



# Identification of drug candidates and repurposing opportunities through compound–target interaction networks

Anna Cichonska MSc , Juho Rousu PhD & Tero Aittokallio PhD

To cite this article: Anna Cichonska MSc , Juho Rousu PhD & Tero Aittokallio PhD (2015) Identification of drug candidates and repurposing opportunities through compound–target interaction networks, Expert Opinion on Drug Discovery, 10:12, 1333-1345, DOI: 10.1517/17460441.2015.1096926

To link to this article: <http://dx.doi.org/10.1517/17460441.2015.1096926>



Published online: 01 Oct 2015.



Submit your article to this journal [↗](#)



Article views: 321



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 3 View citing articles [↗](#)

# EXPERT OPINION

1. Introduction
2. Experimental approaches to ligand–target interaction mapping
3. Computational prediction of compound–target interaction networks
4. Future directions: incorporating genetic variation into the pharmacogenomic models
5. Conclusions
6. Expert opinion

## Identification of drug candidates and repurposing opportunities through compound–target interaction networks

Anna Cichonska, Juho Rousu & Tero Aittokallio<sup>†</sup>

<sup>†</sup>*University of Helsinki, Institute for Molecular Medicine Finland FIMM, Helsinki, Finland*

**Introduction:** System-wide identification of both on- and off-targets of chemical probes provides improved understanding of their therapeutic potential and possible adverse effects, thereby accelerating and de-risking drug discovery process. Given the high costs of experimental profiling of the complete target space of drug-like compounds, computational models offer systematic means for guiding these mapping efforts. These models suggest the most potent interactions for further experimental or pre-clinical evaluation both in cell line models and in patient-derived material.

**Areas covered:** The authors focus here on network-based machine learning models and their use in the prediction of novel compound–target interactions both in target-based and phenotype-based drug discovery applications. While currently being used mainly in complementing the experimentally mapped compound–target networks for drug repurposing applications, such as extending the target space of already approved drugs, these network pharmacology approaches may also suggest completely unexpected and novel investigational probes for drug development.

**Expert opinion:** Although the studies reviewed here have already demonstrated that network-centric modeling approaches have the potential to identify candidate compounds and selective targets in disease networks, many challenges still remain. In particular, these challenges include how to incorporate the cellular context and genetic background into the disease networks to enable more stratified and selective target predictions, as well as how to make the prediction models more realistic for the practical drug discovery and therapeutic applications.

**Keywords:** cell-based models, drug repositioning, drug–target interactions, machine learning, network pharmacology, phenotypic screening, target validation

*Expert Opin. Drug Discov.* (2015) **10**(12):1333–1345

### 1. Introduction

Traditionally, decisions on adopting chemical lead molecules into drug development and clinical trials were made mainly based on their ability to generate desired physiological changes in pre-clinical models, without paying that much attention to the underlying mechanism of action of the chemical probes. However, the rapidly developing genomic, proteomic, and metabolomic technologies have facilitated the targeted drug discovery approach by enhancing our understanding of the molecular basis of complex disease processes and/or drug action, and thereby enabling us to pinpoint their key molecular players. It is known that drug-like compounds execute their actions mainly by modulating cellular targets, usually proteins, and instead of focusing on physiological or other phenotypic effects, the emphasis in

**Article highlights.**

- System-level network pharmacology models, combined with computational machine-learning algorithms, offer a powerful approach for systematic identification of candidate compounds and their cellular targets.
- One of the biggest computational challenges is how to better use the genetic background information in the pharmacogenomic models to provide more stratified or even personalized predictions in the future.
- Experimental validation using functional, binding, or phenotypic assays is the only way how to really demonstrate the practical utility of the model predictions in practical drug discovery applications.
- Tissue-specific interaction networks provide improved insights into complex molecular interplays behind diseases, critical for modeling and predicting compounds' mode of action and potential side effects.
- Open data and information sharing from compound–target interaction mapping efforts will provide a valuable knowledgebase for many exciting drug discovery and repurposing applications in the future.

This box summarizes key points contained in the article.

target-based approaches was directed toward designing selective chemical agents, each binding with maximal affinity to an individual molecular target involved in a particular disease, the concept often referred to as 'one drug – one target – one disease' paradigm.

However, despite the continuous increase in the number of promising therapeutic targets, the amount of new drugs approved by US FDA per billion dollars (inflation-adjusted) spent on R&D has halved every 9 years since 1950 [1]. A large number of drug candidates fail in clinical trials due to their lack of efficacy and safety. It has been shown in many cases that high binding affinity of a ligand is not – as initially thought – associated with its therapeutic efficacy [2], and drugs acting on single targets are rather an exception [3]. Instead, most of the FDA-approved drug molecules exhibit polypharmacological properties, which may also contribute to their therapeutic efficacy [4]. Although a large part of compounds' adverse side effects can be attributed to their off-targets, it has also become clear that diseases are driven by multiple molecular abnormalities, and thereby modulating single targets is unlikely to result in long-lasting therapeutic effect. Cellular networks are very robust to perturbing their individual components, due to compensatory signaling and metabolic pathways, and therefore complex multifactorial disorders, such as cancers, cardiovascular diseases, and mental disorders, should be ideally treated by simultaneous modulation of multiple targets or pathways. These observations have led to a recent paradigm shift from target-centered to systems-driven drug discovery [5-9]. However, what still remains missing are systematic approaches and practical modeling frameworks to deal with this polypharmacological complexity.

The concept of network pharmacology allows a broader, systems-level perspective on drug mode of action (MoA) by placing molecules in the context of underlying biochemical processes and mechanisms governing interactions between chemical compounds and their cellular targets [10-14]. In this framework, a network can be depicted as a connected graph, where each node represents an individual molecular entity, for instance, a drug compound, its protein target, a modifier molecule within a biological process, or target pathway, whereas an edge models either a direct or indirect interaction between the two molecules. Ultimately, both the efficacy and toxicity of a drug are a consequence of complex interplay among different cellular components, and therefore a system-scale perspective is needed to aid modern drug discovery. For instance, while many off-target effects were initially classified as unwanted adverse reactions, some of them might be beneficial for achieving desired therapeutic effect, perhaps in other than original indication. An illustrative example is the case of thalidomide, which in the 50s was prescribed to treat insomnia and nausea in pregnant women, but was turned out to cause severe skeletal defects in newborns, and was therefore quickly withdrawn from the market; nowadays, however, it is being used in the treatment of leprosy by inhibiting TNF-alpha protein [15].

Network pharmacology models have the potential to help us to better understand the MoA of promiscuous drugs, and to identify new therapeutic uses of approved, abandoned, or preclinical compounds. Such drug repositioning opportunities can greatly reduce the increasing cost, time, and risk associated with the drug discovery and development processes [16]. As an example, a phenotype-based repurposing approach recently identified a VEGFR tyrosine kinase inhibitor axitinib as a selective and effective treatment for patients with BCR-ABL1(T315I)-driven chronic myeloid leukemia [17]. This is a great example how comprehensive drug sensitivity testing of patient-derived cancer cells, combined with their target selectivity information, can lead to unexpected personalized treatment strategies. However, since the experimental mapping of all the possible compound–target interactions is an infeasible task, even with the current technologies, computational approaches are needed to accelerate the experimental work by prioritizing the most promising drugs and targets for further experimental validation.

In the past decade, machine learning models have drawn an increasing attention in the field of drug discovery. In particular, a lot of work has been devoted to methods based on quantitative structure–activity relationship (QSAR), which aim to relate structural properties of the chemical molecules to their bioactivity profiles [18]. Such *in silico* prediction models can be used, among other things, to predict cancer growth inhibition potential of drug compounds [19], or their potency against a given cellular target [20]; for instance, regression-based machine learning algorithms have been used to infer quantitative bioactivity signatures [21], while in a classification setup, the task is to predict whether a particular compound is

active or not [22]. Instead of focusing on single targets or compounds only, we argue that the development of rational, effective, and safe therapies requires a network perspective of the system-wide cellular components, combined with a proteome-wide profile of drug–target interactions (DTIs), including both on- and off-targets. In this framework, given a set of promising therapeutic intervention points in the disease network, *in silico* approaches can then be used to suggest the small molecules that may target these network nodes, together with their potential off-target effects. There exist recent reviews on computational compound–target interaction prediction methods [20,23,24], which give an excellent overall and theoretical view of the different algorithms.

Here, we focus on computational models for drug compound identification using large-scale ligand–target interactions mapping in which the network perspective plays a central role. In contrast to previous works, we place a specific emphasis on describing the practical aspects of how such network pharmacology approaches to predicting compound–target interactions could lead to exciting drug repurposing opportunities, as well as to novel candidates for target-based drug discovery, after a careful experimental validation. Furthermore, we offer our personal view of the current challenges and the key future opportunities in this emerging field, including the use of findings from genome-wide association studies as safe drug target candidates, and the application of new genome editing techniques to testing and further optimizing the drug target inhibition at the pre-clinical level. We also shortly discuss the recent progress made in modeling of tissue-specific networks as well as their important role in therapeutic target identification. We argue that many of the current limitations are related to the question how to make the computational models more realistic for the practical drug discovery and therapeutic applications, and that experimental validation is the only way to really demonstrate their practical utility.

## 2. Experimental approaches to ligand–target interaction mapping

Experimental identification of novel drug candidates and repurposing opportunities typically relies on ligand profiling assays, which can be roughly divided into three main categories: binding, functional, and phenotypic assays. Binding and functional assays are of particular interest here, as they directly facilitate compound–target interaction mappings. In the phenotype-based screening, the measured signal (or readout) corresponds to a phenotypic response to drug perturbation, such as  $IC_{50}$ , which equals to the concentration that results in half-maximal inhibition of cell proliferation or viability.

Binding assays aim to understand a direct physical interaction between two molecules. The readout corresponds to a value characterizing the binding affinity between the ligand and target, for example, dissociation constant  $K_D$ , which

measures the tendency of a larger molecular complex to dissociate reversibly into the component molecules. A functional bioassay, on the other hand, measures the signal induced by an activation or suppression of a pre-defined function. In functional kinase assays, for instance, the activity of a kinase enzyme is a parameter of interest affected by binding of a drug compound. A typical readout is  $K_i$ , the concentration of the inhibitor that causes 50% inhibition of the reaction catalyzed by the enzyme of interest.

The studies by Davis *et al.* [25] and Metz *et al.* [26] are illustrative examples of large-scale binding and functional assays, respectively, both generating kinome-scale compound–target interaction maps. Davis *et al.* profiled 72 clinically relevant compounds against 442 kinases, in their mutated and wild-type forms, providing a broad overview of selectivity of established kinase inhibitors across over 80% of the known catalytically active human kinome. In the study of Metz *et al.*, a larger collection of 3,858 chemical probes was tested against 172 kinases, followed by a statistical analysis of the resulting dataset, which enabled the construction of a comprehensive kinome interaction network. However, in such mechanistic *in vitro* assays, molecules are treated as isolated entities taken away from their cellular context, which may affect compound's potency. For instance, differences in the *in vitro* and *in vivo* concentrations of ATP can lead to unrealistic profiling results of the kinase inhibitors, which are known to be ATP competitive [27].

The work of Taipale *et al.* was one of the first attempts to establish a high-throughput compound–target interaction profiling approach in living cells, utilizing chaperones as thermodynamic sensors of binding strength [27]. This method was initially applied to screen 30 kinase inhibitors against > 300 wild-type and mutated kinases, but it could be extended also to other protein targets that either naturally or when engineered associate with chaperons. Recently, Savitski *et al.* devised a more general approach for proteome-wide monitoring of protein–ligand binding directly in living cells by combining the technique of cellular thermal shift assay with quantitative mass spectrometry (MS) [28]. Their *thermal proteome profiling* method enables the identification of not only direct interaction partners of over 7,000 proteins embedded in the cellular network, but also downstream effectors as potential markers for drug efficacy and toxicity.

Although experimental bioactivity profiling is critical to characterize a compound's MoA, and proteome-wide ligand–target interaction mapping under various conditions has become feasible owing to recent technological advances (see [29] for a comprehensive review), computational approaches can greatly help to accelerate the process of exploring the enormous size of the chemical universe. It is estimated that there are approximately  $10^{20}$  –  $10^{24}$  molecules exhibiting good pharmacological properties [30]. *In silico* methods enable suggesting the most potent interactions that should then be subjected to further experimental validation. To facilitate such computational–experimental approaches, there are

**Table 1. A list of representative compound and target databases.**

Database	Content	Website
ChEMBL [80]	Quantitative bioactivity data for drug-like small molecules	<a href="http://www.ebi.ac.uk/chembl">www.ebi.ac.uk/chembl</a>
DrugBank [81]	Detailed drug data with comprehensive target information	<a href="http://www.drugbank.ca">www.drugbank.ca</a>
BindingDB [82]	Binding affinities of proteins with small drug-like ligands	<a href="http://www.bindingdb.org">www.bindingdb.org</a>
STITCH [83]	Known and predicted compound-protein interactions	<a href="http://stitch.embl.de/">http://stitch.embl.de/</a>
KEGG BRITE [84]	Information on various biological objects, including drug–target relationships	<a href="http://www.genome.jp/kegg/brite.html">www.genome.jp/kegg/brite.html</a>
PubChem [85]	Compound chemical and bioactivity information	<a href="https://pubchem.ncbi.nlm.nih.gov/">https://pubchem.ncbi.nlm.nih.gov/</a>
BRENDA [86]	Comprehensive enzyme information	<a href="http://www.brenda-enzymes.org">www.brenda-enzymes.org</a>
DGIdb [87]	Drug–gene interaction database	<a href="http://dgidb.genome.wustl.edu">http://dgidb.genome.wustl.edu</a>
DrugKINET	Detailed target information on kinase inhibitors	<a href="http://www.drugkinet.ca">www.drugkinet.ca</a>
TTD [88]	Information on the therapeutic targets	<a href="http://bidd.nus.edu.sg/group/ttd">http://bidd.nus.edu.sg/group/ttd</a>
SuperTarget [89]	Drug–target interactions	<a href="http://insilico.charite.de/supertarget">http://insilico.charite.de/supertarget</a>
MATADOR [89]	Data on a subset of more extensively annotated drugs from SuperTarget	<a href="http://matador.embl.de">http://matador.embl.de</a>
PDSP [90]	Binding affinities ( $K_i$ ) of psychoactive compounds	<a href="http://pdsp.med.unc.edu/kidb.php">http://pdsp.med.unc.edu/kidb.php</a>
ZINC [91]	Database of purchasable compounds	<a href="http://zinc.docking.org/">http://zinc.docking.org/</a>

several publicly available databases containing ligand–target interaction data and other resources useful for building computational models (Table 1).

However, it is also important to understand the potential limitations of the drug and target databases for the modeling purposes. For instance, the overlap in compounds and targets between distinct repositories may sometimes be relatively limited, and even for the overlapping compound–target pairs, the types of bioactivity measurements and annotations of the molecules might differ drastically. The bioactivity data are often incomplete, vast, and complex, and some erroneous or contradictory interactions are bound to occur [31]. These issues need to be carefully considered when collecting and preparing the data for computational modeling. We encourage the use of multiple data sources to make the full advantage of the available bioactivity information, as well as tools like UniChem [32], which facilitate cross-referencing of chemical structures and their identifiers between different repositories.

### 3. Computational prediction of compound–target interaction networks

Network-based approaches to compound–target interaction inference hold a great promise to aid modern drug discovery. Computational prediction methods enable large-scale, systematic pre-screening of chemical agents, providing thereby insights into the potency of investigational compounds as well as potential new indications for already approved drugs, examples of which are given later in this section. The network concept allows one to abstract, integrate, and organize the data from large-scale experiments, facilitating the extraction of useful information from complex biological systems. In such analyses, a network graph can represent different types of relationships, depending on the system being modeled, for example, protein–protein interactions (PPIs), gene co-expression interactions, metabolic interplays, or DTIs. Figure 1

illustrates the main components of the DTI prediction methods described in this section (see Table 2 for the summary).

#### 3.1 The use of DTI network topology in the DTI prediction

The topology of the experimentally mapped interaction network can provide important insights into the system under investigation, hence enabling the prediction of new compound–target links. For instance, van Laarhoven *et al.* introduced a Gaussian Interaction Profile (GIP) kernel – a similarity metric defined on binary vectors, each encoding the presence or absence of an interaction of a drug (respectively target) with every target (drug) in the considered DTI network [33]. New compound–target links can be inferred by Regularized Least Squares model using a Kronecker product of constructed drug and target kernels. This study demonstrated that known interactions themselves constitute an important information source for the prediction algorithms. However, employing additionally structural knowledge about the molecules further improved the GIP method's performance [33,34].

A different approach was taken by Cheng *et al.* who combined the principles of recommender systems theory with topological features of drug–target bipartite graph in their network-based inference (NBI) algorithm for predicting new connections between compounds and targets [35]. NBI operates on the bipartite graph, where drug compounds and targets constitute two disjoint sets of nodes, and an edge is drawn between drug and target if they are known to interact with each other. Given a target (respectively drug) node, resources, analogs to the mass in physics, are initially allocated in all drugs (targets) linked with it. Then, predictive scores are calculated for each drug (target), using a network diffusion process, corresponding to the mass diffusion in physics. Finally, the nodes are sorted in descending order based on their scores, thereby forming the recommendation list of

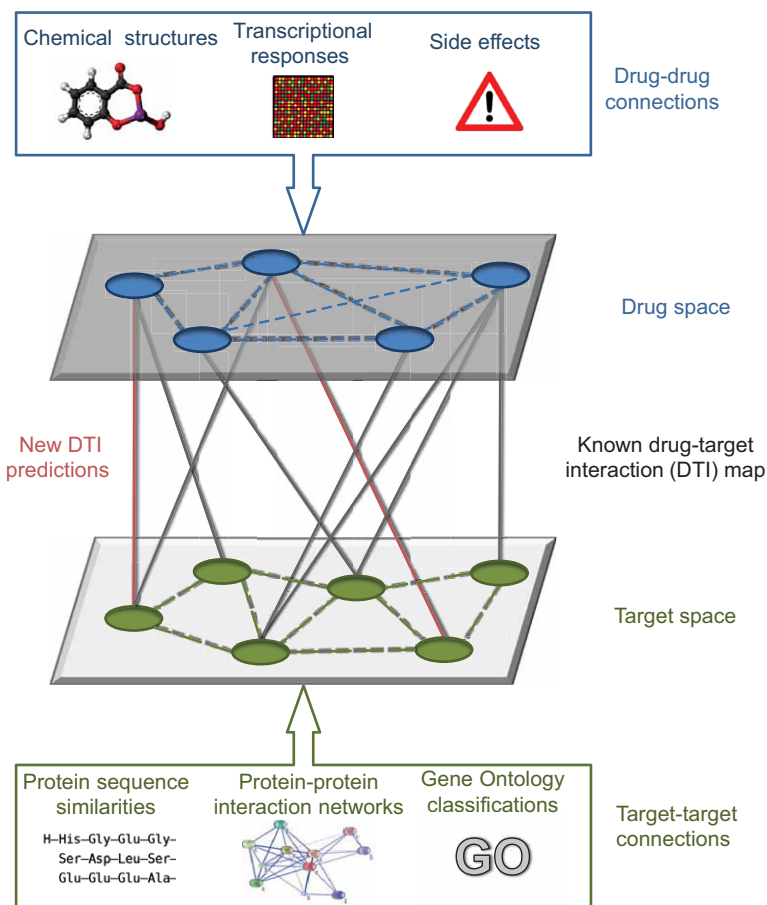


Figure 1. The components of the drug–target interaction prediction methods.

Table 2. Summary of representative *in silico* methods for compound–target interaction prediction described in Section 3.

Method	Information source					
	Compound–target interaction network topology	Compound chemical structures	Target structures	Transcriptional responses	PPI network	Others, for example, Gene Ontology, side effects
GIP [33]	x					
NBI [35]	x					
NWNBI, EWNBI [36]	x					
Wang and Zeng [37]	x					
SEA [38]		x				
Yamanishi <i>et al.</i> [40]	x	x	x			
DT-Hybrid [43]	x	x	x			
NRWRH [44]	x	x	x			
TL_HGBI [45]	x	x	x			x
Campillos <i>et al.</i> [47]		x				x
Ravindranath <i>et al.</i> [50]		x		x		
drugCIPHER [51]	x	x			x	x
Laenen <i>et al.</i> [52]				x	x	
SITAR [53]		x	x	x	x	x

drugs (targets). Drugs (targets) from the top of the list are more likely to interact with a given target (drug), and these are used as new DTI predictions.

NBI was first validated on four popular benchmark datasets of drug–target proteins (enzymes, ion channels, nuclear receptors, and G-protein-coupled receptors), collected from KEGG BRITE, BRENDA, SuperTarget, and DrugBank, where it showed high true positive and low false positive rates. Importantly, experimental validation of the model predictions *in vitro* confirmed montelukast as a new inhibitor of dipeptidyl peptidase-IV, and four other drugs, including simvastatin, diclofenac, ketoconazole, and itraconazole, to effectively bind estrogen receptors ER $\beta$  and/or ER $\alpha$ . Moreover, both ketoconazole and simvastatin showed antiproliferative potency in human MDA-MB-231 triple-negative breast cancer cell line [35]. These experimental validations demonstrated that NBI is useful for extending DTI maps, as well as providing novel repositioning suggestions.

In the follow-up work, Cheng *et al.* introduced node- and edge-weighted versions of the NBI method (termed NWNBI and EWNBI, respectively) [36]. EWNBI allows one to weight links between drugs and targets by their quantitative binding affinity values, such as  $K_i$  (see Section 2), whereas NWNBI provides a more realistic scheme for the resource distribution, compared to the original NBI algorithm, where the information was evenly distributed across all the neighboring nodes.

The above-mentioned approaches can predict only binary DTIs, providing no additional information on their nature. To address this limitation, Wang and Zeng proposed an algorithm based on a two-layer graphical model, known as restricted Boltzmann machine, which is capable of differentiating direct from indirect DTIs, as well as predicting binding, activation, and inhibition types of connections [37]. The adoption of such annotations of the edges in the network makes the method useful not only for compound–target interaction profiling and identifying starting points for drug repositioning, but also for providing practical insights into molecular basis of drug's MoA, for instance, distinguishing direct physical binding of the two molecules from an indirect interaction being a consequence of cellular perturbations caused by the drug molecule.

### 3.2 The use of chemical and genomic profiles in the network-based DTI inference

A drawback of the methods relying merely on topological information of compound–target networks is their inability to predict interactions for such drug candidate compounds that have no known target information in the training data. Supervised approaches that incorporate additional sources of information, such as chemical and genomic profiles of the molecules, can effectively address this problem by introducing drug–drug and target–target inter-relationships into the prediction task. A prime example is the Similarity Ensemble Approach by Keiser *et al.*, in which DTIs are inferred

statistically by comparing structural similarities between drugs and known set of ligands for a given target [38,39]. In one of the first machine learning models utilizing both chemical and genomic profiles, Yamanishi *et al.* proposed a supervised learning framework for binary DTI inference based on simultaneous integration of drug chemical structures, protein target amino acid sequences, and known drug–target bipartite network topology into a unified *pharmacological space* [40]. Motivated by the observed correlations between compound structure similarity, protein sequence similarity, and compound–protein interaction network topology, the modeling framework of supervised bipartite graph inference from chemical, genomic, and pharmacological data has been further developed and applied in the follow-up studies [41,42].

Since the seminal work of Keiser *et al.* [38] and Yamanishi *et al.* [40], the available sources of chemical and genomic information have become more accurate and systematic, and these are being commonly used in the compound–target interaction network predictions. For instance, Alaimo *et al.* extended the NBI recommendation framework by including additional biological knowledge in the form of chemical similarities among drugs and amino acid sequence similarities among targets [43]. These similarities were incorporated into the function that defines how the resource transfer takes place in the drug–target network. The resulting domain tuned-hybrid (DT-Hybrid) method was found to outperform the original NBI algorithm, which utilized only DTI network topology in the inference. However, no novel compound–target interaction predictions were validated experimentally in this computational work.

Chemical and genomic information sources were also used in the method named Network-based Random Walk with Restart on the Heterogeneous network (NRWRH) [44]. Specifically, drug–drug and protein–protein similarity networks, as well as known drug–protein interaction network, were integrated into a heterogeneous graph on which a random walk was applied to infer new DTIs. In contrast to the above methods, NRWRH computes drug–drug (respectively target–target) associations as a weighted sum of chemical structure (protein sequence) similarity matrix and similarity matrix constructed based on the number of known targets (drugs) shared by each pair of drugs (targets). The underlying idea is that molecules sharing many common targets (drugs) should be considered similar, and in some cases this similarity might not be well captured by structural features only, for example, due to the fact that even a minor structural difference between two chemical compounds can cause a dramatic change in their activity. NRWRH demonstrated an excellent performance on the benchmark datasets, but the need for the user to select five parameter values can be considered as practical weaknesses.

It is known that chemical agents achieve their therapeutic effects as a consequence of modulating, either directly or indirectly, the molecular targets relevant to a disease process. Although drug target prediction and drug repurposing are

naturally interconnected, they are typically treated as two subsequent but separate tasks in the prediction approaches. However, Wang *et al.* introduced a network-based method that integrates these two tasks into a unified framework, termed Triple Layer Heterogeneous Graph Based Inference (TL\_HGBI) [45]. This is achieved through a heterogeneous network model, composed of three distinct types of nodes representing targets, drugs, and diseases, while edges capture their inter- and intra-relationships. Drug–target and drug–disease connections are constructed based on the prior knowledge from established databases, such as DrugBank and OMIM [46]. Drug–drug, target–target, and disease–disease edges are derived from chemical structure, amino acid sequence, and phenotypic description similarities, respectively.

In TL\_HGBI, drug repurposing is formulated as missing drug–disease link inference problem, solved using iterative algorithm that propagates information across the three-layer graph. Moreover, TL\_HGBI extends the druggable space, as new drug–target connections are predicted simultaneously. A case study on five disorders (Huntington disease, non-small-cell lung cancer, alcohol dependence, small-cell lung cancer, and polysubstance abuse) demonstrated that TL\_HGBI can be useful in practice for identifying new drug repositioning opportunities, since many of the top-ranked drugs were supported by the recent literature, and some of them have even already been tested *in vivo* in clinical trials. For instance, carboplatin and temozolomide are currently under evaluation for treating small-cell lung cancer ([www.clinicaltrials.org](http://www.clinicaltrials.org)).

### 3.3 Integration of interaction networks for drug candidate and target prediction

Computational methods for compound–target interaction prediction by means of utilizing drug–target network topology as well as additional chemical and genomic profiles of the molecules have become a standard in the field. However, integrating additional network-centric knowledge into the models, for example, in the form of gene co-expression patterns and PPIs, offers even broader perspective for the identification of new potential drug candidates and repositioning opportunities. A well-known example of employing additional pharmacological information to the prediction task is the seminal work of Campillos *et al.*, where the authors used phenotypic side-effect similarities in the inference of compound–target connections, several of which were tested and confirmed with *in vivo* assays [47].

A development of the Connectivity Map (CMap) resource [48], containing a public collection of gene expression profiles from human cell lines treated with over 1,000 diverse compounds, has facilitated the use of transcriptional signatures in the *in silico* drug discovery. For instance, Iorio *et al.* exploited CMap datasets in the approach called Mode of Action by NeTwoRk Analysis, and successfully inferred MoA of nine anticancer compounds. Importantly, they also

experimentally verified a surprising prediction of fasudil, a drug with well-established safety profile, as a promoter of cellular autophagy, thus providing a novel repurposing opportunity [49]. In the more recent work, Ravindranath *et al.* integrated drug target predictions from probabilistic chemical similarity-based algorithm with the gene expression data from CMap in order to provide more in-depth understanding of MoA of compound clusters under the assumption that compounds modulating similar targets will also trigger similar genes and related pathways [50].

The network of PPIs constitutes another comprehensive source of information for *in silico* drug discovery. Zhao and Li used PPIs in their drugCIPHER framework for predicting DTIs on a genome-wide scale [51]. The algorithm models DTIs using linear regression that relates compound space to genomic space. The compound space consists of drugs' therapeutic similarity computed using the Anatomic Therapeutic Chemical (ATC) classification codes, and drugs' structural similarity. The genomic space, on the other hand, is defined as drug–protein closeness established on the basis of the PPI network and known drug–protein interactions. Although the aspects of protein druggability or cellular location were not considered in this work, drugCIPHER has a potential to provide genome-wide insights into novel drug–protein connections, enabling the identification of starting points for drug repositioning and understanding adverse drug reactions, especially with the future improvements in the quality and completeness of the PPI networks.

Toward more multi-omics approach, Laenen *et al.* demonstrated the power of integrating both transcriptional signatures and protein association network in a genome-wide scale DTI inference [52]. Specifically, differential expression values (drug-treated vs non-treated cells) were projected onto the PPI network, and then expression signals were diffused to the subnetwork around each protein. The resulting scores of the nodes allow one to prioritize proteins as potential drug targets. A somewhat different integrative approach was taken by Perlman *et al.*, who also used transcriptional drug response profiles and PPIs, as well as compound structures, ATC codes, side effects, protein amino acid sequences, and Gene Ontology classes to compute multiple drug–drug and target–target similarity measures integrated in the logistic regression-based SITAR (Similarity-based Inference of drug-TARgets) framework for DTI inference [53].

Although the *in silico* approaches described above were shown to have good predictive performance, and therefore may provide novel insights into the identification of drug and target candidates, modeling of tissue-specific networks holds perhaps the greatest promise to aid drug discovery, since many disorders manifest only in certain tissues, whereas inhibiting the same targets in the other tissues may lead to adverse side effects. Such context-specific approaches are needed to better account for the differences in drug's MoA across various tissue and cell types. In one of the first studies, Magger *et al.* used tissue-specific gene expression data to



construct PPI networks, and demonstrated that utilizing these networks, instead of the conventional approach of ignoring differences between tissues, indeed improved the prioritization of candidate disease-causing target genes [54].

Along the same lines, Liu *et al.* built and analyzed 30 tissue-specific networks based on protein expression profiles, and showed that their topological features differ from those of generic networks [55]. Recently, Greene *et al.* developed a Bayesian framework for the construction of tissue-specific protein interaction networks by integrating plethora of heterogeneous datasets, and used these to build 144 tissue- and cell lineage-specific maps, which were demonstrated to be capable of predicting tissue-specific gene functions and responses to pathway perturbations [56]. Moreover, such maps were also shown to improve the identification of disease-gene associations, under the introduced network-wide association study framework, when compared to the standard genome-wide association study (GWAS) alone [56]. It is likely that tissue-specific interaction networks will also prove useful for the computational inference of therapeutic targets and MoA proteins [57], and that new tissue-specific target prediction methods are likely to emerge in the near future.

#### 4. Future directions: incorporating genetic variation into the pharmacogenomic models

Considering the genetic background of the patient subpopulations is an important part of a modern drug development process (so-called precision or stratified medicine). However, incorporating the personal genomic information into the computational network-based approaches has been challenging, given the complexity and different nature of the genetic profiles, compared with the biochemical or cell-based compound and target profiles. There are, however, several emerging approaches that aim to link the genetic information obtained from large-scale patient cohorts to pharmacological information obtained either from clinical or pre-clinical studies, as well as from electronic healthcare records. Such *pharmacogenomic models*, once implemented efficiently, are likely to provide a more comprehensive network view of the molecular mechanisms behind disease processes, toward enabling more accurate predictions of the system-level phenotypic responses to both genetic and chemical perturbations.

##### 4.1 Genome-wide associations for drug repurposing and drug candidate identification

A relatively straightforward approach is to take the top variants from GWAS for a particular disease phenotype, and then use the identified disease susceptibility genes as leads in therapeutic target selection. There exist pilot studies showing the potential of GWAS findings for generating hypotheses about potential drug-disease links for both drug repurposing or target validation applications [56,58-60]. However, adding functional and structural network information is critical to

reduce the false positives from the genetics-only-based approaches when searching for effective and personalized drug candidates or disease-relevant targets [61-63]. Further, since GWAS data are often limited to one or maximally a few disease phenotypes, there has been recent interest in using large-scale phenome-wide association studies (PheWAS) to identify susceptibility genes simultaneously for thousands of disease traits, extracted from electronic medical records in large patient cohorts, thereby expanding the phenotypic space and capturing the pleiotropic effects among the diseases [64,65]. However, regardless of the association approach used, these genetics-based drug repurposing applications rely heavily on the availability of comprehensive and accurate DTI networks for confirming the druggability of the selected targets. Although the initial focus in these pilot studies has been more on drug repositioning applications, it should be possible to predict also not yet approved compounds that target the disease risk-modifying genes, which might offer novel leads for future drug discovery efforts.

Since the published GWAS or PheWAS variants often lack a more detailed functional characterization of their disease biology and toxicity profiles across various tissue types, information on the pathogenesis of a disease and the drug's MoA is required for better understanding both its therapeutic and potential side effects [66]. In particular, neutral or even protective loss-of-function (LoF) variants that can be tolerated in a homozygote state in humans are of particular interest as potential safe targets for therapeutic inhibition. Isolated populations, including Finns, with recent bottlenecks, have been instrumental for discovering low-frequency LoF variants, by taking advantage of such unique population genetic history combined with harmonized data from large population cohorts and national medical records [67]. Beyond genomics only, more comprehensive profiling of the patient cohorts using, for example, transcriptomics, proteomics, metabolomics, or even epigenetics platforms will be needed for more accurate safety and specificity estimates. Also, genome editing tools, such as RNAi or CRISPR/Cas9 systems, facilitate pre-clinical target discovery and validation [68,69], using either comprehensive panels of cell line models across tissue types or patient-derived cell samples in genetically selected patient populations. These exciting genome-engineering tools may even provide effective and safe therapeutic applications in the future, provided that RNAi-based delivery or CRISPR/Cas9 systems can be improved in terms of their fidelity and selectivity to target a specific part and tissue context of the disease network only [70,71].

##### 4.2 Prediction of drug response profiles based on genomic and molecular signatures

A more direct way of utilizing genomic profiles in drug development and treatment selection is to study the influence of genetic variation on the individual drug efficacy and side effect profiles in large cohorts of individuals [72]. Despite

successful examples of such pharmacogenetic findings that have led to clinical applications, perhaps best exemplified in the field of cancer treatment [73], the prediction of responding and non-responding patients is highly challenging, even when using large patient cohorts and high-density panels of genetic variants, not to mention going into more continuous prediction of treatment efficacies or toxicities using a panel of only a few genetic markers. The rather low predictive accuracies seen for many complex diseases may be due to the lack of low-frequency and rare variants in the genomic data, which should carry predictive signal for individual patients or patient subgroups, instead of studying population-level averages. While measuring the whole spectrum of genetic variation, using, for example, next-generation sequencing (NGS)-based genomic profiling technologies, is likely to improve the prediction of the treatment responses, these data pose also many computational challenges. However, once fully implemented and tested in clinical trials, NGS-based clinical genomics should facilitate more systematic and unbiased understanding of the genotype-specific mechanisms behind drug action, resistance, side effects, and other treatment phenotypes at the level of individual patient, hence enabling the translation of the genotype–phenotype relationships into precision medicine.

Pre-clinical models are also increasingly being used to model and predict the response of drugs based on panels of genomic and molecular markers. In addition to animal models, cell-based models, such as those based on established cell lines or patient-derived samples, make it possible to profile the response patterns to a comprehensive set of drug-like compounds. In a recent community-based effort in cancer cell line models, it was shown that modeling of non-linearities was a common component of the top-performing prediction approaches, and that most predictive molecular features were constructed based on gene expression and protein activity profiles [19]. However, there is still much room for improvement as the best prediction accuracies were only slightly better than random guessing. Enhanced knowledge of the compound–target interaction networks, which effectively connect the molecular and pharmacological spaces, should also lead to more predictive models. Indeed, a few theoretical works have already demonstrated the power of compound–target interactions in drug sensitivity prediction [74,75]. We should learn from these lessons when moving toward more individualized treatment prediction, either in patient-derived samples or in the clinical practice. Most likely single markers are not enough in the majority of practical cases to predict drug sensitivity, resistance, or toxicity. Therefore, machine learning algorithms for in-depth mining of the most predictive panels of markers, using, for example, Exome-seq, RNA-seq, or MS-based profiles, should prove beneficial for understanding network-level mechanisms and the key players behind the observed drug sensitivity and resistance patterns for future clinical translation.

## 5. Conclusions

System-level network pharmacology models, combined with computational machine learning algorithms, offer a powerful means for prioritization of the most potent compound–target pairs for further experimental evaluation, thereby reducing the massive search spaces spanned by candidate compounds and their cellular targets. At the moment, these methods are mainly used for providing suggestions for filling the gaps in the experimentally mapped compound–target networks, but once made more realistic and efficient, they may enable prediction of extended target space for a lead compound of interest (phenotype-based drug discovery), or even identification of new candidate compounds that selectively inhibit a particular target under investigation (target-based drug discovery). Mapping of the full spectrum of interactions between compounds and their cellular targets, including both the compound's intended primary targets as well as its secondary or 'off-targets', is a critical part of the drug discovery efforts, as it enables us not only to explore the therapeutic potential of the agents but also to better understand their possible adverse reactions prior to the actual clinical trials, thereby de-risking and accelerating the drug development process.

Currently, due to the lack of better treatment alternatives, there exist many chemotherapies used in the clinical practice, despite their known associations with important off-target toxicities or causing other severe side effects. An illustrative example is methotrexate, a folate analog metabolic inhibitor, which is being used as a predominant first-line therapy for rheumatoid arthritis (RA). Methotrexate has been reported as a highly toxic chemotherapy, which is associated with, among others, hepatotoxicity, lung infections, renal damage, bone marrow changes, and skin tumors [76]. However, since RA is a chronic disease with no cure, the benefit of preventing progressive joint damage prevails the downsides of the methotrexate therapy. This example demonstrates that in many cases, even if a therapeutic option exists, there is an urgent need for more effective and safer therapeutic alternatives, and we anticipate that *in silico* network pharmacology approaches will play an important role in the development of increasingly targeted and selective drug treatments.

One of the current computational challenges is how to better use the available genetics information in the prediction models. In the personalized medicine setting, for instance, we often need to consider also the protein variants, not only the wild-type targets, to fully realize the power of these models. Going back to the motivating example of axitinib, which was shown to inhibit BCR-ABL1(T315I) in a mutation-selective binding mode [17], making predictions for such mutated targets requires more in-depth modeling of the structural and molecular properties of both the compounds and protein variants, respectively, something that still poses challenges to the current QSAR or docking-based approaches. If proven successful, however, such improved machine learning

approaches could greatly aid the experimental phenotype-based drug repositioning efforts, for instance, by proving a straightforward approach to prioritize the most potent and safe compounds that provide effective killing of the cancer cells harboring specific mutations, while having minimal side effects in the other cell types lacking the mutations.

Integrated modeling approaches are being developed for combining functional information obtained either from the large-scale compound screening or from RNAi/CRISPR/Cas9 systems with the complementary information obtained from personal genomic profiling. While still at their early phases, such pharmacogenomics approaches have the potential to improve both the safety and efficacy of the individualized treatment strategies in the future. Beyond monotherapies, network pharmacology models offer effective means also for predicting synergistic drug combinations. In particular, targeted combinatorial therapies may be less vulnerable to drug resistance and even show fewer side effects by targeting disease networks through interacting or complementary pathways [9]. However, due to the exponentially increasing number of combinations, effective computational approaches are required for exploring the most potent multi-target regimens on a network level [11]. Another exciting application area involves multi-targeted mixtures of natural products, where network pharmacology approaches may enable systematic mapping of yet unexplored target space of natural compounds [77].

## 6. Expert opinion

---

Phenotype-based drug discovery using comprehensive compound collections and their systematic target mappings is starting to result in unexpected drug repurposing opportunities and novel lead compound discoveries, thereby effectively complementing the current NGS-driven target-based drug discovery approaches that have their limitations in translating the genomic detections into clinically actionable therapeutic strategies. However, critical improvements are required not only in the compound profiling setups, such as those based on high-content screening and more realistic 3D cell-based or disease-relevant assays, but also in the downstream analyses of the candidate probes that show the desired phenotypes; the key challenges range from the system-level MoA determination using biochemical or cell-based assays, to safety and efficacy testing in pre-clinical models or in patient-derived material, all the way to clinical trials and compound optimization for genetically defined patient populations. Network modeling and machine learning approaches, such as those described in this review, can help in each of these challenges. Nevertheless, despite their great potential, there remain both computational and experimental issues that will need to be addressed so that we can make the most of these emerging approaches toward exciting drug discovery applications.

Many of the current computational limitations are related to the question how to make the model predictions more realistic. We argue that experimental validation is the only way

how to really show the practical utility of the model predictions. Too often the compound–target interaction models are being evaluated under theoretical settings that do not reflect the practical applications, therefore leading to over-optimistic prediction accuracies [78]. We also argue that compound–target prediction should be formulated as a continuous regression problem, to quantify the full activity spectrum of chemical compounds across their potential target space, instead of the standard binary classification setup (i.e., interaction or no interaction). It is also important to acknowledge that each supervised prediction model is limited in applicability by the data used in its training. At the moment, we feel that one can make fairly accurate and realistic predictions within a particular drug and target family only (local models). However, with the increasing coverage of the interaction maps in the future, we may be able to extend these models across multiple compound or target families (global models). Such experimental and computational developments should lead to more comprehensive network models of compounds' MoA, an important step toward safer and more effective therapeutic applications in the future.

From the application point of view, we believe that perhaps the best repurposing and discovery opportunities will originate from pre-clinical 'orphan compounds', that is, de-prioritized targeted agents and tool compounds that currently are not under further development by Pharma or Biotech companies. There already exist a number of pre-competitive, public–private research partnerships, such as the one built on Structural Genomics Consortium, inviting open-source collaboration among academia, industry, hospitals, and patient groups to improve the predictive utility of the current assays based on patient-derived primary cells [79]. We highly encourage such open data and information sharing between all the parties also in the context of compound–target interaction mappings. Along these lines, we have recently initiated a community-driven effort, with the aim to collectively extract, manage, and curate high-quality compound–target bioactivity data from public databases, literature, and other resources. We anticipate that such an open environment, with an increasing number of research groups and pharmaceutical companies joining this crowdsourcing effort, will provide valuable knowledgebase for many exciting applications, including combinatorial target identification, drug discovery, and novel repurposing opportunities using system-level compound–target interaction networks.

## Acknowledgmentst

---

The authors would like to thank Mrs. Janica Wakkinen and Dr. Simon Anders for many useful discussions about different types of experimental assays and computational models.

## Declaration of interest

---

A Cichonska is financially supported by the Biocentrum Helsinki Foundation. T Aittokallio is financially supported

from the Academy of Finland (grants 269862, 272437, 279163 and 292611) and the Cancer society of Finland. The authors have no other relevant affiliations or financial

involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

## Bibliography

Papers of special note have been highlighted as either of interest (●) or of considerable interest (●●) to readers.

1. Scannell JW, Blanckley A, Boldon H, Warrington B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat Rev Drug Discov* 2012;11(3):191-200
2. Mestres J, Gregori-Puigjané E. Conciliating binding efficiency and polypharmacology. *Trends Pharmacol Sci* 2004;30(9):470-4
3. Yıldırım MA, Goh KI, Cusick ME, et al. Drug-target network. *Nat Biotechnol* 2007;25(10):1119-26
4. Peters JU. Polypharmacology—foe or friend? *J Med Chem* 2013;56(22):8955-71
5. Pujol A, Mosca R, Farrés J, Aloy P. Unveiling the role of network and systems biology in drug discovery. *Trends Pharmacol Sci* 2010;31(3):115-23
6. Chandra N, Padiadpu J. Network approaches to drug discovery. *Expert Opin Drug Discov* 2013;8(1):7-20
7. Billur Engin H, Gursoy A, Nussinov R, Keskin O. Network-based strategies can help mono-and poly-pharmacology drug discovery: a systems biology view. *Curr Pharm Des* 2014;20(8):1201-7
8. Nantasenamath C, Prachayasittikul V. Maximizing computational tools for successful drug discovery. *Expert Opin Drug Discov* 2015;10(4):321-9
9. Sun X, Vilar S, Tatonetti NP. High-throughput methods for combinatorial drug discovery. *Sci Translat Med* 2013;5(205):205rv1
10. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol* 2008;4(11):682-90
- **One of the very first papers that discussed the concept of network pharmacology and its potential in drug discovery.**
11. Tang J, Aittokallio T. Network pharmacology strategies toward multi-target anticancer therapies: from computational models to experimental design principles. *Curr Pharm Des* 2014;20(1):23-36
12. Schadt EE, Friend SH, Shaywitz DA. A network view of disease and compound screening. *Nat Rev Drug Discov* 2009;8(4):286-95
- **An early perspective article that already projected the future use of network-centric models in systems medicine.**
13. Boran AD, Iyengar R. Systems approaches to polypharmacology and drug discovery. *Curr Opin Drug Discov Devel* 2010;13(3):297
14. Arrell DK, Terzic A. Network systems biology for drug discovery. *Clin Pharmacol Ther* 2010;88(1):120-5
15. Sampiao EP, Sarno EN, Galilly R, et al. Thalidomide selectively inhibits tumor necrosis factor alpha production by stimulated human monocytes. *J Exper Med* 1991;173(3):699-703
16. Li J, Zheng S, Chen B, et al. A survey of current trends in computational drug repositioning. *Brief Bioinform* 2015. [Epub ahead of print]
- **A comprehensive review of computational drug repositioning strategies.**
17. Pemovska T, Johnson E, Kontro M, et al. Axitinib effectively inhibits BCR-ABL1 (T315I) with a distinct binding conformation. *Nature* 2015;519(7541):102-5
- **An excellent example how phenotype-based compound screening, with compound–target information, can repurpose approved drug for new and personalized use.**
18. Tropsha A. Best practices for QSAR model development, validation, and exploitation. *Mol Inform* 2010;29(6-7):476-88
19. Costello JC, Heiser LM, Georgii E, et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nat Biotechnol* 2014;32(12):1202-12
- **A community-effort to drug sensitivity prediction challenge, as well as to characterizing the factors contributing to the success of the top-performing learning algorithms.**
20. Ding H, Takigawa I, Mamitsuka H, Zhu S. Similarity-based machine learning methods for predicting drug–target interactions: a brief review. *Brief Bioinform* 2014;15(5):734-47
- **A theoretical review of machine learning models for compound–target interaction prediction, as well as of the current computational challenges.**
21. Riddick G, Song H, Ahn S, et al. Predicting in vitro drug sensitivity using random forests. *Bioinformatics* 2011;27(2):220-4
22. Gonen M, Kaski S. Kernelized Bayesian matrix factorization. *Pattern Anal Machine Intell* 2014;36(10):2047-60
23. Yamanishi Y. Chemogenomic approaches to infer drug–target interaction networks. *Methods Mol Biol* 2013;939:97-113
24. Mousavian Z, Masoudi-Nejad A. Drug-target interaction prediction via chemogenomic space: learning-based methods. *Expert Opin Drug Metabol Toxicol* 2014;10(9):1273-87
- **A comprehensive review of the learning-based methods for drug–target interaction prediction including their theoretical aspects.**
25. Davis MI, Hunt JP, Herrgard S, et al. Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol* 2011;29(11):1046-51
26. Metz JT, Johnson EF, Soni NB, et al. Navigating the kinome. *Nat Chem Biol* 2011;7(4):200-2
27. Taipale M, Krykbaeva I, Whitesell L, et al. Chaperones as thermodynamic sensors of drug–target interactions reveal kinase inhibitor specificities in living cells. *Nat Biotechnol* 2013;31(7):630-7
28. Savitski MM, Reinhard FB, Franken H, et al. Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* 2014;346(6205):1255784
29. Schirle M, Bantscheff M, Kuster B. Mass spectrometry-based proteomics in preclinical drug discovery. *Chem Biol* 2012;19(1):72-84
- **A review paper describing the recent advances in experimental approaches for preclinical drug discovery.**
30. Reymond JL, Awale M. Exploring chemical space for drug discovery using the chemical universe database. *ACS Chem Neurosci* 2012;3(9):649-57

31. Tang J, Szwajda A, Shakyawar S. Making sense of large-scale kinase inhibitor bioactivity data sets: A comparative and integrative analysis. *J Chem Inf Model* 2014;54(3):735-43
- **A study of the heterogeneity of bioactivity data across different databases, as well as how to integrate various bioactivity measurements.**
32. Chambers J, Davies M, Gaulton A, et al. UniChem: a unified chemical structure cross-referencing and identifier tracking system. *J Cheminform* 2013;5(1):3
33. van Laarhoven T, Nabuurs SB, Marchiori E. Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics* 2011;27(21):3036-43
34. van Laarhoven T, Marchiori E. Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile. *PLoS ONE* 2013;8(6):e66952
35. Cheng F, Liu C, Jiang J, et al. Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput Biol* 2012;8(5):e1002503
36. Cheng F, Zhou Y, Li W, et al. Prediction of chemical-protein interactions network with weighted network-based inference method. *PLoS ONE* 2012;7(7):e41064
37. Wang Y, Zeng J. Predicting drug-target interactions using restricted Boltzmann machines. *Bioinformatics* 2013;29(13):i126-34
38. Keiser MJ, Roth BL, Armbruster BN, et al. Relating protein pharmacology by ligand chemistry. *Nat Biotechnol* 2007;25(2):197-206
- **A fundamental work that introduced the statistical model SEA for relating proteins based on the chemical similarity of their ligands.**
39. Keiser MJ, Setola V, Irwin JJ, et al. Predicting new molecular targets for known drugs. *Nature* 2009;462(7270):175-81
40. Yamanishi Y, Araki M, Gutteridge A, et al. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 2008;24(13):i232-40
- **A pivotal study on the use of machine learning models to integrate chemical and genomic profiles for drug-target interaction prediction.**
41. Bleakley K, Yamanishi Y. Supervised prediction of drug-target interactions using bipartite local models. *Bioinformatics* 2009;25(18):2397-403
42. Yamanishi Y, Kotera M, Kanehisa M, Goto S. Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics* 2010;26(12):i246-54
43. Alaimo S, Pulvirenti A, Giugno R, Ferro A. Drug-target interaction prediction through domain-tuned network-based inference. *Bioinformatics* 2013;29(16):2004-8
44. Chen X, Liu MX, Yan GY. Drug-target interaction prediction by random walk on the heterogeneous network. *Mol BioSys* 2012;8(7):1970-8
45. Wang W, Yang S, Zhang X, Li J. Drug repositioning by integrating target information through a heterogeneous network model. *Bioinformatics* 2014;30(20):2923-30
- **An interesting computational model for simultaneous drug-target and drug-disease links inference for drug-repurposing applications.**
46. Hamosh A, Scott AF, Amberger JS, et al. Online mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 2005;33(1):514-17
47. Campillos M, Kuhn M, Gavin AC, et al. Drug target identification using side-effect similarity. *Science* 2008;321(5886):263-6
- **A pivotal work on the use of phenotypic side-effect similarities of drugs in the task of drug target identification.**
48. 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(5795):1929-35
49. Iorio F, Bosotti R, Scacheri E, et al. Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc Natl Acad Sci* 2010;107(33):14621-6
- **An important study that exploited the similarities in gene expression profiles to support network pharmacology.**
50. Ravindranath AC, Perualila-Tan N, Kasim A, et al. Connecting gene expression data from connectivity map and in silico target predictions for small molecule mechanism-of-action analysis. *Mol BioSys* 2015;11(1):86-96
51. Zhao S, Li S. Network-based relating pharmacological and genomic spaces for drug target identification. *PLoS One* 2010;5(7):e11764
52. Laenen G, Thorrez L, Börnigen D, Moreau Y. Finding the targets of a drug by integration of gene expression data with a protein interaction network. *Mol BioSys* 2013;9(7):1676-85
53. Perlman L, Gottlieb A, Atias N, et al. Combining drug and gene similarity measures for drug-target elucidation. *J Comput Biol* 2011;18(2):133-45
54. Magger O, Waldman YY, Ruppin E, Sharan R. Enhancing the prioritization of disease-causing genes through tissue specific protein interaction networks. *PLoS Comput Biol* 2012;8(9):e1002690
55. Liu W, Wang J, Wang T, Xie H. Construction and analyses of human large-scale tissue specific networks. *PLoS ONE* 2014;9(12):e115074
56. Greene CS, Krishnan A, Wong AK, et al. Understanding multicellular function and disease with human tissue-specific networks. *Nat Gen* 2015;47(6):569-75
- **A recent, excellent work on modeling and utilizing tissue-specific networks for disease target identification and function prediction.**
57. Woo JH, Shimoni Y, Yang WS, et al. Elucidating compound mechanism of action by network perturbation analysis. *Cell* 2015;162(2):441-51
- **A recent network-based approach for elucidating genome-wide MoA proteins for given compounds using tissue-specific networks.**
58. Sanseau P, Agarwal P, Barnes MR, et al. Use of genome-wide association studies for drug repositioning. *Nat Biotechnol* 2012;30(4):317-20
- **A pilot study showing the applicability of GWAS in the drug repositioning.**
59. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. *Nat Rev Drug Discov* 2013;12(8):581-94
60. Nelson MR, Tipney H, Painter JL, et al. The support of human genetic evidence for approved drug indications. *Nat Gen* 2015;47(8):856-60
61. Wang H, Gu Q, Wei J, et al. Mining drug-disease relationships as a complement to medical genetics-based drug repositioning: Where a

- recommendation system meets genome-wide association studies. *Clin Pharmacol Ther* 2015;97(5):451-4
62. Taşan M, Musso G, Hao T, et al. Selecting causal genes from genome-wide association studies via functionally coherent subnetworks. *Nat Methods* 2015;12(2):154-9
63. Xie L, Ge X, Tan H, et al. Towards structural systems pharmacology to study complex diseases and personalized medicine. *PLoS Comput Biol* 2014;10(5):e1003554
64. Denny JC, Bastarache L, Ritchie MD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol* 2013;31(12):1102-11
65. Rastegar-Mojarad M, Ye Z, Kolesar JM, et al. Opportunities for drug repositioning from phenome-wide association studies. *Nat Biotechnol* 2015;33(4):342-5
66. Wang ZY, Zhang HY. Rational drug repositioning by medical genetics. *Nat Biotechnol* 2013;31(12):1080-2
67. Lim ET, Würtz P, Havulinna AS, et al. Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet* 2014;10(7):e1004494
68. Barrangou R, Birmingham A, Wiemann S, et al. Advances in CRISPR-Cas9 genome engineering: lessons learned from RNA interference. *Nucl Acids Res* 2015;43(7):3407-19
69. Shi J, Wang E, Milazzo JP, et al. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol* 2015;33(6):661-7
70. Chen S, Feng J, Ma L, et al. RNA interference technology for anti-VEGF treatment. *Expert Opin Drug Del* 2014;11(9):1471-80
71. Liang P, Xu Y, Zhang X, et al. CRISPR/Cas9-mediated gene editing in human triploid zygotes. *Protein Cell* 2015;6(153):363-72
72. Harper AR, Topol EJ. Pharmacogenomics in clinical practice and drug development. *Nat Biotechnol* 2012;30(11):1117-24
- **A perspective paper that discussed the use of genomic information in the various steps of the drug development process.**
73. Wheeler HE, Maitland ML, Dolan ME, et al. Cancer pharmacogenomics: strategies and challenges. *Nat Rev Genet* 2013;14(1):23-34
74. Berlow N, Haider S, Wan Q, et al. An integrated approach to anti-cancer drug sensitivity prediction. *Comput Biol Bioinform* 2014;11(6):995-1008
75. Ammad-ud-din M, Georgii E, Gonen M, et al. Integrative and personalized QSAR analysis in cancer by kernelized bayesian matrix factorization. *J Chem Inf Model* 2014;54(8):2347-59
76. National center for biotechnology information. PubChem compound database; CID=126941. Available from: <https://pubchem.ncbi.nlm.nih.gov/compound/126941> [Last accessed 3 September 2015]
77. Kibble M, Saarinen N, Tang J, et al. Network pharmacology applications to map the unexplored target space and therapeutic potential of natural products. *Nat Product Reports* 2015;32(8):1249-66
78. Pahikkala T, Airola A, Pietilä S, et al. Toward more realistic drug-target interaction predictions. *Brief Bioinform* 2015;16(2):325-37
79. Edwards AM, Arrowsmith CH, Bountra C, et al. Preclinical target validation using patient-derived cells. *Nat Rev Drug Discov* 2015;14(3):149-50
80. Bento AP, Gaulton A, Hersey A, et al. The ChEMBL bioactivity database: an update. *Nucl Acids Res* 2014;42(D1):D1083-90
81. Law V, Knox C, Djoumbou Y, et al. DrugBank 4.0: shedding new light on drug metabolism. *Nucl Acids Res* 2014;42(D1):D1091-7
82. Liu T, Lin Y, Wen X, et al. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucl Acids Res* 2007;35(1):D198-201
83. Kuhn M, Szklarczyk D, Pletscher-Frankild S, et al. STITCH 4: integration of protein-chemical interactions with user data. *Nucl Acids Res* 2014;42(D1):D401-7
84. Kanehisa M, Goto S, Sato Y, et al. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucl Acids Res* 2014;42(D1):D199-205
85. Kim S, Thiessen PA, Bolton EE, Bryant SH. PUG-SOAP and PUG-REST: web services for programmatic access to chemical information in PubChem. *Nucl Acids Res* 2015;43(W1):W605-11
86. Chang A, Schomburg I, Placzek S, et al. BRENDA in 2015: exciting developments in its 25th year of existence. *Nucl Acids Res* 2015;43(D1):D439-46
87. Griffith M, Griffith OL, Coffman AC, et al. DGIdb: mining the druggable genome. *Nat Methods* 2013;10(12):1209-10
88. Qin C, Zhang C, Zhu F, et al. Therapeutic target database update 2014: a resource for targeted therapeutics. *Nucl Acids Res* 2014;42(D1):D1118-23
89. Günther S, Kuhn M, Dunkel M, et al. Supertarget and matador: resources for exploring drug-target relationships. *Nucl Acids Res* 2008;36(1):D919-22
90. Roth BL, Lopez E, Patel S, Kroeze WK. The multiplicity of serotonin receptors: uselessly diverse molecules or an embarrassment of riches? *Neuroscientist* 2000;6(4):252-62
91. Irwin JJ, Sterling T, Mysinger MM, et al. ZINC: a free tool to discover chemistry for biology. *J Chem Inf Model* 2012;52(7):1757-68

#### Affiliation

Anna Cichonska<sup>1,2</sup> MSc, Juho Rousu<sup>3</sup> PhD & Tero Aittokallio<sup>†4,5</sup> PhD

<sup>†</sup>Author for correspondence

<sup>1</sup>University of Helsinki, Institute for Molecular Medicine Finland FIMM, Helsinki, Finland

<sup>2</sup>HIIT/FIMM-EMBL PhD student, Aalto University, Helsinki Institute for Information Technology HIIT, Department of Computer Science, Espoo, Finland

<sup>3</sup>Professor at Aalto University, Aalto University, Helsinki Institute for Information Technology HIIT, Department of Computer Science, Espoo, Finland

<sup>4</sup>EMBL-FIMM Group Leader, University of Helsinki, Institute for Molecular Medicine Finland FIMM, Helsinki, Finland  
Tel: +358 5 03 18 24 26;

E-mail: [tero.aittokallio@fimm.fi](mailto:tero.aittokallio@fimm.fi)

<sup>5</sup>Professor, University of Turku, Department of Mathematics and Statistics, Turku, Finland