Drug Discovery Today • Volume 00, Number 00 • September 2012 PERSPECTIVE

feature

In silico drug repositioning – what we need to know☆

Zhichao Liu[1], Hong Fang[1], Kelly Reagan[2], Xiaowei Xu[2,3], Donna L. Mendrick[2], William Slikker Jr[2] and Weida Tong[2,*], weida.tong@fda.hhs.gov

Drug repositioning, exemplified by sildenafil and thalidomide, is a promising way to explore alternative indications for existing drugs. Recent research has shown that bioinformatics-based approaches have the potential to offer systematic insights into the complex relationships among drugs, targets and diseases necessary for successful repositioning. In this article, we propose the key bioinformatics steps essential for discovering valuable repositioning methods. The proposed steps (repurposing with a purpose, repurposing with a strategy and repurposing with confidence) are aimed at providing a repurposing pipeline, with particular focus on the proposed Drugs of New Indications (DNI) database, which can be used alongside currently available resources to improve in silico drug repositioning.

Introduction

Despite an enormous increase in R&D spending, the number of new drugs being brought to the market has been falling [1]. Mergers and acquisitions can enrich the drug pipeline in the short term but their success has been limited, and in some cases has disrupted R&D activity [2]. In addition, there has been a significant investment on the part of pharmaceutical companies to optimize the drug discovery pipeline using advanced techniques such as structure-based drug design, combinatorial chemistry, HTS and 'omics' technologies. However, the impact of these innovations is not likely to be felt within the foreseeable future [3]. Drug repositioning, or drug repurposing, is 'the process of finding new uses outside the scope of the original medical indications for existing drugs or compounds' and represents a new and promising direction [1].

Candidates for repositioning are usually either marketed drugs or drugs that have been discontinued in clinical trials for reasons other than safety concerns. Because the safety profiles of these drugs are known, clinical trials for alternative indications are cheaper, potentially faster and carry less risk than *de novo* drug development [4]. Among the 51 new medicines and vaccines that were brought to market in 2009, new indications, new formulations and new combinations of previously marketed products accounted for more than 30% [5,6]. Drug repositioning has drawn widespread attention from the pharmaceutical industry, government agencies and academic institutes [7].

Current successes in drug repositioning have primarily been the result of serendipity or clinical observation, such as the observed usefulness of sildenafil for erectile dysfunction and pulmonary arterial hypertension [8], as well as the new indications, including leprosy [9] and multiple myeloma [10], for thalidomide [4,11]. Systematic approaches seem more reasonable and feasible to explore repositioning opportunities. For example, phenotypic screens, which are familiar to drug development researchers, can be used for systematic repurposing [12], however, this approach also requires the additional wet bench work of developing appropriate screening assays for each disease being investigated. Bioinformatics and chemoinformatics offer an unprecedented opportunity to transform this one-drug-at-a-time serendipitous process into a rational and exhaustive exploration of all possible repositioning opportunities for most drugs based on available data sources. This *in silico* drug repositioning applies various strategies [5,13–16] to retrieve, integrate and analyze datasets systematically from diverse sources. One of the key challenges in this approach is

☆ *Disclaimer*: The views presented in this article are those of the authors and do not necessarily reflect those of the FDA.

1359-6446/06/$ - see front matter. Published by Elsevier B.V. http://dx.doi.org/10.1016/j.drudis.2012.08.005

Please cite this article in press as: Z.. Liu, et al., *In silico* drug repositioning – what we need to know, Drug Discov Today (2012), http://dx.doi.org/10.1016/j.drudis.2012.08.005

determining whether a bioinformatics workflow is successful and should be used as a general practice.

The prospective value of repositioning candidates is typically assessed by experimental verification for a few promising compounds using a method such as binding affinity assays. This validation is absolutely essential to the repositioning process but it is time- and resource-intensive and, thus, it makes most sense to use it as a way to validate *in silico* findings rather than as an adjunct tool for optimizing the *in silico* repositioning pipeline in the early discovery stage, which often involves exploring various bioinformatics parameters and methodologies.

The choice of bioinformatics methods is dependent upon the intended purpose of repositioning. In this review, we divide the *in silico* repositioning process into three interconnected steps, as illustrated in Fig. 1. We first describe the key questions in drug repositioning (repurposing with a purpose). Then, we discuss several strategies commonly used for repositioning studies (repurposing with a strategy). Lastly, and most importantly, we emphasize the need for a 'basic truth' database that can be incorporated into the *in silico* repositioning process (repurposing with confidence).

Repurposing with a purpose

As illustrated in the first step in Fig. 1, there are two general approaches to drug repositioning: discovering new indications for an existing drug (drug-centric) and identifying effective drugs for a disease (disease-centric). In the drug-centric space pharmaceutical companies mainly focus on drug candidates demonstrated to be safe in Phase I clinical trials but that have failed owing to efficacy issues in subsequent clinical trials (Phases II and III). It is difficult to estimate how successful this approach is, largely because of the fact that a successful repurposing does not always translate into a marketed drug. Bringing a promising repurposed drug to market is dependent on many factors such as the value of the new indication in terms of market competition and cost:profit tradeoff. Owing to the proprietary nature of drugs that fail in clinical trials, efforts led by most academic institutes and government agencies focus on marketed drugs including prescription, off-patent and over-the-counter (OTC) drugs. Several benefits could arise from repurposing marketed drugs, such as finding new therapies for unmet medical needs, finding more efficacious therapies, replacing expensive with inexpensive drugs, substituting safer drugs for drugs with unwanted effects and broadening the application of efficacious drugs into a broader population. For many complex diseases, such as HIV and many cancers, therapies with only a limited efficacy are available and new approaches are therefore needed.

In the disease-centric space, repositioning studies usually focus on specific diseases, particularly those chronic diseases that lack safe and effective therapeutic options for long-term treatment and disease stabilization, such as inflammatory bowel disease [17]. Another

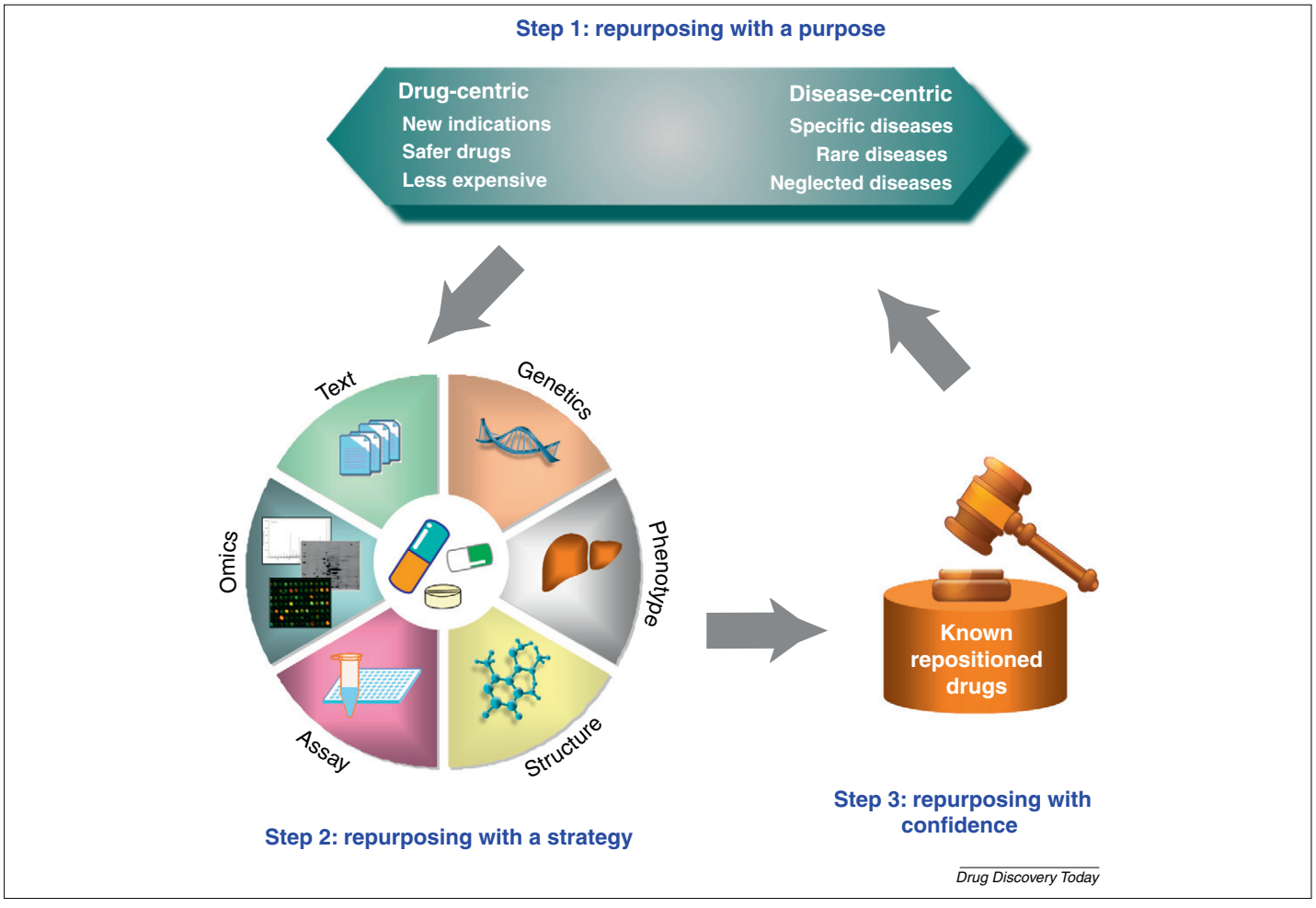

FIGURE 1

The key steps in process of the bioinformatics approach to drug repositioning.

Please cite this article in press as: Z.. Liu, et al., *In silico* drug repositioning – what we need to know, Drug Discov Today (2012), http://dx.doi.org/10.1016/j.drudis.2012.08.005

endeavor in this space is to treat rare and neglected diseases [5]. Rare diseases impact very small populations (less than 200,000 in the USA), whereas neglected diseases are usually tropical infectious diseases affecting developing regions. Because of the low return on investment, these areas have been historically neglected by pharmaceutical companies so the US government agencies are making efforts to address these unmet needs. For example, the FDA has approved a Phase I safety trial of a protease inhibitor (K777) for Chagas disease this year [18]. The US National Center for Advancing Translational Science (NCATS) also includes drug repositioning for rare and neglected diseases as part of its mission [19].

Repurposing with a strategy

The underlying hypotheses in drug-centric and disease-centric repositioning are different. The former hypothesizes that 'similar drugs' have the same therapeutic effects and are equally effective for a disease, whereas the latter assumes that 'similar diseases' need the same therapies and can thus be treated with the same drugs. Although a drug-centric and disease-centric focus affects the choice of bioinformatics approaches and strategies, both encounter the challenge of assessing 'similarity' between drugs or between diseases. Different bioinformatics and chemoinformatics approaches have been explored for this purpose, including docking and SAR as well as emerging approaches such as network pharmacology and systems biology.

Network modeling has a predominant presence in drug repositioning [20], as illustrated in Box 1. A typical network modeling based drug repositioning is illustrated by Nacher and Schwartz [21]. They constructed a drug–therapy network based on the association between drugs and their known therapeutic applications. They assumed that, if two drugs shared the same therapeutic indication, a link (or edge) was established between them. Subsequently, a drug-based network (i.e. drugs as nodes) was established. Using the same approach, but examining whether two disease therapies associated with the same drug, a disease-based network (i.e. diseases as nodes) was also developed. By using network topological measures, such as betweenness centrality and closeness, they found that drugs that were involved in multiple therapies usually have a high centrality value in the drug–therapy network and act on multiple molecular targets in the human system. This approach can be extended to use other types of data (e.g., transcriptomics, phenotypic data) to construct a network.

BOX 1

Network modeling in the context of drug repositioning

Network modeling links repositioning objects in a network format. The network consists of nodes, edges, hubs, modules and outliers (see Figure below), and its biological relevance is measured by the purity of identified modules and topological parameters such as betweenness centrality and closeness. The network can be used to predict novel repositioning opportunities.

Nodes and edges

Nodes can be any one of the repositioning objects including drugs, diseases, targets and modes of action (MoAs). The functional connection between the nodes (named edges) reflects the specific physiological relationship between the nodes. For example, in a drug network where nodes are drugs, the edges represent similarity between drug parameters such as chemical structure or transcriptional responses.

Hubs and outliers

Nodes possessing a high number of functional connections are named hubs (Figure I, dark blue circles). Hubs can be inside the modules or serve to link several modules. In a disease network, for example if one disease is a hub that connects with multiple diseases, it might indicate that the different diseases share commonality (e.g., pathogenesis, genetic mutations). By contrast, outliers are defined as nodes possessing few connections with other nodes. For instance, in a protein–protein network a target is highly connected (a hub) and its inhibition might be involved in multiple biological processes. Thus, a drug that interacts with the target has a potential to be repurposed. A less connected target (an outlier) could be specific for a particular disease, and thus its corresponding drug might have the least potential to be repositioned via this target.

Modules

A module contains a set of nodes that are highly interlinked. It is assumed that the nodes in the same module possess similar biological properties as defined by the module. For example, if a drug module is enriched for a specific therapeutic category all the drugs in the module could be applied for this therapeutic use. The purity is a measure of enrichment.

Betweenness centrality

This statistical measure quantifies the relative importance of a node in the network. For example, in a drug network a drug with a high betweenness centrality value tends to be involved in multiple therapeutic usage and thus has a high potential to be repositioned.

Predict repositioning opportunities

There are two ways to predict repositioning opportunities using network modeling. One method is to establish the previously unknown edge (new functional connection) based on the topological relationships of nodes, shown as a broken red line in Figure I. The other method is to position a new objective (e.g., drug or disease) in the network.

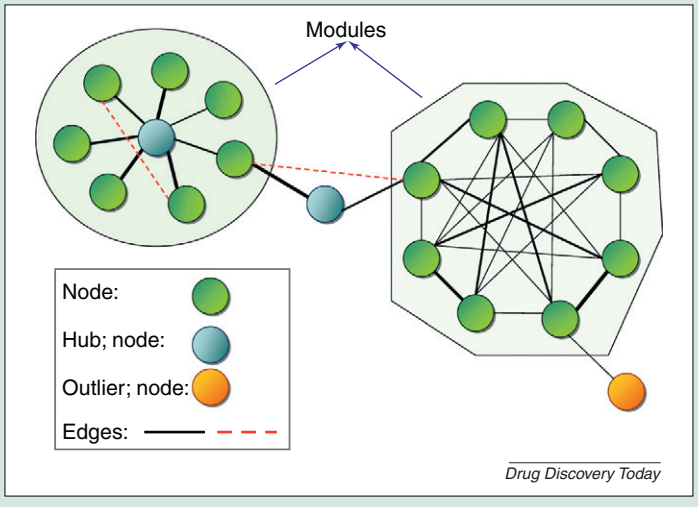

FIGURE I

A diagram of a simple network with two modules.

Please cite this article in press as: Z.. Liu, et al., *In silico* drug repositioning – what we need to know, Drug Discov Today (2012), http://dx.doi.org/10.1016/j.drudis.2012.08.005

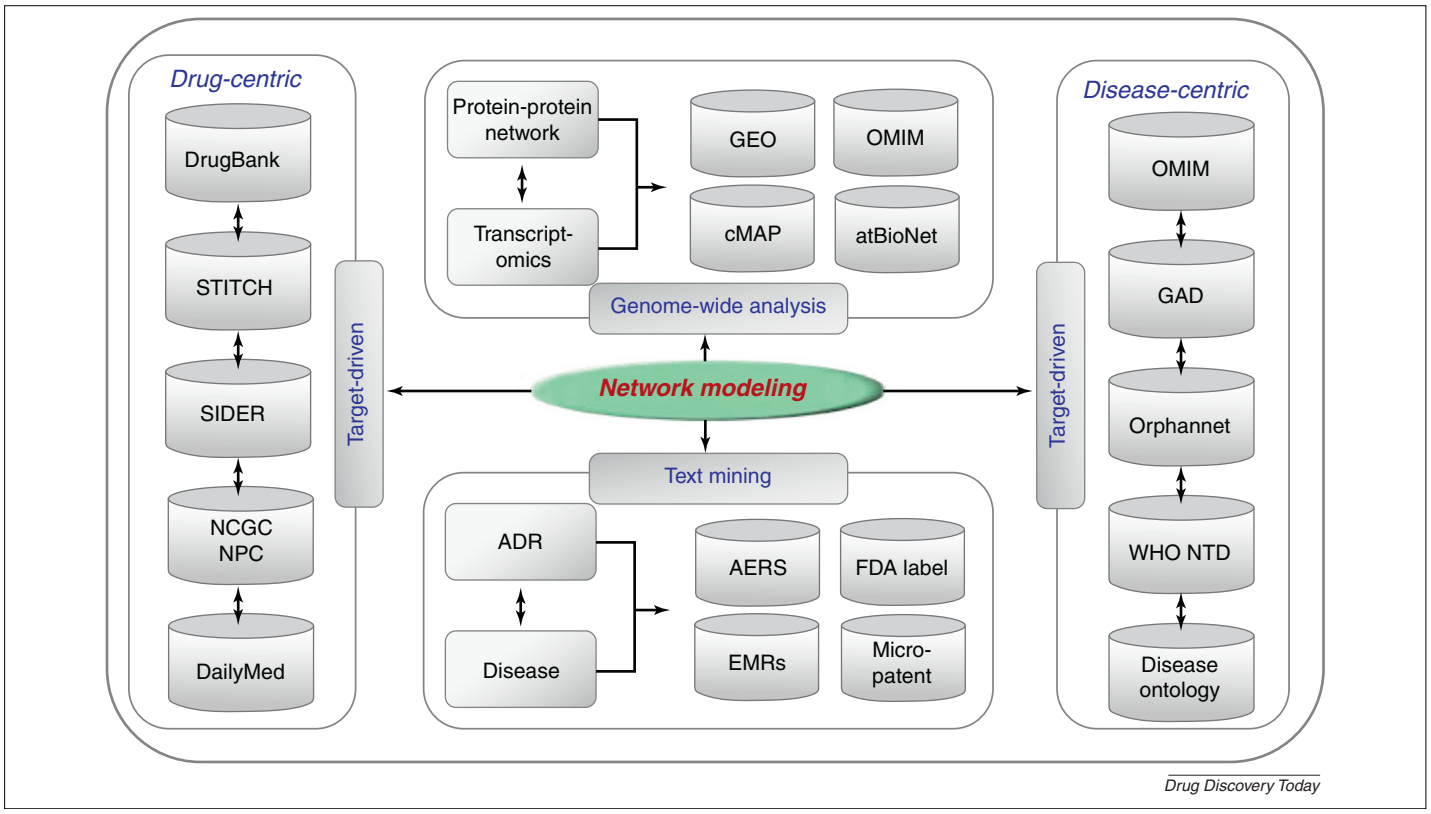

FIGURE 2

Selected resources categorized by strategic contribution to drug repositioning. A brief description and citation of each resource is available from Supplementary Table S1.

In addition to the complexity of choosing an appropriate bioinformatics workflow, there is also a minimum of six different types of data that can be used individually and in combination to measure the similarity of different biological modes, chemical properties and phenotypic information, as shown in the second step of Fig. 1. The available permutations of bioinformatics approaches and different data types yield a complex *in silico* repurposing space. Figure 2 groups the key data sources according to different purposes with different strategies. The domain-specific databases are characterized by three main strategies: target-driven repositioning, genome-wide repositioning and text mining for drug repositioning. Target-driven repositioning databases are subdivided further by their focus on either drug-centric or disease-centric spaces. All the sources, described in more detail in Supplementary Table S1, are organically integrated and organized by using network approaches. The details of strategies with different data sources are summarized below.

Target-driven repositioning

A biological target is the direct link between a drug and a disease. Off-target interactions can be driven by the presence of the known target on other cell types or by drug promiscuity [22]. The binding of a drug to more than one target is a key aspect of drug repositioning. There are several ways to generate a view of target-centric drug promiscuity. One aspect of target-driven repositioning relies on the chemoinformatics approaches such as molecular docking, molecular dynamics and QSAR modeling, which aim to detect the off-target effects of existing drugs and compounds. Currently, an enormous amount of HTS assays, such as the NCGC pharmaceutical collection (NPC) (see Supplementary Table S1), have been made publicly available. These resources offer opportunities to mine and detect drug repositioning candidates with the assistance of chemoinformatics approaches [23]. The most direct way is to explore the global drug–target space using DrugBank [24]. Alternatively, a drug–target network can be built by integrating diverse information such as target protein sequences and drug chemical structures, or by using more-advanced off-target identification methods [25]. However, only a limited number (~320) of drug targets have been reported, so this limits the effectiveness of indication discovery solely based on known drug–target interactions [26]. One way to expand the drug–target interactome is to screen drugs or compounds against multiple types of cultured cells [27]. *In silico* modeling is another way to survey the drug–target interactome. For example, molecular docking methodologies, which take advantage of the large number of protein structures available from the Protein Data Bank, have been applied to drug repositioning [28] to yield information about off-target drug effects and their associated phenotypic outcomes [29–31]. Furthermore, the similarity ensemble approaches (SEA) used to detect the off-target for specific side effects or disease can also be helpful to the systematic identification of repositioning opportunities for existing drugs [32].

Genome-wide repositioning

Most disease-centric repositioning discovery approaches use genome-wide metrics to assess the similarity between diseases. These studies largely rely on the Gene Expression Omnibus (GEO) and the Online Mendelian Inheritance in Man (OMIM). GEO is the largest public data repository for transcriptomic data, whereas OMIM provides a comprehensive collection of Mendelian disorders and their associated genes. Both databases enable a systematic survey of disease similarity in the context of the genome [33,34]. Protein–protein interaction (PPI) networks represent another domain of genome-wide data for disease-centric repositioning

Please cite this article in press as: Z.. Liu, et al., *In silico* drug repositioning – what we need to know, Drug Discov Today (2012), http://dx.doi.org/10.1016/j.drudis.2012.08.005

studies because disease pathways can be constructed from PPI network analysis. Drug targets are involved in multiple disease pathways and can be identified using PPI networks. Integration of the numerous available PPI databases that are experimentally generated and manually curated might enhance the accuracy of a PPI network [35]. Butte *et al*. provide concrete examples (e.g., new indications for cancer and inflammatory bowel disease) of how to reinterpret and compare genome-wide metrics to explore which drugs can be repurposed for alternative diseases [17,36].

Text mining for drug repositioning

Literature-driven repositioning has the potential to have a broad impact on drug-centric and disease-centric approaches because its techniques, possibly combined with other tools in bioinformatics and chemoinformatics, can be used to develop *de novo* discoveries in a systematic manner [37]. There exists a broad range of text-based databases that can be used for indication discovery, including OMIM and PubMed. Given the fact that over half a terabyte of data including PubMed, electronic medical records (EMRs) and patent filings are available for use in repositioning, we will see an increased application of text mining in these studies.

Repurposing with confidence

Among current practices, validation of a repurposing finding is predominantly dependent on wet-lab experiments including *in vitro* and *in vivo* assays as well as controlled population studies. There is no doubt that such a validation strategy is essential for confirmation. Unfortunately, the variation in bioinformatics methods used and the choice of different data types available present a huge number of possible strategies to explore the repositioning space, making a comprehensive experimental validation strategy difficult. There is a pressing need in this field to have a 'basic truth' approach that can be used to help select plausible repositioning candidates before experimental validation takes place, as illustrated in the third step in Fig. 1.

We have developed a database for this purpose called Drugs of New Indications (DNI), as shown in Supplementary Table S2. The data are mainly obtained from literature review [4], the drug repositioning Wiki [38] and the Rare Disease Repurposing Database (RDRD). Currently, the DNI contains 237 drugs with original and new indications along with additional information (e.g., therapeutic use, chemical data, target information) extracted from DrugBank v3.0 [39]. The database is currently under development for release to the public. Other sources, such as the CDD database [40] which contains more than 100 repositioning candidates identified by HTS methods, will be considered for inclusion in DNI in the future. The caution for such an approach comes from the reliance on published reports originating from many investigators and sites without overall quality control. However, the large volume of data and repetition of studies might enable the true connections to be identified.

Concluding remarks

Drug repositioning has economic and public health benefits for drug makers, regulatory agencies, patients and taxpayers. A rational way to search for repositioning opportunities is an important step in optimizing the drug repositioning pipeline. Enormous amounts of data generated by various techniques, in different formats and from diverse domains are available to researchers who need strategies and tools to retrieve, organize and mine these resources effectively for drug repositioning possibilities. The bioinformatics task is to identify potential repositioning candidates systematically. Here, we have proposed the crucial steps involved in an *in silico* drug repurposing pipeline – to help direct the pipeline and improve the success rate in this area.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Acknowledgements

ZL and XX are grateful to the FDA National Center for Toxicological Research (NCTR) for postdoctoral support through the Oak Ridge Institute for Science and Education (ORISE).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.drudis.2012.08.005.

References

1 Chong, C.R. and Sullivan, D.J. (2007) New uses for old drugs. *Nature* 448, 645–646
2 Shibayama, S. *et al.* (2008) Effect of mergers and acquisitions on drug discovery: perspective from a case study of a Japanese pharmaceutical company. *Drug Discov. Today* 13, 86–93
3 Usdin, S. (2002) Industry development: pipeline or flatline? *BioCentury* 1
4 Ashburn, T.T. and Thor, K.B. (2004) Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* 3, 673–683
5 Sardana, D. *et al.* (2011) Drug repositioning for orphan diseases. *Brief. Bioinform.* 12, 346–356
6 Graul, A.I. *et al.* (2010) The year's new drugs & biologics – 2009. *Drug News Perspect.* 23, 7–36
7 Oprea, T.I. *et al.* (2011) Drug repurposing from an academic perspective. *Drug Discov. Today: Ther. Strat.* 8, 61–69
8 Galiă, N. *et al.* (2005) Sildenafil citrate therapy for pulmonary arterial hypertension. *N. Engl. J. Med.* 353, 2148–2157
9 Silverman, W.A. (2002) The schizophrenic career of a "monster drug". *Pediatrics* 110, 404–406
10 Singhal, S. *et al.* (1999) Antitumor activity of thalidomide in refractory multiple myeloma. *N. Engl. J. Med.* 341, 1565–1571
11 Yang, T. and Liang, H. (1963) Thalidomide and congenital abnormalities. *Lancet* 1, 552–553
12 Reaume, A.G. (2011) Drug repurposing through nonhypothesis driven phenotypic screening. *Drug Discov. Today: Ther. Strat.* 8, 85–88
13 Ekins, S. *et al.* (2011) *In silico* repositioning of approved drugs for rare and neglected diseases. *Drug Discov. Today* 16, 298–310
14 Deftereos, S.N. *et al.* (2011) Drug repurposing and adverse event prediction using high-throughput literature analysis. *Wiley Interdiscip. Rev.: Syst. Biol. Med.* 3, 323–334
15 Dudley, J.T. *et al.* (2011) Exploiting drug–disease relationships for computational drug repositioning. *Brief Bioinform.* 12, 303–311
16 Loging, W. *et al.* (2011) Cheminformatic/bioinformatic analysis of large corporate databases: application to drug repurposing. *Drug Discov. Today: Ther. Strat.* 8, 109–116
17 Dudley, J.T. *et al.* (2011) Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Sci. Transl. Med.* 3 96ra76
18 Leslie, M. (2011) Drug developers finally take aim at a neglected disease. *Science* 333, 933–935
19 Collins, F.S. (2011) Mining for therapeutic gold. *Nat. Rev. Drug Discov.* 10, 397
20 Spiro, Z. *et al.* (2008) Drug–therapy networks and the prediction of novel drug targets. *J. Biol.* 7, 20
21 Nacher, J. and Schwartz, J.-M. (2008) A global view of drug–therapy interactions. *BMC Pharmacol.* 8, 5
22 Hopkins, A.L. (2009) Drug discovery predicting promiscuity. *Nature* 462, 167–168
23 Swamidass, S.J. (2011) Mining small-molecule screens to repurpose drugs. *Brief Bioinform.* 12, 327–335
24 Yildirim, M.A. *et al.* (2007) Drug–target network. *Nat. Biotechnol.* 25, 1119–1126
25 Campillos, M. *et al.* (2008) Drug target identification using side-effect similarity. *Science* 321, 263–266
26 Overington, J.P. *et al.* (2006) How many drug targets are there? *Nat. Rev. Drug Discov.* 5, 993–996
27 Huang, R. *et al.* (2011) The NCGC Pharmaceutical Collection: a comprehensive resource of clinically approved drugs enabling repurposing and chemical genomics. *Sci. Transl. Med.* 3 80ps16
28 Kinnings, S.L. *et al.* (2011) A machine learning-based method to improve docking scoring functions and its application to drug repurposing. *J. Chem. Inform. Model.* 51, 408–419
29 Li, H. *et al.* (2011) Fragment-based drug design and drug repositioning using multiple ligand simultaneous docking (MLSD): identifying celecoxib and template compounds as novel inhibitors of signal transducer and activator of transcription 3 (STAT3). *J. Med. Chem.* 54, 5592–5596

Please cite this article in press as: Z.. Liu, et al., *In silico* drug repositioning – what we need to know, Drug Discov Today (2012), http://dx.doi.org/10.1016/j.drudis.2012.08.005

30 Li, Y.Y. *et al.* (2011) A computational approach to finding novel targets for existing drugs. *PLoS Comput. Biol.* 7, e1002139
31 Xie, L. *et al.* (2011) Drug discovery using chemical systems biology: weak inhibition of multiple kinases may contribute to the anti-cancer effect of nelfinavir. *PLoS Comput. Biol.* 7
32 Lounkine, E. *et al.* (2012) Large-scale prediction and testing of drug activity on side-effect targets. *Nature* 486, 361–367
33 Hu, G.H. and Agarwal, P. (2009) Human disease–drug network based on genomic expression profiles. *PLoS ONE* 4, 11
34 Goh, K.I. *et al.* (2007) The human disease network. *Proc. Natl. Acad. Sci. U. S. A.* 104, 8685–8690
35 Cerami, E.G. *et al.* (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.* 39, 685–690
36 Sirota, M. *et al.* (2011) Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci. Transl. Med.* 3 96ra77
37 Agarwal, P. and Searls, D.B. (2009) Can literature analysis identify innovation drivers in drug discovery? *Nat. Rev. Drug Discov.* 8, 865–878
38 Drug repositioning Wiki. Available at: http://www.linkedin.com/news?viewArticle=&articleID=382887904&gid=3705627&type=member&item=44408536&articleURL=http%3A%2F%2Fwww%2Edrugrepurposing%2Einfo&urlhash=tsrx&goback=%2Egmp_3705627%2Egde_3705627_member_44408536.
39 Knox, C. *et al.* (2011) DrugBank 3.0: a comprehensive resource for 'Omics' research on drugs. *Nucleic Acids Res.* 39, 1035–1041
40 Hohman, M. *et al.* (2009) Novel web-based tools combining chemistry informatics, biology and social networks for drug discovery. *Drug Discov. Today* 14, 261–270

Zhichao Liu[1]
Hong Fang[1]
Kelly Reagan[2]
Xiaowei Xu[2,3]
Donna L. Mendrick[2]
William Slikker Jr[2]
Weida Tong[2]
[1]*ICF International at FDA National Center for Toxicological Research, 3900 NCTR Rd, Jefferson, AR 72079, USA*
[2]*National Center for Toxicological Research, US Food and Drug Administration, 3900 NCTR Road, Jefferson, AR 72079, USA*
[3]*Department of Information Science, University of Arkansas at Little Rock, 2801 S. University Ave., Little Rock, AR 72204-1099, USA*

Please cite this article in press as: Z.. Liu, et al., *In silico* drug repositioning – what we need to know, Drug Discov Today (2012), http://dx.doi.org/10.1016/j.drudis.2012.08.005