

REVIEW

Open Access



A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions

Tamer N. Jarada¹, Jon G. Rokne¹ and Reda Alhajj^{1,2*}

Abstract

Drug repositioning is the process of identifying novel therapeutic potentials for **existing drugs** and discovering therapies for **untreated diseases**. Drug repositioning, therefore, plays an important role in optimizing the pre-clinical process of developing novel drugs by **saving time** and **cost** compared to the traditional de novo drug discovery processes. Since drug repositioning relies on **data** for existing drugs and diseases the enormous growth of publicly available large-scale biological, biomedical, and electronic health-related data along with the **high-performance computing capabilities** have accelerated the development of computational drug repositioning approaches. Multidisciplinary researchers and scientists have carried out numerous attempts, with different degrees of efficiency and success, to computationally study the potential of repositioning drugs to identify alternative drug indications. **This study** reviews **recent advancements in the field of computational drug repositioning**. **First**, we highlight different drug repositioning **strategies** and provide an overview of frequently used resources. Second, we summarize computational **approaches** that are extensively used in drug repositioning studies. Third, we present different computing and experimental **models** to validate computational methods. Fourth, we address prospective **opportunities**, including a few target areas. Finally, we discuss **challenges and limitations** encountered in computational drug repositioning and conclude with an outline of further research directions.

Keywords: Computational drug repositioning, Drug repositioning strategies, Data mining, Machine learning, Network analysis

Introduction

Drug repositioning has attracted considerable attention due to the potential for discovering new uses for existing drugs and for developing new drugs in pharmaceutical research and industry, due to its efficiency in **saving time and cost** over the traditional de novo drug development approaches [1, 2]. Drug repositioning is also known as **drug repurposing**, **drug reprofiling**, **drug redirecting**, **drug retasking**, and **therapeutic switching**.

At the present time, the drug repositioning approach has taken on a new urgency due to the worldwide Coronavirus disease (COVID-19) epidemic, which originated in China. The rapid onset of the epidemic and its potential for infecting large numbers of people (the reproduction number R_0 is greater than 1 in the absence of social distancing and other countermeasures) has led to an urgency for developing new drugs for dealing with this disease. The status of drug and vaccine development for COVID-19 is, therefore, rapidly changing and almost every day, there is an update of the state of the developmental effort [3]. Because of the urgency in developing new drugs and treatments **traditional drug development** is **too slow** and the faster repositioning approach has,

*Correspondence: rsalhajj@gmail.com

² Department of Computer Engineering, Istanbul Medipol University, Istanbul, Turkey

Full list of author information is available at the end of the article



© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

therefore, attracted great interest due to its potential for finding drugs that could be used to combat the effects of the virus infection [4–6].

Generally speaking, **traditional drug** repositioning studies focus on **uncovering drug effect** and mode of action (MoA) similarities [7], revealing novel drug indications by screening the current pharmacopeia against new targets [8], investigating prevalent characteristics between drug compounds such as chemical structures and side effects [9], or discovering the relationships between drugs and diseases [10].

The **explosive growth** of large-scale biomedical and electronic health-related data such as microarray gene expression signatures, pharmaceutical databases, and online health communities that are publicly available along with high-performance computing has **empowered** the development of computational drug repositioning approaches that generally include data mining, machine learning, and network analysis [11]. Investigating the relationship between different biomedical entities forms a vital part of most recent studies in the drug repositioning field. These biomedical entities include drugs, diseases, genes, and adverse drug reactions (ADRs), etc.

In this survey paper, we **detail recent trends** related to computational drug repositioning from various points of view. **First**, we recap different drug repositioning **strategies** and the corresponding data sources that are widely used. **Second**, we identify the computational approaches that are frequently used in drug repositioning studies. **Third**, we address computing and experimental validation models in computational drug repositioning research. Finally, we outline **prospective opportunities**, including a few target areas, and conclude with a summary of the outstanding complications and issues in drug repositioning. Figure 1 summarizes the **workflow** of computational drug repositioning studies, which, as shown, mainly comprise four main steps.

Drug repositioning strategies

There are generally **two fundamental drug repositioning principles**. First, drugs related to a specific disease may also work on other diseases due to the **interdependence** between these different diseases. Second, a drug can be associated with **various targets and pathways** since drugs are confounding by nature [1]. Hence, drug repositioning studies could be classified into two categories based on



where the findings **originate** from: (i) drug-based strategies where discovery originates from knowledge related to **drugs** and (ii) disease-based strategies where discovery originates from knowledge related to **diseases**.

Drug-based strategies

Drug-based strategies depend on **data related to drugs** such as chemical, molecule, biomedical, pharmaceutical, and genomics information as the foundation for predicting therapeutic potentials and novel indications for existing drugs. Drug-based strategies are used where there is either **substantial** drug-related data accessible or **significant motivation** for studying how pharmacological characteristics can contribute to drug repositioning [10]. The vast majority of studies under this category share the hypothesis that if **two drugs**, R_1 and R_2 have similar profile and mode of action, and drug R_1 is used to cure disease D , then drug R_2 can be considered as a strong candidate for treating disease D . The two main strategies that represent this category are the **genome strategy** [7, 12–29], and the **chemical structure and molecule information strategy** [9, 30–39].

Genome strategy

A genome is a term that is used to describe **all genes concerning a specific organism**. In other words, biological data stored in a genome is represented by its DNA and is divided into separate units called genes [40]. **The introduction of the human genome sequencing project** [41] marked a **turning point** in the acquisition of knowledge at a molecular level about how living organisms function and revolutionized drug repositioning studies. More specifically, by finding the **genes or proteins** that perform a significant role in drugs and diseases' molecular actions, the human genome sequencing project initiative has allowed a better understanding of drugs and diseases' mode of actions. These genes and proteins have become enticing targets for governments and the pharmaceutical industry, which led to having this field of science as **one of the most intensely studied research areas at research labs around the world**.

The enormous volumes of publicly available genomic and transcriptomic **data** generated for disease samples, as well as clinical databases, provide a **unique opportunity** for understanding the disease and drug mechanisms of actions and discovering new uses for existing drugs. However, due to the tremendous complexity of biological systems, the comprehensive understanding of such systems is **still incomplete**. As a result, the research into a molecular explanation of biological systems is still pursued extensively.

It is noteworthy that the **microarray gene expression profile** is the **most widely used** transcriptomic

profile among the genetic profile methods that have been explored for drug repositioning. Unlike most traditional molecular biology tools that allow the studying of a single gene or a small set of genes, microarray gene expression profiling captures the **dynamic properties** of a cell and measures **all the transcriptional activity of thousands of genes at the same time**, leading to a **revolution** in the molecular biology research field. The application of microarray gene expression profiling has, therefore, received considerable attention for its vital role in understanding how genes act at the same time and under the same conditions.

Computational drug repositioning studies using gene regulatory data **presume** that drugs target the **same proteins** with comparable gene expression profiles. This understanding has led to the discovery of a tremendous number of novel and unexpected functional gene interactions, the detection of novel disease subtypes, and the identification of underlying mechanisms of disease or drug responses [42–45].

The Connectivity Map (CMap) project and its extended Library of Integrated Network-Based Cellular Signatures (LINCS) are considered to be a key concept behind various well-recognized drug repurposing studies. The Connectivity Map can be defined as a combination of genome-wide transcriptional expression data that helps in revealing functional connections between drugs, genes, and diseases [12]. The extended project of the CMap produced large-scale gene expression profiles from human cancer cells that were targeted by various drug compounds in different environments [24, 28]. Lamb et al. [12] used microarray gene expression data to build a connectivity map that is used to discover relationships between the list of genes related to a specific disease or drug, called a query signature, and a set of gene expression profiles called the reference database. Expression profiles that are highly positively correlated to the query signature are considered to have a very similar mode-of-action to the query signature. Expression profiles that are highly **negatively correlated** to the query signature are considered for **further treatment investigation**.

Iorio et al. [7] developed an automatic approach that takes advantage of the **similarity in gene expression** profiles in order to discover drugs that have a shared effect and mode of action. Initially, the authors built a drug network where nodes represent drugs, and edges indicate similarities between a pair of drugs. Then, they used graph techniques for detecting drug communities. Drugs in each of these communities have a similar mode of action. Hu and Agarwal [13] conducted an extensive analysis of human drug perturbation and disease gene expression based on a negative correlation to construct a disease-drug network for predicting new

applications for already approved drugs. Sirota et al. [14] performed a comprehensive systematic analysis of gene expression profiles for different diseases and drugs that led to discover new drug repositioning candidates.

CMap has gained considerable attention in drug repositioning since its introduction. It has shown promises in uncovering paths for drug repositioning for a variant group of diseases by identifying and suggesting new indications for existing drugs. Numerous researches have been conducted by integrating CMap data sources with other functional genomics databases such as the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) [46] to discover associations among genes, drugs, and diseases.

Jiang et al. [16] used CMap data sources to determine relationships between small molecules and miRNAs in human cancers in order to come up with therapeutic potentials and new indications for existing drugs. Jahchan et al. [21] used gene expression profiles to identify drug molecules for the treatment of small-cell lung cancer, which has not had effective treatments. Huang et al. [27] introduced a new connectivity map called (DMAP) that overcome the CMap data limitation by proposing a drug-protein connectivity map. DMAP consists of directed drug-to-protein effects and their scores. All previously-observed relationships between the associated drug and protein in various data sources were used to calculate effect scores from all database entries between the drug and protein as well as the confidence level of the quality of these calculated effect scores.

The massive amounts of publicly available gene expression profiles datasets have encouraged researchers to consider the guilt by association [47] concept to investigate drug-drug and drug-disease associations for identifying therapeutic indications for existing drugs. Iorio et al. [20] adopted the guilt by association concept to compare different drugs in order to identify any transcriptional responses similarity assuming that these drugs would share a similar mode-of-action (MoA).

Recently, microRNAs (miRNAs) have received considerable attention in biological and biomedical studies for their roles in regulating different types of cell activities [25, 26]. Hence, miRNAs have become key players in identifying drug repositioning therapeutic targets since miRNAs are vital for homeostasis of cells and active in many disease stages [48].

Jiang et al. [16] also used miRNAs along with small molecules, as potential drugs, to build networks for different types of cancer in order to identify small molecule-miRNA associations for drug repositioning based on miRNA regulations and transcriptional responses.

There have been several attempts to build public repositories aiming to elevate the development of small molecule-based miRNA therapeutics.

Liu et al. [15] manually curated scientific literature looking for how small molecules affected miRNA expressions and developed a database (SM2miR) in order to capture existing small molecule-miRNA associations aiming. Li et al. [22] manually retrieved experimentally supported miRNA-disease associations from scientific articles and built the Human MicroRNA Disease Database (HMDD v2.0) to facilitate data exploration. Huang et al. [29] introduced (HMDD v3.0) by adding a significant number of miRNA-disease associations to (HMDD v2.0) and improving the accuracy of these associations based on literature-based evidence. Rukov et al. [19] established a database (PharmacomiR) to identify miRNA-gene-drug triplet set associations by combining data on miRNA targeting and protein-drug interactions.

Meanwhile, as most of the genome-based studies have focused on using gene expression profiles as a valuable source for discovering therapeutic indications for existing drugs, some studies have focused on other different types of genomic profiles such as genome-wide association studies (GWAS). GWAS follows the phenotype-to-genotype concept, where it starts with a specific genotype and checks for associations with genetic variants across the genome [49].

Sanseau et al. [17] filtered published GWASs catalog of disease-associated genes to come up with a list of GWAS-associated genes that were then evaluated against targets of drugs under clinical and preclinical investigation for potential novel indications and drug repositioning opportunities. Okada et al. [23] developed a new in silico approach by conducting a multi-stage GWAS analysis of targeted disease patients to uncover a set of unknown risk loci related to the targeted disease and further identify a set of biological candidate genes that are targeted by already approved drugs. Also, a collection of existing drugs approved for other indications was identified and linked to the studied disease for potential drug repositioning chances.

Garnett et al. [18] conducted a large-scale multivariate analysis of genetic of cancer cell lines and drugs in pharmaceutical pipeline projects to unveil new biomarkers of sensitivity and resistance to cancer therapeutics. To a certain degree, mutated genes demonstrate the molecular activity of drugs and can be considered as drug biomarkers during the drug repositioning process. A few mutated cancer cell line genes were found to be associated with drug sensitivity, which may serve as potential biomarkers for drug repositioning.

Chemical structure and molecule information strategy

As the genome drug-based strategy assumes drugs share common indications because of having similar profiles, chemical structure and molecule information of drugs are also considered to be **worthwhile sources** of pointing towards any transcriptional responses similarity between drugs for repositioning opportunities as these drugs usually affect genes, proteins, and other biological entities in similar forms [33, 34, 36]. **Chemical structure similarity** can be **measured** in various ways, such as two-dimensional (2-D) topological fingerprints and three-dimensional (3-D) conformational fingerprints [50].

Keiser et al. [9] proposed a systematic chemical structure similarity approach to screen compounds of existing and in-process drugs against hundreds of ligands that bind protein targets. Chemical structure similarity between drug compounds and ligand targets revealed thousands of **unforeseen associations**, some of which were tested and confirmed experimentally. The proposed approach can explain some of the side effects of existing drugs, and may also contribute to the identification of new repositioning applications for existing drugs.

Swamidass [33] suggested using chemical structures to determine which drug targets would modulate disease-relevant phenotype. Such a tactic would give indications of how other drugs, with **similar chemical structures**, modulate disease-relevant phenotype and hence treat the disease.

Most frequently, chemical structure similarity is **incorporated** with molecular activity and other biological information to identify new associations and potential off-target effects for approved and investigational drugs. Yamanishi et al. [30] developed a **supervised learning model** for a bipartite graph to identify possible drug-target interactions. The authors integrated drug chemical structure information, protein-protein interaction network, and drug-target interaction network to predict therapeutic potentials and unveil drug repositioning applications. Kinnings et al. [31] used drug chemical interactions under different environment variables to build a drug similarity **network** where drugs are defined as nodes, and an edge is drawn between two drugs when they have a high similarity score. Then, the authors analyzed the drug network to detect different drug communities and investigated drugs within each community for potential drug repositioning. Bleakley et al. [32] proposed a **statistical method** to predict drug-target interactions using chemical structure information and genomic sequence information. The authors built a **supervised learning bipartite graph model** based on independent local supervised learning problems to predict target proteins of a given drug and then to predict drugs targeting a given protein.

Li and Lu [35] **combined** drug chemical structure information with drug targets and interactions information to develop a novel bipartite graph model to calculate drug pairwise similarity. The results significantly enriched both the biomedical literature and clinical trials when compared to a control group of drug uses. The developed approach **outperformed** other approaches that **only** use drug target profiles and captured the implicit information between drug targets. Wang et al. [37] integrated drug chemical structure data along with molecular activity and drug side effect data to check for drug similarity and predict drug-diseases interactions.

Tan et al. [38] came up with a new form of “expression profile” based on 3-D drug chemical structure information, gene semantic similarity information, and drug-target interaction networks. The authors gave consensus response scores (CRS) between each drug and protein and used the absolute value of correlation coefficient between every two drugs as their degree similarity to build a drug similarity network (DNS), which led to identifying new drug indications. The proposed approach took into consideration the 3-D drug chemical structure information to overcome the instability of gene expression profiles acquired from different experiments due to experimental conditions such as environment and patient age.

While most of the available drug repositioning approaches that use chemical structure strategy focus on predicting direct or indirect drug interactions on a **small scale**, Zheng et al. [39] conducted a **large scale** study on drug-target relationships and introduced a new algorithm called Weighted **Ensemble** Similarity (WES). The authors identified the key ligand structural features of a protein as a set named ensemble. Rather than comparing two compounds to determine their similarity, each compound was compared to ensembles in order to calculate the overall ensemble similarity instead of using a single ligand similarity because ensembles usually represent smaller chemical structure features. The whole ensemble similarity scores were normalized and used to predict direct interactions of drugs and targets.

A **further** molecular repurposing strategy is provided by the geometry of a drug molecule as expressed by the **proteomic signature**. That is, repurposing candidates are identified by their proteomic signature similarities. This approach is exploited by Mangione and Samudrala in the paper [51] which describes a simulation system for drug molecule docking interactions applied to the repurposing of drugs. The **shapes** of molecules are in general determined using **X-ray diffraction** techniques and recent advances in the type of molecular docking required for the repurposing are discussed by Yan et al. [52] where the

general limitations of the approacher for molecular shape determinations are also outlined.

Disease-based strategies

Disease-based strategies depend on data related to diseases such as phenotypic traits information, side effects, and indications information as the foundation to predict therapeutic potentials and novel indications for existing drugs. Disease-based strategies are used when there is either **insufficient** drug-related data available or when the motivation in studying how pharmacological characteristics can contribute to drug repositioning effort concentrated on a particular disease [1]. The studies under this category share the hypothesis that if **two diseases**, D_1 and D_2 have a similar profile and indications, and drug R is used to cure disease D_1 , then drug R can be considered as a strong candidate for curing disease D_2 . The primary strategy that represents this category is the **phenome strategy** [10, 53–60].

Phenome strategy

The phenome is described as the **overall set of phenotypic traits information**, and it has arisen as a **new strategy** to connect drugs with clinical effects for drug repositioning due to the argument that it represents the **unwitting effects** of a drug and defines the physiological consequences of its biological activities. Moreover, the phenotypic expression of a drug's side effect may be closely related to the phenotypic expression of a disease, which suggests that both the **drug** and the **disease** may share **similar underlying pathways** [10].

Clinical side effects and **unexpected activities** derived from off-targets have been shown to have the ability of profiling human phenotypic traits related to drugs and may ultimately help unveil potential therapeutic uses for these drugs. Campillos et al. [53] proposed a side-effect similarity measure based on the strong correlation between targeted portion binding profiles and side-effect similarity and experimentally verified that side-effect similarity indicates novel therapeutic uses for existing drugs. Yang and Agarwal [54] demonstrated that clinical side effects could be used to build a phenotypic profile of drugs and identify potential new disease indications. A side effects-drugs relationship dataset was integrated with a drug-disease relationship dataset to derive side effects-disease relationships. Then, side effects were used as features for building a prediction model for disease indications.

Ye et al. [57] constructed a drug–drug similarity network based on clinical side effects assuming that drugs with similar side effects may share similar therapeutic indications. Novel drug indications were identified in addition to already known indications. Bisgin et al. [58]

used side effect information to build a model for predicting new therapeutic indications for existing drugs. It is worth mentioning that a profound background in molecular mechanisms is required for using phenotypic traits information in predicting new drug indications. While most of the phenotypic based research is **leveraging data from clinical studies and drug labels**, Nugent et al. [59] used side-effect data mined from **social media** to identify novel therapeutic indications in addition to previously identified indications.

Eventually, phenotypic traits information can be **integrated** with **other data sources** such as genome for therapeutic potentials and novel drug indications. Hoehndorf et al. [55] used phenotypic similarity to identify genotype-disease associations which were later combined with genotype-disease association data to predict novel drug-disease associations. Such a model can be considered as an introduction to an integrated system to identify drug-disease associations for diseases with an unknown molecular basis. Gottlieb et al. [56] developed a model using various drug–drug similarity measures, including phenome-based similarity, to predict novel drug–drug interactions and severity level associated with each of these interactions. Sridhar et al. [60] integrated different drug–drug similarity measures, including phenotypic similarity with already known drug–drug interactions to unveil drug–drug interactions, including several novel interactions.

Data resources

The advanced technologies nowadays have produced a massive amount of data (e.g., gene expression, drug-disease associations, drug chemical structure profiles, drug targeted proteins, phenotypic traits), which has supported the enormous effort that has been devoted towards developing fascinating drug repositioning strategies. A list of the widely used data resources and their drug repositioning strategies classification is summarized in Table 1.

Computational drug repositioning approaches

A significant **challenge** in drug repositioning is to distinguish between the molecular targets of a drug and the hundreds to thousands of additional gene products that respond indirectly to changes in the activity of the targets. **Unfortunately**, **classical statistical approaches** are **ineffective** for detecting the molecular targets of a drug among the vast amount of genes. Moreover, **conventional statistical methods** use **small datasets** and biological networks that are coming from experiments on different platforms and environments, which might lead to inconsistent findings reported by some studies. Also, when the data used to conduct such studies is limited, or

Table 1 Data resources widely used in drug repositioning research

Drug repositioning strategy	Data resource	Description (as of January 2020)
Genome	ArrayExpress [61]	A repository of high-throughput functional genomics experiments data
	Cancer Cell Line Encyclopedia (CCLE) [62]	A combination of gene expression, chromosomal copy number, and massively parallel sequencing data from human cancer cell lines
	Database for Annotation, Visualization and Integrated Discovery (DAVID) [63]	A comprehensive set of functional annotation tools
	Drug versus Disease (DvD) [64]	A platform for comparing drug and disease gene expression profiles retrieved from publicly available microarray data resources
	Expression Atlas [65]	A tool to explore gene expression data across species and different biological conditions
	Gene Expression Omnibus (GEO) [66]	A repository of gene expression profiles
	Gene Set Enrichment Analysis (GSEA) [67]	A tool for interpreting gene expression data
	Gene Signature Database (GeneSigDB) [68]	A repository of gene signatures data reported in the literature
	Gene Ontology (GO) [69]	A repository of functional genomics
	International Cancer Genome Consortium [70]	A repository of genomic data for many cancer types
	Kyoto Encyclopedia of Genes and Genomes (KEGG) [71]	A repository of genome sequencing and high-throughput functional genomics experiments molecular datasets
	Library of Integrated Network-based Cellular Signatures (LINCS) [72]	A catalog of gene expression data and how human cells respond to different genetic and environmental conditions
	Molecular Signature Database (MsigDB) [73]	A repository of annotated gene sets and expression profiles
	The Cancer Genome Atlas (TCGA) [74]	A repository of genomic data for more than 30 cancer types
	The Connectivity Map (CMap) [75]	A collection of genome-wide transcriptional expression data
	The Universal Protein Resource (UniProt) [76]	A repository of protein sequence and functional information
Phenome	ClinicalTrials.gov [77]	A repository of publicly and privately funded clinical studies from around the world
	Side Effect Resource (SIDER) [78]	A repository of adverse drug reactions related to drugs. Information includes drug side effects and side effects classifications
Chemical structure	ChEMBL [79]	A repository of drug structural information, such as 3D structures, and abstracted biological activities, extracted from the scientific literature
	Chemicalize [80]	A repository of chemical structure information
	ChemSpider [81]	A repository of chemical structure information
	DrugBank [82]	A repository of drug-related information, such as chemical structure for drugs
	Drug Central [83]	A repository of drug-related information such as chemical structure and biological activities
	PubChem [84]	A repository of chemical substances and their biological activities
	Protein Data Bank (PDB) [85]	A repository of 3D shape of proteins and nucleic acids information
	SWEETLEAD [36]	A repository of 2D and 3D drug chemical structure information, including approved and illegal drugs
	The NCGC Pharmaceutical Collection (NPC) [86]	A collection of chemical structure information related to approved and investigational drugs
	Therapeutic Target Database (TTD) [87]	A repository of drug-related information such as 3D structure, therapeutic class, and clinical development status

Table 1 (continued)

Drug repositioning strategy	Data resource	Description (as of January 2020)
Phenome/genome	repoDB[88]	A repository of approved and failed drug-disease associations
	Online Mendelian Inheritance in Man (OMIM)[89]	A repository of human genes and genetic phenotype information
	The Pharmacogenetics and Pharmacogenomics Knowledge Base (PharmGKB) [90]	A repository of drug-related information such as drug labels, drug-gene associations, and genotype-phenotype relationships
Phenome/chemical structure	Drugs@FDA Database[91]	A repository of FDA approved drugs and related information

the biological network is small, the proposed approaches might recover only partial knowledge of a living system. As a result, some approaches that claim inferences and discoveries may not be replicated.

The amount of publicly available large-scale biomedical and pharmaceutical data is growing exponentially, and computational drug repositioning approaches using data mining, machine learning, and network analysis become ever more critical when it comes to systematic drug repositioning due to the **ability to overcome classical statistical approaches limitations and unreliable conclusions**.

The drug repositioning field can benefit from **new computational methods** in detecting relationships among different types of biological entities such as genes, portions, diseases, and drugs and identify therapeutic potentials and novel indications for existing drugs. Such findings would help to treat cancer and other incurable illnesses, which eventually require the necessary and sufficient data to undertake the intended research. Table 2 presents an overview of computational drug repositioning studies, the adopted strategies, computational approaches, main techniques, data sources, key findings, and evaluation metrics.

Data mining

The tremendous amount of genes, drugs, and diseases related information stored in databases in addition to the vibrant literature grown by the rapid increase in the number of the biological, biomedical, and pharmaceutical studies have led to the need for data mining where researchers can discover a tremendous amount of information hidden in the literature [92, 93]. The majority of studies adopting the data mining approach use **text mining techniques**.

Text mining

Text mining as applied to the drug repositioning problem is typically used to find data related to a **particular gene, disease**, or **drug specified** and then classify the relevant

entities or knowledge from the retrieved data based on the co-occurrence between the relevant entities or by using **natural language processing**. For instance, if drug *R* is connected with gene *G*, and gene *G* is related to disease *D*, then drug *R* may have a new connection with disease *D*. Generally, text mining includes **four steps** which are: (1) Information retrieval (IR), (2) Entity recognition (NER), (3) Information extraction (IE), and (4) Knowledge discovery (KD) [94].

Cheng et al. [95] developed a web-based text mining system for extracting relationships between different biological terms such as diseases, tissues, genes, proteins, and drugs by using a variety of text mining and information retrieval techniques over a massive set of existing biological databases in order to identify, highlight and rank informative abstracts, paragraphs or sentences. Li and Lu [96] introduced a model to identify clinical pharmacogenomics (PGx) gene-drug-disease relationships from clinical trial data. The authors determined **text of interest in clinical trial records** retrieved from ClinicalTrials.gov [77] and used a dictionary to identify PGx concepts. Then, they considered the co-occurrence of PGx concepts in each clinical trial to define gene-drug-disease relationships. Finally, they indexed each clinical trial using its identified gene-drug-disease relationships. Therefore, given a PGx gene, the introduced model can identify related diseases and drugs within the corresponding clinical trials. Likewise, given a pair of PGx gene-drug or gene-disease, the introduced model can return clinical trials in which the PGx pair is or has been studied.

Leaman et al. [97] built a tool for recognizing disease entities mentioned in literature. The authors used disease corpus from the National Center for Biotechnology Information (NCBI)[46] and the MEDIC vocabulary [98] to single out diseases mentioned in PubMed abstracts and subsequently handle abbreviations. Afterward, they used pairwise learning to rank, which has proven to be successful in information retrieval, for normalizing

Table 2 An overview of computational drug repositioning studies, their adopted strategies, computational approaches, main techniques, data sources, key findings, and evaluation metrics

Study	Strateg(ies)		Computational approach(es)			Main technique(s)	Data source(s)	Evaluation criteria	Key finding(s)
	Genome	Chemical structures	Phenome	Data mining	Machine learning				
Hu et al.[123]	✓	✓		✓		CNN	KEGG and DrugBank	CV, ROC and Comp(Acc, Sen, F1, and AUC)	R-TIs
Han et al.[99]	✓		✓	✓		TM, GCN, and MF	OMIM	CV and Comp(MAP, AUC, and NDCG)	D-GAs
Zeng et al. [106]	✓		✓	✓		MM-AE and V-AE	DrugBank, repoDB, and ClinicalTrials.gov	CV, AUC, AUC-PR, and Comp(Rec-TK)	R-DIs
Yang et al. [127]	✓		✓		✓	ARM and LP	PharmGKB, SIDER, and MedHelp [128]	CV, AUC, AUC-PR, and Comp(Sen, Spc, and F1)	R-DIs
Ozsoy et al.[119]	✓	✓	✓	✓		SM, PDo, and CF	PubChem, UniProt, SIDER, and TIRs [35]	CV and Comp(Prec, Rec, F1, Spc, and AUC)	R-DIs
Segler et al. [124]		✓		✓		RNN	SMILES	EOR	R-DIs
Altae-Tran et al. [122]		✓	✓		✓	CCN and LSTM	SIDER and Tox21 [129]	Acc and AUC	R-DIs
Papanikolaou et al. [105]		✓			✓	NER	DrugBank	—	R-RAs
Brown et al. [104]		✓		✓		SM and Clust	MEDLINE and DrugBank	CSs	R-DIs
Nugent et al. [59]			✓	✓		SCM	Twitter	Comp(Acc)	R-DIs
Aliper et al. [121]	✓			✓		DNN	LINCS	CV, F1, NNCM, Spr, and Comp(Acc)	R-RAs
Lim et al. [118]	✓	✓		✓		SM and CF	ZINC [130]	CV and Comp(Acc, and Rec)	R-TIs
Sridhar et al. [60]	✓	✓	✓	✓		SM and PSL	DrugBank	Comp(AUC, AUC-PR, F1) and CSs	R-RIs
Zheng et al. [39]		✓		✓	✓	SM	BindingDB [131]	CV, Acc, Spc, Sen, AUC, Prec, F1, Comp(AUC, Sen, and Spc), and CSs	R-TIs
Rastegar-Mojarad et al. [103]	✓			✓	✓	TM and SM	SemMedDB [132], UMLS [133], CTD [134], and Medline	Comp(CoS)	R-DIs
Rakshit et al. [126]	✓			✓		TM and Genotator [135], PolySearch [95], Pescador [136], and DrugBank	Comp(Acc)	R-TIs	
Huang et al. [27]	✓			✓		SM and BGL	CMap	CV, AUC, Comp(AUC), and CSs	R-TIs
Zhu et al. [108]	✓	✓		✓		SM and WOL	PharmGKB	CSs	R-DIs

Table 2 (continued)

Study	Strateg(ies)		Computational approach(es)				Main technique(s)	Data source(s)	Evaluation criteria	Key finding(s)
	Genome	Chemical structures	Phenome	Data mining	Machine learning	Network analysis				
Tari et al. [102]	✓	✓		✓			TM and LFRs	Medline, GO, UniProt, NCBI, CancerQuest [137], and DrugBank	Comp(Rec) and CSs	R-DIs
Okada et al. [23]	✓	✓		✓			GWAS	GWASs	P and CSs	R-DIs
Yang et al. [117]	✓	✓			✓		CI-PMF	DrugBank, BioCarta [138], and CTD [134]	Comp(AUC and Prec) and CSs	R-DIs
Zhang et al. [116]	✓	✓	✓		✓		SM and BCD	DrugBank, NDF-RT [139], and HPRD [140]	CV, Comp(AUC), and CSs	R-DIs
Bisgin et al. [58]			✓		✓		SM and LDA	SIDER	Acc and CSs	R-TIs
Tan et al. [38]	✓	✓				✓	SM and Clust	PubChem and DrugBank	Comp(Acc)	R-TIs
Ye et al. [57]			✓			✓	SM	SIDER	CV, Acc, and CSs	R-TIs
Menden et al. [114]	✓	✓			✓		FF-PNN and RF	CCL [18]	CV, Acc, R_p , R^2 , RMSE, and CSs	R-TIs
Napolitano et al. [115]	✓	✓			✓		SVM and CF	CMap, NCBI, DrugBank, and ATC [141]	CV, Acc, Comp(AUC)	R-DIs
Wang et al. [37]	✓	✓	✓		✓		SM, KF, and SVM	PubChem, KEGG, BRENDA [142], Super-Target [143], DrugBank, and SIDER	CV, Comp(AUC, Acc, Sens, Spc, Prec, and F1), \$ CSs	R-DIs
Gottlieb et al. [56]	✓	✓	✓	✓			SM	DrugBank, Drugs.com, and SIDER	CV, Acc, AUC, and CSs	R-TIs
Wu et al. [125]	✓	✓	✓	✓		✓	SM and Clust	KEGG	CV, Comp(Acc), and CSs	R-DIs
Chen et al. [107]	✓		✓	✓	✓		SM and LP	Chem2Bio2RDF [144], DrugBank, and Mator [143]	Acc, AUC, Copm(Acc and AUC), and CSs	R-TIs
Li and Lu [96]	✓			✓			TM	ClinicalTrials.gov and PharmGKB	Acc, Comp(Acc), and CSs	PGx
Li and Lu [35]	✓	✓	✓	✓		✓	SM and BGL	DrugBank, NDF-RT [139], HPRD [140], and ClinicalTrials.gov	CV (Sen, Spc, and AUC), Comp(Spc, Sen, and AUC), and CSs	R-TIs
Hoehndorf et al. [55]	✓		✓		✓		SM	PhenomeNET [145] and PharmGKB	AUC and CSs	R-TIs
Gottlieb et al. [113]	✓	✓	✓	✓			SM and LR	DrugBank, SMILES, Mator [143], DCDB [146], KEGG, DailyMed [147], SIDER, GO, UniPort, and ClinicalTrials.gov	CV, Acc, P, AUC, and Comp(Acc and AUC)	R-DIs

mentioned text and identifying MEDIC concepts for the disease entities mentioned in PubMed abstracts.

Text mining has also been widely used successfully for discovering relationships between genes, diseases, and drug [99], investigating gene-gene interactions [100], and building a heterogeneous network of genes, diseases, and drugs [27]. Li et al. [101] proposed a new approach that integrates literature text-mining data with protein interaction networks to build a drug-protein connectivity map for a specific disease. The authors used Alzheimer's disease (AD) as a case study and showed that their approach outperformed curated drug-target databases and conventional information retrieval systems and also suggested two existing drugs as candidate drugs for AD treatment.

Unlike common text mining approaches where biological networks are built based on the co-occurrence of biological entities, Tari et al. [102] introduced a novel approach that considered interaction types, interaction type directions, and drug mechanism representation. The authors used text mining to obtain data from publicly available sources that then used to produce a set of logical facts. Then, the set of logical facts was used along with logical rules that represent drug mechanism properties to build an automated reasoning model for identifying therapeutic potentials and novel indications for existing drugs. Rastegar-Mojarad et al. [103] used text mined data in order to identify drug-gene and gene-disease semantic predictions, which then were utilized to compile a list of potential drug-disease pairs. Finally, the authors ranked the drug-disease pairs using the predicates between drug-gene and gene-disease pairs, evaluated their model against two different datasets, and concluded that the combination of drug-gene and gene-disease predicates could eventually be used to highlight the drugs in the top-ranked drug-disease pairs as drug repositioning candidates.

Brown et al. [104] proposed a web-based text mining system for drug repositioning. The authors used the number of shared indications across drug-drug pairs to disclose similarity among these drug-drug pairs and then clustered drugs based on their similarity, which revealed both known and novel drug indications. Papanikolaou et al. [105] applied text mining on the DrugBank database's text attributes to identify drug-drug associations. The authors used Name Entity Recognition (NER) to identify biological entities (proteins, genes, diseases, etc.) in the DrugBank's description, indication, pharmacodynamics, and mode-of-action text fields. Then, they used an algorithm to eliminate any insignificant terms and created a binary vector representing each DrugBank record. Finally, they clustered DrugBank records using several clustering algorithms and similarity measures. Such an approach can

facilitate the retrieval of novel drug-drug associations, which may significantly contribute to new drug repositioning applications.

Recently, Zeng et al. [106] introduced a deep-learning approach where they retrieved data from various publicly available sources to build ten heterogeneous networks to identify potential drug-disease associations. The proposed approach outperformed conventional approaches in discovering novel drug-disease associations when its findings were examined using cross-validation, external validation, and case studies. Moreover, the approach suggested several potential drug repositioning candidates for Alzheimer's and Parkinson's diseases. Han et al. [99] leveraged text mining of OMIM phenotypes to construct a phenotype network and used Graph Convolutional neural Network (GCNN) to identify disease-gene interactions by focusing on non-linear disease-gene correlations. The authors found out that their approach surpassed all other state-of-the-art methods on the majority of metrics.

Semantic technologies

Semantic technologies have allowed to easily combine data from different sources to predict therapeutic potentials and novel indications for existing drugs. For example, Chen et al. [107] proposed a statistical model based on the network's topology and semantics of the sub-network between a drug and a target to predict drug-target associations in a linked heterogeneous network composed, semantically, of annotated data obtained from various publicly available sources, including protein-protein, drug-drug, and drug-side effects, etc. The model successfully differentiated between already known direct drug-target associations and random drug-target associations with high accuracy and identified indirect drug-target associations. Moreover, a drug similarity network signalled that drugs with very different indications from different disease areas are clustered with each other, which may suggest therapeutic potentials and new indications for these drugs.

Zhu et al. [108] used clinical pharmacogenomics (PGx) data, including relations among drugs, diseases, genes, pathways, and single nucleotide polymorphisms (SNPs), and Semantic Web to generate pharmacogenomics Web Ontology Language (WOL) profiles and identify pharmacogenomics associations for FDA approved breast cancer drugs. The authors evaluated their approach using several case studies and indicated that leveraging semantic web technology while studying pharmacogenomics data could lead to higher standard findings of novel drug-disease associations and drug indications.

Machine learning

Computational drug repositioning has evolved over the past two decades from naïve drug similarity attempts, which often used a single source of biological or biomedical data, into an innovative application domain for machine learning approaches. Similar to machine learning models in other domains, computational drug repositioning models require an extensive amount of data to train these models and come up with robust decision rules, aiming to reveal the underlying associations between biological and biomedical entities. The tremendous growth in the volume of publicly available biological and biomedical data and the valuable advancement resulting from machine learning models in other disciplines has assisted the considerable effort in the creation, study, and use of machine learning methods for discovering novel drug-disease associations and drug repositioning applications. Such methods used Naïve Bayesian, k-nearest neighbors (kNN) [109], random forest [110], support vector machines (SVM) [111], and more recently deep neural networks [112] for binary classification, multiclass classification, and values prediction.

Classification

Gottlieb et al. [113] leveraged various data sources to predict drug-disease associations. The authors used drug-drug (e.g., chemical structure, side effects, etc.) and disease-disease (e.g., gene expressions, phenotype, etc.) similarity measures as classification features. Then, they applied a logistic regression classifier to distinguish between true and false drug-disease associations and eventually predict novel drug-disease associations.

Menden et al. [114] developed machine learning models to predict the reactions of cancer to drug treatment using the combination of cell lines genomics and drug chemical structures. The authors integrated both data sources to build a feed-forward perceptron neural network model and a random forest regression model and then validated their findings by cross-validation and an independent blind test. They claimed that the utilization of such models could go further than virtual drug screening since it systematically tested drug efficiency and thus identified potential drug repositioning applications and ultimately could be useful for personalized medicine by linking the cell lines genomics to drug intolerance.

Collaborative filtering

It is noteworthy that several studies based on machine learning have applied collaborative filtering, which depends on historical trends such as gene expression in different samples, to predict novel drug indications and drug-disease associations. Napolitano et al. [115] used several drug-related similarity datasets as features

to predict the therapeutic class of FDA-approved compounds and intentionally considered any mismatches between known and predicted drug classifications as potential alternative therapeutic indications. The authors combined three drug-drug similarity datasets, based on gene expression signatures, chemical structures, and molecular targets, into one drug similarity matrix, which was used as a kernel to train a multi-class Support Vector Machine (SVM) classifier. Afterward, they utilized collaborative filtering techniques to predict novel drug-disease indications.

Zhang et al. [116] introduced a unified computational framework for integrating numerous biological and biomedical sources in order to infer novel drug-drug similarities as well as disease-disease similarities. The authors incorporated drug similarities (e.g., target proteins, side effects, and chemical structure), disease similarities (e.g., gene-disease associations and disease phenotype), and known drug-disease associations datasets to build a drug-disease network. The drug-disease network was treated as an optimization problem, which was solved using block coordinate descent (BCD) strategy. The results demonstrated that such a framework could be useful in finding novel drug-disease associations and identifying new drug repositioning opportunities.

Yang et al. [117] presented a causal inference-probabilistic matrix factorization (CI-PMF) approach to identify and classify drug-disease associations. The authors used several biological and biomedical sources (e.g., drug targets, pathways, pathway-related genes, and disease-gene associations) to build a causal network that linking drug, target, pathway, gene, disease entities together in order to rank drug-disease associations. Furthermore, they leveraged known drug-disease associations to form a probabilistic matrix factorization (PMF) model, which was used to construct a PMF model to classify constructed drug-disease associations into different classes. Finally, they exploited drug-disease association ranking scores and predicted classes to identify novel drug-disease association.

Lim et al. [118] conducted a large-scale study to infer off-target drug interactions and identify novel drug repositioning candidates. The authors used drug chemical structures and protein targets data to build a dual regularized one-class collaborative filtering model that surpassed the previously introduced state-of-the-art models. Ozsoy et al. [119] treated the drug repositioning process as a recommendation process and utilized Pareto dominance and collaborative filtering to identify drug-disease associations. The authors integrated multisource drugs data (protein targets, chemical structures, and side effects) and applied a variety of similarity measures to calculate drug-drug similarities and then used a Pareto

dominance model to identify neighbor drugs. Finally, they used diseases that are shared among neighbor drugs to infer potentials and novel indications for existing drugs.

Deep learning

With the significant growth in publicly available datasets and rapid increase in computational power, deep learning (DL), or neural network (NN), has gained considerable attention. As an inspiring machine learning division, deep learning has given a significant boost and emerged as the leading technique for drug discovery and development in the most recent published studies [112, 120].

Deep learning, a notion closely linked to artificial neural networks (ANNs), can be defined as the learning from nonlinear processing of interconnected neurons layers. It has attracted researchers for its architecture's flexibility, which enables the development of single task or multi-task machine learning models for identifying potential therapeutic applications and predicting drug-disease interactions. Although deep learning has been utilized to develop up-and-coming models in the drug repositioning field, it is worth emphasizing that the full-power employment of deep learning still has some limitations. For instance, deep neural network models need to be adjusted to fit the data used in training these models, which takes substantial time and effort. Additionally, the selection of which machine learning technique or similarity measure to use with each dataset in the deep neural network layers is not straightforward and somehow depending on the used datasets. Neural networks can be mainly classified, based on network's architecture, into (1) fully-connected deep neural network (DNN), (2) convolutional neural network (CNN), (3) recurrent neural network (RNN), (4) autoencoder (AE) [112].

Aliper et al. [121] employed a fully-connected deep neural network to predict the pharmacological properties of drugs and identifying therapeutic potentials and novel drug indications. The authors used gene expression signatures data and pathways data to build deep neural networks models which outperformed support vector machine model and achieved high classification accuracy in predicting drug indications and, hence such deep neural networks could be useful for drug repurposing. Furthermore, they proposed using deep neural net confusion matrices for drug repositioning.

Altae-Tran et al. [122] integrated a standard one-shot learning paradigm with a convolutional neural network to come up with an iterative refinement long short-term memory (LSTM) learning model. The authors adopted the standard one-shot learning paradigm to enhance the learning of meaningful distance metrics over small-molecules in new experiment systems. When evaluated

against two different related datasets, the proposed one-shot model achieved remarkable success in identifying molecular behaviour in low-data drug discovery experiments.

Hu et al. [123] introduced a convolutional neural network model to unveil drug-target interactions. The authors used drug chemical structures and protein sequences data to construct their convolutional neural network classifier that showed superior performance in comparison with other state-of-the-art models. The proposed model inferred drug-target associations in the case of having multiple target proteins interacting with multiple chemical molecules, which demonstrate the potential of such a model in identifying therapeutic novel indications and drug repositioning opportunities.

Segler et al. [124] proposed a recurrent neural network model to generate novel molecules for drug repositioning applications. The authors used drug structures and drug-target interactions data to train their recurrent neural network classifier to produce new molecules that are strongly associated with the desired biological targets. The proposed model was evaluated against two different known drug-target association datasets and performed fairly well. However, the introduced model mimicked the complete de novo drug design cycle and generated large sets of novel molecules when it was integrated with a scoring function.

Zeng et al. [106] used multi-modal deep autoencoder and variational autoencoder models to discover drug-disease associations. The authors integrated various drug-related datasets (drug-disease associations, drug-target associations, drug-drug associations, and drug side effects) to train a multi-modal deep autoencoder and then define high-level drug features. After that, they encoded and decoded the combination of high-level drug features and clinically reported drug-disease associations using variational autoencoder to identify novel therapeutic indications in addition to already identified indications. The findings were validated against a well-known dataset of drug-disease associations and surpassed the previous state-of-the-art machine learning models. Furthermore, the authors reported drug repositioning candidates for Alzheimer's and Parkinson's diseases.

Network analysis

Networks and their analysis have been excessively used in the field of computational drug repositioning as they can provide considerable insight into drug mode-of-action and indications and how drug targets work and, therefore, identify therapeutic potentials and unveil drug repositioning applications. Networks are an excellent way of modelling biological and biomedical entities and their interactions and relationships. Such models can, in turn,

be used to discover informative relationships by leveraging graph theory concepts, statistical analysis, and computational models. In such networks, nodes are used to represent genes, proteins, molecules, phenotypes, or any other biological or biomedical entities, and edges are used to represent functional similarities, mode-of-actions, underlying mechanisms, or any other relationships. Additionally, nodes and edges can be weighted to represent specific attributable information. Moreover, integrating different entities/relationships in a network result in a heterogeneous network while focusing on one entity class or relationship produces a homogeneous network.

Like other computational drug repositioning approaches, drug-based strategy studies, as well as disease-based studies, have also benefited from the network analysis approach to infer drug-target associations and identify novel drug repositioning candidates. Studies based on network analysis can be categorized, according to their data sources, into categories: gene regulatory networks, metabolic networks, protein-protein interaction networks, drug-target interaction networks, drug-drug interaction networks, drug-disease association network, drug-side effect association networks, disease-disease interaction networks, and integrated heterogeneous networks.

Bipartite graph

Yamanishi et al. [30] proposed a bipartite graph supervised learning model to infer novel drug-target interactions. The authors combined protein-protein interaction information with drug chemical structure information and drug-target interaction network to predict different drug-target interaction classes, which could significantly help in improving drug repositioning research productivity. Kinnings et al. [31] built a drug-drug interaction network to unveil drug communities within the network and eventually identify therapeutic potentials and novel indications for existing drugs. The authors represented drugs as nodes and used drug chemical structure information and drug-target interactions similarity to draw edges between drugs. Afterward, they studied the drug-drug interaction network and came up with drug repositioning candidates that were validated using case studies.

Hu and Agarwal [13] constructed a disease-drug network to identify drug repositioning applications and discover drug side effects. The authors used microarray gene expression profiles to build a disease-drug network, which they then enriched using CMap data. The proposed model was validated against gold-standard data and showed high potential in identifying novel therapeutic indications for existing drugs. Li and Lu [35] develop a novel bipartite graph model to infer drug-target

indications based on drug pairwise similarity. The authors used drug chemical structure information along with drug-targets interactions information to build their supervised learning bipartite graph model, which captured the implicit information between drug targets and surpassed other state-of-the-art models.

Clustering

Wu et al. [125] built a weighted drug-disease heterogeneous network and applied network clustering to identify potential drug repositioning candidates within closely connected network modules. The authors used disease-gene associations and drug-target interactions to construct their weighted heterogeneous network where drugs and diseases were defined as nodes, edges were drawn when a pair of nodes share genes, targets, biological processes, pathways, phenotypes, or a combination of these features, and edges were weighted using Jaccard coefficient similarity. Subsequently, they used two network clustering algorithms to cluster nodes into modules and then assembled all potential drug-disease pairs within each of these modules. Finally, they treated drug-disease pairs suggested by the two network clustering algorithms as drug repositioning candidates and performed literature validations and presented several case studies in support of their proposed model.

Tan et al. [38] built a drug-drug interaction network in order to identify novel drug target indications. The authors utilized drug chemical structure information, gene semantic similarity information, and drug-target interaction networks to calculate the degree of drug similarity which then used to construct a drug-drug interaction network, neighbor drugs by clustering the drug-drug interaction network into modules based on mode-of-action, and finally propose new drug therapeutic indications. The proposed model showed high accuracy when validated using the literature.

Network centrality measures

Rakshit et al. [126] developed a novel network-based bidirectional top-down and bottom-up approaches to predict potential drug repositioning applications for a specific disease. The authors used disease-specific (Parkinson's disease) target information and drug-target indications to construct two networks. Subsequently, they utilized several network centrality measures to identify genes and drugs of interests in both networks and used them as an input for the top-down and bottom-up models. The introduced models identified a set of drug repositioning candidates to be investigated for Parkinson's disease treatment, which was validated against a well-known drug-target indications data source.

Yang et al. [127] proposed a new systematic model to identify therapeutic potentials and drug repositioning candidates in heterogeneous networks. The authors combined molecular data, side effects, and online health community information to construct a heterogeneous network that consists of drugs, diseases, and adverse drug reactions as intermediates. Subsequently, they applied several path-based heterogeneous network mining models to identifying and drug repositioning candidates and literately validated their models and concluded that the more data sources used for constructing such heterogeneous networks, the better for predicting models.

Validation of computational drug repositioning models

Ideally, computational drug repositioning studies are conducted to identify new uses for already existing drugs and optimize the pre-clinical process of developing new drugs by saving time and cost compared to the traditional de novo drug discovery and development approach. Researchers validate/evaluate their findings and conclude their models by recommending a set of drug repositioning candidates.

However, validation/evaluation models might differ, in contexts, from the proposed computational models, or specific validation models might not be accurate and trustworthy. Thus, comprehending and picking out suitable validation models is highly crucial for the success of the proposed computational models. Furthermore, selecting the right set of drug repositioning candidates for validation is crucial too due to different factors, such as high price, high level of toxicity, and reduced bioavailability, and due to certain drugs having been abandoned or not preferred by physicians or biologists. Therefore, it is essential that all interested parties are deeply engaged in the process of drug repositioning to boost the conducted research in this field.

Practically speaking, validation/evaluation models vary from one study to another and can depend, up to a certain extent, on the nature of desired outcomes. These models can be classified into (1) in vitro experiments (2) in vivo experiments (3) electronic health records (4) leave-one-out and cross-validation (5) benchmarking against previous models (6) case studies (7) literature cross-referencing, and (8) domain experts consultation.

Despite some well-known drawbacks, in vitro and in vivo experimental validation models have been widely used to validate drug repositioning candidates. In vitro and in vivo validation models refer to performing experiments in a controlled environment outside of a living organism (e.g., cellular biology studies outside of organisms or cells) and in a whole living body (e.g., animal studies and clinical trials) respectively. For example, Lim

et al. [118] identified albendazole as a drug repositioning candidate for anti-cancer effects and presented in vitro and in vivo pieces of evidence in support of using it to treat liver cancer and ovarian cancer.

In order to evaluate the efficiency of potential repositioned drugs, Rakshit et al. [126] introduced a metric called On-Target Ratio (OTR) which is the ratio between the number of drug targets in their proposed disease-specific genes network to the total number of interactions of the same drug in the DrugBank database. Moreover, Ozsoy et al. [119] evaluated their results against ClinicaTrials.gov, which is a collection of publicly and privately funded clinical studies from around the world. The authors also performed a leave-one-out test and benchmarked their model against state-of-the-art models.

Yang et al. [127] used scientific articles published by PubMed as a medical literature cross-referencing model to evaluate the performance of their proposed models. Furthermore, the authors consulted medical experts to evaluate their findings and guarantee the accuracy of their proposed model. The medical experts indicated that the repositioning drugs candidates identified by the proposed model offered significant benefit in filtering and reducing the number of drugs that can be possibly used for the suggested indications. In addition to using an electronic health records validation model, Zeng et al. [106] presented two case studies to validate their proposed deep learning model, which identifies potential drug-disease associations. The authors used Alzheimer's disease and Parkinson's disease to showcase how robust their proposed model is and suggested approved drugs for Alzheimer's disease (e.g., risperidone and aripiprazole) and Parkinson's disease (e.g., methylphenidate and pergolide).

It is noteworthy that literature-based validation models have been widely adopted in recent studies as literature mining approaches have snowballed. Additionally, K-fold cross-validation is often used to train models in machine learning-based studies to overcome the over-optimistic estimation of model performance, which can also be tackled using a new testing dataset independent of the training set, assuming that such information is available.

Current and prospective drug repositioning applications

As a result of reviewing a number of computational drug repositioning studies and zooming in into their findings, we have identified a set of disease areas and related therapeutics that have benefited from drug repositioning applications. When drug repositioning started to get the scientific community attention, a number of studies were conducted to learn about mode-of-action for antidepressant, neurological, and non-neurological drugs. These

Table 3 Examples of drug repositioning applications in various disease areas and related therapeutics

Disease area	Drug name/active ingredients	Original indication	New indication(s)	New indication(s) status
Depression	Duloxetine hydrochloride	Major Depressive Disorder (MDD)	Neuropathic pain, generalized anxiety disorder (GAD), osteoarthritis, and stress incontinence	Approved
	Fluoxetine hydrochloride	Major Depressive Disorder (MDD)	Premenstrual dysphoric disorder (PMDD)	Approved
Neurology	Sibutramine hydrochloride	Major Depressive Disorder (MDD)	Obesity	Approved
	Atomoxetine hydrochloride	Parkinson's disease (PD)	Attention-deficit hyperactivity disorder (ADHD)	Approved
Non-neurology	Ropinirole hydrochloride	Hypertension (HTN)	Parkinson's disease (PD)	Approved
	Minoxidil	Hypertension (HTN)	Hair loss	Approved
	Finasteride	Benign prostatic hyperplasia (BPH)	Hair loss	Approved
	Zidovudine	Failed clinical trials for cancer	Human immunodeficiency virus (HIV)	Approved
	Sildenafil	Angina	Erectile dysfunction (ED) and pulmonary arterial hypertension (PAH)	Approved
Cancer	Auranofin	Rheumatoid arthritis (RA)	Gastrointestinal stromal tumor (GIST)	Approved
	Imatinib	Chronic myeloid leukemia (CML)	Gastrointestinal stromal tumors (GIST)	Approved
	Irinotecan hydrochloride	Colorectal cancer	Pancreatic cancer	Approved
	Nelfinavir	Human immunodeficiency virus 1 (HIV-1)	Colorectal cancer, lung cancer, cervical cancer, pancreatic cancer, ovarian cancer, metastatic cancer	Investigational
	Metformin hydrochloride	Type 2 diabetes (T2DM)	Breast cancer, pancreatic cancer, endometrial cancer, colorectal cancer, and esophageal cancer	Investigational
	Trastuzumab	Human epidermal growth factor receptor 2 (HER2)-positive breast cancer	Metastatic breast cancer, gastric cancer, and early breast cancer	Approved
	Sunitinib	Renal cell carcinoma (RCC) and Gastrointestinal stromal tumor (GIST)	Pancreatic neuroendocrine tumors (PNETs)	Approved
	Crizotinib	Clinical trials for anaplastic large cell lymphoma (ALCL)	Non-small cell lung cancer (NSCLC)	Approved
Infectious	Thalidomide	Morning sickness (withdrawn)	Erythema nodosum leprosum (Leprosy)	Approved
	Everolimus	Immunosuppressant	Pancreatic neuroendocrine tumors (PNETs), renal cell carcinoma (RCC), and subependymal giant cell astrocytoma (SEGA)	Approved
	Sirolimus	Organ rejection in patients receiving renal transplants	Malaria	Investigational
Rare and orphan	Alefacept	Chronic plaque psoriasis	Rejection in patients receiving allogeneic solid organ transplants	Investigational
	Indium In-111 pentetreotide	Agent for the scintigraphic localization of primary and metastatic neuroendocrine tumors	Neuroendocrine tumors (NETs)	Investigational

studies have successfully unveiled new indications for already approved drugs as well as drugs in the pipeline.

In Table 3, a number of successes are listed with their original indication as well as the new and in most of the cases approved indication. There are five drugs with the original indication being for aspects of the nervous system (depression and neurology). The new indications are also for aspects of the nervous system. The new indications included a new medicine to treat **obesity**, a disease of plenty. As reported by the World Health Organization in 2020, 650 million people are obese worldwide [154]. The new indication for treating obesity is, therefore, a significant step forwards.

Cancer is another area where a number of new drug indications have been found. Cancer is also a disease that is on the rise. Globally an estimated 10 million people die of cancer each year [155], and many more continue a reduced lifestyle as they are combatting the effects of cancer. Furthermore, the incidence of this disease increases with age. Since the average age of populations is on the rise, it is expected that this will lead to a higher number of people with cancer. The new drug indications for various cancers are, therefore, of the highest importance. Pessetto et al. [156] conducted a high-throughput screening study on FDA-approved drugs and found that *auranofin* could be repositioned for the treatment of gastrointestinal stromal tumours. Stenvang et al. [157] applied a biomarker-guided repurposing approach on genome information and clinical studies and proposed *irinotecan* for the treatment of breast cancer.

Infectious diseases are caused by pathogenic microorganisms such as bacteria and viruses. Multi-drug resistance and extensively antibiotic drug-resistant microbes threaten the treatment of such diseases and require new processes of treatment. Drug repositioning has led to success in combating infectious diseases. Ng et al. [158] proposed an integrated chemical genomics and structural systems biology approach which identified *plasmodium falciparum* targets of drug-like active compounds from the malaria box, and suggested that several approved drugs may be active against malaria.

Rare and orphan diseases affect a small proportion of the world's population. A key motivator behind developing a treatment for an incurable disease is the potential market size for the treatment. As a result, thousands of rare and orphan diseases lack treatments because of the **insignificant potential market size for these treatments**. Drug repositioning has gained some attention in identifying therapeutics for orphan and rare diseases. Molineris et al. [159] utilized several resources (e.g., OMIM, DrugBank, CMap) to conduct a systematic analysis of gene co-expression and successfully identified HDAC1 and TSPO as two significant targets for epileptic syndromes.

Xu et al. [160] used FDA orphan designation database and FDA-approved drugs to establish the Rare Disease Repurposing Database (RDRD). RDRD provides a comprehensive resource for developing targeted effective therapies for rare disease patients.

The drawn-out traditional de novo drug development process, the success and high potential of computational drug repositioning, and the strong demand and the need to treat cancer, infectious, orphan, and rare diseases have, therefore, motivated researchers from different disciplines to unify forces in searching for therapeutic potentials and novel indications for existing drugs, which have already been approved for human use and are safer than products that are still being developed to treat cancer and other incurable diseases. Moreover, approved drugs are already optimized to target specific proteins, which could be highly useful if there is another disease that shares the same targets. Lastly, **utilizing different sources of biological and biomedical data** in developing computational drug repositioning models could be a **promising tactic** towards personalized treatment. Table 3 provides examples of drug repositioning applications in various disease areas and related therapeutics retrieved from Drugs@FDA database [91] and DrugBank [82].

Drug repositioning opportunities

Drug repositioning is a highly promising technique that has attracted growing attention from governments and pharmaceutical companies for its key role in reducing time, cost, and risk in the process of developing drugs for cancers and other incurable illnesses. As this technique emerged, teams of multidisciplinary researchers and scientists carried out numerous attempts, with different degrees of efficiency and success, to computationally study the potential of repositioning drugs to treat other diseases and identify alternative indications regardless of the status of the investigated drug, whether it is approved, withdrawn, in clinical trials, or failed in clinical trials. Although drug repositioning is a quite up-and-coming technique, the **traditional, costly, failure-prone de novo drug development process** is still **essential** for discovering and testing new drugs; however, adopting some computational drug repositioning models within this process can help to push drugs steps forward in the development pipeline and eventually improve drug efficiencies in clinical trials.

The **opportunities** provided by drug repositioning to develop the urgently needed drugs to treat the current coronavirus epidemic cannot be underestimated. The general search for coronavirus effective drugs is reviewed on a weekly basis by Nature Medicine, the latest of which is [161]. The specific opportunities for repurposing drugs for coronavirus infections are reviewed in [5].

Discussion and conclusion

After surveying various avenues in which computational drug repositioning strategies have been adopted, and models have been introduced to identify novel therapeutic interactions, we can conclude that **each strategy and approach has its advantages and limitations** and also that **combining different strategies and approaches often achieve a higher success rate**.

Despite having some outstanding computational drug repositioning models, developing robust models is still a complex process that comes with a few challenges. One of the main challenges is the difficulty of **putting theoretical computational approaches into action**; because of the complexity of mapping such theoretical approaches to simulate living organism's behaviour and other obstacles such as **missing, biased, and inaccurate data**. For instance, reliable gene expression signature profiles may be hard to define due to several reasons such as variations in experimental conditions (e.g., environment variables and patient age) across different experiments, which may result in a data discrepancy in gene expression signatures, contributing to having biased data. Also, there may not always be significant changes in gene expressions when these genes are used as drug targets, which can lead to having inaccurate data. Moreover, **the lack of high-resolution structural data for drug targets** makes it hard to identify potential drug-target interactions when following the chemical structure and molecule information strategy. Another challenge facing computational drug repositioning models is the lack of **trusted gold-standard datasets** that can be used to evaluate the performance of such models.

Researchers, therefore, either have to build their own gold-standard dataset and subsequently use prevalent evaluation metrics (e.g., accuracy, recall, sensitivity, specificity, F1 score, and area under the receiver operating characteristic curve) to compare and evaluate their proposed models or they have to split their data into training, testing, and validating sets and then utilize K-fold cross-validation and prevalent evaluation metrics combined to avoid ending up with an over-fitted model.

Despite all the challenges encountered in computational drug repositioning studies, we envision that integrating multi-source data related to drugs (e.g., chemical structures), diseases (e.g., phenotypic information), and how these drugs and diseases affect human body (e.g., gene expression signature profiles and side effects) is crucial to enrich computational drug repositioning models and improve their performance and thus take them up to the next level. Furthermore, there is a significant number of **diseases that still lack treatments to slow, stop, or reverse their courses**, which motivates and inspires multidisciplinary researchers and scientists to carry out

studies, especially in compacting different cancers and thousands of orphan and rare diseases.

In summary, we strongly believe that computational drug repositioning can be of enormous benefit to humanity by discovering **new indications** for approved drugs, **speeding up** the process of developing new drugs, and giving a **second chance** to withdrawn and failed drugs. While governments and pharmaceutical companies are directing more support towards computational drug repositioning ventures, researchers and scientists should pick up the ball and make further efforts to come up with creative state-of-the-art models towards novel findings and significant breakthroughs.

Acknowledgements

No acknowledgement to report.

Authors' contributions

All three authors initiated the study and put together the research plan. TJ wrote the draft. All three authors went through a sequence of versions of the manuscript with RA and JR providing guidance and feedback which were incrementally integrated in the manuscript. All authors read and approved the final manuscript.

Funding

No funding information to declare

Availability of data and materials

No data to share as this is a review paper.

Competing interests

No competing interest to declare

Author details

¹ Department of Computer Science, University of Calgary, Calgary, Alberta, Canada. ² Department of Computer Engineering, Istanbul Medipol University, Istanbul, Turkey.

Received: 19 February 2020 Accepted: 13 July 2020

Published online: 22 July 2020

References

- Ashburn TT, Thor KB (2004) Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov* 3(8):673–683
- Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, Doig A, Guillems T, Latimer J, McNamee C et al (2019) Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov* 18(1):41–58
- Ledford H (2020) Dozens of coronavirus drugs are in development—what happens next? *Nature*
- Serafin MB, Bottega A, Foletto VS, da Rosa TF, Hörner A, Hörner R (2020) Drug repositioning an alternative for the treatment of coronavirus COVID-19. *Int J Antimicrob Agents* 105:969
- Harris M, Bhatti Y, Buckley J, Sharma D (2020) Fast and frugal innovations in response to the COVID-19 pandemic. *Nat Med* 1:4
- Guy RK, DiPaola RS, Romanelli F, Dutch RE (2020) Rapid repurposing of drugs for COVID-19. *Science* 368(6493):829–830
- Iorio F, Bosotti R, Scacheri E, Belcastro V, Mithbaokar P, Ferriero R, Murino L, Tagliaferri R, Brunetti-Pierri N, Isacchi A et al (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc Natl Acad Sci* 107(33):621–626
- Gloekner C, Garner AL, Mersha F, Oksov Y, Tricoche N, Eubanks LM, Lustigman S, Kaufmann GF, Janda KD (2010) Repositioning of an existing drug for the neglected tropical disease onchocerciasis. *Proc Natl Acad Sci* 107(8):3424–3429

9. Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, Hufeisen SJ, Jensen NH, Kuijter MB, Matos RC, Tran TB et al (2009) Predicting new molecular targets for known drugs. *Nature* 462(7270):175
10. Dudley JT, Deshpande T, Butte AJ (2011) Exploiting drug-disease relationships for computational drug repositioning. *Brief Bioinf* 12(4):303–311
11. Li J, Zheng S, Chen B, Butte AJ, Swamidass SJ, Lu Z (2015) A survey of current trends in computational drug repositioning. *Brief Bioinf* 17(1):2–12
12. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet J-P, Subramanian A, Ross KN et al (2006) The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313(5795):1929–1935
13. Hu G, Agarwal P (2009) Human disease-drug network based on genomic expression profiles. *PLoS ONE* 4(8):e6536
14. Sirota M, Dudley JT, Kim J, Chiang AP, Morgan AA, Sweet-Cordero A, Sage J, Butte AJ (2011) Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci Transl Med* 3(96):96ra77–96ra77
15. Liu X, Wang S, Meng F, Wang J, Zhang Y, Dai E, Yu X, Li X, Jiang W (2012) SM2miR: a database of the experimentally validated small molecules' effects on microRNA expression. *Bioinformatics* 29(3):409–411
16. Jiang W, Chen X, Liao M, Li W, Lian B, Wang L, Meng F, Liu X, Chen X, Jin Y et al (2012) Identification of links between small molecules and miRNAs in human cancers based on transcriptional Responses. *Sci Rep* 2:282
17. Sanseau P, Agarwal P, Barnes MR, Pastinen T, Richards JB, Cardon LR, Mooser V (2012) Use of genome-wide association studies for drug repositioning. *Nat Biotechnol* 30(4):317
18. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, Greninger P, Thompson IR, Luo X, Soares J et al (2012) Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 483(7391):570
19. Rukov JL, Wilentzik R, Jaffe I, Vinther J, Shomron N (2013) PharmacomiR: linking microRNAs and drug effects. *Brief Bioinf* 15(4):648–659
20. Iorio F, Rittman T, Ge H, Menden M, Saez-Rodriguez J (2013) Transcriptional data: a new gateway to drug repositioning? *Drug Discov Today* 18(7–8):350–357
21. Jahchan NS, Dudley JT, Mazur PK, Flores N, Yang D, Palmerton A, Zmoos A-F, Vaka D, Tran KQ, Zhou M et al (2013) A drug repositioning approach identifies tricyclic antidepressants as inhibitors of small cell lung cancer A Drug repositioning approach identifies tricyclic antidepressants as inhibitors of small cell lung cancer and other neuroendocrine tumors. *Cancer Discov* 3(12):1364–1377
22. Li Y, Qiu C, Tu J, Geng B, Yang J, Jiang T, Cui Q (2013) HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res* 42(D1):D1070–D1074
23. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, Yoshida S et al (2014) Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 506(7488):376
24. Vidović D, Koletić A, Schürer SC (2014) Large-scale integration of small molecule-induced genome-wide transcriptional responses, kinome-wide Binding Affinities and Cell-growth Inhibition Profiles Reveal Global Trends Characterizing Systems-level Drug Action. *Front Genet* 5:342
25. Ding X-M (2014) MicroRNAs: regulators of cancer metastasis and epithelial-mesenchymal transition (EMT). *Chin J Cancer* 33(3):140
26. Wen X, Deng F-M, Wang J (2014) MicroRNAs as predictive biomarkers and therapeutic targets in prostate cancer. *Am J Clin Exp Urol* 2(3):219
27. Huang H, Nguyen T, Ibrahim S, Shantharam S, Yue Z, Chen JY (2015) DMAP: a connectivity map database to enable identification of novel drug repositioning candidates. *BMC Bioinf* 16(13):54
28. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, Gould J, Davis JF, Tubelli AA, Asiedu JK et al (2017) A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 171(6):1437–1452
29. Huang Z, Shi J, Gao Y, Cui C, Zhang S, Li J, Zhou Y, Cui Q (2018) HMDD v3.0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res* 47(D1):D1013–D1017
30. Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M (2008) Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 24(13):i232–i240
31. Kinnings SL, Liu N, Buchmeier N, Tonge PJ, Xie L, Bourne PE (2009) Drug discovery using chemical systems biology: repositioning the safe medicine to treat multi-drug and Extensively Drug Resistant Tuberculosis. *PLoS Comput Biol* 5(7):e1000423
32. Bleakley K, Yamanishi Y (2009) Supervised prediction of drug-target interactions using bipartite local models. *Bioinformatics* 25(18):2397–2403
33. Swamidass SJ (2011) Mining small-molecule screens to repurpose drugs. *Brief Bioinf* 12(4):327–335
34. Pihan E, Colliandre L, Guichou J-F, Douguet D (2012) e-Drug 3D: 3D structure collections dedicated to drug repurposing and fragment-based drug design. *Bioinformatics* 28(11):1540–1541
35. Li J, Lu Z (2012) A new method for computational drug repositioning using drug pairwise similarity. In: 2012 IEEE international conference on bioinformatics and biomedicine. IEEE, pp. 1–4
36. Novick PA, Ortiz OF, Poelman J, Abdulhay AY, Pande VS (2013) SWEET-LEAD: an in silico database of approved drugs, regulated chemicals, and herbal isolates for computer-aided drug discovery. *PLoS ONE* 8(11):e79568
37. Wang Y, Chen S, Deng N, Wang Y (2013) Drug repositioning by Kernel-based integration of molecular structure, molecular activity, and phenotype data. *PLoS ONE* 8(11):e78518
38. Tan F, Yang R, Xu X, Chen X, Wang Y, Ma H, Liu X, Wu X, Chen Y, Liu L et al (2014) Drug repositioning by applying 'expression profiles' generated by integrating chemical structure similarity and gene semantic similarity. *Mol BioSyst* 10(5):1126–1138
39. Zheng C, Guo Z, Huang C, Wu Z, Li Y, Chen X, Fu Y, Ru J, Shar PA, Wang Y et al (2015) Large-scale direct targeting for drug repositioning and discovery. *Sci Rep* 5:11970
40. Lewin B (2004) *Genes VIII*. Pearson Prentice Hall, Upper Saddle River, p 4
41. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA et al (2001) The sequence of the human genome. *Science* 291(5507):1304–1351
42. Slonim DK, Yanai I (2009) Getting started in gene expression microarray analysis. *PLoS Comput Biol* 5(10):e1000543
43. Lobo I (2008) Environmental influences on gene expression. *Nat Educ* 1(1):39
44. Wall ME, Rechtsteiner A, Rocha LM (2003) Singular value decomposition and principal component analysis. A practical approach to microarray data analysis. Springer, Berlin, pp 91–109
45. Hunter L, Taylor RC, Leach SM, Simon R (2001) GEST: a gene expression search tool based on a novel Bayesian similarity metric. *Bioinformatics* 17(1):S115–S122
46. Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau W-C, Ledoux P, Rudnev D, Lash AE, Fujibuchi W, Edgar R (2005) NCBI GEO: mining millions of expression profiles—database and tools. *Nucleic Acids Res* 33(1):D562–D566
47. Quackenbush J (2003) Microarrays—guilt by association. *Science* 302(5643):240–241
48. Xing Z, Li D, Yang L, Xi Y, Su X (2014) MicroRNAs and anticancer drugs. *Acta Biochim Biophys Sin* 46(3):233–239
49. Hebringer SJ (2014) The challenges, advantages and future of phenome-wide association studies. *Immunology* 141(2):157–165
50. Rognan D (2007) Chemogenomic approaches to rational drug design. *Br J Pharmacol* 152(1):38–52
51. Mangione W, Samudrala R (2019) Identifying protein features responsible for improved drug repurposing accuracies using the CANDOR platform: implications for drug design. *Molecules* 24(1):167
52. Yan Y, Huang S-Y (2019) Pushing the accuracy limit of shape complementarity for protein-protein docking. *BMC Bioinf* 20(25):696
53. Campillos M, Kuhn M, Gavin A-C, Jensen LJ, Bork P (2008) Drug target identification using side-effect similarity. *Science* 321(5886):263–266
54. Yang L, Agarwal P (2011) Systematic drug repositioning based on clinical side-effects. *PLoS ONE* 6(12):e28025
55. Hoehndorf R, Oellrich A, Rebholz-Schuhmann D, Schofield PN, Gkoutos GV (2012) Linking pharmGKB to phenotype studies and animal models of disease for drug repurposing Biocomputing 2012. World Scientific, Singapore, pp 388–399
56. Gottlieb A, Stein GY, Oron Y, Ruppin E, Sharan R (2012) INDI: a computational framework for inferring drug interactions and their associated recommendations. *Mol Syst Biol* 8:1

57. Ye H, Liu Q, Wei J (2014) Construction of drug network based on side effects and its application for drug repositioning. *PLoS ONE* 9(2):e87864
58. Bisgin H, Liu Z, Fang H, Kelly R, Xu X, Tong W (2014) A phenotype-guided drug repositioning through a latent variable model. *BMC Bioinform* 15(1):267
59. Nugent T, Plachouras V, Leidner JL (2016) Computational drug repositioning based on side-effects mined from social media. *PeerJ Comput Sci* 2:e46
60. Sridhar D, Fakhraei S, Getoor L (2016) A probabilistic approach for collective similarity-based drug-drug interaction prediction. *Bioinformatics* 32(20):3175–3182
61. Athar A, Füllgrabe A, George N, Iqbal H, Huerta L, Ali A, Snow C, Fonseca NA, Petryszak R, Papatheodorou I et al (2018) ArrayExpress update-from bulk to single-cell expression data. *Nucleic Acids Res* 47(D1):D711–D715
62. Barretina J, Caponigro G, Stransky N, Vekatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D et al (2012) The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483(7391):603
63. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* 4(9):R60
64. Pacini C, Iorio F, Gonçalves E, Iskar M, Klabunde T, Bork P, Saez-Rodriguez J (2012) DvD: an R/Cytoscape pipeline for drug repurposing using public repositories of gene expression data. *Bioinformatics* 29(1):132–134
65. Papatheodorou I, Fonseca NA, Keays M, Tang YA, Barrera E, Bazant W, Burke M, Füllgrabe A, Fuentes AM-P, George N et al (2017) Expression atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res* 46(D1):D246–D251
66. Edgar R, Domrachev M, Lash AE (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30(1):207–210
67. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* 102(43):15 545–15 550
68. Culhane AC, Schwarzl T, Sultana R, Picard KC, Picard SC, Lu TH, Franklin KR, French SJ, Papenhausen G, Correll M et al (2009) GeneSigDB—a curated database of gene expression signatures. *Nucleic Acids Res* 38(1):D716–D725
69. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al (2000) Gene ontology: tool for the unification of biology. *Nat Genet* 25(1):25
70. Zhang J, Baran J, Cros A, Guberman JM, Haider S, Hsu J, Liang Y, Rivkin E, Wang J, Whitty B et al (2011) International cancer genome consortium data portal—a one-stop shop for cancer genomics data. *Database* 2011:1
71. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28(1):27–30
72. Keenan AB, Jenkins SL, Jagodnik KM, Koplev S, He E, Torre D, Wang Z, Dohlmán AB, Silverstein MC, Lachmann A et al (2018) The library of integrated network-based cellular signatures NIH program: system-level cataloging of human cells response to perturbations. *Cell Syst* 6(1):13–24
73. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27(12):1739–1740
74. Hutter C, Zenklusen JC (2018) The cancer genome atlas: creating lasting value beyond its data. *Cell* 173(2):283–285
75. Lamb J (2007) The connectivity map: a new tool for biomedical research. *Nat Rev Cancer* 7(1):54
76. Consortium TU (2016) UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 45(D1):D158–D169
77. Gillen JE, Tse T, Ide N C, McCray A T (2004) Design, Implementation and Management of a Web-based data entry system for ClinicalTrials.gov. In: *Medinfo*, pp. 1466–1470
78. Kuhn M, Letunic I, Jensen LJ, Bork P (2015) The SIDER database of drugs and side effects. *Nucleic Acids Res* 44(D1):D1075–D1079
79. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, Mutowo P, Atkinson F, Bellis LJ, Cibrián-Uhalte E et al (2016) The ChEMBL database in 2017. *Nucleic Acids Res* 45(D1):D945–D954
80. Swain M (2012) Chemicalize.org
81. Pence H E, Williams A (2010) ChemSpider: an Online Chemical Information Resource
82. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z et al (2017) Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic Acids Res* 46(D1):D1074–D1082
83. Ursu O, Holmes J, Knockel J, Bologa CG, Yang JJ, Mathias SL, Nelson SJ, Oprea TI (2016) DrugCentral: online drug compendium. *Nucleic Acids Res* 26:993
84. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA et al (2015) Pubchem substance and compound databases. *Nucleic Acids Res* 44(D1):D1202–D1213
85. Kouranov A, Xie L, de la Cruz J, Chen L, Westbrook J, Bourne PE, Berman HM (2006) The RCSB PDB information portal for structural genomics. *Nucleic Acids Res* 34(1):D302–D305
86. Huang R, Southall N, Wang Y, Yasgar A, Shinn P, Jadhav A, Nguyen D-T, Austin CP (2001) The NCGC pharmaceutical collection: a comprehensive resource of clinically approved drugs enabling Repurposing and Chemical Genomics. *Sci Transl Med* 3(80):80ps16
87. Wang Y, Zhang S, Li F, Zhou Y, Zhang Y, Wang Z, Zhang R, Zhu J, Ren Y, Tan Y et al (2019) Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res* 48:31–41
88. Brown AS, Patel CJ (2017) A standard database for drug repositioning. *Sci Data* 4:170029
89. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA (2005) Online mendelian inheritance in man (OMIM), a knowledge-base of human genes and genetic disorders. *Nucleic Acids Res* 33(1):D514–D517
90. Hernandez-Boussard T, Whirl-Carrillo M, Hebert JM, Gong L, Owen R, Gong M, Gor W, Liu F, Truong C, Whaley R et al (2007) The pharmacogenetics and pharmacogenomics knowledge base: accentuating the knowledge. *Nucleic Acids Res* 36(1):D913–D918
91. FDA. (2020, January) Drugs@FDA. <http://www.fda.gov/drugsatfda>
92. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci* 98(9):5116–5121
93. Lu Z (2011) PubMed and Beyond: a Survey of Web Tools for Searching Biomedical Literature. *Database*, vol. 2011
94. Zhu F, Patumcharoenpol P, Zhang C, Yang Y, Chan J, Meechai A, Vongsangnak W, Shen B (2013) Biomedical text mining and its applications in cancer research. *J Biomed Inf* 46(2):200–211
95. Cheng D, Knox C, Young N, Stothard P, Damaraju S, Wishart DS (2008) PolySearch: a web-based text mining system for extracting relationships between human diseases, genes, mutations, drugs and metabolites. *Nucleic Acids Res* 36(2):W399–W405
96. Li J, Lu Z (2012) Systematic identification of pharmacogenomics information from clinical trials. *J Biomed Inf* 45(5):870–878
97. Leaman R, Islamaj Doğan R, Lu Z (2013) DNorm: disease name normalization with pairwise learning to rank. *Bioinformatics* 29(22):2909–2917
98. Davis AP, Wiegiers TC, Rosenstein MC, Mattingly CJ (2012) MEDIC: a practical disease vocabulary used at the comparative toxicogenomics database. *Database* 2012:1
99. Han P, Yang P, Zhao P, Shang S, Liu Y, Zhou J, Gao X, Kalnis P (2019) GCN-MF: Disease-gene Association Identification by Graph Convolutional Networks and Matrix Factorization. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, pp. 705–713
100. Gong L, Huang D, Sun S, Gao Z, Pan C, Yang R, Li Y, Yang G (2018) Extraction of interactions of Genes2Genes related to breast cancer. In: *2018 IEEE 16th international conference on software engineering research, management and applications (SERA)*. IEEE, pp. 108–112
101. Li J, Zhu X, Chen JY (2009) Building disease-specific drug-protein connectivity maps from molecular interaction networks and PubMed abstracts. *PLOS Comput Biol* 5(7):e1000450
102. Tari LB, Patel JH (2014) Biomedical literature mining. *Systematic drug repurposing through text mining*. Springer, Berlin, pp 253–267
103. Rastegar-Mojarad M, Elayavilli R K, Li D, Prasad R, Liu H (2015) A new method for prioritizing drug repositioning candidates extracted by

- literature-based discovery. In: 2015 IEEE international conference on bioinformatics and biomedicine (BIBM). IEEE, pp. 669–674
104. Brown AS, Patel CJ (2016) MeSHDD: literature-based drug–drug similarity for drug repositioning. *J Am Med Inf Assoc* 24(3):614–618
 105. Papanikolaou N, Pavlopoulos GA, Theodosiou T, Vizirianakis IS, Iliopoulos I (2016) DrugQuest—a text mining workflow for drug association discovery. *BMC Bioinf* 17(5):182
 106. Zeng X, Zhu S, Liu X, Zhou Y, Nussinov R, Cheng F (2019) deepDR: a network-based deep learning approach to in silico drug repositioning. *Bioinformatics* 35(24):5191–8
 107. Chen B, Ding Y, Wild DJ (2012) Assessing drug target association using semantic linked data. *PLOS Comput Biol* 8(7):e1002574
 108. Zhu Q, Tao C, Shen F, Chute CG (2014) Exploring the pharmacogenomics knowledge base (PharmGKB) for repositioning breast cancer drugs by leveraging web ontology language (OWL) and cheminformatics approaches. *Biocomputing 2014*. World Scientific, Singapore, pp 172–182
 109. Shen M, Xiao Y, Golbraikh A, Gombar VK, Tropsha A (2003) Development and validation of K-nearest-neighbor QSPR models of metabolic stability of drug candidates. *J Med Chem* 46(14):3013–3020
 110. Susnow RG, Dixon SL (2003) Use of Robust Classification Techniques for the Prediction of Human Cytochrome P450 2D6 Inhibition. *J Chem Inf Comput Sci* 43(4):1308–1315
 111. Cristianini N, Shawe-Taylor J (2004) Support vector machines and other Kernel-based learning methods. Cambridge
 112. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T (2018) The rise of deep learning in drug discovery. *Drug Discov Today* 23(6):1241–1250
 113. Gottlieb A, Stein GY, Ruppin E, Sharan R (2011) PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mole Syst Biol* 7:1
 114. Menden MP, Iorio F, Garnett M, McDermott U, Benes CH, Ballester PJ, Saez-Rodriguez J (2013) Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS ONE* 8(4):e61318
 115. Napolitano F, Zhao Y, Moreira VM, Tagliaferri R, Kere J, D'Amato M, Greco D (2013) Drug repositioning: a machine-learning approach through data integration. *J Cheminf* 5(1):30
 116. Zhang P, Wang F, Hu J (2014) Towards drug repositioning: a unified computational framework for integrating multiple aspects of drug similarity and disease similarity. In: *AMIA annual symposium proceedings*, vol. 2014. American Medical Informatics Association, p. 1258
 117. Yang J, Li Z, Fan X, Cheng Y (2014) Drug-disease association and drug-repositioning predictions in complex diseases using causal inference-probabilistic matrix factorization. *J Chem Inf Model* 54(9):2562–2569
 118. Lim H, Poleksic A, Yao Y, Tong H, He D, Zhuang L, Meng P, Xie L (2016) Large-scale off-target identification using fast and accurate dual regularized one-class collaborative filtering and its application to drug repurposing. *PLOS Comput Biol* 12(10):e1005135
 119. Ozsoy MG, Özyer T, Polat F, Alhajj R (2018) Realizing drug repositioning by adapting a recommendation system to handle the process. *BMC Bioinf* 19(1):136
 120. Korotcov A, Tkachenko V, Russo DP, Ekins S (2017) Comparison of deep learning with multiple machine learning methods and metrics using diverse drug Discovery Data Sets. *Mol Pharm* 14(12):4462–4475
 121. Aliper A, Plis S, Artemov A, Ulloa A, Mamoshina P, Zhavoronkov A (2016) Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol Pharm* 13(7):2524–2530
 122. Altae-Tran H, Ramsundar B, Pappu AS, Pande V (2017) Low data drug discovery with one-shot learning. *ACS Central Sci* 3(4):283–293
 123. Hu S, Zhang C, Chen P, Gu P, Zhang J, Wang B (2019) Predicting drug-target interactions from drug structure and protein sequence using novel convolutional Neural Networks. *BMC Bioinf* 20(25):1–12
 124. Segler MH, Kogej T, Tyrchan C, Waller MP (2017) Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Central Sci* 4(1):120–131
 125. Wu C, Gudivada RC, Aronow BJ, Jegga AG (2013) Computational drug repositioning through heterogeneous network clustering. *BMC Syst Biol* 7(5):S6
 126. Rakshit H, Chatterjee P, Roy D (2015) A bidirectional drug repositioning approach for Parkinson's disease through network-based inference. *Biochem Biophys Res Commun* 457(3):280–287
 127. Yang CC, Zhao M (2019) Mining heterogeneous network for drug repositioning using phenotypic information extracted from social media and pharmaceutical databases. *Artif Intell Med* 96:80–92
 128. MedHelp. (2020, January) MedHelp. <https://www.medhelp.org/>
 129. NIEHS. (2020, January) Tox21. [Online]. Available: https://ntp.niehs.nih.gov/whatwestudy/tox21/index.html?utm_source=direct&utm_medium=prod&utm_campaign=ntpgolinks&utm_term=tox2
 130. Irwin and Shoichet Laboratories. (2020, January) ZINC. <https://zinc.docki ng.org/>
 131. ChEMaxon. (2020, January) BindingDB. <https://www.bindingdb.org/bind/index.jsp>
 132. Kilicoglu H, Shin D, Fiszman M, Rosembat G, Rindflesch TC (2012) SemMedDB: a PubMed-scale repository of biomedical semantic predictions. *Bioinformatics* 28(23):3158–3160
 133. Bodenreider O (2004) The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Res* 32(1):D267–D270
 134. Davis AP, Grondin CJ, Lennon-Hopkins K, Saraceni-Richards C, Sciaky D, King BL, Wiegers TC, Mattingly CJ (2014) The comparative toxicogenomics database's 10th year anniversary: update 2015. *Nucleic Acids Res* 43(D1):D914–D920
 135. Wall DP, Pivovarov R, Tong M, Jung J-Y, Fusaro VA, DeLuca TF, Tonellato PJ (2010) Genotator: a disease-agnostic tool for genetic annotation of disease. *BMC Med Genom* 3(1):50
 136. Barbosa-Silva A, Fontaine J-F, Donnard ER, Stussi F, Ortega JM, Andrade-Navarro MA (2011) PESCADOR, a web-based tool to assist text-mining of biointeractions extracted from PubMed queries. *BMC Bioinf* 12(1):435
 137. Emory University. (2020, January) CancerQuest. <https://www.cancerquest.org/>
 138. Darryl Nishimura. (2020, January) BioCarta. <https://omictools.com/biocarta-tool>
 139. NDF-RT. (2020, January) National drug file—reference terminology (NDF-RT). <https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/NDFRT/index.html>
 140. Keshava Prasad T, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A et al (2008) Human protein reference database—2009 Update. *Nucleic Acids Res* 37(1):D767–D772
 141. Darryl Nishimura. (2020, January) WHO. https://www.whocc.no/atc_ddd_index/
 142. Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D (2004) BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res* 32(1):D431–D433
 143. Günther S, Kuhn M, Dunkel M, Campillos M, Senger C, Petsalaki E, Ahmed J, Urdiales EG, Gewiss A, Jensen LJ et al (2007) SuperTarget and matador: resources for exploring drug-target relationships. *Nucleic Acids Res* 36(1):D919–D922
 144. Chen B, Dong X, Jiao D, Wang H, Zhu Q, Ding Y, Wild DJ (2010) Chem2Bio2RDF: a semantic framework for linking and data mining chemogenomic and systems chemical biology data. *BMC Bioinf* 11(1):255
 145. Hoehndorf R, Schofield PN, Gkoutos GV (2011) PhenomeNET: a whole-phenome approach to disease gene discovery. *Nucleic Acids Res* 39(18):e119–e119
 146. Liu Y, Hu B, Fu C, Chen X (2009) DCDB: drug combination database. *Bioinformatics* 26(4):587–588
 147. NIH. (2020, January) DailyMed. <http://dailymed.nlm.nih.gov>
 148. Brown KR, Jurisica I (2005) Online predicted human interaction database. *Bioinformatics* 21(9):2076–2082
 149. Schuffenhauer A, Zimmermann J, Stoop R, van der Vyver J-J, Lecchini S, Jacoby E (2002) An ontology for pharmaceutical ligands and its application for in silico screening and library design. *J Chem Inf Comput Sci* 42(4):947–955
 150. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, Abeyasinghe S, Krawczak M, Cooper DN (2003) Human gene mutation database (HGMD®): 2003 update. *Hum Mutat* 21(6):577–581
 151. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S et al (2007) HMDB: the human metabolome database. *Nucleic Acids Res* 35(1):D521–D526

152. Roth BL, Lopez E, Patel S, Kroeze WK (2000) The multiplicity of serotonin receptors: uselessly diverse molecules or an embarrassment of riches? *The Neurosci* 6(4):252–262
153. Mering Cv, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B (2003) STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* 31(1):258–261
154. W. H. Organization. (2020, June) Obesity and Overweight. <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>
155. W. H. Organization. (2020, June) Obesity. <https://www.who.int/news-room/fact-sheets/detail/cancer>
156. Pessetto ZY, Weir SJ, Sethi G, Broward MA, Godwin AK (2013) Drug repurposing for gastrointestinal stromal tumor. *Mol Cancer Ther* 12(7):1299–1309
157. Stenvang J, Kümler I, Nygård SB, Smith DH, Nielsen D, Brünner N, Moreira JMA (2013) Biomarker-guided repurposing of chemotherapeutic drugs for cancer therapy: a novel strategy in drug development. *Front Oncol* 3:313
158. Ng C, Hauptman R, Zhang Y, Bourne PE, Xie L (2014) Anti-infectious drug repurposing using an integrated chemical genomics and structural systems biology approach. *Biocomputing 2014*. World Scientific, Singapore, pp 136–147
159. Molineris I, Ala U, Provero P, Di Cunto F (2013) Drug repositioning for orphan genetic diseases through conserved antioexpressed gene clusters (CAGCs). *BMC Bioinf* 14(1):288
160. Xu K, Cote TR (2011) Database Identifies FDA-approved Drugs with Potential to be Repurposed for Treatment of Orphan Diseases. *Briefings in bioinformatics* 12(4):341–345
161. Carvalho T (2020) COVID-19 Research in Brief: 30 May to 5 June, 2020. *Nature Medicine*

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

