be useful to reposition drugs for these diseases. In one such study using phenotypic screening, Chong *et al.* screened 1,937 FDA-approved drugs and 750 drugs

that were either approved for use abroad or undergoing phase II clinical trials for inhibition of *Plasmodium falciparum* growth. They identified a drug, astemizole, as an antimalarial agent by testing the drug using two mouse models of malaria [41].

Example pipeline 2: distinct breast cancer metastases + knowledge-based methods + pathway- or network-based methods + drug libraries + *in vitro* and *in vivo* validations + clinical trials ⇒ sunitinib for brain metastasis

We remain unclear about the pathways or mechanisms responsible for breast cancer metastasis to brain, bone and lung, leading to challenges in repositioning drugs for these cancer subtypes. Knowledge-based methods alone cannot solve this repositioning issue because these methods provide only general or canonical breast cancer signaling pathways instead of those specific to various types of metastasis. In a recent study, knowledge-based and network-based methods were combined to reconstruct the signaling networks for these metastatic breast cancer subtypes so that drug repositioning for each type was feasible to implement [17]. The knowledge-based method in this study deployed newly discovered signaling network elements, called cancer signaling bridges [18], to identify general known signaling information for breast cancer, whereas the network-based method used a mathematical model to address the specific signaling networks

for subtypes of metastatic breast cancer. By checking the known targets in a recently developed drug database, DrugMap Central [42], 15, nine and two drug candidates were repositioned for brain, lung and bone metastases, respectively. For breast cancer brain metastasis, *in vitro* and *in vivo* validations were used to test the efficacies of these 15 drug candidates and two drugs were identified (sunitinib and dasatinib); the efficacy of sunitinib for breast cancer brain metastasis is now being tested in a phase II clinical trial (ClinicalTrials.gov ID: NCT00570908).

Example pipeline 3: treatment omics data (164 drug compounds) + disease omics data (100 diseases) + signature-based methods + in vivo validations \Rightarrow cimetidine for lung adenocarcinoma

To test a large number of diseases for a specific drug or a large number of drugs for a specific disease, it is difficult to unify the needed computational approaches because the available information for different diseases or drugs varies. For example, to use target-based methods to reposition drugs for 100 diseases, one would have to know the biomarkers or available pathways for each of these diseases. The knowledge needed for this type of drug repositioning might be unavailable or difficult to derive from the literature or available databases. However, one can derive the gene signatures for these diseases from the publicly available genomics data. Together with the drug signatures identified from treatment omics data, Sirota *et al.* considered signature-based methods to evaluate the drug–disease scores [37]. On the basis of the drug–disease score map, they validated cimetidine for lung adenocarcinoma using tumor xenograft experiments.

Example pipeline 4: treatment omics data (>1000 drug–dose pairs) + targeted mechanism-based methods + in vitro validations \Rightarrow repositioned drugs with targeted mechanisms

In example pipeline 3, there is an issue in the gene signatures for 164 drugs: the treatment gene signatures for the 100 diseases are only based on treatments on three to five cancer cell lines in the CMAP database. Thus, when applying these drug gene signatures to other diseases, it is hard to ensure that the disease–drug scores are not biased by the limited treatment information. Another issue encountered with signature-based methods is in evaluating the accuracy of the large number of predictions on disease–drug associations. By integrating the targeted mechanisms into the drug-repositioning process, new computational approaches for predicting the efficacies of repositioned drugs were tested [18]. It was confirmed that the analysis could accurately predict clinical responses to more than 90% of drugs approved by the FDA and more than 75% of experimental clinical drugs that were tested. The high accuracy of prediction ensures more favorable repositioning results for disease–drug associations. To keep the accuracies of treatment information, only drugs for three cancer cell lines were repositioned that were used for treatments in CMAP. Moreover, the importance of the identified targeted pathways was addressed by explaining the differences in treatment responses.

Concluding remarks

Drug-repositioning studies are dependent on the prior knowledge and available information from specific studies to select and

determine appropriate repositioning methods. Establishing accurate and efficient drug-repositioning pipelines for specific studies requires the prioritization of existing computational methods based on the available knowledge or the development of new computational methods. In this review, we described the available drug-repositioning methods according to the categorization of their integrated knowledge and information. We introduced the LLF to characterize existing repositioning methods and presented them in a generalized drug-repositioning pipeline (Fishbone flow-chart). These flowcharts are powerful tools to understand the existing drug-repositioning methods and customize them into new drug-repositioning pipelines for specific studies.

Still, many challenges remain for cost-effective drug-repositioning studies. Not every existing drug-repositioning study can be generalized to a new study, especially those including computational methods. One has to evaluate carefully the available drug-repositioning methods according to the prior knowledge and available information of the study of interest and determine which is the best for that study. Another issue for the available drug-repositioning studies for diseases with low knowledge and low complexity-of-mechanism is that they have relatively low success rates. The complexities and richness of information available to drug-repositioning studies largely determine their success rates; obviously, knowledge-based and signature-based methods are more likely to identify more successful repurposed drugs than are blinded search or screening methods. As an example, we mentioned drug repositioning for pediatric population in the ‘Knowledge-based drug-repositioning methods’ section. The existing studies did not consider the blinded search or screening methods for drug repositioning for pediatric diseases. Instead, the drug knowledge from *HLH*, PubMed and Google, was used in the drug repositioning targeting this special population [29], because knowledge-based methods consider the factor of patient variability in the drug-repositioning process.

In the era of precision medicine, it is important to delineate disease mechanisms, such as signaling pathways, or treatment mechanisms, such as off-targets and targeted pathways, to explain the mechanisms of action of drugs. This also leads to the application of drug repositioning to new indications for individual patients. Mechanism-based repositioning approaches are able to consider fully the heterogeneity and complexity of patients while reducing the inefficacy and toxicity caused by patient variability [43]. We would like to emphasize that drug-repositioning studies have to be solidly grounded on science to be successful. Toward better drug repositioning, the field needs better development of more in-depth mechanistic computational methods or models that can readily be customized into drug-repositioning pipelines that integrate computational and experimental methods seamlessly to ensure high success rates of repositioned drugs.

Acknowledgements

This research is funded by the National Institutes of Health (NIH) U54 CA149196, National Cancer Institute (NCI) CA121225, a John S. Dunn Research Foundation grant and a T.T. & W.F. Chao Foundation grant to S.T.C.W. The authors would like to thank Rebecca Danforth for proofreading the manuscript.

References

- 1 Ashburn, T.T. and Thor, K.B. (2004) Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* 3, 673–683
- 2 Novac, N. (2013) Challenges and opportunities of drug repositioning. *Trends Pharmacol. Sci.* 34, 267–272
- 3 Pfister, D.G. (2012) Off-label use of oncology drugs: the need for more data and then some. *J. Clin. Oncol.* 30, 584–586
- 4 Burstein, H.J. (2013) Off-label use of oncology drugs: too much, too little, or just right? *J. Natl. Compr. Canc. Netw.* 11, 505–506
- 5 Swinney, D.C. and Anthony, J. (2011) How were new medicines discovered? *Nat. Rev. Drug Discov.* 10, 507–519
- 6 Hurle, M.R. *et al.* (2013) Computational drug repositioning: from data to therapeutics. *Clin. Pharmacol. Ther.* 93, 335–341
- 7 Swamidass, S.J. (2011) Mining small-molecule screens to repurpose drugs. *Brief. Bioinform.* 12, 327–335
- 8 Crisman, T.J. *et al.* (2007) Understanding false positives in reporter gene assays: in silico chemogenomics approaches to prioritize cell-based HTS data. *J. Chem. Inf. Model.* 47, 1319–1327
- 9 Feng, B.Y. *et al.* (2007) A high-throughput screen for aggregation-based inhibition in a large compound library. *J. Med. Chem.* 50, 2385–2390
- 10 Yildirim, M.A. *et al.* (2007) Drug–target network. *Nat. Biotechnol.* 25, 1119–1126
- 11 Zhao, S. and Li, S. (2010) Network-based relating pharmacological and genomic spaces for drug target identification. *PLoS ONE* 5, e11764
- 12 Cheng, F. *et al.* (2012) Prediction of drug–target interactions and drug repositioning via network-based inference. *PLoS Comput. Biol.* 8, e1002503
- 13 Alaimo, S. *et al.* (2013) Drug–target interaction prediction through domain-tuned network-based inference. *Bioinformatics* 29 (16), 2004–2008
- 14 Kinnings, S.L. *et al.* (2009) Drug discovery using chemical systems biology: repositioning the safe medicine to treat multi-drug and extensively drug resistant tuberculosis. *PLoS Comput. Biol.* 5, e1000423
- 15 Ideker, T. *et al.* (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929–934
- 16 Segal, E. *et al.* (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.* 34, 166–176
- 17 Hong, Z. *et al.* (2013) Novel modeling of cancer cell signaling pathways enables systematic drug repositioning for distinct breast cancer metastases. *Cancer Res.* 73 (20), 6149–6163
- 18 Jin, G.X. *et al.* (2012) A novel method of transcriptional response analysis to facilitate drug repositioning for cancer therapy. *Cancer Res.* 72, 33–44
- 19 Jin, G.X. *et al.* (2011) An enhanced Petri-net model to predict synergistic effects of pairwise drug combinations from gene microarray data. *Bioinformatics* 27, I310–I316
- 20 Iskar, M. *et al.* (2013) Characterization of drug-induced transcriptional modules: towards drug repositioning and functional understanding. *Mol. Syst. Biol.* 9, 662
- 21 Stephen, R. *et al.* (2009) Off-label use of oncology drugs in a community oncology EMR database. *Pharmacoepidemiol. Drug Safety* 18, S62
- 22 Ekins, S. *et al.* (2007) In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Br. J. Pharmacol.* 152, 9–20
- 23 Doman, T.N. *et al.* (2002) Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J. Med. Chem.* 45, 2213–2221
- 24 Kolb, P. *et al.* (2009) Docking and chemoinformatic screens for new ligands and targets. *Curr. Opin. Biotechnol.* 20, 429–436
- 25 Yang, L. and Agarwal, P. (2011) Systematic drug repositioning based on clinical side-effects. *PLoS ONE* 6, e28025
- 26 Campillos, M. *et al.* (2008) Drug target identification using side-effect similarity. *Science* 321, 263–266
- 27 Bisgin, H. *et al.* (2012) Investigating drug repositioning opportunities in FDA drug labels through topic modeling. *BMC Bioinformatics* 13 (Suppl. 1), S6
- 28 An, S.M. *et al.* (2013) Stem cell signaling as a target for novel drug discovery: recent progress in the WNT and Hedgehog pathways. *Acta Pharmacol. Sin.* 34, 777–783
- 29 Blatt, J. and Corey, S.J. (2013) Drug repurposing in pediatrics and pediatric hematology oncology. *Drug Discov. Today* 18, 4–10
- 30 Haeblerle, H. *et al.* (2012) Identification of cell surface targets through meta-analysis of microarray data. *Neoplasia* 14, 666–669
- 31 Sanseau, P. *et al.* (2012) Use of genome-wide association studies for drug repositioning. *Nat. Biotechnol.* 30, 317–320
- 32 Lamb, J. *et al.* (2006) The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929–1935
- 33 Lamb, J. (2007) The connectivity map: using gene-expression profiling to identify new therapeutics and potential adverse drug effects. *Chem. Res. Toxicol.* 20, 2018
- 34 Qu, X.Y.A. and Rajpal, D.K. (2012) Applications of Connectivity Map in drug discovery and development. *Drug Discov. Today* 17, 1289–1298
- 35 Dudley, J.T. *et al.* (2011) Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Sci. Transl. Med.* 3, 96ra76
- 36 Lussier, Y.A. and Chen, J.L. (2011) The emergence of genome-based drug repositioning. *Sci. Transl. Med.* 3, 96ps35
- 37 Sirota, M. *et al.* (2011) Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci. Transl. Med.* 3, 96ra77
- 38 Barretina, J. *et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607
- 39 Liu, Z.C. *et al.* (2013) In silico drug repositioning: what we need to know. *Drug Discov. Today* 18, 110–115
- 40 Ekins, S. *et al.* (2011) In silico repositioning of approved drugs for rare and neglected diseases. *Drug Discov. Today* 16, 298–310
- 41 Chong, C.R. *et al.* (2006) A clinical drug library screen identifies astemizole as an antimalarial agent. *Nat. Chem. Biol.* 2, 415–416
- 42 Fu, C. *et al.* (2013) DrugMap Central: an on-line query and visualization tool to facilitate drug repositioning studies. *Bioinformatics* 29, 1834–1836
- 43 Li, Y.Y. and Jones, S.J. (2012) Drug repositioning for personalized medicine. *Genome Med.* 4, 27
- 44 Xia, X. *et al.* (2010) Image-based chemical screening identifies drug efflux inhibitors in lung cancer cells. *Cancer Res.* 70, 7723–7733
- 45 Kondoh, Y. and Osada, H. (2013) High-throughput screening identifies small molecule inhibitors of molecular chaperones. *Curr. Pharm. Des.* 19, 473–492
- 46 Carnero, A. (2006) High throughput screening in drug discovery. *Clin. Transl. Oncol.* 8, 482–490
- 47 Toledo-Sherman, L.M. and Chen, D. (2002) High-throughput virtual screening for drug discovery in parallel. *Curr. Opin. Drug Discov. Dev.* 5, 414–421
- 48 Iorio, F. *et al.* (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc. Natl. Acad. Sci. U. S. A.* 107, 14621–14626
- 49 Wishart, D.S. *et al.* (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 36 (Database issue), D901–D906
- 50 Zhu, F. *et al.* (2012) Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. *Nucleic Acids Res.* 40 (Database issue), D1128–D1136
- 51 Hewett, M. *et al.* (2002) PharmGKB: the Pharmacogenetics Knowledge Base. *Nucleic Acids Res.* 30, 163–165
- 52 Gunther, S. *et al.* (2008) SuperTarget and Matador: resources for exploring drug–target relationships. *Nucleic Acids Res.* 36 (Database issue), D919–D922
- 53 Kuhn, M. *et al.* (2012) STITCH 3 zooming in on protein–chemical interactions. *Nucleic Acids Res.* 40 (Database issue), D876–D880
- 54 Okuno, Y. *et al.* (2006) GLIDA: GPCR-ligand database for chemical genomic drug discovery. *Nucleic Acids Res.* 34 (Database issue), D673–D677
- 55 Roth, B.L. *et al.* (2000) The multiplicity of serotonin receptors: uselessly diverse molecules or an embarrassment of riches? *Neuroscientist* 6, 252–262
- 56 Nicola, G. *et al.* (2012) BindingDB: a protein–ligand database for drug discovery. *Protein Sci.* 21, 208
- 57 Kuhn, M. *et al.* (2010) A side effect resource to capture phenotypic effects of drugs. *Mol. Syst. Biol.* 6, 343
- 58 Schaefer, C.F. *et al.* (2009) PID: the Pathway Interaction Database. *Nucleic Acids Res.* 37, D674–D679
- 59 Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30
- 60 Cerami, E.G. *et al.* (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.* 39, D685–D690
- 61 Keshava Prasad, T.S. *et al.* (2009) Human Protein Reference Database: 2009 update. *Nucleic Acids Res.* 37 (Database issue), D767–D772
- 62 Chatr-Aryamontri, A. *et al.* (2013) The BioGRID interaction database: 2013 update. *Nucleic Acids Res.* 41 (Database issue), D816–D823
- 63 Franceschini, A. *et al.* (2013) STRING v9.1: protein–protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41 (Database issue), D808–D815
- 64 Pagel, P. *et al.* (2005) The MIPS mammalian protein–protein interaction database. *Bioinformatics* 21, 832–834
- 65 Kerrien, S. *et al.* (2012) The IntAct molecular interaction database in 2012. *Nucleic Acids Res.* 40 (Database issue), D841–D846
- 66 Xenarios, I. *et al.* (2002) DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* 30, 303–305
- 67 McKusick, V.A. (2007) Mendelian Inheritance in Man and its online version OMIM. *Am. J. Hum. Genet.* 80, 588–604