Chapter 6

Network-Based Drug Repositioning: Approaches, Resources, and Research Directions

Salvatore Alaimo and Alfredo Pulvirenti

Abstract

The wealth of knowledge and omic data available in drug research allowed the rising of several computational methods in drug discovery field yielding a novel and exciting application called drug repositioning. Several computational methods try to make a high-level integration of all the knowledge in order to discover unknown mechanisms. In this chapter we present an in-depth review of data resources and computational models for drug repositioning.

Key words Drug repositioning, Network-based drug repurposing, Drug-target interaction prediction, Precision medicine, Interaction networks

1 Introduction

Drug design and development is a complex, costly, and time-consuming process. Market-ready drug release usually takes from 10 to 15 years and needs more than 1 billion dollars [1]. Furthermore, the success rate for a new drug is very low, usually only 10% per year of drugs succeeds the FDA evaluation and therefore can be used as actual therapy [2]. As a result, pharmaceutical research faces a decreasing productivity in drug development and a persistent gap between therapeutic needs and available treatments [3].

On the other hand, thanks to the advances in genomics and computational methods, our capability of accumulating omics data has rapidly increased. Data such as gene expression, drug-target interactions, protein networks, electronic health records, clinical trial reports, and drug adverse event reports has become accessible in standardized forms.

Such knowledge, although often high-dimensional and noisy, has raised new challenges and fascinating opportunities to design computational methods capable to integrate these data accelerating drug discovery and generating novel insights surrounding drug mechanisms, side effects, and interactions. Following this trend, a

Quentin Vanhaelen (ed.), *Computational Methods for Drug Repurposing*, Methods in Molecular Biology, vol. 1903, https://doi.org/10.1007/978-1-4939-8955-3_6,

very attractive drug discovery technique is drug repositioning [4]. The usage of known drugs for new therapeutic scopes represents a fast and cost-effective strategy for drug discovery. The prevalence of studies has raised a wide variety of models and computational methods to identify new therapeutic purposes for drugs already on the market and sometimes even in disuse. Computational methods try to make a high-level integration of all the knowledge to discover any unknown mechanisms. In [3, 5, 6] a compressive survey on the techniques and models is given.

In this chapter we review the state of the art in drug repositioning methods focusing on network-based approaches and recommendation models. We also provide a comprehensive and up-to-date view of all the data resources available.

2 Strategies for Computational Drug Repositioning

Computational techniques for drug repositioning (or repurposing) are being employed for identifying novel uses for existing, newly developed, and shelved drugs. Indeed, many drugs share multiple protein targets [7, 8], and many complex pathologies share common traits [3, 9] (mutations, pathways, clinical manifestations). For this reason, a drug that acts on these common factors can, in principle, be useful for several diseases. There are many examples of successfully repositioned drugs: from *Minoxidil*, developed for hypertension and now indicated for hair loss [10], to *Sildenafil*, developed for individuals with heart problems and repositioned for erectile dysfunction [11]. However, these examples are based on observations of secondary effects. In this context, computational techniques help to systematically evaluate all possible repurposing candidates, providing high-quality assumptions for the subsequent experimental phases. The aim is identifying drugs that are more efficient and cost-effective than current ones. In this section we will report the main data sources analyzed by repositioning techniques and their computational strategies.

2.1 Representation of Drug Data and Availability

To develop efficient repositioning strategies, it is necessary to appropriately represent compound properties and their interactions. Encoding such information properly allows the development of several analysis and prediction strategies. There are many drug data sources available to date: from chemical structures to interactions with proteins and their perturbations and from the effects on the phenotype to various classifications.

The developments of simulation techniques, such as molecular docking, rely on the efficient representation of molecule chemical structure. Examples of notations include the linear one used by *SMILES* [12] and *InChI* [13]. These notations were designed to be human-readable. However, their analysis can be complicated by

factors such as variable length. This led to the development of fingerprinting techniques [14, 15], where factors such as the presence of recurrent patterns (i.e., specific atoms, atomic groups) are represented through a bit vector. The major databases that collect such information are *PubChem* [16], *ChEMBL* [17], and *DrugBank* [18, 19].

Many experimental techniques can determine drug-target interactions [20]. This information can be represented in binary form (presence/absence) or through numerical indicators (i.e., effective half-maximal concentration EC50, inhibitory half-maximal concentration IC50). The binary representation can give rise to bipartite interaction networks. The numerical one can provide details on the binding affinity, although effective/inhibitory concentrations are often dependent on experimental conditions [21]. Databases such as *DrugBank* [18, 19], *DT-Web* [22], and *STITCH* [23] provide details only in binary form, while *ChEMBL* [17], *PubChem* [16], and *BindingDB* [24] give a more detailed view with data from interaction quantification experiments.

Evaluating compounds effect can also occur through mRNA expression data. Indeed, by comparing pre- and post-administration expression values of a small molecule, perturbation patterns can be determined. Such patterns can be employed as a compound molecular signature. Recently, thanks to experimental costs reduction, many databases have been created to collect such details. The Connectivity Map (*Cmap*) [25] and its recent update [26] provide expression measurement on thousands of cell lines perturbed by small molecule in vitro. *GEO* [27] and *ArrayExpress* [28] are public repository of expression experiments and can also be used in this context. Such data are fundamental to determine the effect of compounds without in vivo experiments.

Beyond the molecular level, phenotypic effects are important to understand drugs functioning. However, in vitro assays do not always give the complete picture. Phenotypic screens are a methodology that can elucidate drug effects, allowing an easily translation in clinic environment [29]. Several databases integrate results coming from multiple phenotypic screens conducted on diverse conditions. *PubChem* contains screens for more than 1.2 million small molecules. *ChEMBL* has more than 14 million assays.

All the abovementioned resources use several ontologies for drug classification. Such ontologies are based, for example, on therapeutic use or pharmacological action. A drug classification provides an additional level of information that can improve analysis by stratifying drugs into hierarchical structures. Drugs in the same level share the same common characteristic. Many examples of drug ontologies are available. *ATC* (anatomical therapeutic class) classifies drug active ingredients based on their chemical and pharmacological characteristics. *ChEBI* [30] ontologies are divided into four sub-ontologies that classify drugs on molecular structures (i.e.,

organic or inorganic), chemical roles (i.e., inhibitors or ligands), biological role (i.e., antibiotic, antiviral agent), and application (i.e., antirheumatic).

Additional metadata are provided by drugs indications and side effects. DrugBank, Pharos [31], and *PharmGKB* [32] are the main sources of therapeutic indications, while *SIDER* (Side Effect Resource) [33] and *FAERS* (FDA Adverse Event Reporting System) are the main databases of side effects and adverse reaction events.

These databases are crucial since the integration of multiple reliable sources can lead to better computational prediction. A brief review of the abovementioned resources is provided in Table 1.

2.2 Computational Techniques

In the past few years, several types of drug repositioning algorithms have been developed. The primary purpose of these techniques is to make systematic the processes that led to the past serendipitous observations in drug discovery. These methods are especially important for rare diseases, where the development of new molecules is not economically sustainable. It is estimated that about 10% of the world population is affected by one of ~7000 rare diseases for which a treatment is not yet available [34]. For these reasons, effective repositioning tools are becoming a pressing need.

Repurposing approaches can be divided into four main categories: (1) target-based, (2) side-effect-based, (3) expression-based, and (4) similarity-based.

Target-centric approaches leverage on the concept of repositioning a drug by exploiting the role of its targets in diseases. Therefore, given a pathology for which a list of relevant targets is known, through the use of drug-target databases, all the possible drugs acting on such a list are evaluated as candidates for repositioning. In [35], for example, authors drew up a list of 15 high-priority genes for *Leishmaniasis major*. Thus, using drug-target interactions from *DrugBank* and *STITCH*, 254 potential FDA-approved drugs were found. Ten had been independently screened against *L. major* through laboratory assay.

Side-effect-based methods are focused on the idea that they can provide clues to new therapeutic applications. For example, patients with benign prostatic hyperplasia treated with finasteride showed unexpected hair growth. This led to the repositioning of such a drug in patients with androgenetic alopecia. In [36], authors built a side-effect-disease association dataset by merging side effects reported in *SIDER* with *PharmGKB* drug-target interactions. This dataset was used to train a Naïve Bayes classification model that predicted possible drugs for 145 diseases using side effects as features. Through a validation process based on a tenfold cross validation procedure, authors showed an AUC higher than 0.8. However, since side effects are only available for drugs at clinical level, approaches using only side-effect details will not be applicable for early-stage assets.

Table 1
List of drug-related resources, their types, and a brief description

Resource	Type	Description	URL
PubChem	General resource	Database of ~96 million compounds, structures, chemical features, bioactivity, and other details	https://pubchem.ncbi.nlm.nih.gov/
ChEMBL	General resource	Database of ~2 million compounds, structures, chemical features, bioactivity, and other details	https://www.ebi.ac.uk/chembl/
DrugBank	General resource	Database of ~10,000 compounds and their DTIs	https://www.drugbank.ca/
DT-Web	DTIs	Resource of predicted DTIs and drug-related tools	https://alpha.dmi.unict.it/dtweb/
STITCH	DTIs	Archive of ~1.6 billion chemical-proteins interaction	http://stitch.embl.de/
BindingDB	DTIs	Database of ~236,000 drug-target binding measurements	https://www.bindingdb.org/
Cmap	Expression data	Expression data of five cancer cell lines exposed to ~1000 compounds	https://portals.broadinstitute.org/cmap/
LINCS	Expression data	Resource of one million expression profiles of drug-perturbed cell lines	http://www.lincsproject.org/
TCGA	Expression data	An archive of RNA-seq, microarray, and other molecular details of over 30 cancer types	https://cancergenome.nih.gov/
GEO	Expression data	A public repository of expression data	https://www.ncbi.nlm.nih.gov/geo/
ArrayExpress	Expression data	A public repository of expression data	https://www.ebi.ac.uk/arrayexpress/
ChEBI	Ontology	Dictionary of molecular entities focused on small molecules	https://www.ebi.ac.uk/chebi/
Pharos	Drug-disease associations	Resource integrating several data sources to shed light on unstudied and understudied drug targets	https://pharos.nih.gov/
PharmGKB	Drug-disease associations	Resource encompassing drug clinical information	https://www.pharmgkb.org/
SIDER	Drug-side-effects associations	A database of side effects and adverse events	http://sideeffects.embl.de/
FAERS	Drug-side-effects associations	Adverse event and medication error reports of the FDA	https://open.fda.gov/data/faers/

Expression-based methods do not suffer from these problems. Indeed, expression profiles can provide details on cellular state in response to a biological perturbation (drug treatment or disease) without any prior knowledge. Moreover, expression profiles can give an unbiased view of the entire coding genome, limiting side effects. The key concept behind these techniques is called *signature reversion* or *signature matching*. A repositioning is performed if a drug-disease pairs has anticorrelated expression profiles. If a gene is perturbed as a result of a disease, a drug that pushes such a gene in the opposite direction could be a therapeutic. In [37], a similar approach was developed. Authors compare expression profiles of 164 small molecules from Cmap with 100 disease signatures derived from GEO datasets. Over 1000 repurposing predictions were produced of which two were experimentally tested in animal model [37, 38]. Although expression-based approaches are more unbiased, several drawbacks can be found. If a drug or a disease does not produce a strong perturbation on gene expression, noisy profiles will be generated, leading to higher false positives. Moreover, *signature reversion* principle might fail if the observed alterations are a result of the disease instead of a cause.

Nowadays, much knowledge about drugs is described in the biomedical literature, frequently providing only indications on drugs applied to specific conditions, without any further detail. Previously described methods cannot be used in such conditions. By exploit the *guilt-by-association* (GBA) principle, this shortcoming can be overcome. In these methods, if two pathologies share at least one common treatment, then some non-shared medication might be therapeutic for both diseases [39]. This approach has been further extended in [40] by adding similarity measures which modulate the strength of the connection between diseases and drugs. Drug similarity can be assessed using chemical structure or known targets or common side effects, while disease similarity can be defined, for example, using ontologies. The approach thus defined is more accurate since it uses multiple data sources on both drugs and diseases but at the same time can be employed when such details are absent.

A missing piece is bridged by electronic health records (EHRs). They offer a promising resource for both generating new hypotheses and building validation cohorts. However, to date these opportunities are not sufficiently explored, since a need for standardization of these data is still needed. Indeed, by analyzing EHRs, an observational study could be performed by extracting unexpected effects, leading to novel therapeutic indications for existing drugs. Moreover, the vast scale of EHRs data could enable large number of parallel drug repositioning tests, without any need of recruiting specific patients. To date however, no EHR-based repositioning studies have been published [3].

3 Network-Based Drug Repositioning

Networks are simple and versatile data structures on which associations can be inferred through many statistical and computational approaches. In biology, the concept of interaction network is heavily used. In such networks, nodes represent components (genes, proteins, complexes), while edges represent interactions between them. Many different relationships between two nodes can be represented simultaneously. Moreover, edges and nodes can be annotated with quantitative information (weights) derived from high-throughput experiments.

The efficacy of such approaches has been proved several times with drug-target interaction prediction. However, these methods are affected by the incompleteness of current knowledge on molecular interactome, leading to noisy results.

Network-based drug repositioning methods can be grouped into categories based on their main source of biological data: (1) gene regulatory networks, (2) metabolic networks, and (3) drug interaction networks. Additionally, integrated approaches, using multiple data sources simultaneously, can be added as a fourth category.

3.1 Gene Regulatory Networks

Expression data can capture information on molecular perturbations that occur due to drug administration or disease. Such data can be exploited to build gene regulatory networks, or to prioritize nodes in existing networks, selecting candidate genes for drug repositioning.

In [41], authors extract possible candidates, starting from disease/control expression data, by exploiting a known regulatory network. The algorithm prioritizes network nodes by combining four different scores using logistic regression. Then, a source of validated drug-target interactions is employed to look for possible repositioning candidates that targets the prioritized genes. The four metrics are *neighborhood scoring*, *interconnectivity*, *random walk*, and *network propagation*. *Neighborhood scoring* evaluates a node on its fold-change and the fold-change of its neighborhood. *Interconnectivity* orders candidates based on their connection to differentially expressed nodes. Given a node, its score is calculated by summing the size of the common neighborhood between the node and differentially expressed genes (DEGs) that are connected to it. *Random walk* is an iterative process that evaluates a node by estimating the probability that it can be reached in a random visit of the network (the initial probabilities are set only on the DEGs). *Network propagation* exploits the concept of resource flow within a network to define a scoring. An initial score is defined (1 for DEGs, 0 otherwise). Then, through an iterative process, initial score is

distributed in the network until the algorithm stabilizes. The result is an evaluation of nodes importance based on network connectivity.

In [42], authors define a method for determining drug targets that may have a strong influence on a disease using a network flow technique. The network is built by merging several protein-protein interaction (PPI) sources with regulatory interactions (gene-transcription factor). A weight is given to each edge in the network by computing the absolute Pearson correlation coefficient of the user-supplied expression data. Then, the algorithm calculates the amount of resource flow that passes between a set of druggable proteins and user-chosen disease genes. The weight on each edge is used as a flow-limiting capacity. Finally, a subset of druggable nodes, which maximized the flow, is used to determine candidate drugs.

Chen et al. [43] developed a method based on *Functional Linkage Network* (FLN) to find inversely correlated drug-disease modules. An FLN is a network where nodes (proteins or genes) are connected by weighted edges measuring the probability of sharing a common biological function. The network is constructed by exploiting different sources of biological information (e.g. mutations, transcript levels) that act as features for a Bayesian classifier, which compute the likelihood for each edge. The FLN is filtered by removing all genes that are not within a user-specified distance from disease-mutated genes and show a differential expression below some threshold. Starting from the filtered FLN, two subnetworks are extracted for each drug: the subnetwork of upregulated disease genes, which are downregulated by the drug, and the network of downregulated disease genes, which are upregulated by the drug. Such networks are processed to determine how much the drug and the disease genes are correlated to extrapolate possible candidates for repositioning.

Although these methods are effective, many limitations make them difficult to use. Firstly, defining a signature for a disease or a drug is not always possible due to noise or weak signals. Furthermore, drug-target genes do not always show altered expression levels and may not be detected. Therefore, expression data should be augmented with additional molecular details to make these methods more robust.

3.2 Metabolic Networks

A different perspective is provided by metabolic networks. A metabolic network is composed by nodes representing chemical compounds and metabolites. Its edges identify reactions that can be catalyzed by one or more enzymes. Commonly, directed edges indicate irreversible reactions while indirect ones reversible reactions. In this representation, an excessive concentration of a compound, due to an enzyme, can result in pathology. Thus, these enzymes can be considered as targets for possible therapies. The

technique typically used for the analysis of such networks is *flux balance analysis* (FBA). FBA uses linear programming to optimize a constrained objective function predicting essential metabolites for disease progression. An appropriate definition of the objective function and its constraints is crucial to correctly model the system to be simulated. FBA is commonly used for diseases caused by pathogens. A common choice of objective function in this context is the estimation of biomass production by a set of essential metabolites.

In [44], authors developed a two-stage FBA model. The first stage finds reactions optimal fluxes and metabolites mass flows in the disease state. The second stage evaluates fluxes and flows in the medication state. Drug targets are identified by comparing the fluxes in both stages.

In [45], authors devised a large-scale FBA model of cancer metabolism to detect the main alterations across many cancer types. Their strategy integrates the human metabolic model with cancer expression data to find a core set of enzyme-coding genes highly expressed across several cancer cell lines. FBA is used to evaluate the impact of such core set on cell proliferation. To predict the final list of drug targets, a greedy search approach is used on the final metabolic network.

3.3 DTI Networks

A common class of repositioning methodologies is based on drug-target interaction (DTI) prediction. Indeed, many drugs frequently show additional targets than designed ones. For this reason, effectively and accurately predicting drug target could show new unintended uses. However, using experimental techniques is an expensive and time-consuming process, so developing reliable computational techniques is of paramount importance.

Typically, DTI prediction algorithms represent the interaction network through a bipartite graph where nodes are drugs or targets and edges are experimentally validated interactions. The purpose of the algorithm is, therefore, predicting novel edges. Sometimes information on known DTIs is aggregated with similarity measures between target pairs or drug pairs to make prediction more accurate. Yamanishi et al. [46] have shown that if two drugs have a similar structure, they will tend to target similar proteins. Likewise, if two target proteins have a similar sequence, they will likely interact with similar drugs.

There are several approaches to predict novel DTIs. For example, in [47] and [48], first-order logic rules are used to determine new predictions. In [46, 49–52], supervised learning methods are applied to learn a DTI model on the whole interaction network augmented with several additional data, such as similarity. *BLM* [53] and its extension *BLM-NII* [54] train classifiers on each drug or target to make local predictions using drug chemical

similarity and sequence similarity to targets. Gonen et al. [55] propose a Bayesian formulation of the problem to predict DTI interaction networks using only similarity information.

Although these methods are efficient, they suffer from some significant limitations. The main one is the inability to make accurate predictions for new drugs (or targets), that is, drugs (targets) with unknown interacting targets (drugs). Furthermore, the lack of experimentally validated negative DTI example often leads to the prediction of a huge number of false positives. The first problem was partially addressed in [6] proposing the use of chemical similarity measures to produce an initial set of target candidates. However, a thresholding problem is still present. The second issue could be solved by randomly choosing negative examples from all non-validated DTIs. However, there is a risk of including, among the negative cases, some undiscovered DTIs, leading to higher false-negative rates.

3.4 Other Network-Based Approaches

Other repositioning approaches based on several molecular networks are available. However, they show limited applicability.

For example, the concept of drug similarity could be exploited to hypothesize new repositioning. The principle is based on the hypothesis that molecules with a similar chemical structure could influence similar proteins. The degree of similarity can therefore be exploited to propose new uses for a drug. *SITAR* [56], for example, uses a logistic regression classifier trained on various similarity measures to predict drug-drug interactions. It builds an interaction network from which repositioning hypotheses are extracted. Similarity measures are computed on compounds' chemical structure, side effects, gene expression profiles, and *ATC* classification. *MANTRA* [57] exploits databases such as DrugBank to build a drug-drug network. Possible similarities, and therefore repositioning, are obtained by identifying communities. However, these approaches are limited by the unreliability of chemical structures and the fact that physiological effect of a drug cannot always be predicted from it.

Other methodologies use associations with side effects to produce new hypotheses. It is well known that all drugs generate side effects because of off-targets. These off-targets can be used to suggest new possible uses or to suggest similar mechanisms of action between multiple drugs. Indeed, drugs with a similar side-effect profile may share the same therapeutic properties. *PREDICT* [58] uses a logistic classifier trained on side-effect similarity using the data in *SIDER*. However, side effects are better detailed only for thoroughly studied drugs. Moreover, having such details for new molecules may take several years. For this reason, *PREDICT* uses other similarity measures based on chemical structure, targets sequence, and proximity of the target in the PPI network.

3.5 Integrated Approaches

Previously described methods represent drug and disease knowledge through different networks. However, each link in the network represents only a partial vision of the biological system. For example, PPI networks identify potential interactions between proteins but do not capture reactions to stimuli. Expression data accurately capture stimuli reactions, but extracting potential interactions from them is difficult, due to noise. For this reason, the integration of heterogeneous data types and sources is necessary to build a complete view of a biological system, resulting in more accurate predictions.

A widely used model is the *ABC* model. Generally, suppose we know through a data source that a disease C has a certain characteristic B (i.e., disease C is caused by a downregulation of gene B) and that a compound A has some effect on B (i.e., drug A restores the expression of B). Then, we can infer that A will influence C (i.e., drug A is a repositioning candidate for disease C). Multiple relationships between A, B, and C give rise to natural way of measuring interaction strength between A and C. Methodologies like *CoPub* [59] and *Yang* et al. [60] are examples of *ABC* model for drug repositioning.

TL_HGBI [61] is a three-layer heterogeneous network repositioning method. The three layers are drugs, targets, and diseases, and their connection is obtained from several databases such as OMIM and DrugBank. Within each layer, interactions are computed using similarity measures. A repositioning is computed by using the flow of information from the drug layer to the disease layer.

SLAMS [62] uses drug-drug, target-target, and side-effects similarity to compute scores between a drug and a disease. Therefore, using a weighted variant of the k-nearest neighbor algorithm, predictions are computed.

PreDR [63] characterizes drugs by chemical structure, target protein similarity, and side-effect similarity. These measures are used to define a kernel function correlating drugs with diseases. Then, a support vector machine (SVM) is trained to predict novel drug-disease interactions.

NRWRH [64] uses a DTI network enriched with drug-drug and target-target interactions computed by using structural and sequence similarity. Therefore a random walk algorithm is applied to predict novel interactions.

In [65], a multilayer network is built, and network projection is used to determine scores for novel DTIs. The layers represent drugs, targets, and target families. Interactions within each layer are computed using similarities.

4 Recommendation Techniques for Drug Repositioning

Recommendation systems are information filtering algorithm developer to infer user preferences for some objects mainly in the field of e-commerce and content delivery. These methods use the *GBA* principle, where users are considered similar if they share common objects. Therefore, products are recommended by using other product from a set of similar users. In the past few years, recommendation systems have been successfully applied for DTI prediction [22, 66, 67] and, more generally, in bioinformatics [68].

A recommender system consists of users and objects. Users collect objects, for which they have a degree of preference, sometimes unknown. The algorithm should be able to infer preference for objects not yet owned, giving higher rating to the ones which will likely appeal the user.

Formally, we denote objects as $O = \{o_1, o_2, \ldots, o_n\}$ and users as $U = \{u_1, u_2, \ldots, u_m\}$. The initial knowledge can be described as a bipartite graph $G(U, O, E, w)$, where E is the set of known user-object interactions and $w : U \times O \rightarrow \mathbb{R}$ is a weight function, representing a score for such pairs.

For each user, the recommendation system will produce object lists, sorted by a scoring function, where higher values correspond to greater probability that the user will like the object.

The principle behind these models can easily be transported to DTI prediction. Objects are replaced by targets and users by drugs. The set of interactions will therefore represent known DTIs. In this representation, the weight function is usually omitted. A recent review on these methods is available in [6].

The idea that drug similarity can be inferred through common targets, using a GBA approach, is a strength of recommendation systems. In [66], authors used the network-based inference (NBI) recommendation algorithm to infer novel DTIs. Given a set of drugs $D = \{d_1, d_2, \ldots, d_n\}$ and a set of targets $T = \{t_1, t_2, \ldots, t_m\}$, the known DTI network can be represented in an adjacency matrix $A = \{a_{ij}\}_{m \times n}$, where $a_{ij} = 1$ if d_j interacts with t_i, $a_{ij} = 0$ otherwise. First, NBI computes weight matrix $W = \{w_{pq}\}_{m \times n}$, where w_{pq} measures target similarity though common drugs. To compute such a value, network projection is employed as:

$$w_{pq} = \frac{1}{k(t_q)} \sum_{l=1}^{n} \frac{a_{pl} a_{ql}}{k(d_l)},$$

where $k(x)$ is the degree of node x in the DTI network. Then, recommendations are computed as $R = W \cdot A$.

However, this approach is not always accurate since it does not consider structural information. Indeed, two drugs might target the same proteins although they are not structurally similar. The

same could happen for targets. In [67], authors present *DT-Hybrid*, a recommendation approach which extends NBI by considering both drugs chemical similarity and targets sequence similarity. The algorithm combines the two measures modulating the results to reduce false predictions. Let $S = \{s_{ij}\}_{n \times n}$ be a target similarity matrix and $S^1 = \left\{ s''_{ij} \right\}_{m \times m}$ a drug structural similarity matrix. To introduce such a similarity in the recommender model, DT-Hybrid builds a processed similarity matrix $S^2 = \left\{ s''_{ij} \right\}_{n \times n}$, where each element measures target similarity through the average similarity of their interacting drugs. In other words, if two targets are linked by many highly similar drugs, then their similarity will be high. S^2 can be computed as:

$$s''_{ij} = \frac{\sum_{k=1}^{m} \sum_{l=1}^{m} \left(a_{il} a_{jk} s'_{lk} \right)}{\sum_{k=1}^{m} \sum_{l=1}^{m} \left(a_{il} a_{jk} \right)}.$$

Matrices S and S^2 can be therefore combined in a final similarity matrix $S^{(1)} = \{s^{(1)}\}_{n \times n}$ as:

$$S^{(1)} = \alpha \cdot S + (1 - \alpha) \cdot S^2,$$

where α is a tuning parameter. Finally, the weight matrix $W = \{w_{pq}\}_{m \times m}$ is computed as:

$$w_{pq} = \frac{S^{(1)}_{pq}}{k(t_q)^{1-\lambda} k(t_p)^{\lambda}} \sum_{l=1}^{n} \frac{a_{pl} a_{ql}}{k(d_l)},$$

where λ is a fundamental parameter that mediates between two different resource distribution processes: an equal distribution among neighbors (as *NBI*) and a nearest-neighbor averaging process. This aspect has been added to *DT-Hybrid* to ensure greater reliability in the presence of very sparse networks, for which less conservative predictions are desired.

Recommendation systems can be employed for drug repurposing in several ways. A first approach has already been explored in both [66] and [6]. Initially, novel DTIs are predicted from known ones. Then, by reasoning on which targets are associated to a disease, prediction can be made.

Another approach, developed in [69], uses collaborative filtering on a drug-disease network to infer novel repositioning. To predict the similar drugs in the collaborative filtering scheme, the algorithm uses similarity measures computed on several data sources, such as drug chemical structure, drug target proteins, and drug-disease associations. Given a candidate drug a for disease q, its predicted result x^*_{aq} combining K similarity sources can be computed as:

$$p^*_{aq} = \sum_{k=1}^{K} \omega_k \times p^k_{aq},$$

where ω_k is a weight measuring reliability of data source k and p_{aq}^k is the prediction based on kth data source. The kth data source prediction can be evaluated as:

$$p_{aq}^k = \bar{s}_a + \frac{\sum_{d \in \mathrm{NN}_a} \mathrm{sim}_{ad}^k \times \left(s_{dq} - \bar{s}_d\right)}{\sum_{d \in \mathrm{NN}_a} \mathrm{sim}_{ad}^k},$$

where NN_a are the top k-nearest neighbor of drug a, s_{dq} is the score of a drug-disease pair in the network, $\underline{s}_x$ is the average score of element x, and sim_{ad}^k is the kth similarity measure for a drug-disease pair.

A third approach applies the *ABC* principle presented in Subheading 3.5 to infer novel associations. The knowledge is represented as a tripartite network and a modified recommendation algorithm is employed to predict associations.

In [68], authors present *ncPred*, a novel recommendation methodology for tripartite networks. Although the algorithm has been applied for noncoding RNA-disease interaction prediction, it lays the foundation for applications to drug repositioning. For example, our knowledge could be represented as a drug-target-disease network. Then, a set of candidate drug-disease indications could be inferred by tripartite recommendation. *ncPred* uses a multilevel resource allocation process, which can be summarized in a cascaded application of *DT-Hybrid*. Thus, such algorithm synthesizes both *GBA* and *ABC* models in a single reasoning. Furthermore, since *DT-Hybrid* is the base on which the *ncPred* is built, further domain-specific knowledge such as drug structural similarity, target sequence similarity, disease similarity through side effects, or ontologies can be easily plugged into the model to make more accurate predictions.

Acknowledgments

This work has been done within the research project "Marcatori molecolari e clinico-strumentali precoci, nelle patologie metaboliche e cronico-degenerative" founded by the Department of Clinical and Experimental Medicine of University of Catania.

References

1. Emmert-Streib F, Tripathi S, de Matos Simoes R et al (2013) The human disease network. Sys Biomed 1:20–28
2. Weng L, Zhang L, Peng Y, Huang RS (2013) Pharmacogenetics and pharmacogenomics: a bridge to individualized cancer therapy. Pharmacogenomics 14:315–324
3. Hodos RA, Kidd BA, Shameer K et al (2016) In silico methods for drug repurposing and pharmacology. Wiley Interdiscip Rev Syst Biol Med 8:186–210
4. Dovrolis N, Kolios G, Spyrou G, Maroulakou I (2017) Laying in silico pipelines for drug repositioning: a paradigm in ensemble analysis for

neurodegenerative diseases. Drug Discov Today 22:805–813
5. Lotfi Shahreza M, Ghadiri N, Mousavi SR et al (2017) A review of network-based approaches to drug repositioning. Brief Bioinform. https://doi.org/10.1093/bib/bbx017
6. Alaimo S, Giugno R, Pulvirenti A (2016) Recommendation techniques for drug-target interaction prediction and drug repositioning. Methods Mol Biol 1415:441–462
7. Paolini GV, Shapland RHB, van Hoorn WP et al (2006) Global mapping of pharmacological space. Nat Biotechnol 24:805–815
8. Koch U, Hamacher M, Nussbaumer P (2014) Cheminformatics at the interface of medicinal chemistry and proteomics. Biochim Biophys Acta 1844:156–161
9. Piro RM (2012) Network medicine: linking disorders. Hum Genet 131:1811–1820
10. Bradley D (2005) Why big pharma needs to learn the three "R"s. Nat Rev Drug Discov 4:446–446
11. Ghofrani HA, Osterloh IH, Grimminger F (2006) Sildenafil: from angina to erectile dysfunction to pulmonary hypertension and beyond. Nat Rev Drug Discov 5:689–702
12. Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Model 28:31–36
13. Heller SR, McNaught A, Pletnev I et al (2015) InChI, the IUPAC international chemical identifier. J Chem 7:23
14. Xue L, Godden JW, Stahura FL, Bajorath J (2003) Design and evaluation of a molecular fingerprint involving the transformation of property descriptor values into a binary classification scheme. J Chem Inf Comput Sci 43:1151–1157
15. Durant JL, Leland BA, Henry DR, Nourse JG (2002) Reoptimization of MDL keys for use in drug discovery. J Chem Inf Comput Sci 42:1273–1280
16. Bolton EE, Wang Y, Thiessen PA, Bryant SH (2008) PubChem: integrated platform of small molecules and biological activities. Ann Rep Comput Chem 4:217–241
17. Gaulton A, Hersey A, Nowotka M et al (2017) The ChEMBL database in 2017. Nucleic Acids Res 45:D945–D954
18. Wishart DS, Feunang YD, Guo AC et al (2017) DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 46: D1074–D1082
19. Wishart DS, Knox C, Guo AC et al (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 34:D668–D672
20. Keiser MJ, Roth BL, Armbruster BN et al (2007) Relating protein pharmacology by ligand chemistry. Nat Biotechnol 25:197–206
21. Yung-Chi C, Prusoff WH (1973) Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. Biochem Pharmacol 22:3099–3108
22. Alaimo S, Bonnici V, Cancemi D et al (2015) DT-Web: a web-based application for drug-target interaction and drug combination prediction through domain-tuned network-based inference. BMC Syst Biol 9(Suppl 3):S4
23. Kuhn M, von Mering C, Campillos M et al (2008) STITCH: interaction networks of chemicals and proteins. Nucleic Acids Res 36: D684–D688
24. Liu T, Lin Y, Wen X et al (2007) BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res 35:D198–D201
25. Lamb J (2006) The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313:1929–1935
26. Duan Q, Flynn C, Niepel M et al (2014) LINCS Canvas Browser: interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures. Nucleic Acids Res 42:W449–W460
27. Edgar R (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30:207–210
28. Kolesnikov N, Hastings E, Keays M et al (2015) ArrayExpress update--simplifying data submissions. Nucleic Acids Res 43:D1113–D1116
29. Zheng W, Thorne N, McKew JC (2013) Phenotypic screens as a renewed approach for drug discovery. Drug Discov Today 18:1067–1073
30. Degtyarenko K, de Matos P, Ennis M et al (2008) ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res 36:D344–D350
31. Nguyen D-T, Mathias S, Bologa C et al (2017) Pharos: collating protein information to shed light on the druggable genome. Nucleic Acids Res 45:D995–D1002
32. Whirl-Carrillo M, McDonagh EM, Hebert JM et al (2012) Pharmacogenomics knowledge for personalized medicine. Clin Pharmacol Ther 92:414–417
33. Kuhn M, Campillos M, Letunic I et al (2010) A side effect resource to capture phenotypic effects of drugs. Mol Syst Biol 6:343

34. Denis A, Mergaert L, Fostier C et al (2010) A comparative study of European rare disease and orphan drug markets. Health Policy 97:173–179
35. Chavali AK, Blazier AS, Tlaxca JL et al (2012) Metabolic network analysis predicts efficacy of FDA-approved drugs targeting the causative agent of a neglected tropical disease. BMC Syst Biol 6:27
36. Yang L, Agarwal P (2011) Systematic drug repositioning based on clinical side-effects. PLoS One 6:e28025
37. Sirota M, Dudley JT, Kim J et al (2011) Discovery and preclinical validation of drug indications using compendia of public gene expression data. Sci Transl Med 3:96ra77
38. Dudley JT, Sirota M, Shenoy M et al (2011) Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. Sci Transl Med 3:96ra76
39. Chiang AP, Butte AJ (2009) Systematic evaluation of drug-disease relationships to identify leads for novel drug uses. Clin Pharmacol Ther 86:507–510
40. Zhang P, Wang F, Hu J (2014) Towards drug repositioning: a unified computational framework for integrating multiple aspects of drug similarity and disease similarity. AMIA Annu Symp Proc 2014:1258–1267
41. Emig D, Ivliev A, Pustovalova O et al (2013) Drug target prediction and repositioning using an integrated network-based approach. PLoS One 8:e60618
42. Yeh S-H, Yeh H-Y, Soo V-W (2012) A network flow approach to predict drug targets from microarray data, disease genes and interactome network - case study on prostate cancer. J Clin Bioinform 2:1
43. Chen H-R, Sherr DH, Hu Z, DeLisi C (2016) A network based approach to drug repositioning identifies plausible candidates for breast cancer and prostate cancer. BMC Med Genomics 9:51
44. Li Z, Wang R-S, Zhang X-S (2011) Two-stage flux balance analysis of metabolic networks for drug target identification. BMC Syst Biol 5 (Suppl 1):S11
45. Folger O, Jerby L, Frezza C et al (2011) Predicting selective drug targets in cancer through metabolic networks. Mol Syst Biol 7:501
46. Yamanishi Y, Araki M, Gutteridge A et al (2008) Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. Bioinformatics 24:i232–i240
47. Fakhraei S, Huang B, Raschid L, Getoor L (2014) Network-based drug-target interaction prediction with probabilistic soft logic. IEEE/ACM Trans Comput Biol Bioinform 11:775–787
48. Fakhraei S, Raschid L, Getoor L (2013) Drug-target interaction prediction for drug repurposing with probabilistic similarity logic. In: Proceedings of the 12th International Workshop on Data Mining in Bioinformatics - BioKDD '13. ACM, New York, NY
49. Chen H, Zhang Z (2013) A semi-supervised method for drug-target interaction prediction with consistency in networks. PLoS One 8: e62975
50. van Laarhoven T, Nabuurs SB, Marchiori E (2011) Gaussian interaction profile kernels for predicting drug-target interaction. Bioinformatics 27:3036–3043
51. van Laarhoven T, Marchiori E (2013) Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile. PLoS One 8:e66952
52. Xia Z, Wu L-Y, Zhou X, Wong STC (2010) Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. BMC Syst Biol 4(Suppl 2):S6
53. Bleakley K, Yamanishi Y (2009) Supervised prediction of drug–target interactions using bipartite local models. Bioinformatics 25:2397–2403
54. Mei J-P, Kwoh C-K, Yang P et al (2013) Drug-target interaction prediction by learning from local information and neighbors. Bioinformatics 29:238–245
55. Gönen M (2012) Predicting drug-target interactions from chemical and genomic kernels using Bayesian matrix factorization. Bioinformatics 28:2304–2310
56. Perlman L, Gottlieb A, Atias N et al (2011) Combining drug and gene similarity measures for drug-target elucidation. J Comput Biol 18:133–145
57. Iorio F, Bosotti R, Scacheri E et al (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. Proc Natl Acad Sci U S A 107:14621–14626
58. Gottlieb A, Stein GY, Ruppin E, Sharan R (2011) PREDICT: a method for inferring novel drug indications with application to personalized medicine. Mol Syst Biol 7:496
59. Frijters R, van Vugt M, Smeets R et al (2010) Literature mining for the discovery of hidden connections between drugs, genes and diseases. PLoS Comput Biol 6:e1000943
60. Yang H-T, Ju J-H, Wong Y-T et al (2017) Literature-based discovery of new candidates for drug repurposing. Brief Bioinform 18:488–497

61. Wang W, Yang S, Zhang X, Li J (2014) Drug repositioning by integrating target information through a heterogeneous network model. Bioinformatics 30:2923–2930
62. Zhang P, Agarwal P, Obradovic Z (2013) Computational drug repositioning by ranking and integrating multiple data sources. In: Blockeel H, Kersting K, Nijssen S, Železný F (eds) Machine learning and knowledge discovery in databases. Springer, Berlin, pp 579–594
63. 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:e78518
64. Chen X, Liu M-X, Yan G-Y (2012) Drug-target interaction prediction by random walk on the heterogeneous network. Mol Biosyst 8:1970–1978
65. Berenstein AJ, Magariños MP, Chernomoretz A, Agüero F (2016) A multilayer network approach for guiding drug repositioning in neglected diseases. PLoS Negl Trop Dis 10:e0004300
66. Cheng F, Liu C, Jiang J et al (2012) Prediction of drug-target interactions and drug repositioning via network-based inference. PLoS Comput Biol 8:e1002503
67. Alaimo S, Pulvirenti A, Giugno R, Ferro A (2013) Drug-target interaction prediction through domain-tuned network-based inference. Bioinformatics 29:2004–2008
68. Alaimo S, Giugno R, Pulvirenti A (2014) ncPred: ncRNA-disease association prediction through tripartite network-based inference. Front Bioeng Biotechnol 2:71
69. Zhang J, Li C, Lin Y et al (2017) Computational drug repositioning using collaborative filtering via multi-source fusion. Expt Syst Appl 84:281–289