Exploiting drug–disease relationships for computational drug repositioning

Joel T. Dudley, Tarangini Deshpande and Atul J. Butte

Abstract

Finding new uses for existing drugs, or drug repositioning, has been used as a strategy for decades to get drugs to more patients. As the ability to measure molecules in high-throughput ways has improved over the past decade, it is logical that such data might be useful for enabling drug repositioning through computational methods. Many computational predictions for new indications have been borne out in cellular model systems, though extensive animal model and clinical trial-based validation are still pending. In this review, we show that computational methods for drug repositioning can be classified in two axes: drug based, where discovery initiates from the chemical perspective, or disease based, where discovery initiates from the clinical perspective of disease or its pathology. Newer algorithms for computational drug repositioning will likely span these two axes, will take advantage of newer types of molecular measurements, and will certainly play a role in reducing the global burden of disease.

Keywords: bioinformatics; drug repositioning; drug development; microarrays; gene expression; systems biology; genomics

The number of new drugs approved per dollar spent or pharmaceutical research and development productivity has significantly declined in recent decades [1]. By conservative estimates, it now takes ~15 years [2] and $800 million to $1 billion to bring a single drug to market [3]. There are two major reasons for this decline in the total number of safe and effective new drugs reaching the market. The first is that prevalent drug development strategies within pharmaceutical companies remain conservative, typically oriented on discovery of a new therapeutic target combined with a search for a novel therapeutic compound that modulates the activity of the identified target. This is followed by a slow, costly and risky process of experimental and clinical validation.

The second major reason for reduced productivity is the lack of systematic evaluation of additional indications that each drug can target, both during the drug's development phase and subsequent to its arrival on the market. Some of the most profitable and successful pharmaceuticals did not begin development for their current indications, but instead were re-purposed or repositioned for new uses [4]. Accidental discovery, unintended side effects or obvious follow on indications have led to new uses of such drugs. Classic examples include Minoxidil (originally tested for hypertension; now indicated for hair loss), Viagra (originally tested for angina; now indicated for erectile dysfunction and pulmonary hypertension), Avastin (originally indicated for metastatic colon cancer and nonsmall-cell lung cancer; later approved for metastatic breast cancer) and Rituxan (originally indicated for non-Hodgkin's Lymphoma; later approved for chronic lymphocytic leukemia and rheumatoid arthritis).

Revenues generated by repositioned drugs can exceed billions: sales of thalidomide, repositioned for multiple myeloma, reached US $271 million in 2003 alone [5]; sildenafil, repositioned for erectile dysfunction, had annual sales of US $1.88 billion in 2003 [4]. While the revenues generated by repositioned drugs have been substantial, the real incentive

Corresponding author. Atul Butte, MD, PhD, Pediatrics/Systems Medicine, Stanford University School of Medicine, 251 Campus Drive Room X-163, Stanford, CA 94305-5415, USA, Tel: +650-723-3465; Fax: +650-723-7070; E-mail: abutte@stanford.edu

Joel Dudley is a Bioinformatics Specialist at Stanford, developing genomics-based classifications of human disease and studying medical evolution. Joel was previously at Arizona State University, involved with developing MEGA.

Tarangini Deshpande holds a PhD from Purdue University, and is a founder of NuMedii, a company using a comprehensive genomics computational platform for new indications discovery.

Atul Butte is an Assistant Professor and Chief of the Division of Systems Medicine in the Department of Pediatrics at Stanford, and has authored over 100 publications.

for repositioning is the clear benefit for patients. For example, thalidomide's antiangiogenic properties have provided therapeutic benefits to multiple myeloma patients, who otherwise had few treatment options for their disease, while the central dopamine agonist properties of bromocriptine recent led to its approval in the USA for a new indication of Type 2 diabetes [6].

Drug repositioning or repurposing (i.e. finding a new use for an existing drug) can provide solutions to the problems facing pharmaceutical companies. Such efforts have spanned the spectrum from traditionally blind screening methods of chemical libraries against specific cell lines [7] or cellular organisms [8, 9], to serial testing of animal models [10], to the newer data-driven approaches involving computational methods. The latter category typically takes advantage of the fact that a single molecule can act on multiple targets and could be beneficial to indications where the additional targets are relevant (the known compound–new target approach). In fact, there is strong evidence that such off-target interactions, or polypharmacology, are common among many approved drugs compounds [11]. Additionally, repositioning efforts also leverage the fact that mechanisms and targets are shared between diseases or biological processes enabling drugs that work on a target in one work for the other (known targets in a new indication). Drug repositioning has several advantages compared to a traditional approach to the development of a drug *de novo*.

The drug development cycle for a repositioned drug can be as short as 3–12 years compared to the traditional 10–17 years required to bring a new chemical entity to market [12]. This is due to the fact that several steps of the drug development pipeline can be eliminated during repurposing efforts. However, the discovery of a new use of a drug for a new condition can be a haphazard process, as illustrated by the examples given above, where new indications were found through side effects, or through exploiting useful properties of these drugs [13], demonstrated in the utility of arsenic for acute promyelocytic leukemia [14], amphotericin B for leishmaniasis [15] and thalidomide for multiple myeloma [16]. While physicians, pharmaceutical and biotechnology companies have manual methods and prior knowledge that enable repositioning of drugs through clinical trials, these occurrences are often serendipitous and rare. One challenge in drug

repositioning, therefore, lies in predicting and choosing new therapeutic indications to prospectively test for a drug of interest. This review will focus on computational approaches for guiding and selecting new indications for drugs.

COMPUTATIONAL REPOSITIONING STRATEGIES

Many computational strategies for drug repositioning have been published. One way to classify these methods is by categorizing them as either 'drug based', where discovery of repositioning opportunities initiates from the chemical or pharmaceutical perspective, or 'disease based', where discovery initiates from the perspective of disease management, symptomatology or pathology (Figure 1). Drug-based approaches might be preferred, if there is interest or expertise in modeling or understanding more precise pharmacological properties leading to repurposing opportunities, or if rich pharmacological or chemical data for drugs is available. Disease-based approaches may be preferred to overcome missing knowledge in the pharmacology of a drug (e.g. unknown or additional targets), or if repositioning efforts are to be focused on a specific disease or therapeutic category. While each of these approaches present unique informatics challenges, successful repositioning strategies often incorporate elements from both drug- and disease-based methods. Here, we discuss several specific examples of computational approaches developed from these perspectives.

DRUG-BASED COMPUTATIONAL APPROACHES
Chemical similarity

The structure and chemical properties of a drug compound itself are evidently associated with its ultimate effective use as a therapy. It is, therefore, possible to explore repositioning opportunities for a drug compounds based on shared chemical characteristics. The rational basis for this approach is rooted in known quantitative relationships between chemical structures and biochemical activity (QSAR). Although similar structures do not always behave the same in biological systems, the degrees of similarity that exist can be exploited using computational approaches for drug repositioning. The computational basis of chemical similarity approaches is to extract a set of chemical features for each drug in a
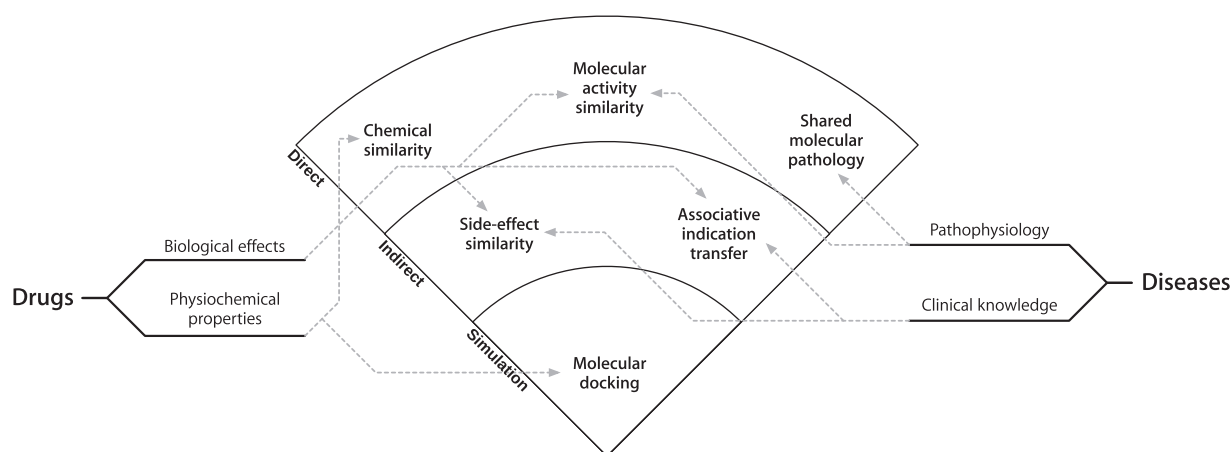
Figure 1: Published drug repositioning strategies described in this review are organized according to their primary mode of inference. The dashed arrows connect high-level informational aspects of drugs and diseases with the methods that incorporate these types of information in their approach. Methods are generally categorized as focusing largely on either (i) 'Direct' inference, where established or directly measured biomolecular or chemical properties are used to infer therapeutic relationships between drugs and diseases, (ii) 'Indirect' inference, where related or higher level data or representations of drugs and diseases is used to infer therapeutic relationships between drugs and diseases or (iii) 'Simulation', where therapeutic interactions are inferred through simulation of interactions between drugs and diseases rather than through direct or indirect measurement of their salient properties. We predict that newer methods will move toward integrating multiple forms of therapeutic inference incorporating many forms of both drug- and disease-based data and knowledge to enable the discovery of new uses for drugs—as some of the methods described in this review have taken steps toward.

set of drugs, and then to relate the drugs directly to each other by clustering or constructing networks based on the extracted features [17]. Drug repositioning opportunities can then be inferred by simple chemical association, or by looking for particular biological features, such as known drug targets, enriched in the resulting relationships.

As an example, to discover novel targets for metabotropic glutamate receptor (mGluR) antagonists, Noeske *et al.* [18] extracted standardized pharmacophore descriptors for a collection of known mGluR and also for a broader set of known drug compounds with diverse known drug targets. These descriptors were then used to project the compounds on to a self-organizing map (SOM), which revealed distinct subclusters of mGluR antagonists, and also overlapping localization with ligands known to bind to histamine (*H1R*), dopamine (*D2R*) and several other targets. These predicted interactions were subsequently confirmed by experimental validation, which showed weak but significant binding affinities between mGluR antagonists predicted off-targets.

Keiser *et al.* [11] took an integrated chemical similarity approach to drug repositioning that incorporated both the structural similarity between drug compounds, as well as knowledge of established compound–target relationships. In this approach, drug targets were represented by the set of ligands which were known to bind to them. To evaluate a possible novel association between an established drug (query compound) and an off-target, a score was derived by calculating the sum of the structural similarity between the query compound and each member of the set of ligands known to bind to the target. The pair-wise similarity score was computed as the Tanimoto coefficient, representing the structural similarity between two compounds based on 2D chemical fingerprint descriptors of their chemical structure. A statistical model based on the extreme value distribution was then used to determine a significant score, and scores surpassing the significance threshold indicated a probable association between the query drug and the target. Several of the off-target interactions predicted by this method were experimentally confirmed by binding assays, and a predicted interaction between N,N-dimethyl-tryptamine and serotonergic receptors was confirmed in a knockout mouse model. Other sources of chemical information can also be used to make associations [19].

Limitations of the chemical similarity approach for drug repositioning largely stem from the fact that

many structures [20] and other chemical properties of known drug compounds contain errors, or are withheld as proprietary information. Furthermore, many physiological effects cannot be predicted by chemical properties alone, because drugs undergo complex, and largely uncharacterized, metabolic transformations and other pharmacokinetic transformations as they are metabolized and physiologically distributed.

Molecular activity similarity

Another way in which drugs can be related to other drugs and disease states for the purpose of repositioning is by computational assessment of similarities in molecular profiles. When a pharmacologically active compound is exposed to a biological system, the result is a perturbation of the biological system through the compound's mechanism of action (MOA). Although the precise MOA is not well-understood for many approved compounds, high-throughput molecular measurement techniques, such as gene expression microarrays, can be used to measure and represent the total molecular activity of a compound in a biological system. In this way, it becomes possible to construct a 'signature' of the molecular activity of a drug compound acting in a biological system. These signatures of molecular activity can then be compared to establish therapeutic relationships between drugs and diseases even in cases when a drug's MOA or even primary target is unknown.

One of the most comprehensive and systematic approaches toward leveraging the molecular activity approach for drug repositioning is the Connectivity Map project [21]. The Connectivity Map currently provides a reference collection of gene expression based molecular activity profiles for 1309 compounds, which were obtained by systematically exposing the compounds to a few key cancer cell lines and measuring the genome-wide transcriptional response. The molecular activity profile of each drug in the reference collection contains, for each gene measured, a rank-based measure of the change in transcriptional activity after exposure to the drug compound. These profiles can be used as the basis of comparison to connect drugs to other drugs and diseases based on shared molecular activity.

As one example of this utility, Iorio *et al.* [23] computed the pair-wise similarity between the molecular activity signatures of all drug compounds represented in the Connectivity Map using a novel, rank-based metric based on Gene Set Enrichment Analysis (GSEA)[22]. Drugs were then organized into a network using the resulting similarity scores, and a network partitioning strategy based on 'affinity propagation' [24] was applied to cluster the network into coherent 'communities' of drugs. The resulting drug communities comprised drugs with similar MOAs, which often shared canonical targets and pathways. Through this approach, repositioning opportunities are revealed by co-location of drugs within the network clusters, which suggests a shared molecular activity with other drugs in the cluster. The authors used this approach to infer previously unknown cellular autophagy activity for the rho-kinase inhibitor Fasudil, which was further supported by experimental validation of predicted target levels.

Another way the Connectivity Map could be used for drug repositioning is to directly compare the molecular activity signatures of drugs with those of a disease state. Since disease pathology can similarly be viewed as a perturbation of a biological state, the same approach can be applied to measure the genome-wide transcriptional changes in the disease condition and generate a signature or profile of its molecular activity. The resulting disease signature provides a common basis by which the molecular activity of a disease can be compared to the molecular activity of a drug to identify novel therapeutic opportunities. Wei *et al.* [25] successfully applied this approach to identify the mTOR inhibitor rapamycin as a modulator of glucocorticoid (GC) resistance in acute lymphoblastic leukemia (ALL). They first derived a gene expression signature GC resistance by comparing gene expression profiles of GC-resistant and GC-sensitive ALL samples. They then compared the GC-resistance signature with the drug compound molecular profiles in the Connectivity Map using the nonparametric GSEA method [22], which revealed a significant enrichment of concordant gene expression changes between gene transcripts the GC-resistance signature and those in the molecular activity profile of rapamycin.

Strategies for drug repositioning based on molecular activity similarity are not limited to analysis of transcriptional response, and may incorporate any number of high-dimensional molecular characterizations of drug effect; such as chemical screening assays or high-throughput gene knockout assays. Chen *et al.* [26] used data from the PubChem bioassay repository to create molecular activity profiles for

represented drug compounds, which were organized into similarity networks based on their bioassay activity profiles. The resulting bioassay network was then mapped on to biological networks constructed from metabolic pathways and protein–protein interactions using a bipartite mapping scheme that considered the sequence similarity between protein targets in the assay nodes and the protein sequence nodes in the biological network. The resulting mappings between the PubChem bioassay and biological networks provide representation and interpretation of the biological activity of compounds within the context of biological systems, with the goal of understanding how the compounds might perturb a biological system toward efficacy or adverse effects.

The primary limitation of drug repositioning strategies based on molecular activity similarity is their heavy reliance on the quality and assumptions of the means used to derive molecular activity profiles. For example, the Connectivity Map is derived by exposing the whole drug compound to isolated cell lines, which may not accurately reflect the biological activity of the drug in a complete physiological system. Many drugs undergo chemical transformations after they are metabolized *in vivo*, and in fact the drug metabolites often provide the eventual therapeutic effect. Furthermore, the pathology of many disease conditions, including metabolic diseases such as Type 2 diabetes, spans multiple tissues and organ systems; therefore, it might be difficult to represent and compare such diseases on the basis of a single molecular activity signature.

Molecular docking

Molecular docking comprises a set of computational methods that aimed at discovering novel relationships between chemical ligands and targets through use of simulation and modeling of their direct physical interaction [27]. Molecular docking methods can enable drug repositioning by attempting to predict physical interactions between existing compounds and novel therapeutic targets. If a drug is predicted to physically interact with a previously unknown target, then the drug might be considered as a possible repositioning candidate for disease conditions in which the predicted target is known or suspected to play a role, or perturb the molecular pathology of the disease. Molecular docking approaches can be used on a target-by-target basis to look for repositioning strategies for a particular target of interest, or to establish networks of ligand–target interactions to

explore drug repositioning opportunities across systems of predicted drug–target interactions.

Zahler *et al.* [28] describe a virtual 'inverse' screening approach to identify novel targets of the kinase inhibitor indirubin. Beginning with indirubin as a chemical of interest, they sequentially screened its ligand against a database of kinase receptor structures through molecular docking to discover and validate a novel interaction between an indirubin derivative and *PDK1*. Kinnings *et al.* [29] employed molecular docking in a chemical systems biology approach to reposition Entacapone, a catechol-*O*-methyl transferase (*COMT*) inhibitor used to treat Parkinson's disease, as a treatment for multi-drug resistant (MDR) tuberculosis. They began by extracting target-binding sites for approved drugs from the 3D structures of their known targets, and then performed a computational search to identify putative off-target proteins with similar ligand binding sites. They then used molecular docking to evaluate candidates from the binding site similarity search for physical interaction with the associated drug ligand, retaining predicted interactions as putative novel off-targets of the drug. Focusing on off-target predictions for the protein enoyl–acyl carrier protein reductase (InhA), which is involved in synthesis of the bacterial cell wall, they used this approach to identify and validate antagonistic ligands of a binding site extracted from *COMT*. Taking this approach further, Chang *et al.* [30] have recently shown that structure-based predictions can also be filtered based on other data- and knowledge sources, such as metabolic networks, tissue localization and gene expression patterns.

There are some potential limitations to the use of molecular docking in drug repositioning. Foremost, the approach typically requires that the 3D structure of both the chemical ligand and the protein target are well resolved. At present, the structures of many physiologically important proteins are not fully resolved; including whole families of G-protein coupled receptors (GPCRs), which are favored as drug targets for many approved drugs. Additionally, the results of molecular docking are known to incur high false-positive rates, due to errors in resolved protein structures and incomplete modeling of atomic and molecular interactions [31]. Despite these challenges, it is clear that there are tens of thousands of drugs waiting for an indication with available information [32], and multiple drug-based approaches are likely to be fruitful.

DISEASE–BASED COMPUTATIONAL APPROACHES

Associative indication transfer

One approach to leveraging disease-based information for drug repurposing is to utilize knowledge of drug indications for disease. Chiang *et al.* [33] used a 'guilt by association' approach to discover novel drug indications based on the similarity of their efficacious indications. Diseases were deemed as similar to each other if they already shared a significant number of therapies. Across each pair of similar diseases, those remaining drugs that were currently used against only one of the pair were then considered as logical candidates as drugs for the other disease in the pair. Novel drug-indication associations could then be inferred by associating drugs with novel indications by expanding from simple pairs into network clusters.

However, repositioning strategies based on the associative transfer of indications are limited by the varied and complex relationships that associate a drug as an indication for a particular disease condition. For example, many drugs are indicated as palliative treatments for various cancers, which are difficult to discern from those drugs indicated as primary chemotherapeutic treatments. At present, there is no comprehensive, systematic representation of known drug indications that would enable such fine-scale delineation of types of drug–disease relationships; however, efforts to construct systematized drug ontologies and other resources are underway [34].

Shared molecular pathology

An implicit assumption in drug repositioning is that a drug can be repositioned from one indication to another because the two indications share some aspect of underlying molecular pathophysiology that is responsive to the therapeutic effect of a drug. Therefore, computational strategies for assessing molecular relationships between distinct disease pathologies can serve as a means for drug repositioning. In this approach, repositioning opportunities exist when diseases are found to exhibit similarity at the molecular level (even without similarity at the phenotypic or clinical level), suggesting that drugs might be shared among diseases with high degrees of molecular similarity.

Hu and Agarwal [35] created a disease-similarity network using publicly available gene expression profiles, and integrated this network with molecular profiles and knowledge of drugs and drug targets to infer drug repositioning opportunities and suggest molecular targets and mechanisms underlying drug effects. They began by identifying and acquiring disease-related experiments in the NCBI Gene Expression Omnibus (GEO) and computed differential gene expression profiles between classes represented in the experiments (e.g. affected versus healthy). The resulting disease profiles were compared using a correlation-based similarity metric and organized into a network to reveal novel disease relationships based on genome-wide transcriptional response. This network was further integrated with drug molecular profiles derived from the Connectivity Map to create a drug–disease network where clusters of drugs and diseases suggest shared drug mechanisms and molecular disease pathology. Similar efforts using the genetics and known pathways involved in these diseases have also been successful [36, 37].

Suthram *et al.* [38] integrated a larger set of molecular profiles of diseases with protein–protein interaction (PPI) data, to infer protein functional modules and networks that were shared among many diseases. Using a human PPI that was organized into 'modules' of functionally interacting proteins, a statistical approach was used to evaluate the molecular signatures of diseases for gene functional module activity, whereby the module activity was determined as the mean normalized transcriptional activity of its component genes in the disease molecular profile. A network of disease–disease relationships was then formed on the basis of functional module activity shared between diseases. Within these networks that she discovered as common across diseases were multiple known drug targets. Curiously, drugs that hit these targets were already known to be effective against multiple indications, compared to drugs hitting other targets in other networks, a property she called pluripotent drug targets. This work also leveraged publicly available experiments from GEO to create molecular profiles of disease, which were restricted to include only those experiments comparing disease affected tissues to healthy controls to create standardized disease signatures.

The limitations of using shared molecular pathology approaches are similar to those implicit in the shared molecular activity approach for drug-based repositioning. The utility of these approaches will be limited by the ability measure and represent the molecular pathology underlying the disease.

As disease pathology often incorporates a multitude of molecular entities, tissues and organ systems, it can be quite challenging to model the molecular state of diseases such that they can be easily compared using computational approaches. Promising network-based approaches to overcoming the challenge of modeling and comparing complex molecular disease states has been proposed by Schadt *et al.* [39–41].

Side effect similarity

Another way to connect drugs to clinical effects for the drug repurposing is through the side effects of drugs, which represent unintended consequences of the drug action. Side effects provide a means to connect drugs to diseases, because they encode the physiological consequence of a drug compound's biological activity. Furthermore, the phenotypic expression of a side effect can be similar to that of a disease, implying that the underlying pathways or physiological systems may be similarly perturbed by both the drug and the disease condition. This provides the basis to relate drugs to other drugs or diseases by side effect profiles, even in cases where the precise pharmacological mechanism facilitating the side effect is unknown.

Campillos *et al.* [42] performed a systematic analysis to identify novel drug–target relationships for 746 approved drugs using a side effect similarity approach. For each drug, they extracted side effects from the drug package insert and mapped them to standardized medical symptom terms using the Unified Medical Language System (UMLS) Metathesaurus [43]. The side effect terms were given weights using a scheme that incorporated their frequency and correlation across all drugs in the set, and similarity scores were computed between a pair of drugs based on the sum of the respective weights of their common side effects. A randomization approach was used to establish the significance of the side effect similarity scores, which were further incorporated with a measure of the structural similarity between drugs to increase predictive power. The resulting drug–drug relationships were shown to recapitulate many shared target relationships between drugs, and several predicted novel drug–target relationships were experimentally confirmed.

The most apparent limitation of the side effect similarity approach is the necessity for having well-defined side effect profiles for a drug. Despite rigorous preclinical assessment, the side effect profile for a newly approved drug may only be fully discerned after years of clinical use and postmarket surveillance. In addition, the assumption that similar phenotypic expression of a drug side effect implies a common pathophysiological basis may not always hold. For example, the side effect of 'hair loss' can arise when a drug interferes with hormonal systems that regulate hair growth, or alternatively disrupt immune function in a manner that causes harm to the cells comprising the hair follicle.

CONCLUSIONS

With continued difficulties in accelerating the growth rate of new chemical entities reaching regulatory approval, it appears certain that the strategy of repositioning of existing pharmaceuticals will only gain in acceptance. Computational methods for repositioning are probably the most efficient way to yield novel indications for these drugs, and the power of these methods will only increase as more molecular measurements of increasingly different types become available. However, in general, extensive animal model and clinical trials have not yet been launched based on these computational predictions; these studies are now needed to fully demonstrate the utility of computational drug repositioning.

Regardless of whether computational methods become the standard for drug repositioning, it is clear that many other undiscovered uses of drugs do exist. Finding these new uses is an important and necessary step towards reducing the burden of disease.

Key Points

- There are likely many undiscovered uses for existing drugs.
- Dozens of computational methods have been published that enable the discovery of new uses for established drugs.
- Computational methods for drug repositioning can be classified in two axes: drug- or disease-based.
- Using computational methods to find new uses for drugs is an important and necessary step toward reducing the burden of disease.

References

1. Booth B, Zemmel R. Prospects for productivity. *Nat Rev Drug Discov* 2004;**3**:451–6.

2. Dimasi JA. New drug development in the United States from 1963 to 1999. *Clin Pharmacol Ther* 2001;**69**:286–96.

3. Adams CP, Brantner VV. Estimating the cost of new drug development: Is it really 802 million dollars? *Health Aff* 2006;**25**:420–8.

4. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov* 2004;**3**:673–83.

5. Anand G. 'How drug's rebirth as treatment for cancer fueled price rises: once-demonized thalidomide boosts cel-gene's sales; patients see costs soar' 2004. New York: The Wall Street J, New York, 2004;**15**:A1, A18.

6. Pijl H, Ohashi S, Matsuda M, *et al*. Bromocriptine: a novel approach to the treatment of Type 2 diabetes. *Diabetes Care* 2000;**23**:1154–61.

7. Weinstein JN, Myers TG, O'Connor PM, *et al*. An information-intensive approach to the molecular pharmacology of cancer. *Science* 1997;**275**:343–9.

8. Hughes T, Andrews B, Boone C. Old drugs, new tricks: using genetically sensitized yeast to reveal drug targets. *Cell* 2004;**116**:5–7.

9. Lum PY, Armour CD, Stepaniants SB, *et al*. Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes. *Cell* 2004;**116**:121–37.

10. Kinnamon KE, Rothe WE. Biological screening in the US Army antimalarial drug development program. *Am J Trop Med Hyg* 1975;**24**:174–8.

11. Keiser MJ, Setola V, Irwin JJ, *et al*. Predicting new molecular targets for known drugs. *Nature* 2009;**462**:175–81.

12. Elvidge S. Getting the drug repositioning genie out of the bottle. *Life Sci Leader* 2010;14–8.

13. Aronson JK. Old drugs—new uses. *Br J Clin Pharmacol* 2007;**64**:563–5.

14. Soignet SL, Maslak P, Wang ZG, *et al*. Complete remission after treatment of acute promyelocytic leukemia with arsenic trioxide. *N Engl J Med* 1998;**339**:1341–8.

15. Yardley V, Croft SL. Activity of liposomal amphotericin B against experimental cutaneous leishmaniasis. *Antimicrob Agents Chemother* 1997;**41**:752–6.

16. Singhal S, Mehta J, Desikan R, *et al*. Antitumor activity of thalidomide in refractory multiple myeloma. *N Engl J Med* 1999;**341**:1565–71.

17. Eckert H, Bajorath J. Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. *Drug Discov Today* 2007;**12**:225–33.

18. Noeske T, Sasse BC, Stark H, *et al*. Predicting compound selectivity by self-organizing maps: cross-activities of meta-botropic glutamate receptor antagonists. *Chem Med Chem* 2006;**1**:1066–8.

19. Qu XA, Gudivada RC, Jegga AG, *et al*. Inferring novel disease indications for known drugs by semantically linking drug action and disease mechanism relationships. *BMC Bioinformatics* 2009;**10(Suppl 5)**:S4.

20. Fourches D, Muratov E, Tropsha A. Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J Chem Inf Model* 2010;**50**:1189–204.

21. Lamb J, Crawford ED, Peck D, *et al*. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006;**313**:1929–35.

22. Subramanian A, Tamayo P, Mootha VK, *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005;**102**:15545–50.

23. Iorio F, Bosotti R, Scacheri E, *et al*. Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc Natl Acad Sci USA* 2010;**107**:14621–6.

24. Frey BJ, Dueck D. Clustering by passing messages between data points. *Science* 2007;**315**:972–6.

25. Wei G, Twomey D, Lamb J, *et al*. Gene expression-based chemical genomics identifies rapamycin as a modulator of MCL1 and glucocorticoid resistance. *Cancer Cell* 2006;**10**:331–42.

26. Chen B, Wild D, Guha R. PubChem as a source of polypharmacology. *J Chem Inf Model* 2009;**49**:2044–55.

27. Ekins S, Mestres J, Testa B. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Br J Pharmacol* 2007;**152**:9–20.

28. Zahler S, Tietze S, Totzke F, *et al*. Inverse in silico screening for identification of kinase inhibitor targets. *Chem Biol* 2007;**14**:1207–14.

29. Kinnings SL, Liu N, Buchmeier N, *et al*. Drug discovery using chemical systems biology: repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. *PLoS Comput Biol* 2009;**5**:e1000423.

30. Chang RL, Xie L, Bourne PE, *et al*. Drug off-target effects predicted using structural analysis in the context of a metabolic network model. *PLoS Comput Biol* 2010;**6**:e1000938.

31. Lorber DM, Shoichet BK. Flexible ligand docking using conformational ensembles. *Protein Sci* 1998;**7**:938–50.

32. von Eichborn J, Murgueitio MS, Dunkel M, *et al*. PROMISCUOUS: a database for network-based drug-repositioning. *Nucleic Acids Res* 2010;**39**(Database issue):D1060–6.

33. Chiang AP, Butte AJ. Systematic evaluation of drug-disease relationships to identify leads for novel drug uses. *Clin Pharmacol Ther* 2009;**86**:507–10.

34. Solomon WD, Wroe CJ, Rector AL, *et al*. A reference terminology for drugs. *Proc Amia Symp* 1999;152–5.

35. Hu G, Agarwal P. Human disease-drug network based on genomic expression profiles. *PLoS One* 2009;**4**:e6536.

36. Li Y, Agarwal P. A pathway-based view of human diseases and disease relationships. *PLoS One* 2009;**4**:e4346.

37. Agarwal P, Searls DB. Can literature analysis identify innovation drivers in drug discovery? *Nat Rev Drug Discov* 2009;**8**:865–78.

38. Suthram S, Dudley JT, Chiang AP, *et al*. Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. *PLoS Comput Biol* 2010;**6**:e1000662.

39. Segal E, Friedman N, Koller D, *et al*. A module map showing conditional activity of expression modules in cancer. *Nat Genet* 2004;**36**:1090–8.

40. Yang X, Deignan JL, Qi H, *et al*. Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks. *Nat Genet* 2009;**41**:415–23.

41. Dudley JT, Schadt E, Sirota M, *et al*. Drug discovery in a multidimensional world: systems, patterns, and networks. *J Cardiovasc Transl Res* 2010;**3**:438–47.

42. Campillos M, Kuhn M, Gavin AC, *et al*. Drug target identification using side-effect similarity. *Science* 2008;**321**:263–6.

43. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res* 2004;**32**:D267–70.