Molecular BioSystems

RSC Publishing

REVIEW

View Article Online
View Journal | View Issue

Published on 26 February 2013. Downloaded by Technical University of Munich on 12/1/2022 11:55:56 AM.

Network-based drug repositioning

Cite this: *Mol. BioSyst.*, 2013, **9**, 1268

Zikai Wu,[a] Yong Wang[b] and Luonan Chen*[cd]

Network-based computational biology, with the emphasis on biomolecular interactions and omics-data integration, has had success in drug development and created new directions such as drug repositioning and drug combination. Drug repositioning, *i.e.*, revealing a drug's new roles, is increasingly attracting much attention from the pharmaceutical community to tackle the problems of high failure rate and long-term development in drug discovery. While drug combination or drug cocktails, *i.e.*, combining multiple drugs against diseases, mainly aims to alleviate the problems of the recurrent emergence of drug resistance and also reveal their synergistic effects. In this paper, we unify the two topics to reveal new roles of drug interactions from a network perspective by treating drug combination as another form of drug repositioning. In particular, first, we emphasize that rationally repositioning drugs in the large scale is driven by the accumulation of various high-throughput genome-wide data. These data can be utilized to capture the interplay among targets and biological molecules, uncover the resulting network structures, and further bridge molecular profiles and phenotypes. This motivates many network-based computational methods on these topics. Second, we organize these existing methods into two categories, *i.e.*, single drug repositioning and drug combination, and further depict their main features by three data sources. Finally, we discuss the merits and shortcomings of these methods and pinpoint some future topics in this promising field.

Received 19th September 2012,
Accepted 26th February 2013

DOI: 10.1039/c3mb25382a

www.rsc.org/molecularbiosystems

1 Introduction

The high incidence of complex diseases, for example, cancer, cardiovascular disease, and diabetes, poses a challenge to the pharmaceutical industry nowadays. One evidence is that the number of new drugs approved by regulatory agents has declined steadily for decades, though time and expenditure on drug Research & Development (R&D) increased annually.[1] Even worse, the recurrent emergence of drug resistance significantly reduces the efficacy of drugs and confound the efforts combating complex diseases. Thus it is in pressing need to develop more effective drugs with limited cost and time range to fill in the gap of new drug's productivity. Drug repositioning is one of the promise frameworks to circumvent this situation. It is also known as drug repurposing, drug reprofiling, drug redirecting, drug re-tasking, and therapeutic switching.[2,3] The basic idea for drug repositioning is to find new uses for existing or failed drugs, which have a well-known safety profile and pharmacokinetic profile.[2] This valuable safety information highlights drug repositioning's biggest advantage of mitigating the costs and risks associated with the early development stage, and shortening routes to approval for therapeutic indications. Motivated by this significant benefit, drug repositioning has been a promising alternative drug discovery approach for decades. Successful examples of drug repositioning include the use of sildenafil for erectile dysfunction and pulmonary hypertension, thalidomide for severe erythema nodosum leprosum, and retinoic acid for acute promyelocytic leukemia.[4,5] More information about these repositioned drugs can be found in the literature.[2,3,6,7] We believe that the initialization of the drug repositioning program will be further advocated by industry, government, and academia in the future, and the repositioning strategies are expected to increase unprecedentedly in the area of drug discovery.[8,9]

In recent years, systems biology continues to make significant progress to address fundamental questions in biology and leading to practical applications in medicine and drug discovery.[10] Network-based computational systems biology emphasizes the interactions among biomolecules and highlights the network concept. Inspired by this, we extend single

[a] *Business School, University of Shanghai for Science and Technology, Shanghai, China*
[b] *Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China*
[c] *Key Laboratory of Systems Biology, SIBS-Novo Nordisk Translational Research Centre for PreDiabetes, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, China. E-mail: lnchen@sibs.ac.cn; Fax: +86 21-5497-2551; Tel: +86 21-6436-5937*
[d] *Collaborative Research Center for Innovative Mathematical Modelling, Institute of Industrial Science, The University of Tokyo, Japan*

This journal is © The Royal Society of Chemistry 2013

Published on 26 February 2013. Downloaded by Technical University of Munich on 12/1/2022 11:55:56 AM.

drug repositioning to repositioning the relationships among drugs, *i.e.*, the drug interactions. In such a sense, drug repositioning actually includes the efforts of combining drugs with distinct indications, which are able to reduce the drug resistance and to combat complex disease more effectively. For example, successful stories include the advent of cocktail treatment against HIV infection and other drug combinations for various complex diseases.

However, most of the successful examples of drug repositioning are from serendipitous clinical observations. For many cases, the underlying molecular mechanisms are often not clear. This makes it difficult to reposition drugs in a large scale manner. Therefore, computational methods are expected to effectively reposition drugs against various diseases. Furthermore, drug repositioning is also spurred by the following observations: additional targets of many drugs have not been discovered; such drugs are likely to exert additional effects by direct activation/inhibition on their targets.[11,12]

On a molecular level, drug repositioning is grounded in two facts. Firstly, complex diseases are generally caused by the collective abnormalities of a number of correlated genes. The dysfunction of these genes will propagate and perturb certain biological processes by the interactions among molecules, leading to the onset of diseases. Secondly, mediated by drug–target binding,[13,14] some biological processes will be perturbed due to the manifestation of the drug's effect. Therefore, the common biological processes perturbed under disease state or under drug administration may imply potential drug repositioning. Collectively, drug repositioning's main challenge is to unveil the underlying responsive biological processes, and further bridge the molecular details and phenotype under specific conditions.

Therefore, the mechanism-guided drug re-usage calls for analysis of biomolecular networks. Network-based computational biology, a booming area studying various biomolecular networks, is a multidisciplinary intersection of mathematics, computer science, and biology.[15–19] It aims to organize the relationships among biological molecules in the form of networks, to find newly emerged properties at a network level, and to investigate how cellular systems induce different biological phenotypes under different conditions. This final goal perfectly matches the challenge of repositioning drugs rationally in a large scale manner. Additionally, a huge amount of high-throughput data related to drugs at various levels have been rapidly accumulated. This enables us to model cellular systems and further uncover the mechanisms underlying manifested phenotypes in a network framework. Therefore, computationally repositioning drugs has been a hot topic, and so far many computational approaches have been proposed to reposition drugs against various diseases.

Inspired by the significance of drug repositioning and abundant investigations devoted to it, here we provide a brief review on network-based studies. The flowchart of this paper is illustrated in Fig. 1. Firstly, we comprehensively summarize the available genome-wide data for drugs, biomolecules, and diseases. Then the individual networks are constructed to describe the relationships among drugs, interactions among genes and proteins, and the associations among diseases at different levels. Furthermore, these three networks are further integrated into a single three-layer drug–gene–disease network by considering the drug–target and the disease–gene interactions. Within this integrated network, we then focus on several popular topics regarding repositioning drugs from the perspective of networks, such as network-based single drug repositioning and drug interaction repositioning. With the different data sources utilized, these two classes of topics are further categorized into several subclasses. In addition, some works closely related to drug repositioning, such as drug–target prediction, mode of action identification, and side effect prediction are also partially touched upon in this review. Further research trends and future topics are discussed from the computational perspective. Finally, we make several general remarks on drug repositioning in a network framework to conclude this paper.

2 Available data and biomolecular networks for drugs, genes, and diseases

Typically, a network comprises a set of nodes and edges, and is described by graph theory in a mathematical manner.[15] A node can be a biological molecule taken from a cell, *e.g.* gene, RNA, protein, metabolite, or compound. A node can also be at the phenotype level such as a disease. An edge can represent the complex interplay between two nodes that coordinate their activities in biology. Depending on the heterogeneity of the node and level at which the nodes exist, a homogeneous or heterogeneous network with various scales could be constructed, and new properties will emerge at different network scales.

The accumulation of various high-throughput data has made the reconstruction of biomolecular and cellular networks possible. Depending on the cellular level, the reconstructed networks form a hierarchical structure. At the genomic level, transcription regulatory networks or gene regulatory networks model the molecular relationships to regulate gene expression. It could be reconstructed by leveraging direct binding data of transcription factors and target genes, microRNA and gene associations, or being indirectly inferred from microarray gene expression data. At the proteomic level, direct (or physical) protein–protein interaction (PPI), functional associations from co-membership of protein complexes and pathways could be combined to capture the interplay among proteins. At the metabolic level, the co-membership of biochemical reactions could be used to unveil a cellular metabolic network. Finally, a more complete heterogeneous cellular interaction network could be obtained by integrating those biomolecular networks at all levels.

Complex diseases generally originate from the abnormalities of multiple genes. These dysfunctional genes and their products may further perturb other biological molecules, activate/inactivate the interplay among these molecules, and rewire the whole cellular network. The transcriptional or proteomics data can measure the biomolecule activities and genetic risk

Published on 26 February 2013. Downloaded by Technical University of Munich on 12/1/2022 11:55:56 AM.

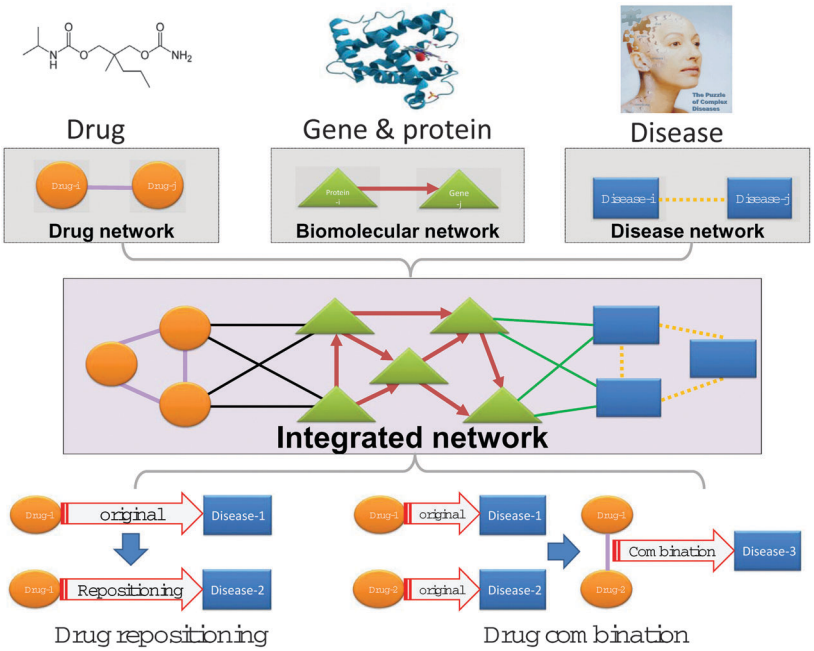

Fig. 1 A flowchart to illustrate the network-based drug repositioning in this paper. Firstly, the available genome-wide data are summarized for drugs, biomolecules, and diseases. Then the individual networks are constructed to describe the relationships among drugs, interactions among genes and proteins, and the associations among diseases at different levels. Then these three networks are further integrated into a single three-layer drug–gene–disease network by considering the drug target and the disease gene interactions. Within this integrated network, repositioning drugs can be naturally decomposed as single drug repositioning and drug interaction repositioning, *i.e.*, drug combination.

factors. In addition, the rewired biomolecular network under disease state could be uncovered through extracting subnetworks with high aggregated activities and relevance from a general background interaction network.[20] These responsive subnetworks under the disease state could be used further to decipher a disease's mechanism of action. Similarly by binding with target proteins, a drug's intake will also perturb other biological molecules, activate/inactivate the interplay among these molecules, and also rewire the whole cellular network. The rewired biomolecular networks or responsive subnetworks induced by the target proteins can be further unveiled by transcriptional or proteomic data. To some extent, these responsive subnetworks under drug treatment can be regarded as the surrogate of mechanism of drug's action. The proximity of those mechanisms may imply associations among diseases and drugs.

Additionally, disease and drug associations could also be inferred from similarity of molecular basis, chemical substructure, and phenotype. By combining the inferred associations with the confirmed associations, such as known drug–indication relations, co-membership in drug combinations, co-morbidity of diseases and so on, a comprehensive heterogeneous disease–molecule–drug network can be constructed. These reconstructed networks at distinct levels and on different scales collectively present a systemic view on the relation of diseases and drugs, and may provide fruitful opportunities for drug repositioning.

As we mentioned, diseases generally originate from the abnormalities of multiple correlated factors while drug action always originates from the drug–target protein binding. These direct effects can be deemed as the molecular origin of diseases and drugs development. Subsequently, microarray, proteomic data, and other omics data can record the molecular activities induced by diseases and drugs, which can be viewed as the intermediate effect of disease development and drug's mode of action. Moreover, drugs' indications and side effects and other phenotypes are the terminal effect caused by drug intake. The cause–effect relationship among them is summarized in Fig. 2. To this end, data about diseases and drugs clearly originate from three distinct levels, that is, molecular origin, molecular activity, and phenotype. Nowadays, various sources of data are available for reconstructing networks characterizing the relation of diseases and drugs. To facilitate the efforts of constructing networks and repositioning drugs, we list and briefly describe the various data sources in Table 1.

3 Repositioning drugs against various diseases

The accumulation of various data in Table 1 sparked many computational methods for single drug repositioning. As available data about diseases and drugs clearly originate from three distinct levels, we can categorize these methods by the data sources. In the following, we will discuss each class of methods in detail.

3.1 Repositioning drugs by molecular origin profile

Diseases are a problem of the corresponding system, and highly related to the abnormalities of multiple correlated molecules.

 This journal is © The Royal Society of Chemistry 2013

Published on 26 February 2013. Downloaded by Technical University of Munich on 12/1/2022 11:55:56 AM.

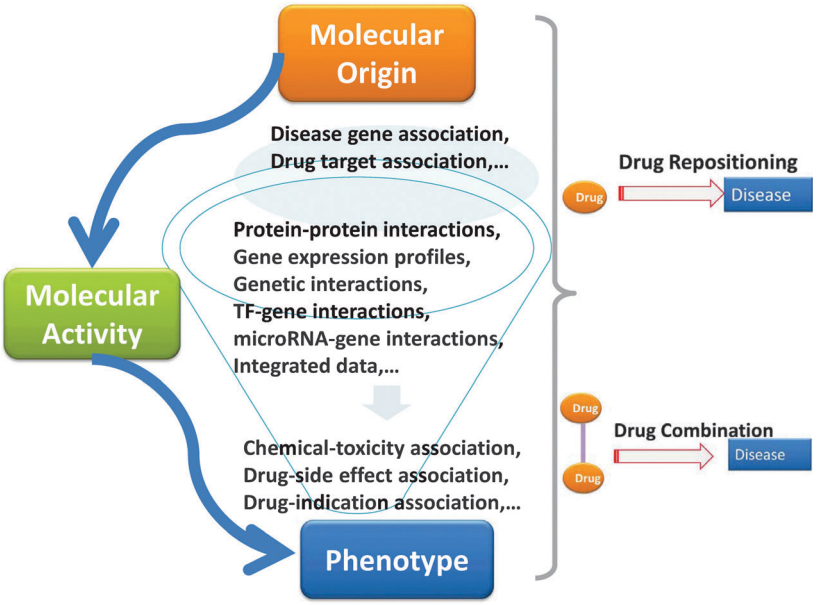

Fig. 2 The cause–effect relationship between molecular origin, molecular activity and phenotype.

Table 1 Various databases related to diseases, genes, proteins, and drugs. These databases are used for drug repositioning

Database	URL	Data type	Data level
OMIM[21]	http://www.ncbi.nlm.nih.gov/omim	Gene–disease association	Molecular origin
miR2disease[22]	http://www.mir2disease.org/	miRNA–disease association	Molecular origin
STITCH[23]	http://stitch.embl.de/	Chemical–Protein interaction	Molecular origin
HPRD[24]	http://hprd.org/	Protein–protein interaction(PPI)	Molecular activity
BioGRID[25]	http://wiki.thebiogrid.org	PPI/genetic interaction	Molecular activity
DIP[26]	http://dip.doe-mbi.ucla.edu/dip/Main.cgi	PPI	Molecular activity
TRED[27]	http://rulai.cshl.edu/cgi-bin/TRED/tred.cgi?process=home	TF–gene interaction	Molecular activity
TRANSFAC[28]	http://www.gene-regulation.com/pub/databases/transfac/doc/	TF–gene interaction	Molecular activity
miRBase[29]	http://www.mirbase.org/	MicroRNA–gene interaction	Molecular activity
TargetScan[30]	http://www.targetscan.org/	MicroRNA–gene interaction	Molecular activity
Signaling Network[31]	NA	Human signaling network	Molecular activity
GO[32]	http://www.geneontology.org/	Functional annotation	Molecular activity
CMAP[33]	http://www.broadinstitute.org/cmap/	Microarray(special to drug)	Molecular activity
Arrayexpress[34]	http://www.ebi.ac.uk/arrayexpress/	Microarray	Molecular activity
GEO[35]	http://www.ncbi.nlm.nih.gov/geo/	Microarray	Molecular activity
Drugs@FDA	http://www.accessdata.fda.gov/scripts/cder/drugs/atfda/	Drug–disease association	Phenotype
ClinicalTrial.gov	http://clinicaltrials.gov/	Drug–disease association	Phenotype
Pubchem[36]	http://pubchem.ncbi.nih.gov/	Bioassay of compound	Phenotype
SIDER[37]	http://sideeffects.embl.de/	Drug–side effect association	Phenotype
ACToR[38]	http://actor.epa.gov/	Chemical–toxicity association	Phenotype
DCDB[39]	http://www.cls.zju.edu.cn/dcdb/	Drug combination–disease association	Phenotype
KEGG[40]	http://www.genome.jp/kegg/	Pathway,disease, drug	Integrated data
DrugBank[41]	http://www.drugbank.ca/	Integrated data of drug	Integrated data
CTD[42]	http://ctdbase.org/	Chemical–disease–gene association	Integrated data
PharmGKB[43]	http://www.pharmgkb.org/	Integrated data of drug	Integrated data
NPC[44]	http://tripod.nih.gov/npc/	Integrated data of drug	Integrated data

These dysfunctional molecules, *e.g.*, genes, can be viewed as the cause of the disease phenotype or cellular changes at a molecular level. Additionally, relevant diseases indeed have relevant or shared aberrant genes.[45,46] This shows that the shared molecular origin always pinpoints relevant or shared etiology. Drug intervention aims to restore the etiology of a disease. Therefore, therapies and medications could be transferred among diseases with relevant or shared molecular origins. On the other hand, a drug's effect on a living organism is always resulted from drug–target protein binding. Thus, the drug–protein binding profile can be viewed as the cause of phenotype change or cellular changes at a molecular level after drug intake. To this extent, drugs with relevant or shared targets may behave similarly in biological systems and can target at a similar or even the same disease. These facts indicate possible repositioning opportunities. Given the above two points, many

Published on 26 February 2013. Downloaded by Technical University of Munich on 12/1/2022 11:55:56 AM.

network-based methods have been developed to reposition drugs against various diseases, by exploiting the relevance or overlap of diseases' and drugs' molecular origins.

However, the known molecular origins of diseases and drugs are far from complete. This limits the efficacy to quantify the relevance or overlap of molecular origins. To overcome this problem, some network-based computational methods have been proposed to predict potential genetic risk factors for diseases.[17,47,48] The general scheme is summarized as follows. Firstly, the known disease genes are mapped to a biomolecular interaction network and highlighted as seed genes. Secondly, the topological distance between the remaining genes and set of the seed genes is evaluated and ranked. Lastly, the top ranked genes are prioritized as new disease genes by equalling topological proximity with relevance of diseases.

Drugs are generally non-specific and show reactivity to additional targets besides primary targets. Accurate prediction of drug targets helps the follow-up drug repositioning task. Hence, many computational methods have been developed for uncovering potential drug–target interaction. Cheng *et al.* built a bipartite graph composed of the approved drugs and proteins linked by drug target binary associations, and relied on a supervised network-based inference method to predict drug–target interactions.[49] On the other hand, Chen *et al.* constructed a general heterogeneous network which comprises drugs and proteins linked by protein–protein sequence similarity, drug–drug chemical similarity, and the known drug–target interaction.[50] Subsequently, the method of random walk was employed to prioritize potential targets by simulating a random walker's transition in the heterogeneous network starting at the highlighted drug seed nodes. Different from the above two methods, Yeh *et al.* developed a network flow approach for identifying potential target proteins, which have a strong influence on disease genes in the context of biomolecular networks, that is weighted by degree of co-expression of interacting protein pair.[51] Also, Wang *et al.* showed that data integration further improved the drug target prediction.[13,14]

Based on the verified molecular origins augmented with the predicted molecular origins, many methods are ready to be employed to reveal a drug's new roles. The common idea is that drugs with similar molecular origins could combat diseases with similar molecular origins. Therefore, indications transferred between drugs and medications could also be transferred among diseases. From such a viewpoint, the key for computational methods is how to measure the similarity at a molecular origin level from the data sources at hand. In particular, for network-based methods, the key becomes how to capture similarities of molecular origins among drugs and diseases at a network level.

DRAR-CPI is one of the first online servers to reposition drugs in a network's framework.[52] It firstly constructed a comprehensive drug–protein interaction network by combining primary drug–target interactions with off drug–target associations inferred by the embedded target prediction tool. Then the known indications of one drug were transferred to other drugs having similar protein association profiles. Edberg *et al.* further proposed to transfer indications among drugs based on relevance of their molecular origin profiles, which were formed by augmenting target proteins with proteins that interact with them.[53] On the other hand, pathways enriched in drug targets and disease genes were utilized to infer drug–disease associations in Ye *et al.*'s work.[54] Lee *et al.* made several improvements in terms of diversity of data sources and algorithms. Their work started with mapping drug target proteins, drug metabolizing proteins, and protein products of disease genes onto the protein–protein interaction network.[55] In more detail, the drugs, diseases, and proteins in the resulting tripartite network were linked by drug–target binary associations, drug–metabolizing enzyme associations, disease–protein(gene) binary associations, drug–disease associations, and protein–protein interactions. Finally, a shared neighborhood scoring algorithm was developed to predict potential links between drugs and diseases. The basic idea was that if non-linked drug–disease pairs had some common neighbor nodes, these drug–disease pairs could be linked with the confidence quantified by the number of the shared neighbors and the relationship strengths among neighbors. As the shared neighbors may be drugs or proteins, actually, the inference is based on the shared molecular origins between two drugs or the shared molecular origins between drugs and diseases.

Previous researches have shown that a biomolecular network always has a hierarchical structure.[16] The building blocks named communities or modules, with high intra-connectivity, topologically and functionally bridge the individual molecules into a global network. Molecules in the same module are more likely to be functionally related. Based on this insight, Daminelli *et al.* proposed a method for drug repositioning by identifying modules from heterogeneous association networks.[56,57] Daminelli *et al.*'s work started with the construction of a drug–target–disease network, in which drugs, targets, and diseases are linked by drug–target associations and drug–disease associations. Subsequently, the technique of power graphs was introduced to identify actually incomplete cliques in the network. Given that the network is the superposition of a bipartite drug–target network and a drug–disease network, the bi-modules that are connected by common drugs are detected. Finally, resultant bi-modules' completion introduces novel, predicted links from drugs to targets and diseases, which means that this method can reposition drugs and predict a drug's off targets simultaneously.

Zhao *et al.* developed a Bayesian partition method to identify drug–gene–disease co-modules.[57] In this method, a comprehensive protein–protein interaction network was assembled firstly by integrating data from HPRD[24] and other databases. Then disease-related genes from OMIM[21] and drug targets from DrugBank[41] were mapped onto the protein–protein interaction network. Subsequently, a gene–drug closeness profile was computed to reflect the aggregated network distance between a gene and each drug's targets. Similarly, gene–disease closeness profile was determined to reflect the aggregated network distance between a gene and each disease related genes. Finally, a Bayesian partition method was developed to identify

 This journal is © The Royal Society of Chemistry 2013

Published on 26 February 2013. Downloaded by Technical University of Munich on 12/1/2022 11:55:56 AM.

drug–gene–disease co-modules underlying the gene closeness data. Therefore, the method can infer drug–disease associations and pinpoint the molecular basis underlying these associations simultaneously.

Some non-network-based methods have also been published for repositioning drugs by mining molecular origin, such as the chemical similarity search based method proposed by Keiser *et al.*[58] Collectively, the limitations of these molecular origin based repositioning strategies largely stem from the following facts. Firstly, current profiles of molecular origin for diseases and drugs are incomplete and noisy. Therefore, they cannot predict the whole outcome of disease and drug accurately and comprehensively. Secondly, there is no simple mapping between casual molecular origin and living organism's response. It is the interplay between physiological environment and casual molecular origin that finally determines the outcomes of the disease development and drug treatment. Therefore, similar molecular origins do not correspond to similar outcomes necessarily.

3.2 Repositioning drugs by molecular activity profile

Due to the complex interactions among biological molecules, the disease genes will perturb the remaining molecules' activities through network during the disease development. Similarly, the introduced drug molecule and its targets will also induce the abnormalities of molecules' activities in the organism. Actually, molecular activities can reflect the disease's and drug's effect comprehensively. On the other hand, numerous researches have shown that microarray, proteomic data, and other omics data can indeed measure and represent the total molecular activities of biological systems under specific conditions. Thus, microarray and other molecular activity data have been utilized extensively by various methods for repositioning drugs.

Due to the high dimensionality of molecular activity data, extracting profiles or signatures from the data is the common starting point. Generally, these methods have a similar procedure as follows. Firstly, the indications of drugs are transferred if their molecular activity profiles are similar. Secondly, the medications of diseases are transferred based on similar molecular activity profiles. Lastly, the drug intervention is conducted to restore the disease state into the normal state. If the molecular activity profile of drug anti-relates to that of a disease, the drug will have the potential to cure the disease.

The first systematic approach to reposition drugs by leveraging molecular activity profile is the Connectivity Map project conducted by Lamb *et al.*[33] They created a reference collection of gene expression profiles from cultured human cells treated with 1309 bioactive small molecules, which has become the main data available of molecular activity at the transcription level. Then the top differentially expressed genes were deemed as the molecular activity profile. Subsequently, a measure rooted in gene set enrichment analysis was developed to evaluate the molecular activity profiles' correlation among drugs and anti-correlation between diseases and drugs. Finally, the indications transferring among drugs were based on the correlated molecular activity profiles, and the drug–disease associations were inferred by anti-correlations. Motivated by Lamb *et al.*'s work, Iorio *et al.*, Dudley *et al.*, Sirota *et al.*, Hu *et al.*, and Shigemizu *et al.* separately proposed methods by measuring the correlation between diseases' top differentially expressed genes and that of drugs, in which the resulting measure of the modified gene set enrichment analysis, Pearson correlation coefficient, and size of overlap were employed to evaluate the correlation, respectively.[59–64]

Apart from the above general methods, several methods have also been developed by exploiting molecular activity profile from the network's viewpoint. The specific idea underlying these methods is that the perturbation by diseases' genetic factors or drug–target binding is assumed to always rewire the biomolecular network. Such rewiring coordinates molecules' activities, enables biological processes to function, and further leads to the disease development (or drug intervention). To some extent, these responsive networks could be used as the surrogate of mechanism of action underlying diseases or drugs. Therefore, the relevance of responsive networks inferred by leveraging molecular activity data may imply associations among diseases and drugs. Additionally, the key to distinguish them is how to extract the responsive networks and evaluate the relevance of one responsive network to another.

Many methods have been proposed for extracting responsive networks. Their general procedure is as follows. Firstly, a background interaction network is constructed by integrating protein–protein interaction (PPI), protein–DNA interaction, and other interaction data. Secondly, the background network is transformed into a weighted network by leveraging gene expression and other information. Subsequently, various models and the corresponding algorithms are developed to extract subnetworks with high aggregate weight from the weighted background network. Those subnetworks are then identified as the responsive networks by equalling mathematical maximum with the biological relevance. Guided by this framework, we also proposed several computational methods to detect responsive networks from PPI networks by leveraging microarray data.[20,65–69] Based on global characteristics of interactome coupled with gene expression data, we developed a novel method to detect disease-related gene modules or dysfunctional pathways. In this method, interactions among genes are exploited to define a gene's activity score and responsive networks are inferred by the support vector regression.[65] Besides, extracting responsive networks was formulated as a mixed integer linear programming model and we implemented this method to identify differential pathways of breast cancer metastasis,[66] the critical transitions of liver cancer[70,71] and also colorectal cancer.[68] Moreover, a quadratic programming model and a fast approximate algorithm were proposed and applied to identify the responsive subnetwork of type-2 diabetes.[20] For other approaches, the readers are suggested to refer to recent reviews.[17,67,68,71,72]

On the other hand, more specific methods have been developed to identify responsive networks under drug intervention to decipher a drug's mode of action. As an example, Mitsos *et al.*

Published on 26 February 2013. Downloaded by Technical University of Munich on 12/1/2022 11:55:56 AM.

proposed a method to identify networks relevant to drugs by leveraging phosphoproteomic data.[73] The method reconstructed a generic pathway made by logical gates. Then the task of identifying responsive networks was reduced to the problem of finding subnetworks with which key phopshoprotein signals were fitted optimally. Subsequently, the problem was formulated as a linear programming model and solved by standard solvers. Besides, we also developed a mixed integer linear programming model for identifying responsive networks under drug treatment.[74]

The responsive network concept then can be exploited further to uncover the relevance among diseases and drugs, *i.e.*, reposition drugs against diseases. Apart from these methods, there exist other straightforward methods. As an example, Silberberg *et al.* proposed a method for elucidating drug-response pathways and predicting indications by integrating physical interaction data with transcription data.[75] In this method, the drug specific networks were defined as subnetworks to compactly connect drug targets with the corresponding sets of differentially expressed genes from the CMAP data set.[33] Then it was extracted from an assembled human physical interaction network by a Steiner tree and shortest path approaches.[76] Further, latent responsive pathways were detected from the resulting subnetworks by consecutive steps of dissection and expansion. Finally, a pathway-based similarity assessment method between drugs using the Jaccard score was derived and used for repositioning drugs with convincing accuracy.

Jin *et al.* proposed another network-based method to facilitate drug repositioning for cancer therapy by using transcription data.[77] In their work, the responsive networks under drug treatment were defined as the network motifs that compactly connect the products of disease genes to the known canonical signaling pathways. With the assumption that there were some uncovered driver pathways bridging drug's targets and responsive network, the transcriptional data from the CMAP data set[33] were further decomposed by Bayesian factor regression. And the hidden driver pathways and drug's effect on them were deciphered simultaneously. Subsequently, the effect of a drug on each driver pathway was measured and summarized into a drug-driver pathway interaction profile. Then the association between drug-driver pathway interaction profile and repositioning potentials was mined by support vector regression. Finally, their method accurately predicted clinical responses to more than 90% of drugs approved by the U.S. FDA and more than 75% of experimental clinical drugs.

Despite great successes made by repositioning drugs based on molecular activity profile from the network perspective, there are still some intrinsic limitations hampering their efficacy. Firstly, molecular activity data is noisy, thereby leading to the noisy molecular activity profile. Besides, biomolecular interaction data are also noisy and incomplete. Those limitations make the extracted responsive networks biased. Secondly, the assumption that mathematical optimum of responsive networks equals biologically maximal relevance does not hold necessarily. Lastly but most importantly, there is no simple mapping between a responsive network and living organism's response. On the one hand, the development process of a complex disease always involves the dysfunction at multiple levels and in various cell types. On the other hand, drug's final therapeutic effect always relies on its metabolites' activities and the information propagation way. Hence, the defect at data and underlying assumptions may significantly limit the power of repositioning drugs based on molecular activity profile from the network's perspective.

3.3 Repositioning drugs by phenotype profile

Many drugs induce some unintended effects in the living organism besides the primary desired effects, which constitute a drug's overall effect profile. Those wanted or unwanted behavioral or physiological changes in response to drug treatment can be measured as drugs' indications and side effects. Therefore, drug's indications and side effects could be utilized to infer mechanisms underlying disease development or drug treatment. This may provide additional information to explore the correlation of drugs' mechanism of actions. Given this, phenotypic data, more specifically, indications and side effects were used extensively by various methods for drug repositioning.

We note that current side effect information of drugs is far from complete. Thus, predicting drugs' side effects is necessary. As an example, Lounkine *et al.* developed a computational method to predict side effects based on the predicted off targets, with the assumption that off targets induce side effects.[78] With the same idea, Huang *et al.* further proposed a machine learning method to predict side effects by a multifacet of drug targets.[79] In this method, the relation among the functional subnetworks induced by drug targets was used to predict side effects. Chang *et al.* and Atias *et al.* separately proposed different side effect prediction methods based on biomolecular networks.[80,81] In Chang *et al.*'s work, the functional effects of perturbing off targets in the context of a metabolic network were measured and identified as the side effects induced by targets. Besides, Atias *et al.* combined two algorithms to predict side effects. In the first algorithm, the relation between molecular data and side effect was trained by canonical correlation analysis and used for predicting side effects, while a diffusion process in a side effect similarity network was adopted to rank side effects for a given drug in the second algorithm.

With the augmented profile of drug effect, many methods have been developed for drug repositioning. As an example, Yang *et al.* proposed to reposition drugs based on their side effect profiles.[82] The basic hypothesis was that if a drug and a cluster of drugs prescribed to some disease have a similar side effect profile, then this drug should be evaluated as a candidate for treating that disease. Based on this hypothesis, a disease–side effect association network was constructed by combining drug–disease associations with drug–side effect associations. Then the relations between side effect and disease were trained by a Laplacian-modified Bayesian method and used for predicting new indications for drugs.

Apart from the methods only leveraging phenotype profile, there are other methods which integrate phenotype profile with

 This journal is © The Royal Society of Chemistry 2013

Published on 26 February 2013. Downloaded by Technical University of Munich on 12/1/2022 11:55:56 AM.

molecular origin profile and/or molecular activity profile for drug repositioning. PROMISCUOUS is one of the first online servers to reposition drugs by integrating multiple data sources.[83] The method constructed a heterogeneous network composed of target proteins, general proteins, and drug side effects by integrating drug–target associations, drug–side effect associations, and drug–drug chemical similarity with PPIs. Then the indications among drugs are transferred based on new associations, which are derived by integrating their relevance at chemical structure, molecular origin, and side effect. Similarly, Re *et al.* also constructed a heterogeneous network and transformed it into a drug–drug network by network projection methods.[84] Then a random walk was conducted in the drug–drug network, starting from a set of drugs targeting the same therapeutic category, to prioritize potential drugs. Compared to PROMISCUOUS and Re *et al.*'s work, Gottlieb *et al.* made several improvements in terms of diversity of data sources.[85]

Although great successes have been made in repositioning drugs by phenotype profile, there are still some intrinsic limitations. Firstly, current disease and drug phenotype data are noisy and far from complete. Take side effect as an example. Not all side effects discovered in the preclinical stage will emerge at the clinical stage and *vice versa*. On the other hand, partial side effect information collected by spontaneous reporting systems during post-marketing surveillance is also confused by patient's medication history or trait and other hidden factors. Secondly, there is no simple mapping between phenotype and mechanism of action. The phenotypical outcomes of drug's mode of action rely heavily on living organism's genetic map, medication history, and traits. Therefore, a similar phenotype does not correspond to the similar mode of action necessarily.

The common aim of the above methods is to unravel the relevance and the mechanism of action among diseases and drugs, at the level of molecular origin, molecular activity, or phenotype. From the perspective of computation, the key question is just how to identify the mechanism of action. From this viewpoint, the above mentioned methods could be roughly grouped into two classes, that is, deterministic methods and statistical methods. Deterministic methods mainly rely on optimization techniques to infer the relevant mechanism of action. As an example, Silberberg *et al.* proposed a deterministic method for elucidating drug response pathways.[75] In this method, latent responsive pathways were detected by dissecting the drug-specific networks extracted from a background interaction network by an optimization approach, namely Steiner tree and shortest path approach.[76] The advantage of deterministic methods is relatively fast in terms of computational speed. However, it is generally difficult to deal with large scale networks. On the other hand, a statistical method can always obtain desired numerical results, *e.g.*, by applying Monte Carlo Markov Chain (MCMC) strategies. As an example, a Bayesian method developed by Zhao *et al.* to identify drug–gene–disease co-modules[57] is a statistical method. In this method, to extract drug–gene–disease co-modules from a background interaction network, MCMC strategies were introduced to partition the background network into co-modules that fit the data optimally. However, one disadvantage of statistical methods is their relatively heavy computational burden, although it can algorithmically deal with large scale problems in an easy manner.

4 Identifying effective drug combinations

Recently, the phenomenon that the treated living organism will develop a resistance mechanism after a period of drug intervention draws more and more attention. From the perspective of networks, the emergence of drug resistance stems from two facts. Firstly, complex diseases are often caused by multiple molecular abnormalities, which will induce responsive networks through interactions among molecules. Back-up circuits or fail safe mechanisms will further be developed or appear in the context of responsive networks or subsystems to resist perturbations caused by drugs. Secondly, pharmaceutical industry has been driven by a target-based paradigm in recent decades. Most of the developed drugs only have single primary targets, which cannot overcome the back-up circuits underlying the disease state. Therefore, drugs are greatly needed to target multiple dysfunctional molecules and counter back-up circuits. In addition, drug combination may generate synergistic effects to combat a disease, which is also attractive. Due to the high risk of developing new multi-target drugs, combining existing drugs to combat complex diseases has been a promising and economic strategy.

To meet the challenge, many computational methods have been proposed to identify effective drug combinations. In a similar way to the description of drug repositioning, we also categorize drug combination methods by the data sources used, *i.e.*, molecular origin, molecular activity, and phenotype. Next, we will discuss each class of methods in the corresponding subsections.

4.1 Identifying drug combinations by molecular origin

A drug's general mechanisms and side effects are closely related to the proximity of disease–gene and drug target proteins in a biomolecular network.[86,87] It is possible to explore the biomolecular network to find the drugs which can attack multiple disease–genes collaboratively and also bypass the back-up circuits.

Inspired by this idea, various methods have been designed to explore novel mechanisms of drug combinations by integrating molecular origin data with biomolecular interaction data. As an example, Xu *et al.* proposed a network biology approach to uncover underlying rules of effective drug combinations.[88] They investigated the distance between drug combinations' targets in the context of protein interaction network, human pathway, and functional annotations. Their investigations concluded that effective combination's target proteins tend to be close in a protein interaction network and collectively execute specific biological functions.[88] At the same time, they investigated the distance between drug combinations' targets in the context of a genetic interaction network. They concluded that

Published on 26 February 2013. Downloaded by Technical University of Munich on 12/1/2022 11:55:56 AM.

targets of drug combinations tend to modulate functionally-related pathways.[89] Li *et al.* further explored combination rules of a special class of drug combination, that is, Traditional Chinese Medicine from a network perspective.[90] A herb–herb association network was constructed by measuring the tendency of participating common formulae and was further dissected into modules. Then the common molecular basis of herbs in the same module was delineated. Their investigation showed that the herbs working together are usually prescribed to the diseases with the correlated molecular basis. This implies that correlation among drug's indications is a valuable clue for identifying effective drug combinations.

Although the above reviewed works are not explicitly designed for identifying new drug combinations, the rules deciphered could shed light on the underlying principles of drug combinations and guide efforts of identifying new effective drug combinations. Different from these methods, Li *et al.* proposed a network-based method for identifying effective drug combinations.[91] In line with the rules uncovered by Xu *et al.*, the basic idea of their method was as follows. If the drug targets are located at the central position and close to each other in the disease specific interaction network, these drugs have the potential to attack the disease network with strong effect collaboratively and could be used combinatorially for combating that disease. Moreover, the targets' topological importance and proximity in network were measured and used to identify synergistic drug combinations. As a proof of the concept, it successfully recovered five known synergistic agent pairs that have efficacious effect on a pathological process instanced by angiogenesis.

Although molecular origin based methods successfully recovered some conformed drug combinations and deciphered some underlying mechanisms, their applications are limited due to several facts. Firstly, current methods for exploring drug combinations are limited by the incompleteness of molecular origin data and interaction data. Furthermore, there is a gap between a drug's mechanism of action and its molecular origin. On the other hand, the mechanism of drug's action combinations is very complex and far from the simple superposition of its member's mechanism of actions. Therefore, there will be a larger gap between the molecular origin's topological proximity at the cellular network and mechanism of action of drug combinations.

4.2 Identifying drug combinations by molecular activity profile

By leveraging molecular activity data, many methods have been developed for exploring the mechanism underlying effective drug combinations or identifying new effective drug combinations. As an example, Zhang *et al.* explored the working mechanism of Chinese medicinal formula Realgar-Indigo naturalis on promyelocytic leukemia.[92] Their investigation showed that there are three active compounds in the formula and each hits distinct targets and induces complementary therapeutic effects. At the same time, they also interpreted the synergistic effects of arsenic sulfide and Imatinib in BCR/ABL-associated leukemia from the network perspective.[93] A systematic analysis of dynamic changes of the proteome, phosphoproteome, and transcriptome shows that arsenic sulfide and Imatinib target distinct pathways that synergistically arrests the cell cycle and decreases activity of BCR/ABL. Collectively, these two investigations provide hints that drugs targeting distinct pathways tend to produce synergistic effects.

Some methods have been designed to explore new effective drug combinations. However, the molecular activity data under drug combination intervention are scarce. To meet this challenge, several investigations combine the molecular activities induced by single agents to predict the molecular activities induced by a combinatorial therapy. Geva-Zatorsky *et al.* used a dynamic proteomics approach to accurately follow some key protein levels in human cells in response to single drugs and drug combination.[94] The results show that protein dynamics in response to combination drugs could be described as the weighted sum of their response to individual drugs, in which weights describe the relative impact of each drug on each protein.[94] On the other hand, Nelander *et al.* presented a method to directly predict quantitative outcomes of combinatorial perturbations.[95] In this method, a nonlinear dynamical system was developed to model input–output relations by taking the simplified cellular network structure into account. Finally, the dynamical system that fits molecular activities optimally was obtained by parameter identification. It was used to predict molecular activities induced by new combinatorial perturbations, such as combinatorial drug treatment. Although the rules need further validation at the whole genome level, to some extent, it paves the way for predicting molecular activities induced by combinatorial therapy through activities by single agents.

If molecular activities induced by both single drugs and combinatorial drugs are available, various methods have been proposed to compare the effects of single agents to those of combinatorial therapy. For instance, Jin *et al.* proposed an enhanced Petri-Net model to recognize the synergistic effects of drug combinations by leveraging drug-treated microarray data.[96] In this method, an enhanced Petri-Net model fitting to the drug-treated gene expression data was applied to simulate the propagation process of drug molecule's perturbation in the network. Finally, drug combinations that exert a stronger effect than the summed effect were identified as the effective drug combinations. Different from Jin *et al.*'s work, Iadevaia *et al.* utilized phosphoproteome data to detect effective drug combinations.[97] In this method, a mass action model, that fits to molecular activities of insulin-like growth factor signaling network optimally, was identified by the particle swarm optimization algorithm. Then the mass action model was used to predict the effect of targeting individual signaling proteins that exert on the rest of the network. Finally, the drug combinations that lead to minimal effects to the rest of network were identified as effective drug combinations.

Molecular activity data under combinatorial therapy in large scale are still scarce. To deal with this situation, we proposed a network-based approach to identify effective drug combinations

 This journal is © The Royal Society of Chemistry 2013

Published on 26 February 2013. Downloaded by Technical University of Munich on 12/1/2022 11:55:56 AM.

just by leveraging microarray data of single drug treatment.[74] The basic assumption of the method is that a subnetwork or pathway will be affected in cellular systems after a drug is administrated. Our work started with the construction of a background interaction network by integrating PPIs, protein–DNA interactions, and signaling pathways. As the microarray profile induced by the drug combination is unavailable, a Taylor expansion based scheme, that takes the saturation effect into account, was developed to predict the gene expression profile of the drug combination. Subsequently, we developed a mixed integer linear programming, named network flow model to identify responsive subnetworks under drug treatment from a background interaction network weighted by degree of differential expression. In this model, various constraints were considered to include drug targets and connectivity of the resulting subnetworks. Furthermore, we defined a score to evaluate the overall effect of one drug by maximizing efficacy and minimizing side-effects. Drug combinations, whose responsive subnetwork's score is higher than that of any single drugs, will be identified as effective drug combinations. Actually, a standard linear programming solver was employed to find the optimal solution and the networks relevant to drugs prescribed to type 2 diabetes, such as Metformin and Rosiglitazone, were identified successfully. We successfully recovered the combination of Metformin and Rosiglitazone, which is actually an approved drug called Avandamet.

By leveraging molecular activity data, identifying drug combination and drug repositioning share some common limitations, which stem from the incompleteness of data and gap between the predicted mode of action and the real one. Apart from these shared limitations, there are some other limitations. Molecular activity data under combinatorial therapy in large scale are also scarce. Predicting them through integrating molecular activities induced by single drug is still in its infancy stage. Therefore, mode of action of combinatorial therapy inferred by the predicted molecular activities profile is putative.

4.3 Identifying drug combinations by phenotype profile

When both phenotype data and single agent data are available, drug combination's effects on living organisms could be compared. Drug combinations, that have a stronger effect than summation effects, can be identified as the effective drug combinations. Furthermore, the underlying combination principle could be deciphered computationally and exploited.

Based on the above idea, many computational methods have been proposed. For example, Loewe *et al.*, Bliss *et al.*, and Chou *et al.* separately presented earlier methods for detecting drug combinations by leveraging dose–response curves, in which response is usually measured as the percentage of cells that are killed or inhibited by the drug.[98–100] The common procedure is summarized as follows. Firstly, the relative mechanism is hypothesized that drugs' actions have no interaction. Then the null model, that describes responses of the combination drug whose members have no interaction, is defined. Subsequently, the real dose–response curve of combination drugs, synergism, additive, and antagonism between drugs can be defined based on the comparison between the predicted dose–response curve of the null model. Moreover, Wong *et al.* and Calzolari *et al.* developed some optimization algorithms to identify drug combinations that induce the optimal therapeutic effects.[101,102]

With the verified drug combinations at hand, it is possible to explore the underlying combination rules to identify more effective drug combinations. Xu *et al.* discerned effective drug combinations and constructed a drug–drug association network with therapeutic effects mapped.[103] The enriched patterns in the network show that the agents in an effective combination tend to have more similar therapeutic effects and share more interaction partners.[103] Finally, the rules were coded by a measure and used to predict possible drug combinations by exploiting the topological features of the drug–drug association network.

Zhao *et al.* proposed a machine learning method to predict drug combinations by integrating molecular and pharmacological data.[104] In their work, molecular and pharmacological information associated with components of the approved drug combinations, including their target proteins and corresponding downstream pathways, medical indication areas, therapeutic effects as represented in the Anatomical Therapeutic Chemical (ATC) classification system, and side effects were collected. Each such property of a drug was called a feature, and a drug pair was represented as a features vector composed of feature pairs. Subsequently, features being predictive of the 184 pairwise drug combinations approved by the FDA were identified. Finally, these predictive features were integrated for predicting potential drug combinations. Among their top ranked predictions of effective combinations, 69% are supported by literature, while the others represent novel potential drug combinations.

Although a few drug combinations have been identified by utilizing phenotype data, there are still some factors hampering their further applications. Firstly, there are enormous possible drug combinations considering the exponential number of possible combinations. Therefore, it is a challenging task to detect effective drug combinations in the huge search space of possible combinations. Besides, as the known drug combinations are far from complete, the features enriched are therefore biased and can only predict potential effective drug combinations with a moderate power.

The mechanism of action of drug combination is generally not just the linear superposition of single drugs' mechanisms. As a result, identifying effective drug combinations may be a more challenging task. Therefore, more or less, the above reviewed methods in Section 4 leverage some prior knowledge about the approved drug combination to explore potential drug combinations. Based on the degree of prior knowledge about the known drug combinations utilized, these methods could be grouped into two classes, that is, unsupervised methods and supervised methods. The network-based approach to identify effective drug combinations proposed by us[74] is an unsupervised method. Just based on the transcriptional expression and

Published on 26 February 2013. Downloaded by Technical University of Munich on 12/1/2022 11:55:56 AM.

function of genes in the extracted responsive networks corresponding to single drugs and drug combinations, some effective drug combinations were identified. As less prior knowledge about the known drug combinations is utilized, the accuracy of unsupervised methods is always moderate. However, they can predict some new drug combinations with new rules of combination, which are absent in the known drug combinations. On the other hand, the machine learning method proposed by Zhao *et al.*[104] is an example of a supervised method. In this method, various molecular and pharmacological features of the approved drug combinations were mined and used for predicting new drug combinations. As supervised methods make use of prior knowledge about known drug combinations, the accuracy of these methods is always satisfactory. However, potential drug combinations with unseen rules of combination may be missed.

5 Conclusions

The efforts to reveal new roles for old drugs or failed drugs, called drug repositioning, are increasingly attracting attention in the field of pharmacy and medicine. To accelerate these efforts, numerous computational methods have been proposed to reposition drugs rationally in a large scale manner by utilizing available high throughput data. With the promise of deciphering mechanisms underlying disease development and drug intervention, network-based methods separate themselves from other methods by focusing on the interactions or associations of various factors. Given this, we summarized the recent progresses in repositioning drugs by network concepts in this paper.

We firstly unified the single drug repositioning and drug combinations into one framework. Then we summarized and further described most of the available data related to diseases or drugs. Based on cause–effect relationships, these data were roughly classified into three classes, that is, molecular origin, molecular activity, and phenotype. Correspondingly, we categorized several topics regarding repositioning drugs (both single drug repositioning and drug combination identification) into subclasses based on levels of data sources used. Through summarizing the existing works, we showed that these network-based methods hold great potential for deciphering mechanisms underlying complex diseases, mode of action of drugs, and repositioning drugs against diseases.

We also indicated several obstacles that hamper the computational methods for further applications. In most of the existing methods, various interaction data are utilized to construct a background network model. These interaction data are known for high rate of false positive and false negatives and are also both noisy and incomplete, which makes the inferred relationship inaccurate. There are two ways to tackle this problem. On the one hand, additional biological information such as functional annotation could be leveraged for filtering the interaction data. On the other hand, integrating various interaction data can alleviate the problem of the incompleteness of interaction data to some extent.

Since molecular origin, molecular activity, and phenotype can only capture diseases or drug's molecular basis, intermediate effect, and terminal effect respectively, the measured relevances of a disease and drugs based on one of them are partial and biased. Obviously, the direct approach to address this problem is to integrate data of molecular origin, molecular activity, and phenotype together and to characterize disease and drug's effect more comprehensively. However, integrating these data is not trivial because of their complex nature and complex cause–effect relationships. Therefore, it is a promising research direction to develop data integration methods that can handle multiple data levels differing in statistical power and network coverage. It also needs to take cause–effect relationships among them into account.[105] In addition, when the known drug–disease associations are organized into a drug–disease network and further molecular origin, molecular activity, disease phenotype and drugs are mapped onto the drug–disease association network, the task of drug repositioning can be reduced to a link prediction problem or a personalized recommendation problem in complex network theory. Therefore, it is promising to develop methods for drug repositioning that can borrow ideas from the area of complex network research.

Lastly, designing an appropriate measure for evaluating relevance of disease–disease pair, drug–drug pair, and drug–disease pair also plays a key role in developing promising methods for drug repositioning. For example, a good measure should evaluate the overall relevance in the context of the network. Generally, mechanisms underlying diseases and drugs are identified as a responsive subnetwork. In such a sense, network comparison and query techniques could be borrowed and modified to compare responsive networks of diseases and drugs, and further evaluate the relevant mechanism.[106] It is also a future topic to integrate information of not only network but also dynamics[71,107,108] for improving the effectiveness and efficiency of drug repositioning.

Although computational predictions are far from the animal model and clinical trial, still we believe that network-based methods will eventually revolutionize our understanding of the interaction mechanism between diseases and drugs and lead to practical application in drug discovery.

Acknowledgements

We thank Dr Zhiping Liu from Shanghai Institutes for Biological Sciences for helpful discussions. ZW is supported by PhD start funds of University of Shanghai for Science and Technology under Grant No. 1D-10-303-003, Science Foundation for the Youth Scholars by Educational Commission of Shanghai under Grant No. 51-12-303-105 and partially supported by National Natural Science Foundation of China under Grant No. 11171221. YW is supported by National Natural Science Foundation of China (NSFC) under Grants No. 61171007 and No. 11131009. LC is supported by NSFC under Grants No. 91029301, No. 61134013 and No. 61072149; and by the Chief Scientist Program of Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences with Grant No. 2009CSP002;

 This journal is © The Royal Society of Chemistry 2013

Published on 26 February 2013. Downloaded by Technical University of Munich on 12/1/2022 11:55:56 AM.

by Shanghai Pujiang Program; and by the FIRST program from JSPS initiated by CSTP. The authors thank the reviewers for their helpful comments.

References

1 A. Mullard, *Nat. Rev. Drug Discovery*, 2012, **11**, 91–94.
2 T. T. Ashburn and K. B. Thor, *Nat. Rev. Drug Discovery*, 2004, **3**, 673–683.
3 J. T. Dudley, T. Deshpande and A. J. Butte, *Briefings Bioinf.*, 2011, **12**, 303–311.
4 J. K. Aronson, *Br. J. Clin. Pharmacol.*, 2007, **64**, 563–565.
5 M. Sirota, J. T. Dudley, J. Kim, A. P. Chiang, A. A. Morgan, A. Sweet-Cordero, J. Sage and A. J. Butte, *Sci. Transl. Med.*, 2011, **3**, 96ra77.
6 D. Sardana, C. Zhu, M. Zhang, R. C. Gudivara, L. Yang and A. G. Jegga, *Briefings Bioinf.*, 2011, **12**, 346–356.
7 J. Jia, F. Zhu, X. Ma, Z. Cao, Y. Li and Y. Chen, *Nat. Rev. Drug Discovery*, 2009, **8**, 111–128.
8 S. J. Weir, L. J. DeGennaro and C. P. Austin, *Cancer Res.*, 2012, **72**, 1055–1058.
9 A. Mullard, *Nat. Rev. Drug Discovery*, 2012, **11**, 505–506.
10 Y. Wang, X.-S. Zhang and L. Chen, *BMC Syst. Biol.*, 2012, **6**, S1.
11 J. P. Overington, B. Al-Lazikani and A. L. Hopkins, *Nat. Rev. Drug Discovery*, 2006, **5**, 993–996.
12 K. Henriksen, C. Christiansen and M. A. Karsda, *Drug Discovery Today*, 2011, **16**, 967–975.
13 Y. Wang, C. Zhang, N. Deng and Y. Wang, *Comput. Biol. Chem.*, 2011, **35**, 353–362.
14 Y. Wang, Z. Yang, Y. Wang and N. Deng, *Lett. Drug Des. Discovery*, 2010, **7**, 370–378.
15 A.-L. Barabási and Z. N. Oltvai, *Nat. Rev. Genet.*, 2004, **5**, 101–113.
16 L. Chen, R. Wang and X. Zhang, *Biomolecular networks: methods and applications in systems biology*, John Wiley & Sons Inc, New Jersey, 1st edn, 2009.
17 Z. Liu, Y. Wang, X. Zhang and L. Chen, *IET Syst. Biol.*, 2012, **6**, 22–33.
18 L. Chen, R. Wang, C. Li and K. Aihara, *Modeling biomolecular networks in cells: structures and dynamics*, Springer Verlag, London, 1st edn, 2010.
19 Y. Wang, X. Zhang and L. Chen, *BMC Syst. Biol.*, 2010, **4**, S1.
20 Y. Wang and Y. Xia, *Lect. Notes Oper. Res.*, 2008, **9**, 333–340.
21 A. Hamosh, A. F. Scott, J. S. Amberger, C. A. Bocchini and V. A. McKusick, *Nucleic Acids Res.*, 2005, **33**, D514–D517.
22 Q. Jiang, Y. Wang, Y. Hao, L. Juan, M. Teng, X. Zhang, M. Li, G. Wang and Y. Liu, *Nucleic Acids Res.*, 2009, **37**, D98–D104.
23 M. Kuhn, D. Szklarczyk, A. Franceschini, M. Campillos, C. von Mering, L. J. Jensen, A. Beyer and P. Bork, *Nucleic Acids Res.*, 2010, **38**, D552–D556.
24 T. S. K. Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D. S. Somanathan, A. Sebastian, S. Rani, S. Ray, C. J. H. Kishore, S. Kanth, M. Ahmed, M. K. Kashyap, R. Mohmood, Y. L. Ramachandra, V. Krishna, B. A. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady and A. Pandey, *Nucleic Acids Res.*, 2009, **37**, D767–D772.
25 C. Stark, B.-J. Breitkreutz, A. Chatraryamontri, L. Boucher, R. Oughtred, M. S. Livstone, J. Nixon, K. V. Auken, X. Wang, X. Shi, T. Reguly, J. M. Rust, A. Winter, K. Dolinski and M. Tyers, *Nucleic Acids Res.*, 2011, **39**, D698–D704.
26 L. Salwinski, C. S. Miller, A. J. Smith, F. K. Pettit, J. U. Bowie and D. Eisenberg, *Nucleic Acids Res.*, 2004, **32**, D449–D451.
27 F. Zhao, Z. Xuan, L. Liu and M. Q. Zhang, *Nucleic Acids Res.*, 2005, **33**, D103–D107.
28 V. Matys, E. Fricke, R. Geffers, E. Gößling, M. Haubrock, R. Hehl, K. Hornischer, D. Karas, A. E. Kel, O. V. Kel-Margoulis, D. U. Kloos, S. Land, B. Lewicki-Potapov, H. Michael, R. Münch, I. Reuter, S. Rotert, H. Saxel, M. Scheer, S. Thiele and E. Wingender, *Nucleic Acids Res.*, 2003, **31**, D374–D378.
29 A. Kozomara and S. Griffiths-Jones, *Nucleic Acids Res.*, 2011, **39**, D152–D157.
30 B. P. Lewis, C. B. Burge and D. P. Bartel, *Cell*, 2005, **120**, 15–20.
31 Q. Cui, Y. Ma, M. Jaramillo, H. Bari, A. Awan, S. Yang, S. Zhang, L. Liu, M. Lu, M. O'Connor-McCourt, E. O. Purisima and E. Wang, *Mol. Syst. Biol.*, 2007, **3**, 152.
32 M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock, *Nat. Genet.*, 2000, **25**, 25–29.
33 J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J.-P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander and T. R. Golub, *Science*, 2006, **313**, 1929–1935.
34 H. Parkinson, U. Sarkans, N. Kolesnikov, N. Abeygunawardena, T. Burdett, M. Dylag, I. Emam, A. Farne, E. Hastings, E. Holloway, N. Kurbatova, M. Lukk, J. Malone, R. Mani, E. Pilicheva, G. Rustici, A. Sharma, E. Williams, T. Adamusiak, M. Brandizi, N. Sklyar and A. Brazma, *Nucleic Acids Res.*, 2011, **39**, D1002–D1004.
35 T. Barrett, T. O. Suzek, D. B. Troup, S. E. Wilhite, W.-C. Ngau, P. Ledoux, D. Rudnev, A. E. Lash, W. Fujibuchi and R. Edgar, *Nucleic Acids Res.*, 2005, **33**, D562–D566.
36 Y. Wang, J. Xiao, T. O. Suzek, J. Zhang, J. Wang and S. H. Bryant, *Nucleic Acids Res.*, 2009, **37**, W623–W633.
37 M. Kuhn, M. Campillos, I. Letunic, L. J. Jensen and P. Bork, *Mol. Syst. Biol.*, 2010, **6**, 343.
38 R. Judson, A. Richard, D. Dix, K. Houck, F. Elloumi, M. Martin, T. Cathey, T. R. Transue, R. Spencer and M. Wolf, *Toxicol. Appl. Pharmacol.*, 2008, **233**, 7–13.

Published on 26 February 2013. Downloaded by Technical University of Munich on 12/1/2022 11:55:56 AM.

39 Y. Liu, B. Hu, C. Fu and X. Chen, *Bioinformatics*, 2010, **26**, 587–588.
40 H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono and M. Kanehisa, *Nucleic Acids Res.*, 1999, **27**, 29–34.
41 C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, Y. Djoumbou, R. Eisner, A. C. Guo and D. S. Wishart, *Nucleic Acids Res.*, 2011, **39**, D1035–D1041.
42 A. P. Davis, B. L. King, S. Mockus, C. G. Murphy, C. Saraceni-Richards, M. Rosenstein, T. Wiegers and C. J. Mattingly, *Nucleic Acids Res.*, 2011, **39**, D1067–D1072.
43 M. Hewett, D. E. Oliver, D. L. Rubin, K. L. Easton, J. M. Stuart, R. B. Altman and T. E. Klein, *Nucleic Acids Res.*, 2002, **30**, 163–165.
44 R. Huang, N. Southall, Y. Wang, A. Yasgar, P. Shinn, A. Jadhav, D.-T. Nguyen and C. P. Austin, *Sci. Transl. Med.*, 2011, **3**, 80ps16.
45 K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal and A.-L. Barabási, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 8685–8690.
46 J. Park, D.-S. Lee, N. A. Christakis and A.-L. Barabási, *Mol. Syst. Biol.*, 2009, **5**, 262.
47 X. Wu, R. Jiang, M. Q. Zhang and S. Li, *Mol. Syst. Biol.*, 2008, **4**, 189.
48 B. Jiang, J. Wang, J. Xiao and Y. Wang, *Lect. Notes Oper. Res.*, 2009, **11**, 319–328.
49 F. Cheng, C. Liu, J. Jiang, W. Lu, W. Li, G. Liu, W. Zhou, J. Huang and Y. Tang, *PLoS Comput. Biol.*, 2012, **8**, e1002503.
50 X. Chen, M. Liu and G. Yan, *Mol. BioSyst.*, 2012, **8**, 1970–1978.
51 S.-H. Yeh, H.-Y. Yeh and V.-W. Soo, *J. Clin. Bioinf.*, 2012, **8**, 1.
52 H. Luo, J. Chen, L. Shi, M. Mikailov, H. Zhu, K. Wang, L. He and L. Yang, *Nucleic Acids Res.*, 2011, **39**, W492–W498.
53 A. Edberg, D. Soeria-Atmadja, J. B. Laurila, F. Johansson, M. G. Gustafsson and U. Hammerling, *J. Chem. Inf. Model.*, 2011, **52**, 1238–1249.
54 H. Ye, L. Yang, Z. Cao, K. Tang and Y. Li, *Chin. Sci. Bull.*, 2012, **57**, 2106–2112.
55 H. S. Lee, T. Bae, J.-H. Lee, D. G. Kim, Y. S. Oh, Y. Jang, J.-T. kim, J.-J. Lee, A. Innocenti, C. T. Supuran, L. Chen, K. Rho and S. Kim, *BMC Syst. Biol.*, 2012, **6**, 80.
56 S. Daminelli, V. J. Haupt, M. Reimann and M. Schroeder, *Integr. Biol.*, 2012, **4**, 778–788.
57 S. Zhao and S. Li, *Bioinformatics*, 2012, **28**, 955–961.
58 M. J. Keiser, V. Setola, J. J. I. C. Laggner, A. I. Abbas, S. J. Hufeisen, N. H. Jensen, M. B. Kuijer, R. C. Matos, T. B. Tran, R. Whaley, R. A. G. J. Hert, K. L. H. Thomas, D. D. Edwards, B. K. Shoichet and B. L. Roth, *Nature*, 2009, **462**, 175–181.
59 F. Iorio, R. Tagliaferri and D. D. Bernardo, *J. Comput. Biol.*, 2009, **16**, 241–251.
60 F. Iorio, R. Bosotti, E. Scacheri, V. Belcastro, P. Mithbaokar, R. Ferriero, L. Murino, R. Tagliaferri, N. Brunetti-Pierria, A. Isacchi and D. di Bernardo, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 14621–14626.
61 J. T. Dudley, M. Sirota, M. Shenoy, R. K. Pai, S. Roedder, A. P. Chiang, A. A. Morgan, M. M. Sarwal, P. J. Pasricha and A. J. Butte, *Sci. Transl. Med.*, 2011, **3**, 96ra76.
62 M. Sirota, J. T. Dudley, J. Kim, A. P. Chiang, A. A. Morgan, A. Sweet-Cordero, J. Sage and A. J. Butte, *Sci. Transl. Med.*, 2011, **3**, 96ra77.
63 G. Hu and P. Agarwal, *PLoS One*, 2009, **4**, e6536.
64 D. Shigemizu, Z. Hu, J.-H. Hung, C.-L. Huang, Y. Wang and C. Delisi, *PLoS Comput. Biol.*, 2012, **8**, e1002347.
65 Y. Qiu, S. Zhang, X. Zhang and L. Chen, *BMC Bioinformatics*, 2010, **11**, 26.
66 Y. Qiu, X. Zhang and L. Chen, *IET Syst. Biol.*, 2008, **3**, 475–486.
67 Z. Wen, Z. Liu, Y. Yan, G. Piao, Z. Liu, J. Wu and L. Chen, *PLoS One*, 2012, **7**, 7.
68 Z. Wen, Z. Liu, Z. Liu, Y. Zhang and L. Chen, *J. Am. Med. Informat. Assoc.*, 2012, DOI: 10.1136/amiajnl-2012-001168.
69 W. Song, J. Wang, Y. Yang, N.-H. Jing, X.-S. Zhang, L. Chen and J.-R. Wu, *J. Mol. Cell Biol.*, 2012, **4**, 197–206.
70 D. He, Z. Liu, M. Honda, S. Kaneko and L. Chen, *J. Mol. Cell Biol.*, 2012, **4**, 140–152.
71 L. Chen, R. Liu, Z. Liu, M. Li and K. Aihara, *Sci. Rep.*, 2012, **2**, 342.
72 Z. Wu, X. Zhao and L. Chen, *Mol. Cells*, 2009, **27**, 271–277.
73 A. Mitsos, I. N. Melas, P. Siminelakis, A. D. Chairakaki, J. Saez-Rodriguez and L. G. Alexopoulos, *PLoS Comput. Biol.*, 2009, **5**, e1000591.
74 Z. Wu, X. Zhao and L. Chen, *BMC Syst. Biol.*, 2010, **4**, S7.
75 Y. Silberberg, A. Gottlieb, M. Kupiec, E. Ruppin and R. Sharan, *J. Comput. Biol.*, 2012, **19**, 163–174.
76 N. Yosef, L. Ungar, E. Zalckvar, A. Kimchi, M. Kupiec, E. Ruppin and R. Sharan, *Mol. Syst. Biol.*, 2009, **5**, 248.
77 G. Jin, C. Fu, H. Zhao, K. Cui, J. Chang and S. T. C. Chang, *Cancer Res.*, 2012, **72**, 33–44.
78 E. Lounkine, M. J. Keiser, S. Whitebread, D. Mikhailov, J. Hamon, J. L. Jenkins, P. Lavan, E. Weber, A. K. Doak, S. Côté, B. K. Shoichet and L. Urban, *Nature*, 2012, **486**, 361–367.
79 L.-C. Huang, X. Wu and J. Y. Chen, *BMC Bioinformatics*, 2011, **12**, S11.
80 R. L. Chang, L. Xie, L. Xie, P. E. Bourne and B. Ø. Palsson, *PLoS Comput. Biol.*, 2010, **6**, e1000938.
81 N. Atias and R. Sharan, *J. Comput. Biol.*, 2011, **18**, 207–218.
82 L. Yang and P. Agarwal, *PLoS One*, 2009, **6**, e1000938.
83 J. von Eichborn, M. S. Murgueitio, M. Dunkel, S. Koerner, P. E. Bourne and R. Preissner, *Nucleic Acids Res.*, 2011, **39**, D1060–D1066.
84 M. Re and G. Valentini, *Lect. Notes Comput. Sci.*, 2012, **7292**, 225–236.
85 A. Gottlieb, G. Y. Stein, E. Ruppin and R. Sharan, *Mol. Syst. Biol.*, 2011, **7**, 496.
86 M. A. Yildirim, K.-I. Goh, M. E. Cusick, A.-L. Barabási and M. Vidal, *Nat. Biotechnol.*, 2007, **25**, 1119–1126.

 This journal is © The Royal Society of Chemistry 2013

Published on 26 February 2013. Downloaded by Technical University of Munich on 12/1/2022 11:55:56 AM.

87 J. Wang, Z. Li, C. Qiu, D. Wang and Q. Cui, *Briefings Bioinf.*, 2012, **13**, 377–382.
88 K. Xu, J. Song and X. Zhao, *Lect. Notes Oper. Res.*, 2010, **13**, 347–354.
89 Y. Wang, K. Xu, J. Song and X. Zhao, *BMC Bioinformatics*, 2012, **13**, S7.
90 S. Li, B. Zhang, D. Jiang, Y. Wei and N. Zhang, *BMC Bioinformatics*, 2010, **11**, S6.
91 S. Li, B. Zhang and N. Zhang, *BMC Syst. Biol.*, 2011, **5**, S10.
92 L. Wang, G. Zhou, P. Liu, J. Song, Y. Liang, X. Yan, F. Xu, B. Wang, J. Mao, Z. Shen, S. Chen and Z. Chen, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 4826–4831.
93 Q. Zhang, J. Mao, P. Liu, Q. Huang, J. Lu, Y. Xie, L. Weng, Y. Zhang, Q. Chen, S. Chen and Z. Chen, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 3378–3383.
94 N. Geva-Zatorsky, E. Dekel, A. A. Cohen, T. Danon, L. Cohen and U. Alon, *Cell*, 2010, **140**, 643–651.
95 S. Nelander, W. Wang, B. Nilsson, Q. She, C. Pratilas, N. Rosen, P. Gennemark and C. Sander, *Mol. Syst. Biol.*, 2008, **4**, 216.
96 G. Jin, H. Zhao, X. Zhou and S. T. C. Wong, *Bioinformatics*, 2011, **27**, i310–i316.
97 S. Iadevaia, Y. Lu, F. C. Morales, G. B. Mills and P. T. Ram, *Cancer Res.*, 2010, **70**, 6704–6714.
98 S. Loewe, *Ergeb. Physiol.*, 1928, **27**, 47–187.
99 C. I. Bliss, *Ann. Appl. Biol.*, 1939, **26**, 585–615.
100 T.-C. Chou and P. Talalay, *Trends Pharmacol. Sci.*, 1983, **4**, 450–454.
101 P. K. Wong, F. Yu, A. Shahangian, G. Cheng, R. Sun and C.-M. Ho, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 5105–5110.
102 D. Calzolari, S. Bruschi, L. Coquin, J. Schofield, J. D. Feala, J. C. Reed, A. D. McCulloch and G. Paternostro, *PLoS Comput. Biol.*, 2008, **4**, e1000249.
103 K. Xu, J. Song and X. Zhao, *BMC Syst. Biol.*, 2012, **6**, S5.
104 X. Zhao, M. Iskar, G. Zeller, M. Kuhn, V. van Noort and P. Bork, *PLoS Comput. Biol.*, 2011, **7**, e1002323.
105 Y. Wang, X. Zhang and L. Chen, *OMICS*, 2009, **13**, 313–324.
106 S. Zhang, G. Jin, X. Zhang and L. Chen, *Proteomics*, 2007, **7**, 2856–2869.
107 R. Liu, M. Li, Z. Liu, J. Wu, L. Chen and K. Aihara, *Sci. Rep.*, 2012, **2**, 813.
108 J. Wang, Y. Sun, S. Zheng, X.-S. Zhang, H. Zhou and L. Chen, *Sci. Rep.*, 2013, **3**, 1097.