

Chapter 2

Networks and Pathways in Systems Pharmacology

Abstract This chapter presents an extensive overview of aspects involved in the thriving field of systems pharmacology. The three main directions along which network- and pathway-based analysis methods can contribute in systems pharmacology are spotlighted. Current approaches for the characterization of drugs mechanism of action, including the elucidation of mechanisms through which disease phenotypes dysregulate biological processes are first discussed. Subsequently, the latest research work done in systems pharmacology and polypharmacology toward the identification of novel drug targets, as well as in optimizing drug combinations for more efficient therapies, is surveyed. Within this context, the benefits of integrating evidence from multiple biological scales are examined, and the most popular databases used to store various biological data are provided. Drug repositioning is another direction along which pathway analysis is bound to bring significant contributions. An overview of drug repositioning approaches based on molecular and phenotypic profiles is presented. Subsequently, the main aspects involved in systems pharmacology applications for in silico drug side effect modeling and prediction are reviewed. Finally, current challenges and future considerations for pathway analysis and systems pharmacology are discussed.

Keywords Pathway analysis • Sub-pathways • Systems pharmacology • Drug mechanism of action • Polypharmacology • Drug repositioning • Drug targets • Drug safety • Drug side effect modeling • Heterogeneous data integration • Databases

2.1 Introduction

In the context of continuously surging drug development and healthcare costs, with the cost of developing a new drug being recently estimated at \$2.8 billion, a more than 145% increase within the past decade only (DiMasi et al. 2016), and with global annual spending on prescription medication forecasted to reach \$1.8 trillion, it becomes increasingly clear that the conventional reductionist *one drug, one target, one disease* paradigm which has been traditionally driving pharmacology

needs radical rethinking. Even with the significant raise in R&D expenditure by big pharma companies, the mean time between synthesis to approval surpasses 120 months, while the number of newly approved molecular compounds annually is $\sim 20\text{--}30$, not significantly different from what it was half a century ago (DiMasi et al. 2016; Csermely et al. 2013). Systems pharmacology brings the promise of revolutionizing the drug discovery process, while at the same time catalyzing the translation of pharmacogenomics applications to clinical environment, which has been lagging behind despite the recent wave of groundbreaking research on genomics implications in disease.

In this context, methods for modeling and analysis of molecular interaction networks, which have recently found extensive application in systems biology, are able to provide a theoretical platform for systems pharmacology. Studies on gene regulatory networks, protein–protein interaction networks, metabolic networks and other types of molecular interaction networks, provided significant insight into cellular organization and behavior, and shed light on specific biological processes, as well as disease processes and pathophysiology (Rual et al. 2005; Jeong et al. 2000; Ideker et al. 2002; Maraziotis et al. 2006, 2007; Bezerianos and Maraziotis 2008; Glaab et al. 2010). Consequently, based on this new network-based paradigm, new areas of translational research have emerged, and new terms have been coined, such as network physiology, network medicine, and network pharmacology (Barabasi et al. 2011; Hopkins 2008; Bashan et al. 2012).

Analysis of molecular interaction networks in systems pharmacology holds the promise of contributing along three main directions (Fig. 2.1):

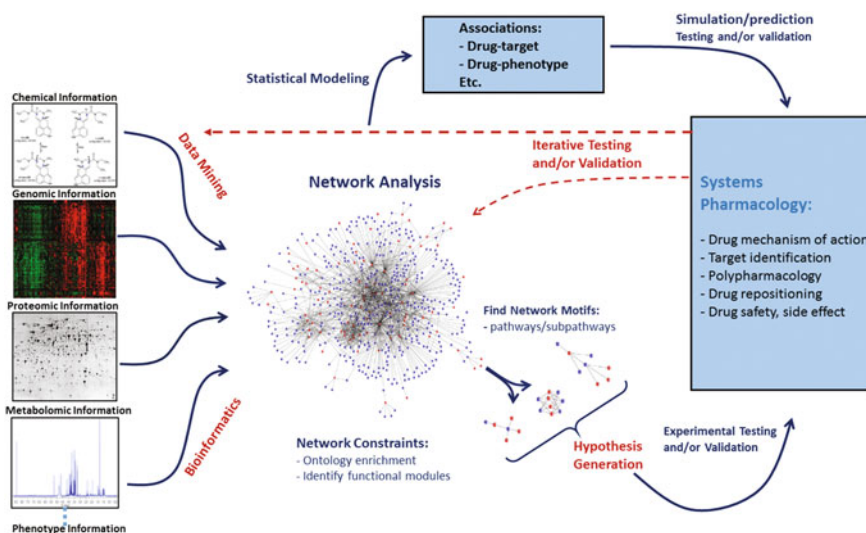


Fig. 2.1 Overview of network-based analysis in systems pharmacology [adapted from Arell and Terzic (2010)]

- (i) Allowing for the identification of new putative drug targets relevant to specific diseases, through a better characterization of what makes an optimal target. In this context, pathway-based analysis allows a more mechanistic characterization of drugs *mechanism of action*, including the characterization of response to treatment, challenging the traditional way drug action was viewed: act on a specific target and observe the modulating effects downstream of that target. Current view is that a drug hits several targets (including off-targets) co-existing in a complex interacting network which is perturbed by disease, and the therapeutic effect of the drug aims to re-establish homeostasis (Berger and Iyengar 2009; Xie et al. 2009; Arell and Terzic 2010; Woo et al. 2015). Often times such approaches combine pathway and network analyses with pharmacokinetic and pharmacodynamic models to incorporate data from multiple biological scales, striving to build advanced quantitative and predictive models of therapeutic efficacy. As a corollary to achieving (i) follows the improved ability to predict effective drug combinations and the possibility to investigate mechanisms underlying drug resistance (Boran and Iyengar 2010; Zhao et al. 2013; Reddy and Zhang 2013; Lazar et al. 2014; Hwang et al. 2016).
- (ii) *Drug repositioning* or *drug repurposing* is another direction in which systems pharmacology is making significant impact. Motivated by the success stories of several drugs with different initial indications, such as sildenafil (initially developed to treat hypertension and angina pectoris, eventually used to treat erectile dysfunction after clinical trial observations), or monoclonal antibody bevacizumab (originally developed to treat colon cancer and non-small cell lung cancer, currently used in treatment of macular degeneration disease), drug repositioning significantly shortens the path for approval of normal drugs and reduces the R&D expenditure (Van Eichborn et al. 2011; Wu et al. 2013; Pan et al. 2014; Li et al. 2016). Almost 20% of new drugs introduced to market in 2013 were actually new indications for existing drugs (Li et al. 2016). Originally based on serendipitous clinical observations, drug repositioning is picking up significant interest recently due to the increased understanding of the underlying molecular processes, drugs mechanisms of action, as well as the availability of advanced computational models for network and pathway-based analysis.
- (iii) Another direction in which significant research efforts in systems pharmacology are focusing is that of *drug safety* and prediction of *drug toxicity* and *side effects*. Drug safety is a major source of drug attrition and of vital interest for pharmaceutical companies in their efforts to reduce drug development cost, while increasing efficiency (Hutchinson and Kirk 2011; Waring et al. 2015). Recent high profile failures during clinical trials or even for marketed drugs underline the fact that even efficacious drugs may cause severe side effects with dangerous consequences. Some examples include the cases of rosiglitazone, an antidiabetic drug which was later found to induce significant risk for myocardial infarction, rofecoxib, a pain relief drug recalled from the market after increased risk of stroke was reported, and the BIA-10-2474

(a molecule developed for a range of diseases) clinical trial death cases in 2016 (Graham et al. 2005; Nissen and Wolski 2007; Esserink 2016). It is therefore of paramount importance that molecular mechanisms of drug toxicity are comprehensively evaluated and used for hypothesis generation and testing, having as goal the development of *in silico* models for prediction of side effects. Systems pharmacology provides the framework for augmenting traditional pharmacokinetic and pharmacodynamics models while studying most common scenarios of drug toxicity from a pathway-based perspective: (a) off-target perturbations generating side effects unrelated to on-target effects, (b) side effects caused by pathways downstream of the intended on-target and (c) unrelated pathways generating side effects due to cross-talk with pathways downstream of intended target (Boran and Iyengar 2010; Wallach et al. 2010; Kuang et al. 2014; Lorberbaum et al. 2015; Cao et al. 2015; Trame et al. 2016; Schotland et al. 2016).

A concept with significant overlap to systems pharmacology, in both that it integrates systems biology with drug discovery and in its application areas, is *polypharmacology*. Polypharmacology includes studying the modulation of multiple targets by single drugs, as well as modulation of different targets by multiple drugs, primarily focusing on therapeutic interventions in complex diseases with the goal of identifying less toxic and more effective approaches (Boran and Iyengar 2010; Reddy and Zhang 2013; Anighoro et al. 2014). Another discipline that naturally converges to the more inclusive field of systems pharmacology is *pharmacogenomics*. Pharmacogenomics is defined by its search for variation in the human genome that explains inter-individual drug response variability (Antman et al. 2012). Currently in its incipient stage, with few genotype-drug response associations identified and finding their way into clinical practice by means of biomarkers present on drug labeling (FDA: Table of Pharmacogenomic Biomarkers in Drug Labeling 2016), translation of pharmacogenomic associations into clinical practice is still slowed by inconsistent findings and below par predictive power. Since these limitations are largely due to the complex interactions between drug-specific molecular response and environmental factors, systems pharmacology holds the promise to facilitate pharmacogenomics in unraveling the mechanisms behind the drug response variability. Rather than just identify mutations associated to diseases (e.g., genome-wide association studies), or perform statistical correlation type analysis between genetic signatures and patient phenotype, network- and pathway-based approaches of systems pharmacology allow integration of additional information for a better understanding of the bases of inter-individual variation, and in conjunction with pharmacogenomics, eventually lead toward the overarching goal of precision medicine (Turner et al. 2015).

The rest of this chapter is structured as follows: Sect. 2.2 describes current approaches in network and pathway-based characterization of drugs mechanism of action, Sect. 2.3 presents latest research work done in systems pharmacology and polypharmacology toward the identification of new drug targets, Sect. 2.4 provides an overview of systems pharmacology approaches in drug repositioning, Sect. 2.5

presents systems pharmacology applications for in silico drug side effect modeling and prediction. Final section presents current challenges and future considerations for pathway analysis and systems pharmacology.

2.2 Network- and Pathway/Sub-pathway-Based Characterization of Drugs Mechanism of Action

Initial efforts deviating from the traditional *one drug-one target-one disease* paradigm, and the related search for highly selective ligands that dominated the past decades, were triggered by the recognition that pharmacological compounds modulate the activity of targets in complex networks of deregulations underlying disease phenotypes (Gardner et al. 2003; Ambesi-Impiombato and Di Bernardo 2005; Hopkins 2008; Turner et al. 2015). These observations and the ensuing endeavor for investigating the compounds mechanisms of action (MoA) were only possible with the advent of high throughput technologies which started generating wreaths of data and with the concomitant rise of the new field of systems biology (Ideker et al. 2001).

The elucidation of mechanisms by which drug compounds affect the deregulated interactions in disease phenotypes is bound to become an essential part of the modern drug discovery process. With this comes an increased need for computational methods to mine large datasets and assist in providing initial hypotheses for further in vitro and in vivo validation studies. About a decade ago, data resources originating from genome-wide transcriptional profiles and containing drug response phenotypes, such as the Connectivity Map (CMap—which contains more than 7000 gene expression profiles obtained in response to treatment with 1309 drug and drug-like small molecules) became available, followed in recent years by similar databases, such as the Library of Integrated Network-based Cellular Signatures project (LINCS) (Lamb et al. 2006; Wang et al. 2016). The use of gene expression data (transcriptional mRNA profiles, initially obtained from microarray experiments, more recently from RNAseq experiments) in investigating drugs' MoA has become norm, as this type of data allows genome-wide investigation of drug response' correlation with disease phenotype. Early work successfully characterized compounds perturbation mechanisms by searching for commonalities in the phenotypic responses based on the simple hypothesis that, if two drugs induce similar transcriptional responses they potentially share a common MoA and a similar therapeutic application, even if they act on different cellular target (Kibble et al. 2016). This idea was adapted from early investigations in genomic data analysis in which it was observed that genes with similar expression profiles are more likely to be involved in common biological processes. Transcriptional response profiles were initially compared using various methods similar to the Gene Set Enrichment Analysis (GSEA), based on the Kolmogorov–Smirnov statistic (Subramanian et al. 2005). Briefly, query signature profiles' similarity to the reference expression

profiles in the CMap database is assessed. Query profiles are usually sets of genes differentially expressed between disease and normal conditions, or sets of up- and/or down-regulated genes. In parallel, genes on the reference CMap arrays (each one corresponding to experiments in which cells are perturbed using a specific drug) are rank-ordered according to their differential expression relative to control. Subsequently, the query signature is compared to every rank-ordered gene list and it is determined whether up-regulated genes in the query tend to be located near the top of the list and down-regulated genes are found toward the bottom of the reference ranked list, or vice versa. The former case denotes a ‘positive connectivity’ and the latter a ‘negative connectivity’ between the query and the respective perturbation instance (array containing the cells gene expressions in response to the drug treatment). Connectivity scores are then computed and used to rank all instances in the database according to their correlation to the query signature. This approach was used by Lamb et al. (2006) to elucidate the MoA of uncharacterized drug compounds, such as gedunin. The mechanism through which gedunin is capable of abrogating the expression of androgen receptor (AR) activation in prostate cancer was determined by finding high connectivity scores of a gedunin signature with multiple instances of three heat shock protein 90 inhibitors (HSP90): geldanamycin, 17-allylamino-geldanamycin, and 17-dimethylamino-geldanamycin (Lamb et al. 2006). It was therefore inferred that gedunin might impinge upon the HSP90 pathway, hypothesis which was subsequently validated experimentally. This hypothesis would not have been warranted by solely studying compounds structures, as gedunin is structurally dissimilar to known HSP90 inhibitors.

Various other approaches based on ranked lists of differentially expressed genes, have been used, such as the MANTRA method (Iorio et al. 2010), which adopts a rank-aggregation procedure to dilute cell-line-specific effects in transcription, as well as experimental batch effects, or different drug concentrations in different treatment instances. Iorio et al. (2010) defined pairwise distances between compounds using ‘enrichment scores’ based on the distribution of optimal gene signatures of each compound (extracted as top and bottom 250 genes in their corresponding ranked lists) within the ranked gene list of the other compound of the pair and vice versa. These distances were used to build a drug network in which nodes correspond to compounds and connecting edges reflect the estimated distances between the compound pairs. This network was subsequently mined via network clustering to identify communities (or modules) of closely interconnected compounds. The retrieved drug modules were found to be highly enriched with common biological pathways and characterized by similar MoAs. The authors have then proceeded to predict MoA for anticancer drugs with profiles not present in the reference CMap database, by estimating the distance of their transcriptional profiles to the drug network modules. Following this framework, PHA-690509, PHA-793887, and PHA-848125 were correctly classified as CDK inhibitors, distinct from the other kinase inhibitors in the CMap database, and were also predicted to have highly similar MoA to Topoisomerase inhibitors. The original method in (Iorio et al. 2010) was recently extended to filter out spurious effects of compounds’ nonspecific secondary effects on transcriptional profiles. To this goal, they use an

iterative supervised approach to refine the original drug network module of a compound of interest while deriving a transcriptional signature representative of the primary MoA (Iorio et al. 2015).

Some studies have argued that methods based solely on differentially expressed sets of genes (i.e., transcriptional profiles) may miss essential knowledge on regulatory influence among genes and their products. Consequently, methods such as the mode-of-action by network identification (MNI), which incorporates differential expression of genes with regulatory information encoded in gene networks structures, have been proposed (Xing and Gardner 2006). In MNI systems of linear differential equations are used initially to build the gene network model and the subsequent inference of network parameters is done based on transcriptional profiles. Once the canonical gene network is created, it is used to filter test transcriptional profiles from drug treatment experiments in order to distinguish genes that are mediators of treatment response from the other genes which exhibit expression changes. This is achieved by searching for genes with changes in their transcriptional profiles that are not in accordance with the canonical gene network, under the assumption that such genes are perturbed by the drug treatment. Significance of the perturbation on these putative molecular targets is quantified using a z -score scheme. MNI was utilized to identify molecular targets of antifungal compounds based on genome-wide transcriptional profiles in yeast.

Recently, it was proposed that data from additional sources, such as signaling and metabolic pathway databases, protein structure databases, compound structure and drug target databases, as well as DNA sequence or functional non-coding RNA, may be incorporated in the analysis. This integrative approach has the potential to enrich the computational model, by making it more biologically plausible, and enhance its predictive power (see Fig. 2.1). Table 2.1 presents some of the most commonly used databases containing data and annotations involved in MoA identification and generally in drug discovery. Within this context, Iskar et al. (2013) used bi-clustering to identify drug-induced transcriptional modules from human and rat transcriptional profiles databases [CMap and DrugMatrix (Ganter et al. 2005)]. The modules conserved across organisms were checked for functional coherence at protein level using information from the STRING database (Szklarczyk et al. 2014) and then connected into a module network. The module network was extensively characterized by annotation with relevant pathways and functional information from KEGG (Kanehisa et al. 2015), BioCarta (Nishimura 2001) and the Gene Ontology (Gene Ontology Consortium 2013) databases, as well as drug structure, target, and side effect information from STITCH and SIDER databases (Szklarczyk et al. 2015; Kuhn et al. 2015). The integrative model thus defined allowed the authors to discover novel MoAs for six drugs, four with cell-line-specific mechanism and two with mechanisms conserved in all modules, using module-based statistical tests and overrepresentation analysis. Specifically, zaprinast, was suggested to be a novel modulator of the PPAR γ receptor in the PC3 cell line, the main target of antidiabetic drugs, a hypothesis subsequently validated with target binding assays experiments. Similarly, nitrendipine was found to be a modulator of estrogen receptor in MCF7 cells, hexetidine and (+)-chelidoniumine were

Table 2.1 Public databases commonly used in systems pharmacology

Database	Web address	Notes
DrugBank	http://www.drugbank.ca/	Richly annotated bioinformatics and cheminformatics resource that combines detailed drug data (e.g., chemical, pharmacological, and pharmaceutical) with comprehensive target information (e.g., sequence, structure and pathway)
TTD	http://bidd.nus.edu.sg/group/cjttd/TTD_HOME.asp	Therapeutic target database (TTD) provides the information about known and explored therapeutic protein and nucleic acid targets, the targeted diseases, pathway information and corresponding drugs directed at each of these targets
SuperTarget	http://insilico.charite.de/supertarget/	SuperTarget is an extensive database for analyzing drug–target interactions. This database allows querying by drugs, targets, drug–target-related pathways, drug–target-related ontologies
STITCH	http://stitch.embl.de/	Database of known and predicted chemical–protein interactions, which integrates the evidence derived from experiments, other databases, and literature
BindingDB	http://www.bindingdb.org/bind/index.jsp	The BindingDB is a binding database of experimentally determined protein–ligand binding affinities among small molecule ligands and protein targets
SIDER	http://sideeffects.embl.de/	Public and computer-readable database that contains information on marketed drugs and their recorded side effects (i.e., adverse drug reactions), including side-effect frequency, drug and side effect classifications as well as drug–target associations
OFFSIDES/TWOSIDES	http://tatonettilab.org/resources/tatonetti-stm.html	Offsides database is a resource of off-label side effects not listed on the FDA's official drug label. Twosides databases is a resource of polypharmacy side effects for pairs of drugs
FAERS	http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm082193.htm	FDA database that contains the information obtained from adverse event and medication error reports submitted to FDA on side effect keywords (adverse event keywords) for drugs

(continued)

Table 2.1 (continued)

Database	Web address	Notes
MATADOR	http://matador.embl.de/	Database resource for protein–chemical interactions, including multiple direct and indirect modes of drug–target interactions. The manually annotated list of direct (binding) and indirect interactions between proteins and chemicals was assembled by automated text mining followed by manual collection
ChEMBL	https://www.ebi.ac.uk/chembl/	Literature-derived database of molecule structures and molecule–protein interactions. This includes a catalog of approved drugs
ChemBank	http://chembank.broadinstitute.org/	Database containing collections of data derived from small molecules and small molecule screens, and resources for studying their properties
PubChem	https://pubchem.ncbi.nlm.nih.gov/	Public repository of biological assay results for hundreds of thousands of molecules
canSAR	https://cansar.icr.ac.uk	Cancer-related resource on biological annotations, chemical screening, RNA interference screening, expression, amplification and 3D structural data
Gene Ontology	http://geneontology.org/	Bioinformatics initiative to unify the representation of gene and gene product attributes across all species. Aims to maintain and develop its controlled vocabulary of gene and gene product attributes, annotate genes and gene products
STRING	http://string-db.org/	Database of known and predicted protein–protein interactions. The interactions include direct (physical) and indirect (functional) associations; they stem from computational prediction, from knowledge transfer between organisms, and from interactions aggregated from other (primary) databases
CMap	http://portals.broadinstitute.org/cmap/	Gene expression profiles of 1309 drugs across human primary cell lines
Gene Expression Omnibus (GEO)	https://www.ncbi.nlm.nih.gov/geo/	NIH public repository of functional genomics data

(continued)

Table 2.1 (continued)

Database	Web address	Notes
LINCS	http://www.lincsproject.org/	Large scale NIH repository containing gene expression profiles of drugs and genetic perturbagens across numerous cell lines
DrugMatrix	https://ntp.niehs.nih.gov/drugmatrix/index.html	Molecular toxicology reference database; is populated with the comprehensive results of thousands of highly controlled and standardized toxicological experiments in which rats or primary rat hepatocytes were systematically treated with therapeutic, industrial, and environmental chemicals at both non-toxic and toxic doses
PharmGKB	https://www.pharmgkb.org/	An interactive tool for researchers investigating how genetic variation effects drug response. Contains corresponding drug–gene relations
KEGG	www.genome.jp/kegg/	Resource for understanding high-level functions and utilities of the biological system from molecular-level information. Contains signaling pathways, drug, structure and chemical information
IntAct	http://www.ebi.ac.uk/intact/main.xhtml	Database of evidence for molecular interactions, and maintains the reference resource for macromolecular complexes
HINT	http://hint.yulab.org/	Database of high-quality protein–protein interactions in different organisms. These have been compiled from different sources and then filtered both systematically and manually to remove erroneous and low-quality interactions. HINT can be used for individual queries as well as for batch downloads
HPRD	http://www.hprd.org/	Database with information on human protein functions including protein–protein interactions, post-translational modifications, enzyme-substrate relationships and disease associations
BIOGRID	https://thebiogrid.org/	Public database that archives and disseminates genetic and protein interaction data from model organisms and humans

(continued)

Table 2.1 (continued)

Database	Web address	Notes
DIP	http://dip.doe-mbi.ucla.edu/dip/Main.cgi	Database which catalogs experimentally determined interactions between proteins. It combines information from a variety of sources to create a single, consistent set of protein–protein interactions. The data stored within DIP have been curated, both manually, by expert curators and automatically, using computational approaches
PathwayCommons	http://www.pathwaycommons.org/	Database of biological pathways and interactions from multiple organisms
BioCarta-PID	http://cgap.nci.nih.gov/Pathways/BioCarta_Pathways	NIH NCI resource on pathways and pathways interaction
Wiki Pathways	http://www.wikipathways.org/index.php/WikiPathways	Open, collaborative platform dedicated to the curation of biological pathways. WikiPathways thus presents a new model for pathway databases that enhances and complements ongoing efforts, such as KEGG, Reactome and Pathway Commons
Reactome	http://www.reactome.org/	Open-source, curated and peer reviewed pathway database
TarBase	http://diana.imis.athena-innovation.gr/DianaTools/index.php?i=tarbase/index	Provides high-quality manually curated experimentally validated miRNA:gene interactions, enhanced with detailed meta-data
miRGen	http://carolina.imis.athena-innovation.gr/diana_tools/web/index.php?i=mirgenv3%2Findex	miRGen is an integrated database of miRNA gene transcripts, transcription factor binding sites, miRNA expression profiles and single nucleotide polymorphisms associated with miRNAs
HMDB	http://www.hmdb.ca/	Database containing small molecule metabolites found in the human body
BRENDA	http://brenda-enzymes.info/	Enzymes database
OMIM	www.omim.org/	Online catalog of human genes and genetic disorders containing gene–phenotype associations
ConsensusPathDB	http://consensuspathdb.org/	Integrates interaction networks in Homo sapiens including binary and complex protein–protein, genetic, metabolic, signaling, gene regulatory and drug–target interactions, as well as biochemical pathways

found and experimentally confirmed to have adrenergic activity. Additionally, the same study identified novel functions for 10 previously poorly characterized genes as modulators of cholesterol homeostasis, based on their strong connections within the transcriptional modules enriched for cholesterol biosynthesis pathways.

Using an approach that attempts to both capture regulatory information encoded in the interaction network and integrate various levels of information (transcriptional, signaling, and protein-level interactions), Woo et al. (2015), extend the cell type-specific approach to a tissue-specific one. To this goal they build lymphoma-specific regulatory networks based on transcriptional profiles on in vivo and in vitro drug perturbations. Their approach incorporated translational level information (protein-protein interaction data) and protein-DNA interaction data to create the contextualized regulatory network. MoAs are characterized by modeling and quantifying compounds' dysregulation of network neighborhoods using a probabilistic framework based on Gaussian kernel smoothing. The approach allows the authors to mechanistically elucidate MoAs, while accounting for differential expression of associated nodes (genes or proteins) from a network-based perspective, rather than a purely statistical one. Their study highlighted key differences in topoisomerase (TOP) inhibitor compounds doxorubicin, camptothecin, and etoposidine, which all have previously known significant common footprint. The identified specific effectors were validated experimentally, confirming the approach's high specificity. The same method was used to identify novel compound effectors and modulators for vincristine (a microtubule formation inhibitor in mitotic spindle), mitomycin C, and altretamine (antineoplastic drugs).

A relatively recent trend in pathway analysis, as highlighted in the previous chapters, is that of sub-pathway-based approaches. Investigating sub-pathways may be more relevant in interpreting the biological processes, since it is known that, frequently, only some regions of pathways are dysregulated by disease, or involved in drug related perturbations. Within this context, Chen et al. (2011) have devised a method to identify sub-pathways involved in dexamethasone (DEX) response in human prostate cancer cell lines. Their approach relied on parsing sub-pathways from the KEGG Pathway database in an exhaustive manner. Sub-pathways were defined as individual paths from start points to end points in a pathway map. Such an approach is biologically relevant, as pathway maps in KEGG database are linear sequences connecting biologically meaningful start nodes (which are commonly membrane receptors or their ligands) to end points which are commonly transcription factors or their targets. The resulting sub-pathways were overlaid with transcriptional profile data of a subset of CMap (instances of DEX treated cells). In order to identify sub-pathways significant for DEX-response, a two-stage approach was followed, by defining aggregate distances between sub-pathway states pre- and post-treatment in terms of their contained genes expression levels, and subsequently identifying through statistical analysis key subsets of genes most perturbed by drug, and therefore deemed top contributors to the sub-pathway state differentiation. Based on this, authors were able to assert that the decrease of VEGFR and EGFR stabilization in order to suppress angiogenesis is a hallmark of DEX-response.

Pritchard et al. (2013) proposed an innovative analysis pipeline based on supervised and unsupervised machine learning methods with the goal of achieving both statistical and biological generalization (predictive accuracy), and at the same time ensure the ability of their framework to recognize novel MoAs for drugs. To this goal they define the drug MoAs in terms of subnetworks consisting of drug nodes and edges representing weighted connections between nodes. The weighted connections correspond to distances in the molecular signature space. Initial subnetwork membership is based on biochemical and genetic evidence encompassing three different types of data: mRNA, chemical interaction and RNAi and each subnetwork corresponds to a drug MoA. The training set corresponds to subnetworks of known drug MoAs. Given a test set of uncharacterized drugs, predictions are made based on a k-nearest neighbors method, and putative MoAs are obtained based on sets of representative features corresponding to subnetworks in the training set. A prediction may interpolate within an existing subnetwork or extrapolate to form a new expanded subnetwork. Detection of new MoAs is warranted when a too large expansion of subnetworks is needed. By using a consensus approach the method identifies new clusters within the training set drugs, based on their molecular features. Subsequently, unsupervised learning (hierarchical clustering) is utilized to identify optimal topological thresholds for the connecting edges within the newly derived subnetworks. The procedure enables the detection of more than mere combinations of existing subnetwork motifs, thus permitting the extension to MoAs underlying entirely distinct biology. Using this subnetwork-based signature, authors confirmed MoA subnetworks for HSP90 and EGFR inhibitors suggested in previous studies. Additionally, they were able to successfully confirm and expand MoA classes including erastin (a Bax/Bak independent death inducing compound), mitochondrial disruptors azide and valinomycin and predict mitoxantrone as a topoisomerase II poison.

A more recent approach which exploits relationships shared between drugs within a network context is presented in (Napolitano et al. 2016). Namely, the method extends the GSEA framework to define enrichment scores for pathways across sets of drugs. It eventually produces ranked lists of drugs highlighting the potential for dysregulation induced in specific pathways by specific sets of drugs. The method, termed drug-set enrichment analysis (DSEA) incorporates pathway information from various related databases to essentially produce a pathway-based connectivity map. This enabled the authors to formulate hypotheses on the MoAs shared by drugs. Thus, DSEA was utilized to identify shared pathways by sets of drugs in five distinct pharmacological classes with known MoA and results were validated by means of gold standard sets of target genes for each class retrieved from molecular databases. Additionally, the method was able to infer a putative MoA for a set of drugs with mild corrective activity in cystic fibrosis, a disorder for which no therapeutic treatment is currently available. The approach has the potential for aiding in the characterization of novel drugs with unknown MoA by simply incorporating related transcriptional profiles into the pipeline.

2.3 Identification of New Drug Targets and Polypharmacology Applications

Systems pharmacology approaches for inferring compound MoA have evolved in the past decade from methods based purely on ranked lists of genes and their transcriptional response to treatment, to gradually incorporate elaborate network and pathway context, as well as various other sources of biological information. The rising interest in understanding compound MoA was accompanied by a simultaneous strive for identifying novel therapeutic targets based on network analysis methods, within the greater context of optimizing the drug discovery process. Computational methods based on network analysis can be used to model the systemic milieu in which putative therapeutic targets are located and consequently identify targets which increase therapeutic efficacy and reduce adverse effects. In order to achieve this goal, the complex relationships between the chemical and genomic factors influencing the interaction between drugs and their targets must be appropriately accounted for.

From this perspective, the concept of similarity among various biological and nonbiological entities (such as compound chemical structure, protein sequence, phenotypic profiles, etc.) is paramount. Similarity is at the base of two important hypotheses in modern drug discovery, in the sense that chemically and pharmacologically similar drugs are targeting similar target proteins (Chen et al. 2012), and that molecularly and clinically related drugs and diseases are likely to share similar phenotypes (Vogt et al. 2014). Additionally, in the context of systems pharmacology, multifaceted similarity metrics can be used to facilitate the integration of heterogeneous data. As in the case of approaches used in MoA identification, networks built for the identification of novel targets have edges representing protein–protein interactions and transcriptional regulation but may also encode drug–target or drug–drug interactions. Commonly, edges are defined based on therapeutic or chemical similarities between two nodes, similarities between proteins sharing associations with diseases, or similarities of diseases based on the shared number of genes/proteins (Zhao and Iyengar 2012). This wide range of possible definitions for network edges, and their underlying similarity metrics, enable networks to model multiple interaction scales, transcending from atomic and molecular level to the phenotypic level of drug–target interactions.

2.3.1 Target Characterization and Identification Using Network Properties of Drug Targets

Since an important part in the process of identification of novel drug targets is the understanding of how signal flow is achieved within molecular pathways, a significant share of research work in this area has been dedicated to studying network topology-based relationships and identification of target-related motifs. Additionally,

concepts such as network paths are important for establishing relations between nodes and network topologies, and formulate biologically relevant constraints in modeling drug perturbation (e.g., start nodes on a path must be receptors, intermediate nodes be specific types of intracellular proteins and end nodes must be transcriptional factors). Such methods rely on interaction networks built from protein–protein data on which drug-related data is overlaid, or on bipartite or multipartite networks used to model drug–target and drug–drug interactions (Yildirim et al. 2007; Yamanishi et al. 2008; Li et al. 2015). Early work focused on formulating network topology criteria which define existing drug targets and, based on these criteria, elaborate methods that would allow the identification of novel targets from the network (Yildirim et al. 2007; Ma’ayan et al. 2007; Yamanishi et al. 2008; Hwang et al. 2008; Nacher and Schwarz 2008; Berger and Iyengar 2009). Yildirim et al. (2007) used a bipartite network based on two projections: in the first, nodes denote drugs which have connecting edges if they share a common target, while in the second projection nodes denote protein targets which are connected if they share a common drug. The analyses of these networks revealed that drug targets tend to have a higher degree (number of connecting edges) than other nodes, and therefore are implicated in more cellular interactions. Additionally, they observed that most new drugs are associated with previously targeted network neighborhoods. Ma’ayan et al. (2007) used a bipartite network connecting drugs and drug targets, overlaid on protein–protein interaction data to show that drug target proteins are primarily located in the cellular membrane. Another important observation derived from the topology-based studies is that network centrality or node degree measures should not be the sole factors for the detection of new target proteins. Although such measures indicate essentiality of respective protein nodes, perturbation induced by drug treatment on the respective protein targets could induce significant undesired effects on the downstream cellular processes. Hwang et al. (2008) instead proposed targeting proteins which are bridging nodes with less regulatory effects on pathways (fewer interacting connections), but located in network positions where their disruptions would result in information flow prevention.

More recently, Mitsopoulos et al. (2015) identified sets of topological and community properties characterizing druggability of target protein nodes and neighborhoods and highlighted differences between cancer and non-cancer drugs. To this goal they used protein–protein interaction data enriched with drug–target information and built sets of predictors based on the network topology descriptors. Machine learning methods such as random forests, gradient boosted machines, and generalized linear models were then utilized to computationally validate their drug–target interaction predictions. In Li et al. (2015) authors define a computational framework based on the guilt by association principle and network topology features, which allows them to identify a large number of potential drug targets, among which some are associated with diseases such as the Torg-Winchester syndrome and rhabdomyosarcoma. Under the guilt by association assumption, a target protein and a drug are likely to interact if the majority of the protein’s neighbors (which share direct interactions with the target protein) in the network can interact with the drug. The authors use a predictive model based on the random forest algorithm and

feature sets consisting from node and edge weights in a bipartite network model (containing protein–protein, drug–target and drug–drug interactions).

2.3.2 Identification of Drug Targets Based on Integrative Network Approaches

The task of identifying drug targets from genome-wide data can be helped by the integration of additional data such as drug chemical structure, target protein sequences, known drug–target interactions, or information about drugs’ side effects. As in the case of MoA characterization, the incorporation of such complementary data can help in adding more biologically plausible context to the models, reduce bias induced by incomplete information and enhance the search space for the computational algorithms deriving the predictive models.

Campillos et al. (2008) proposed a method incorporating information on drugs’ side effects from drug package inserts into a drug–target network in order to define a phenotype-based similarity metric. The side effect similarity metric was combined with a 2D chemical similarity metric based on the Tanimoto coefficient into a probabilistic framework under which to infer the probability of two drugs interacting with the same target. The method was used to derive new targets for existing drugs, and the authors validated using *in vitro* assays 13 drug–target interactions predicted by their method. However, the main limitation of such an approach was that it could only be used on marketed drugs for which side effect information was available.

Based on the same experimentally validated assumption that similar drugs interact with similar target proteins, Chen et al. (2012) integrated a composed drug–drug similarity metric based on drug chemical structure similarity and targets known to be shared by pairs of drugs, a target–target similarity metric based on protein sequence similarity, and a known drug–target interaction network. The authors then implemented a random walk with restart on the resulting bipartite drug–target network to predict potential drug–target interactions. Thus, a target can be predicted even if the investigated drug has no known targets, based on similar drugs and their known targets. The random walk was implemented using transition matrices from target network to drug network and inter-transition matrices indicating the probability of walks from drug to drug (or target to target). Based on this, a probability of finding the walker at node i at step $t + 1$ based on the position at node j at step t can be determined iteratively. The approach was used to predict drug target interactions for four classes of datasets (enzymes, ion channels, G protein coupled receptors and nuclear receptors). Results were validated using gold standard datasets from public databases.

Cheng et al. (2012) combined three supervised inference models to predict drug–target interactions. Namely, the network-based inference (NBI) relying on drug–target bipartite network topological similarity was used in conjunction with a

drug based similarity inference (DBSI), which relies on 2D chemical similarity between drugs and drug–target interaction information, and a target-based similarity inference (TBSI), relying on target sequence similarity and drug–target interaction information, to predict associations between drug–target pairs. DBSI and TBSI incorporated information from the chemical and genomics space, respectively, while NBI was based solely on network topology features. The authors highlighted the performance of NBI inference, superior to the other inference methods. The predicted targets were validated using *in vitro* binding assays. The approach indicated polypharmacological effects on five drugs (montelukast, diclofenac, simvastatin, ketoconazole, and itraconazole) and suggested repositioning potential of these drugs, which was further validated experimentally.

An interesting approach has been recently proposed by Isik et al. (2015), which investigated the transcriptome perturbations in conjunction with functional interaction network information to reveal effects induced by drugs binding to their targets. They derive a new measure for target prioritization, termed local radiality, which is able to identify more diverse targets, with fewer neighbors, and consequently, possibly fewer side effects. They validate the results based on ROC analysis using test datasets from other approaches.

A large number of other network-based and machine learning-based methods have been developed recently, most of them following broadly the same paradigm, as shown in Fig. 2.2: enrich existing networks of known drug–target interactions with information from chemical and/or genomics spaces and learn various

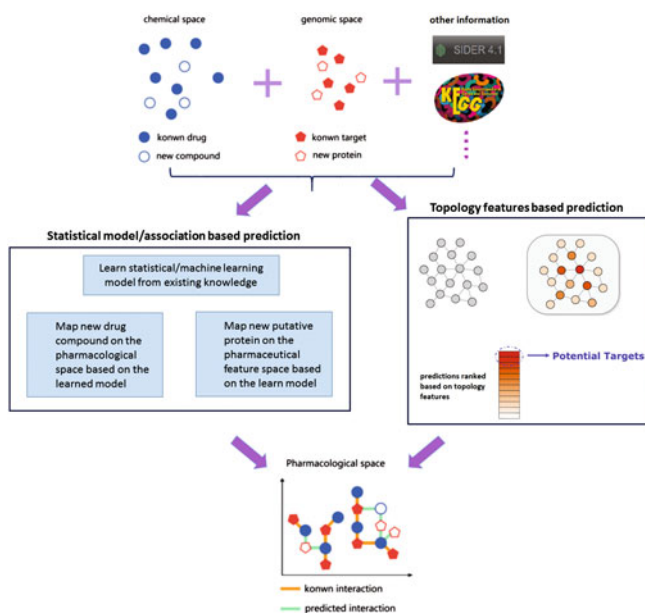


Fig. 2.2 Schematic of target identification approaches in systems pharmacology

supervised or semi-supervised models in order to predict novel interactions. For example Yamanishi et al. (2008) used a kernel regression method to learn chemical and genomic space models and demonstrate correlation with pharmacological space. Yuan et al. (2016) used a similarity approach based on ensemble learning methods to incorporate chemical and genomic space similarity as components into ensembles of learning to rank, while Yamanishi et al. (2014) created a web based engine (DINIES) using supervised learning and relying on similarity matrix kernels (learned from drug, side effects and protein domains) to predict interactions on test sets. Another recently developed web-based tool is TarPred (Liu et al. 2015), which besides predicting targets is also able to provide disease indications and predict side effects.

2.3.3 *Network-Based Polypharmacology*

It is often the case that methods developed for predicting new drug target interactions naturally yield combinations of potential targets (often protein complexes or whole sub-pathways), which naturally classifies them as polypharmacology approaches. Polypharmacology accounts for the important and increasingly accepted concepts that (i) complex diseases tend to be associated with multiple target proteins, and (ii) drugs commonly work by targeting several off-targets, besides the primary target (Xie et al. 2012). Accounting for the polypharmacology properties of drugs has the potential of increasing drug efficacy and overcoming drug resistance and toxicity, thus, the approaches capable of developing multiple target drugs, as well as research in drug combination based on network approaches have received increased attention recently.

An example is the method developed by Yang et al. (2008), which developed a computational framework for inferring multiple targets and suggest optimal combinations of target intervention. Their method, named multiple target optimal intervention (MTOI), searches systematically for effective points of intervention in a disease-based network to restore it to a desired normal state. MTOI relies on a procedure for perturbing the disease network and optimize it toward the desired state based on a Monte Carlo simulated annealing optimization algorithm (MCSA). The disease network is defined as a collection of concentrations of proteins and/or metabolites, or other relevant temporal-based information. Such a network is usually obtained from experimental data on patients or cells in abnormal/disease condition. The desired network is defined as the physiological steady state network. The information and related perturbations are modeled using differential equations and MCSA. The authors applied it to an inflammation based network, the arachidonic acid metabolic network, and derived a combinatorial intervention based on anti-inflammatory drugs.

Other network based polypharmacology studies include that of Cheng et al. (2012b), which extended their previous method (Cheng et al. 2012a) and proposed two different weighted network-based inference methods using four similarity

metrics for predicting multiple chemical–protein interaction. Under this framework they investigated the polypharmacology of five approved drugs: imatinib, dasatinib, sertindole, olanzapine, and ziprasidone. Zhao et al. (2013) used a composite network built from protein–protein interactions and gene regulatory databases onto which Gene Ontology and side effect information was overlaid. Drug–drug pairs, for which the addition of a pair member was reported to result in reduced side effects of the other drug, were exhaustively searched for. Random walk was then used to determine interaction subnetworks between drug pairs, in order to identify nodes that would be preferentially affected by specific interactome perturbations. Following this approach the authors were able to predict drugs which combined with rosiglitazone (an efficacious antidiabetic drug associated with increased myocardial infarction), would mitigate its myocardial infarction risk. Additionally, they predicted that the mitigating effect of exenatide in conjunction with rosiglitazone could occur through clotting regulation. Additional polypharmacology-related approaches are presented extensively in review studies (Reddy and Zhang 2013; Medina-Franco et al. 2013).

It must be noted at this point that usually there are significant overlapping areas between approaches attempting drug–target interaction prediction, polypharmacology-related methods and methods having as goal repositioning strategies for existing drugs. It is often the case that, due to the limited available resources on drugs, target-identification methods are restricted to predict alternate targets for drugs with already known targets, which is essentially a drug repositioning approach. This is the case with methods developed in Campillos et al. (2008) and Cheng et al. (2012a) described above. The same stands for studies investigating drugs MoA, which commonly have as byproduct multiple genes/proteins, often representing entire sub-pathways identified as target of a specific drug (Iskar et al. 2013; Chen et al. 2011), which could be seen as polypharmacology studies. In turn, the search for polypharmacological features naturally leads to new uses for combinations of known drugs, thus providing support in drug repositioning (Chen et al. 2015).

2.4 Network-Based Drug Repositioning

Drug repositioning research has gained significant momentum in recent years due to the pressing needs to reduce costs of developed drugs while increasing efficacy, but also due to large-scale funding programs launched by governmental organizations, such as the National Center for Advancing Translational Sciences and FDA in US, and the Medical Research Council in UK (Li et al. 2016). Drug repositioning is inherently linked to a better understanding of the molecular context underlying specific phenotypes and of the mechanisms of action of drugs, which are additional reasons for drug repositioning approaches to be flourishing with the advent of systems pharmacology.

A ubiquitous feature in network-based drug repositioning is the presence of a disease-related component, since finding associations between drugs and protein targets in a disease context, is the *modus operandi* in such studies (Wu et al. 2013). Therefore, three level drug–target–disease networks are a common occurrence in modern drug repositioning research. Effective network-based approaches typically aim at accurately modeling the cause-effect paradigm which is dominating the current view on disease etiology and drug mechanisms of action: disease originates from abnormalities of one or more (usually genetic) factors and the observed phenotypes are the effect of disease development. Similarly, drug action originates from the drug–target binding and the terminal effect of the drug intake are the drug indications and side effects, which can be seen as drug phenotypes. Along these cause-effect paths, molecular activities induced by drugs and diseases may be observed using high throughput transcriptional and proteomic data, which can be viewed as snapshots of the disease development stages or of drug activity, and consequently be used to model drug–disease associations (Li et al. 2016). From this perspective drug repositioning studies can be categorized as being molecular profile based or phenotype profile based.

2.4.1 Drug Repositioning Based on Molecular Profiles

Generally, drug repositioning approaches based on molecular profiles of drugs and/or diseases rely on the so-called reversed signature hypothesis: if the molecular profile of a drug is opposite to that of the disease, then the drug has the potential to be used in treating that disease (Wu et al. 2013; Li et al. 2016). Work in this area typically follows the now standard procedure in systems pharmacology: first construct a background interaction network from protein–protein interaction databases, pathway databases, protein–DNA interaction databases, and/or other interaction resource available. Then contextualize the initial network, for example by adding weights to the edges leveraging gene expression data from sources such as CMap, LINCS, or GEO, or enriched with data from various other sources (GO, KEGG, etc.). Subsequently, various computational models and algorithms can be used to extract parts of the contextualized networks (response subnetworks or sub-pathways) which maximize the biological relevance related to disease–drug associations (Fig. 2.3).

Following such an approach, Jin et al. (2012) created their interaction network from signaling pathways in PID and BioCarta databases, onto which they overlaid transcriptional data from CMap and subsequently searched for network motifs (sub-pathways) involved in response to cancer drug treatment. These sub-pathways are connecting the disease genes (retrieved from OMIM) to known signaling proteins. They used Bayesian factor regression to uncover such driver sub-pathways bridging drug targets to the disease response signatures. The driver sub-pathways and the drug’s effects on them were found simultaneously. The effect of drugs on each sub-pathway was quantified and summarized into drug–disease signature

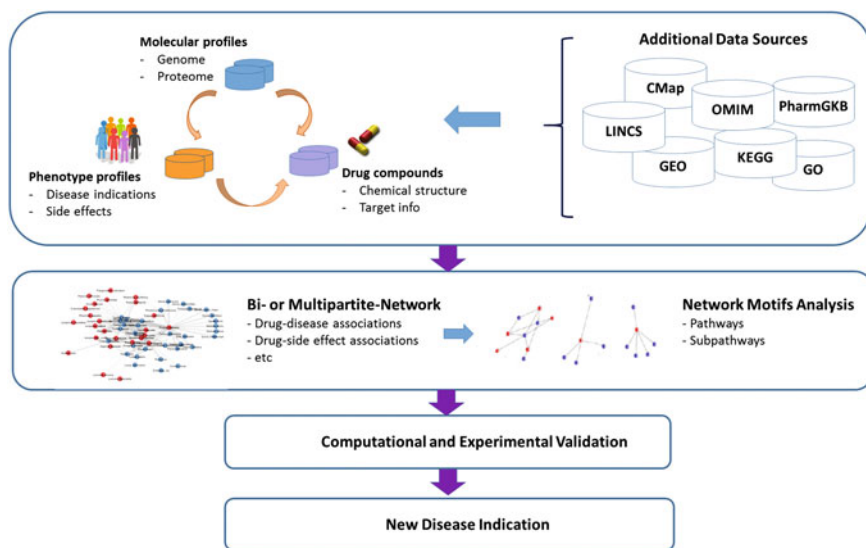


Fig. 2.3 Typical drug repositioning workflow

profiles. Then, ranked repositioning profiles for each of the drugs were created and repositioning potential derived accordingly, using support vector regression. Several high ranking drugs from their analysis were suggested for repositioning in cancer therapy based on the ability to enforce retinoblastoma-dependent repression of important E2F-dependent cell-cycle genes (Jin et al. 2012). Additionally, their method was able to accurately predict responses to more than 90% of the FDA approved drugs and 75% of experimental drugs.

In another study, Gottlieb et al. (2011) utilized multiple heterogeneous sources of evidence which were integrated into a protein–protein interaction network: drug targets, drug side effects, protein sequence and GO annotations, expression profiles and disease phenotype data. They defined several profile-based similarity measures for drugs and diseases: chemical structure based, protein and genetic sequence based, phenotype based, side effect based, network topology based and GO annotation based. The similarities measures were subsequently combined into association scores and used as features for a logistic regression classifier to identify novel drug indications.

Lee et al. (2012) constructed a tripartite drug–protein–disease network based on a large integrative database incorporating drug targets, disease-associated proteins, protein interaction, and pathway data. To explore drug–disease associations within the network they used an in-house algorithm called shared neighborhood scoring. This algorithm allowed them to predict drug–disease pairs based on the guilt by association principle that unlinked pairs which share significant numbers of neighbors with strong relationships between them could be confidently linked. They

used this approach to suggest as repositioning candidate for lung cancer treatment the high blood pressure drug benzthiazide.

Zhao and Li (2012) also used a drug–protein–disease network and developed a Bayesian partition method to retrieve drug–protein–disease modules which were closely connected. The authors started from a comprehensive protein–protein interaction network assembled by integrating data from several databases. Subsequently, information from disease–genes relations from OMIM and drug–target interactions from DrugBank were mapped onto the protein–protein interaction network. Then, gene–drug paths were computed to reflect the network distance between a gene and each drug’s targets. Similarly, gene–disease closeness was estimated to reflect the network distance between a gene and each disease-related genes. Based on these network distances drug–gene–disease modules were identified using a Bayesian partition method. The approach was used to infer drug–disease associations, and suggest new drug applications for anti-asthma drug pranlukast (repositioned for treating cancer metastasis) and cardiovascular stress-testing agent arbutamine (repositioned for treatment of obesity).

Based on the same strategy of searching for closely connected modules (whose members are more likely to be functionally related) within drug–protein–disease networks, Daminelli et al. (2012) implemented a method that searches the network for bi-cliques motifs. In their case bi-cliques are subnetworks in which every drug is linked to every target and disease. They initially built large bipartite networks from various public databases in which drugs, targets, and diseases are linked by drug–target associations and drug–disease associations. Subsequently, network analysis based on power graphs was employed to search for incomplete cliques in the network. Bi-cliques connected by common drugs are thus identified from the bipartite network. Consequently, resulting incomplete bi-cliques’ completion is used to predict novel links from drugs to targets and diseases, respectively, thus allowing the authors to simultaneously suggest reposition for drugs and predict a drug’s off-targets. The approach allowed the authors to suggest and computationally validate repositioning for nine cardiovascular drugs for treating parasitic diseases.

Other approaches on drug repositioning based on molecular profiles are those of Iorio et al. (2010) who, as presented in Sect. 2.2, built a drug–drug network in which drug nodes were linked based on similarity measures derived from ranked gene lists. Their work, developed primarily for MoA discovery, suggested that fasudil, a vasodilator used in stroke, would be effective in treatment of autophagy, which is a major process in cancer. Another work based on molecular profiles and which links MoA to drug repositioning is that of Iskar et al. (2013), also described in Sect. 2.2 above, which identified conserved drug-induced modules from transcriptional profile data and enriched the modules with information from various other databases. Module membership was then used to induce novel indications for existing drugs, predictions which were further validated experimentally. Vasodilator vinburnine, topical antifungal sulconazole, and cardiac stimulant mephentermine were all suggested as candidates as cell-cycle inhibitors in anti-cancer therapy.

In a very recent work Guney et al. (2016), propose an innovative approach which transcends the drug repositioning area, having possible applications also in drug MoA elucidation as well as drug-target identification areas. The authors introduce the concept of drug–disease proximity based on shortest paths between target and disease associated genes within the interactome. They argue that proximity to disease small neighborhoods is a good proxy for describing therapeutic effect and improve the accuracy of drug repositioning predictions. Following this approach they explain why HIV drug plerixafor is repurposed for non-Hodgkin’ lymphoma and provide potential repositioning candidates for rare diseases.

A recent trend is the use of noncoding RNAs, such as miRNA, as therapeutic agents due to their regulation of cellular processes implicated in disease. As a consequence, drug repositioning strategies considering miRNAs are also attracting significant interest. Liu et al. (2014), devised an approach for identifying repositioning candidates for cystic fibrosis based on miRNA-transcription factors feed forward loops. The loops are essentially motifs in a regulatory network with connectivity patterns occurring more frequently than in control networks, and therefore could be seen as response subnetworks. Using GEO expression data, gene-miRNA relationship data, protein interaction, and drug-miRNA interaction data as well as disease-related gene data from public databases, they built regulatory networks which were searched for feed forward loops implicated in cystic fibrosis. They found 48 drugs showing ability to perturb the expression of miRNAs which are part of loops implicated in cystic fibrosis, and which were suggested for repositioning. Similarly, Jiang et al. (2012) have developed a method that searches for modules in a drug-miRNA human cancer network built from CMap data, miRNA target gene databases and enriched with GO annotations. Using hypergeometric tests on the retrieved modules they suggested that 2-deoxy-D-glucose (2DOG) is a candidate for treating thyroid cancers.

2.4.2 Drug Repositioning Based on Phenotypic Profiles

Drug repositioning approaches based on phenotypic profiles typically rely on the principle that, if a drug shares similar side effect profile with a set of drugs prescribed to treat a specific disease, then the respective drug can be considered as a candidate for treating that disease (Wu et al. 2013). Since drug side effects are usually generated when drugs bind to off-targets (known or unknown), and hence perturb metabolic or signaling pathways, it is expected that the side effect profile of drugs may reveal relevant unknown information pertaining their MoA, and hence assist in repositioning.

One of the first works following this principle was that of Campillos et al. (2008) which, as already described in Sect. 2.3.2 used a side effect similarity profile incorporated into a drug–target network to infer probability of two drugs sharing the same target. Based on this, authors identified phenotypic associations between

nootropic drug donepezil and antidepressant venlafaxine and suggested a new market use for donepezil in treating depression.

In another work, Yang and Agarwal (2011) used a disease–side effect network by combining drug–disease associations with drug–side effect associations from PharmGKB and SIDER databases, respectively. Subsequently, they used Naïve Bayes predictors trained on relations between side effect and disease for predicting new indications for drugs. Following this approach they predicted that drugs associated with increased immune response, such as ticlopidine and ACE inhibitors are potential candidates for treating stroke. Ye et al. (2014) used a side effect-based similarity measure to connect drugs into a drug–drug network and searched for subnetwork neighborhoods enriched with drugs having a specific therapeutic indication. They used the guilt by association approach to assign a new indication to drugs present in the same subnetwork. They suggested a number of candidate drugs for repositioning, among which the analgesic drug tramadol and Parkinson’s drug tolcapone in treating depression.

One of the problems related to guilt by association approaches is that they often enforce restrictions on the search space by only considering most similar drug, discarding possible useful information embedded in the whole dataset. Bisgin et al. (2014) used the assumption that all phenotypes in the phenome (both drug indications and side effect) are interconnected with a probabilistic distribution and used a probabilistic generative model for their analysis. They used a Bayesian based model, the Latent Dirichlet Allocation (LDA) to uncover links between drugs and phenotypes, which are actually novel indications. Links are encoded into conditional probabilities. Although their method does not explicitly make use of biological networks, the LDA model they chose can be represented as a tripartite network constructing paths from drugs to phenotypes via connections across latent variables. They suggested new treatment options for all 908 drugs in their study, among which some were confirmed by literature validation, e.g., influenza A drug amantadine’s use for treating epilepsy.

Finally, we must note the development in the recent years of several web servers and open-source packages for the specific goal of drug repositioning, which integrate resources covering both molecular profile and phenotype-based approaches. With some variations, they all rely on the integration of heterogeneous data sources to build the interactome network, and incorporate some of the previously published similarity measures. Among the most popular are the PROMISCUOUS (Van Eichborn et al. 2011), DRAR-CPI (Luo et al. 2011), DMAP (Huang et al. 2015) and ksRepo (Brown et al. 2016). PROMISCUOUS integrates relations between drugs, targets, and side effects and uses drug structural similarity and side effect similarity measures. It allows users to search by single drug ID queries or perform network-based exploration given a set of drugs and targets. DRAR-CPI only uses chemical structure in the chemical-protein interactome to predict network based drug–drug associations and produce lists of drugs which share similar interaction profiles and side effect information with the query drug. DMAP combines both chemical-protein interactome, protein–protein interactions, transcriptional profiles, and phenotype data (disease indications) to build a directional weighted interactome

network. They use already published gene similarity (Iorio et al. 2010) and drug similarity measures to derive a guilt by association model based on the Kolmogorov–Smirnov enrichment (Lamb et al. 2006) to predict novel indications for drugs. ksRepo is a recent open-source software package implemented in R which proposes a generalized methodology enabling integration of transcriptional profiles from various platforms (including RNA-seq). Their method is based on disease transcriptional profiles and gene–drug interactions (available from any user desired source). They implement a variant of the Kolmogorov–Smirnov enrichment to compare single instances (disease transcriptional profile) with multiple drug–gene interaction lists and then derive scores which reflect disease–drug associations based on the transcriptional profiles.

2.5 Network-Based Side Effect Modeling and Prediction

Drug side effects are among the most important factors to be considered in drug design. Recent studies estimated side effects to be the major reason for drug discontinuation in first phase clinical trials and second most common cause of drug attrition overall (Hornberg et al. 2014). Therefore, computational approaches for *in silico* prediction of side effects are highly relevant, and currently under consideration by the pharmaceutical industry in their effort to complement the high throughput *in vitro* screening of newly developed drugs (Bowes et al. 2012).

Side effects are the result of promiscuous binding behavior of the majority of drugs, which in addition to their primary targets can interact with different affinities with many off-targets (Paolini et al. 2006). This way they potentially perturb many signaling and metabolic pathways eliciting both therapeutic effects and unwanted physiological responses. These signaling and metabolic pathways are often partially overlapping, thus producing synergistic or canceling consequences. Currently, there are several important observations and hypotheses which guide research in this area: different drugs can share similar side effect profiles as a result of sharing similar toxicological pathways or networks, which is an extension of the observation that the result of drug on-target and off-target binding behavior is a perturbation that is relayed downstream to partially overlapping (cross-talking) pathways (Bai and Abernethy 2013). This is related also to the principle which states that if a drug shares similar side effect profile with a set of drugs prescribed to treat a specific disease, then the respective drug can be considered as a candidate for treating that disease (Wu et al. 2013). The recent observation that network neighborhood of drug targets is a major determinant of side effect similarity profiles of drugs comes as a corollary to the previously enounced principles (Browers et al. 2011). Consequently, the development of *in silico* methods for side effect prediction is significantly benefiting from the increased interest in the area of drug–target prediction.

The computational approaches based on network analysis aiming at predicting drug side effects and modeling their generation mechanisms can be broadly

categorized as being chemical-based and pathway-based. Both types of approaches heavily use the two important concepts in network modeling: that of network neighborhood (which define areas of the network with inter-related and coherent functional properties), and that of similarity (which is defined on various chemical, genomic or ontology features to reflect proximity between network nodes or neighborhoods).

2.5.1 Approaches Based on Chemical Structure

Chemical-based approaches generally attempt to relate chemical structure of drugs to their side effects, based on the basic observation that similar ligands interact with similar proteins. Thus, based on the backbone consisting of drug chemical structure, protein structure and information on drug–target interactions and incomplete drug–side effect association, models can be built to predict novel drug–side effect associations. Some examples include the work of Schreiber et al. (2009) which developed a method integrating various sources on chemical substructures and information on side effects to find large-scale structure–side effect associations. In their network they linked side effects based on correlations between drug chemical features. Their aim was not a mechanistic understanding of side effect causes but rather drawing a global picture of how different types of side effects may be linked, with the goal of defining possible filters for screening drug compound candidates. Similarly, Pauwels et al. (2011) used sparse canonical correlation analysis (SCCA) to predict side effects and associate them with correlated ensembles formed by chemical substructures. Yamanishi et al. (2010) proposed a unified framework, based on the integration of chemical, genomic, and pharmacological data (and the related similarity measures) with the topology of drug–target interaction networks. Within the framework of supervised bipartite network inference, using a regression approach, they were able to predict the side effect profiles of candidate drug compounds, as well as interpret drug–target interactions. In a subsequent study, they suggested several extensions to the kernel regression model for multiple responses in order to optimally integrate the heterogeneous data sources (Yamanishi et al. 2012). Based on this approach they were able to predict rare side effects for molecules in DrugBank with no available information in SIDER, such as ovarian cyst, breast tenderness, and melisma for synthetic progestational hormone drug levonorgestrel, which were further validated based on literature.

Mizutani et al. (2012) used the co-occurrence of drugs in protein-binding profiles and side effect profiles to extract correlated sets of drug targets and side effects, using SCCA. They used a drug–target interaction network and enrichment analysis, using KEGG and GO data, to show that the retrieved correlated sets were significantly enriched in the same biological pathways, despite having different molecular functions. A biologically relevant interpretation of their results suggests that extracted side effects can be seen as possible phenotypic outcomes of drugs targeting proteins that appear in the same correlated set (i.e., having similar

structures), thus reinforcing the principle mentioned in the previous paragraphs which states that target neighborhood is a predetermining factor for side effects similarity. Their side effect predictions include tremor, constipation, and dry mouth for antihistaminic drug cinnarizine, all of which were confirmed by literature of FDA reports.

Atias and Sharan (2011) combined the SCCA with a diffusion model based on side effect similarity networks. Their approach uses SCCA to project correlated structure-side effect data into a lower dimensional space. This projection is then used to predict side effects. Subsequently, using a query drug and a diffusion model on side effect similarity networks they obtain ranked list of side effects. Their validation scheme was based on a large-scale blind test based on 448 drugs from the Hazardous Substances Data Bank. The approach was able to predict correct side effects in the top 5 ranked predictions for >56% of the drugs in the database.

Lounkine et al. (2012) first used a chemical structure similarity metric, named the similarity ensemble approach (SEA), to predict targets among a set of proteins and subsequently develop a guilt by association metric that links the new targets to the side effects of the related drugs, virtually creating a drug–target–side effect network. For predicting drug target–side effect association they used an enrichment score based on co-occurrence of pairs that were more common than expected by chance, coupled with a statistically significant threshold. Based on this approach, authors predicted epigastralgia as side effect associated with chlorotrianisene, a synthetic non-steroidal estrogen. Interestingly, the off-target protein for this drug, predicted by authors, COX1, bears no sequence or structural similarity with the drug’s primary target (the estrogen nuclear hormone receptor) but cross-activity between the targets is suggested by ligand similarity.

In a recent study, Wang et al. (2016) depart from the target-based approach that currently dominates the drug side effect prediction field. Their approach aims at avoiding the bias induced in the analysis by the incomplete knowledge on drug targets by combining chemical structure information with transcriptional profiles from LINCS database. They use feature sets created from signature transcriptional profiles for each drug instance, cell morphological profiles, drug chemical structure, and enrichment analysis to train a machine learning classifier based on extra trees. The most predictive classifiers are then used to shed light on the mechanisms of side effects.

Interesting insights into the factors contributing to drug side effect resulted from the approach presented in Wang et al. (2013), where authors use a structurally resolved interaction network to systematically examine relationships between drug associated side effects and drug targets. They use a generalized linear regression model and show that it is the number of essential targets (proteins which are critical for cellular survival), and not the total number of targets, that determines the side effects of drugs. Additionally, they highlight several key network topology characteristics of drug targets that are highly correlated with increased side effects profiles. They noted that high node degree (number of interactions for a target) and betweenness (the number of shortest paths between other proteins in the network

passing through the target protein), as well as highly shared interaction profiles are more likely to result in an increase in the number of side effects.

2.5.2 Approaches Based on Pathways/Sub-pathways

Pathway-based approaches relate drug side effects to perturbed biological pathways or sub-pathways which contain drug target proteins. Consequently, they train models on the molecular interaction networks built from various data sources (such as drug–target interactions, gene/protein–disease–drug–side effect connections, or drug–drug interactions) in order to predict side effects for unknown drug–side effect associations based on underlying network motifs. The models thus derived are able to provide mechanistic insights into the side effect generation process.

Lee et al. (2011) used an enrichment score to define drug-biological process associations based on CMap transcriptional profiles and GO ontologies and subsequently built multilevel biological process–drug–side effect network to discover relationships between biological processes and side effects, using drug information as a bridge. For this purpose they employed a co-occurrence-based scoring accounting for how many drugs shared the same side effect in a specific biological process. Bauer-Mehren et al. (2012) use a two-step framework for biological annotation of side effects with relevant pathways. They search for drug–target and target–side effect associations and then compare these associations to derive drug–side effect links. In a subsequent step they substantiate the found associations using pathway information from Reactome database.

Li et al. (2012) used a bipartite drug-metabolic sub-pathway network build after identifying sets of drug-induced differentially expressed genes from CMap and pathway enrichment analysis. By analyzing drug–sub-pathway associations they uncovered that drugs share similar indications and side effect if they are associated to same sub-pathways. Additionally, an increase in the number of sub-pathways shared by drugs correlates with increased numbers of common side effects. Overall, their study confirms the idea that important therapeutic and side effect related mechanisms are relayed through sub-pathways, which are smaller regions of pathways, and may be overlooked by whole pathway-based methods. In a related study highlighting the importance of subnetwork-based approaches, Zhao et al. (2013), proposed an approach for identifying drug combinations to mitigate side effects. To this goal they used a human interactome network built from protein–protein interaction databases and then searched for subnetworks enriched with sets of related GO biological processes annotations. Interactions between drug pairs based on their targets were searched using a random walk method and correlated with information on their side effects. As mentioned in Sect. 2.3.3, following this approach they were able to predict the mitigating effect of exenatide on rosiglitazone’s myocardial infarction side effect and explain that this could occur through a clotting regulation mechanism.

Another subnetwork-based approach was followed by Lorberbaum et al. (2015) which also used an initial interactome network created from protein–protein interactions. Their initial network was pruned based on data from several sources and biological levels, such as to highlight subnetwork modules with mechanistic connections to phenotypes. Their subnetworks were enriched in putative side effect mechanistic pathways and, subsequently, drugs were assigned to subnetworks where their targets were present. Then, subnetworks were used as features in a random forest-based classifier trained to predict whether a given drug will cause side effects.

A number of other approaches combine pathway-based analysis with information related to chemical structure of drugs and their target proteins for a holistic view on mechanisms generating side effects. Examples include the works in (Wallach et al. 2010; Fan et al. 2012), which use pathway information and in silico virtual docking to identify off-targets of drugs and link them to biological pathways. In (Liu et al. 2012), authors integrate information on drug chemical structure with pathway information and phenotypic characteristics of drugs including indications and side effect. They used a machine learning-based approach to build and evaluate the side effect prediction model. Similarly, Kuang et al. (2014) used a number of structural features of drugs integrated with network topology features of the drug–side effect association networks (constructed using correlation based methods) to build classifiers able to predict side effects.

Recently, Cao et al. (2015) integrated multiple data sources such as chemical structure, sequence, transcriptional profiles, ontology and pathways and defined multiple similarity measures based on these data types. Additionally, network topology-based similarity measures were defined, including nearest neighbor and path-based measures, using a drug–side effect network. Classification features were constructed from these similarity measures based on collaborative filtering, and a multiple evidence fusion algorithm was used for creating a multiscale predictor for side effects.

As in the case of the other application areas of systems pharmacology, a number of web servers were created for enabling the prediction of drug’s side effects. The most popular among these are: IntSide (Juan-Blanco et al. 2015), which is a hybrid approach incorporating both structural and pathway information to provide mechanistic insights into drug–side effect associations. Dr. Prodis is a structure-based tool which implements several structure–pocket and structure–structure comparison procedures. Besides predicting drug side effects, it produces also drug–target interaction predictions, as well as associations between drugs and diseases (Zhou et al. 2015).

2.6 Current Challenges and Future Considerations

Despite the great promise, systems pharmacology approaches face a number of challenges while scaling from pre-clinical setting into clinical applications. A major hurdle is the bias caused by incomplete knowledge. For example, network-based

models tend to bias to the targets with more known associated drugs, and even if current studies, such as the one in Wang et al. (2016) attempt to balance their models by incorporating sources of evidence from different biological levels, such as chemical structure, lack of adequate high resolution structural data for targets may induce further problems. However, recent progress in both experimental and computational methods in the area of structural genomics holds the promise to significantly improve the structural coverage. Another limitation of almost all network-based approaches, especially those relying on searching paths across the network is that they cannot provide predictions (e.g., for drug–target pairs) when missing information hampers the identification of reachable paths in the network. Network-based methods need to adequately address these aspects in the future.

Another important issue is that of the amounts of data at multiple scales needed to build accurate predictive models in the context complex disease heterogeneity. And whether the incorporation of such specialized data will still produce models with decent generalization performance, given for example an individual with unobserved new mutation. Current approaches treat insufficiently the problem of inter-subject genetic variability, which is a crucial step toward the goal of precision medicine. Among the other challenges worth mentioning are the lack of structured gold standard, especially in the applications related to drug repositioning and side effect prediction. Ideally, *in silico* experimental results should be integrated into the drug design validation pipeline and tested in binding experiments, cellular assays or animal models for not only providing filters for initial candidate lists, but also retrieving false positives that could be further used to refine the algorithms. While in the case of drug-target prediction and MoA characterization, the results provided by predictive models can be easily tested experimentally, for drug repositioning and side effect prediction it is often the case that genomic responses in animal models vary significantly when compared to human models. Therefore, additional care must be taken for thorough training and testing of the predictive models. From this perspective, the availability of extensive secondary use data from patients electronic health records, presents researchers valuable resources for performing ‘retrospective’ experiments on human subjects in clinical settings (Lorberbaum et al. 2016).

Another aspect is that, despite the increased predictive power generated by the incorporation of multiscale heterogeneous data into the network and statistical models, there still is the question of how relevant it is to discover new knowledge from static statistical models, under conditions that are constantly changing. Under drug treatment, a disease state is not static, but evolves through successive states while responding to the drug-induced perturbations. When sufficient data are collected to successfully build a model describing one disease state, the disease may already be in a different state from the one used to build the model. In such a dynamic situation, a data-driven model is essentially retrospective and not prospective (Xie et al. 2014). Presently, very few methods that offer dynamic resolution are used in systems pharmacology. One such approach is that of (Bansal et al. 2006) where an algorithm called Time Series Network Analysis (TSNI) was proposed to infer targets of antibiotic norfloxacin based on time series transcriptional profiles experimental data. As more time course experimental data is bound

to be produced, dynamic methods such as CHRONOS (Vrahatis et al. 2016a), described in Chap. 3 could become prevalent. CHRONOS can be easily tailored to provide a framework for studying sub-pathways activated by drugs, or other therapeutic molecules, at specific drug treatment stages. Additionally, such an approach could be adapted to identify sub-pathways perturbed during disease progression and optimal time points for drug treatment could be inferred.

The evolution of network and pathway-based approaches used in systems pharmacology has followed a journey starting with methods based on overrepresentation, or enrichment analysis, which use statistical tests to find sets of genes in a particular pathway among the (usually differentially expressed) genes under study (Khatri et al. 2012). A second generation of approaches uses gene-level statistics (e.g., identifying individual differentially expressed genes) and then aggregating the gene-level statistics of all genes in a pathway into a pathway-level statistic or score, such as the Kolmogorov–Smirnov statistic (Lamb et al. 2006; Iorio et al. 2010, 2015). The third generation of methods are topology based, thus allowing the incorporation of information from various sources, beyond simple lists of genes (Pan et al. 2014; Wallach et al. 2010; Fan et al. 2012). We believe the trend in recent publications, which consider small regions of pathways (sub-pathways) with specific topological features and which are activated by perturbations cause by disease or drug treatment, is likely to follow and further expand. The development of robust sub-pathway-based approaches able to provide useful insights into time- and condition-specific activated sub-pathways (CHRONOS), and assist in identifying disease perturbed sub-pathways [DEsubs (Vrahatis et al. 2016b), described in Chap. 4] is therefore of utmost importance for future studies in systems pharmacology.

References

- Ambesi-Impiombato A, Diego d Bernardo A (2005) Computational biology and drug discovery: from single-target to network drugs. *Curr Bioinf* 1:3–13
- Anighoro A, Bajorath J, Rastelli G (2014) Polypharmacology: challenges and opportunities in drug discovery: miniperspective. *J Med Chem* 57:7874–7887
- Antman E, Weiss S, Loscalzo J (2012) Systems pharmacology, pharmacogenetics, and clinical trial design in network medicine. *Wiley Interdisc Rev Syst Biol Med* 4:367–383
- Arrell DK, Terzic A (2010) Network systems biology for drug discovery. *Clin Pharmacol Ther* 88:120–125
- Atias N, Sharan R (2011) An algorithmic framework for predicting side effects of drugs. *J Comput Biol* 18:207–218
- Bai JP, Abernethy DR (2013) Systems pharmacology to predict drug toxicity: integration across levels of biological organization. *Annu Rev Pharmacol Toxicol* 53:451–473
- Bansal M, Della Gatta G, Di Bernardo D (2006) Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics* 22:815–822
- Barabási AL, Gulbahce N, Loscalzo J (2011) Network medicine: a network-based approach to human disease. *Nat Rev Genet* 12:56–68

- Bashan A, Bartsch RP, Kantelhardt JW, Havlin S, Ivanov PC (2012) Network physiology reveals relations between network topology and physiological function. *Nat Comm* 3:702
- Bauer-Mehren A, Van Mullingen EM, Avillach P et al (2012) Automatic filtering and substantiation of drug safety signals. *PLoS Comput Biol* 8:e1002457
- Berger SI, Iyengar R (2009) Network analyses in systems pharmacology. *Bioinformatics* 25:2466–2472
- Bezerianos A, Maraziotis IA (2008) Computational models reconstruct gene regulatory networks. *Mol BioSyst* 4:993–1000
- Bisgin H, Liu Z, Fang H, Kelly R, Xu X, Tong W (2014) A phenome-guided drug repositioning through a latent variable model *BMC Bioinformatics* 15:1
- Boran AD, Iyengar R (2010) Systems approaches to polypharmacology and drug discovery. *Curr Opin Drug Discov Devel* 13:297
- Bowes J, Brown AJ, Hamon J, Jarolimek W, Sridhar A, Waldron G, Whitebread S (2012) Reducing safety-related drug attrition: the use of in vitro pharmacological profiling. *Nat Rev Drug Discov* 11:909–922
- Brouwers L, Iskar M, Zeller G, Van Noort V, Bork P (2011) Network neighbors of drug targets contribute to drug side-effect similarity. *PLoS ONE* 6:e22187
- Brown AS, Kong SW, Kohane IS, Patel CJ (2016) ksRepo: a generalized platform for computational drug repositioning. *BMC Bioinf* 17:1
- Campillos M, Kuhn M, Gavin AC, Jensen LJ, Bork P (2008) Drug target identification using side-effect similarity. *Science* 321:263–266
- Cao DS, Xiao N, Li YJ, Zeng WB et al (2015) Integrating multiple evidence sources to predict adverse drug reactions based on a systems pharmacology model. *CPT: Pharmacometr Syst Pharmacol* 4:498–506
- Chen X, Xu J, Huang B, Li J et al (2011) A sub-pathway-based approach for identifying drug response principal network. *Bioinformatics* 27:649–654
- Chen X, Liu MX, Yan GY (2012) Drug–target interaction prediction by random walk on the heterogeneous network. *Mol BioSyst* 8:1970–1978
- Chen X, Yan CC, Zhang X, Zhang X, Dai F, Yin J, Zhang Y (2015) Drug–target interaction prediction: databases, web servers and computational models. *Briefings Bioinf* bbv066
- Cheng F, Liu C, Jiang J, Lu W et al (2012a) Prediction of drug–target interactions and drug repositioning via network-based inference. *PLoS Comput Biol* 8:e1002503
- Cheng F, Zhou Y, Li W, Liu G, Tang Y (2012b) Prediction of chemical–protein interactions network with weighted network-based inference method. *PLoS ONE* 7:e41064
- Csermely P, Korsmaros T, Kiss HJ, London G, Nussinov R (2013) Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol Ther* 138:333–408
- Daminelli S, Haupt VJ, Reimann M, Schroeder M (2012) Drug repositioning through incomplete bi-cliques in an integrated drug–target–disease network. *Integr Biol* 4:778–788
- DiMasi JA, Grabowski HG, Hansen RW (2016) Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econ* 47:20–33
- Esserink M (2016) *Science Magazine*. doi:10.1126/science.aaf4017, <http://www.sciencemag.org/news/2016/02/french-company-bungled-clinical-trial-led-death-and-illness-report-says>. Retrieved Oct 2016
- Fan S, Geng Q, Pan Z, Li X, Tie L, Pan Y, Li X (2012) Clarifying off-target effects for torcetrapib using network pharmacology and reverse docking approach. *BMC Syst Biol* 6:152
- FDA (2016) Table of pharmacogenomic biomarkers in drug labeling. <http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm>. Retrieved 5 Oct 2016
- Ganter B, Tugendreich S, Pearson CI et al (2005) Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J Biotechnol* 119:219–244
- Gardner TS, Di Bernardo D, Lorenz D, Collins JJ (2003) Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* 301:102–105

- Gene Ontology Consortium (2013) Gene ontology annotations and resources. *Nucleic Acids Res* 41(D1):D530–D535
- Glaab E, Baudot A, Krasnogor N, Valencia A (2010) Extending pathways and processes using molecular interaction networks to analyze cancer genome data. *BMC Bioinf* 11:1
- Gottlieb A, Stein GY, Ruppin E, Sharan R (2011) PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol Syst Biol* 7:496
- Graham DJ, Campen D, Hui R et al (2005) Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: nested case-control study. *Lancet* 365:475–481
- Guney E, Menche J, Vidal M, Barabasi AL (2016). Network-based in silico drug efficacy screening. *Nat Commun* 7
- Hopkins AL (2008) Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol* 4:682–690
- Hornberg JJ, Laursen M, Brenden N, Persson M, Thougard AV, Toft DB, Mow T (2014) Exploratory toxicology as an integrated part of drug discovery. Part I: why and how. *Drug Discov Today* 19:1131–1136
- Huang H, Nguyen T, Ibrahim S, Shantharam S, Yue Z, Chen JY (2015) DMAP: a connectivity map database to enable identification of novel drug repositioning candidates. *BMC Bioinf* 16 (Suppl 13):S4
- Hutchinson L, Kirk R (2011) High drug attrition rates—where are we going wrong? *Nat Rev Clin Oncol* 8:189–190
- Hwang WC, Zhang A, Ramanathan M (2008) Identification of information flow-modulating drug targets: a novel bridging paradigm for drug discovery. *Clin Pharmacol Ther* 84:563–572
- Hwang S, Kim CY, Ji SG et al (2016) Network-assisted investigation of virulence and antibiotic-resistance systems in *Pseudomonas aeruginosa*. *Sci Rep* 6
- Ideker T, Galitski T, Hood L (2001) A new approach to decoding life: systems biology. *Annu Rev Genomics Human Genet* 2:343–372
- Ideker T, Ozier O, Schwikowski B, Siegel AF (2002) Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* 18(suppl 1):S233–S240
- Iorio F, Bosotti R, Scacheri E et al (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc Natl Acad Sci* 107:14621–14626
- Iorio F, Shrestha RL, Levin N, Boilot V, Garnett MJ, Saez-Rodriguez J, Draviam VM (2015) A semi-supervised approach for refining transcriptional signatures of drug response and repositioning predictions. *PLoS ONE* 10:e0139446
- Isik Z, Baldow C, Cannistraci CV, Schroeder M (2015) Drug target prioritization by perturbed gene expression and network information. *Sci Rep* 5
- Iskar M, Zeller G, Blattmann P et al (2013) Characterization of drug-induced transcriptional modules: towards drug repositioning and functional understanding. *Mol Syst Biol* 9:662
- Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási AL (2000) The large-scale organization of metabolic networks. *Nature* 407:651–654
- Jiang W, Chen X, Liao M et al (2012) Identification of links between small molecules and miRNAs in human cancers based on transcriptional responses. *Sci Rep* 2:282
- Jin G, Fu C, Zhao H, Cui K, Chang J, Wong ST (2012) A novel method of transcriptional response analysis to facilitate drug repositioning for cancer therapy. *Cancer Res* 72:33–44
- Juan-Blanco T, Duran-Frigola M, Aloy P (2015) IntSide: a web server for the chemical and biological examination of drug side effects. *Bioinformatics* 31:612–613
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2015) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* gkv1070
- Khatri P, Sirota M, Butte AJ (2012) Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol* 8:e1002375
- Kibble M, Khan SA, Saarinen N, Iorio F, Saez-Rodriguez J, Mäkelä S, Aittokallio T (2016) Transcriptional response networks for elucidating mechanisms of action of multitargeted agents. *Drug Discov Today* 21:1063–1075

- Kuang Q, Wang M, Li R, Dong Y, Li Y, Li M (2014) A systematic investigation of computation models for predicting adverse drug reactions (ADRs). *PLoS ONE* 9:e105889
- Kuhn M, Letunic I, Jensen LJ, Bork P (2015) The SIDER database of drugs and side effects. *Nucleic Acids Res* gkv1075
- Lamb J, Crawford ED, Peck D et al (2006) The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313:1929–1935
- Lázár V, Nagy I, Spohn R et al (2014) Genome-wide analysis captures the determinants of the antibiotic cross-resistance interaction network. *Nat Commun* 5
- Lee S, Lee KH, Song M, Lee D (2011) Building the process-drug–side effect network to discover the relationship between biological processes and side effects. *BMC Bioinf* 12:1
- Lee HS, Bae T, Lee JH et al (2012) Rational drug repositioning guided by an integrated pharmacological network of protein, disease and drug. *BMC Syst Biol* 6:80
- Li C, Shang D, Wang Y, Li J et al (2012) Characterizing the network of drugs and their affected metabolic subpathways. *PLoS ONE* 7:e47326
- Li ZC, Huang MH, Zhong WQ, Liu ZQ, Xie Y, Dai Z, Zou XY (2015) Identification of drug-target interaction from interactome network with “guilt-by-association” principle and topology features. *Bioinf* btv695
- Li J, Zheng S, Chen B, Butte AJ, Swamidass SJ, Lu Z (2016) A survey of current trends in computational drug repositioning. *Briefings Bioinf* 17:2–12
- Liu M, Wu Y, Chen Y et al (2012) Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. *J Am Med Inf Assoc* 19:e28–e35
- Liu Z, Borlak J, Tong W (2014) Deciphering miRNA transcription factor feed-forward loops to identify drug repurposing candidates for cystic fibrosis. *Genome Med* 6:1
- Liu X, Gao Y, Peng J et al (2015) TarPred: a web application for predicting therapeutic and side effect targets of chemical compounds. *Bioinformatics* btv099
- Lorberbaum T, Nasir M, Keiser MJ, Vilar S, Hripcsak G, Tatonetti NP (2015) Systems pharmacology augments drug safety surveillance. *Clin Pharmacol Ther* 97:151–158
- Lorberbaum T, Sampson KJ, Woosley RL, Kass RS, Tatonetti NP (2016) An integrative data science pipeline to identify novel drug interactions that prolong the QT interval. *Drug Saf* 39:433–441
- Lounkine E, Keiser MJ, Whitebread S et al (2012) Large-scale prediction and testing of drug activity on side-effect targets. *Nature* 486:361–367
- Luo H, Chen J, Shi L et al (2011) DRAR-CPI: a server for identifying drug repositioning potential and adverse drug reactions via the chemical–protein interactome. *Nucleic Acids Res* gkr299
- Ma’ayan A, Jenkins SL, Goldfarb J, Iyengar R (2007) Network analysis of FDA approved drugs and their targets. *Mount Sinai J Med J Transl Personalized Med* 74:27–32
- Maraziotis I, Dragomir A, Bezerianos A (2006) Gene networks inference from expression data using a recurrent neuro-fuzzy approach. In *IEEE engineering in medicine and biology 27th annual conference 2006*, 17 Jan IEEE, pp 4834–4837
- Maraziotis IA, Dragomir A, Bezerianos A (2007) Gene networks reconstruction and time-series prediction from microarray data using recurrent neural fuzzy networks. *IET Syst Biol* 1:41–50
- Medina-Franco JL, Giulianotti MA, Welmaker GS, Houghten RA (2013) Shifting from the single to the multitarget paradigm in drug discovery. *Drug Discov Today* 18:495–501
- Mitsopoulos C, Schierz AC, Workman P, Al-Lazikani B (2015) Distinctive behaviors of druggable proteins in cellular networks. *PLoS Comput Biol* 11:e1004597
- Mizutani S, Pauwels E, Stoven V, Goto S, Yamanishi Y (2012) Relating drug–protein interaction network with drug side effects. *Bioinformatics* 28:i522–i528
- Nacher JC, Schwartz JM (2008) A global view of drug-therapy interactions. *BMC Pharmacol* 8:5
- Napolitano F, Sirci F, Carella D, Di Bernardo D (2016) Drug-set enrichment analysis: a novel tool to investigate drug mode of action. *Bioinformatics* 32:235–241
- Nishimura D (2001) *BioCarta*. *Biotech Soft Internet Rep Comput Soft J Sci* 2:117–120
- Nissen SE, Wolski K (2007) Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *New Engl J Med* 356:2457–2471

- Pan Y, Cheng T, Wang Y, Bryant SH (2014) Pathway analysis for drug repositioning based on public database mining. *J Chem Inf Model* 54:407–418
- Paolini GV, Shapland RH, van Hoorn WP, Mason JS, Hopkins AL (2006) Global mapping of pharmacological space. *Nat Biotechnol* 24:805–815
- Pauwels E, Stoven V, Yamanishi Y (2011) Predicting drug side-effect profiles: a chemical fragment-based approach. *BMC Bioinf* 12:1
- Pritchard JR, Bruno PM, Hemann MT, Lauffenburger DA (2013) Predicting cancer drug mechanisms of action using molecular network signatures. *Mol BioSyst* 9:1604–1619
- Reddy AS, Zhang S (2013) Polypharmacology: drug discovery for the future. *Expert Rev Clin Pharmacol* 6:41–47
- Rual JF, Venkatesan K, Hao T et al (2005) Towards a proteome-scale map of the human protein–protein interaction network. *Nature* 437:1173–1178
- Scheiber J, Jenkins JL, Sukuru SCK et al (2009) Mapping adverse drug reactions in chemical space. *J Med Chem* 52:3103–3107
- Schotland P, Bojunga N, Zien A, Trame MN, Lesko LJ (2016) Improving drug safety with a systems pharmacology approach. *Eur J Pharm Sci* 94:84–92
- Subramanian A, Tamayo P, Mootha VK et al (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* 102:15545–15550
- Szklarczyk D, Franceschini A, Wyder S et al (2014) STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* gku1003
- Szklarczyk D, Santos A, von Mering C, Jensen LJ, Bork P, Kuhn M (2015) STITCH 5: augmenting protein–chemical interaction networks with tissue and affinity data. *Nucleic Acids Res* gkv1277
- Trame MN, Biliouris, K, Lesko LJ, Mettetal JT (2016) Systems pharmacology to predict drug safety in drug development. *Eur J Pharm Sci* 94:93–95
- Turner RM, Park BK, Pirmohamed M (2015) Parsing interindividual drug variability: an emerging role for systems pharmacology. *Wiley Interdisc Rev Syst Biol Med* 7:221–241
- Vogt I, Prinz J, Campillos M (2014) Molecularly and clinically related drugs and diseases are enriched in phenotypically similar drug–disease pairs. *Genome Med* 6:1
- Von Eichborn J, Murgueitio MS, Dunkel M, Koerner S, Bourne PE, Preissner R (2011) PROMISCUOUS: a database for network-based drug-repositioning. *Nucleic Acids Res* 39 (suppl 1):D1060–D1066
- Vrahatis, AG, Balomenos P, Tsakalidis AK, Bezerianos A (2016a) DEsubs: an R package for flexible identification of differentially expressed subpathways using RNA-seq experiments. *Bioinformatics* btw544
- Vrahatis AG, Dimitrakopoulou K, Balomenos P, Tsakalidis AK, Bezerianos A (2016b) CHRONOS: a time-varying method for microRNA-mediated subpathway enrichment analysis. *Bioinformatics* 32:884–892
- Wallach I, Jaitly N, Lilien R (2010) A structure-based approach for mapping adverse drug reactions to the perturbation of underlying biological pathways. *PLoS ONE* 5:e12063
- Wang X, Thijssen B, Yu H (2013) Target essentiality and centrality characterize drug side effects. *PLoS Comput Biol* 9:e1003119
- Wang Z, Clark NR, Ma’ayan A (2016) Drug induced adverse events prediction with the LINCS L1000 data. *Bioinformatics* btw168
- Waring MJ, Arrowsmith J, Leach AR et al (2015) An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat Rev Drug Discov* 14:475–486
- Woo JH, Shimoni Y, Yang WS et al (2015) Elucidating compound mechanism of action by network perturbation analysis. *Cell* 162:441–451
- Wu Z, Wang Y, Chen L (2013) Network-based drug repositioning. *Mol BioSyst* 9:1268–1281
- Xie L, Li J, Xie L, Bourne PE (2009) Drug discovery using chemical systems biology: identification of the protein–ligand binding network to explain the side effects of CETP inhibitors. *PLoS Comput Biol* 5:e1000387

- Xie L, Xie L, Kinnings SL, Bourne PE (2012) Novel computational approaches to polypharmacology as a means to define responses to individual drugs. *Annu Rev Pharmacol Toxicol* 52:361–379
- Xie L, Ge X, Tan H et al (2014) Towards structural systems pharmacology to study complex diseases and personalized medicine. *PLoS Comput Biol* 10:e1003554
- Xing H, Gardner TS (2006) The mode-of-action by network identification (MNI) algorithm: a network biology approach for molecular target identification. *Nat Protoc* 1:2551–2554
- Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M (2008) Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 24:i232–i240
- Yamanishi Y, Kotera M, Kanehisa M, Goto S (2010) Drug–target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics* 26:i246–i254
- Yamanishi Y, Pauwels E, Kotera M (2012) Drug side-effect prediction based on the integration of chemical and biological spaces. *J Chem Inf Model* 52:3284–3292
- Yamanishi Y, Kotera M, Moriya Y, Sawada R, Kanehisa M, Goto S (2014) DINIES: drug–target interaction network inference engine based on supervised analysis. *Nucleic Acids Res* 42:W39–W45
- Yang L, Agarwal P (2011) Systematic drug repositioning based on clinical side-effects. *PLoS ONE* 6:e28025
- Yang K, Bai H, Ouyang Q, Lai L, Tang C (2008) Finding multiple target optimal intervention in disease-related molecular network. *Mol Syst Biol* 4:228
- Ye H, Liu Q, Wei J (2014) Construction of drug network based on side effects and its application for drug repositioning. *PLoS ONE* 9:e87864
- Yildirim MA, Goh KI, Cusick ME, Barabási AL, Vidal M (2007) Drug–target network. *Nat Biotechnol* 25:1119–1126
- Yuan Q, Gao J, Wu D, Zhang S, Mamitsuka H, Zhu S (2016) DrugE-Rank: improving drug–target interaction prediction of new candidate drugs or targets by ensemble learning to rank. *Bioinformatics* 32:i18–i27
- Zhao S, Iyengar R (2012) Systems pharmacology: network analysis to identify multiscale mechanisms of drug action. *Annu Rev Pharmacol Toxicol* 52:505
- Zhao S, Li S (2012) A co-module approach for elucidating drug–disease associations and revealing their molecular basis. *Bioinformatics* 28:955–961
- Zhao S, Nishimura T, Chen Y et al (2013) Systems pharmacology of adverse event mitigation by drug combinations. *Sci Transl Med* 5:206ra140–206ra140
- Zhou H, Gao M, Skolnick J (2015) Comprehensive prediction of drug–protein interactions and side effects for the human proteome. *Sci Rep* 5