Expert Opinion

Building a drug–target network and its applications

Soyoung Lee, Keunwan Park & Dongsup Kim†
KAIST, Department of Bio and Brain Engineering, 335 Gwahak-ro, Yuseong-gu, Daejeon, 305-701 Korea, Republic of Korea

Background: One of the most recent and important developments in drug discovery is a new drug development approach of building and analyzing networks that contain relationships among drugs and targets, diseases, genes and other components. These networks and their integrations provide useful information for finding new targets as well as new drugs. *Objective*: This review article aims to review recent developments in various types of networks and suggest the future direction of these network studies for drug discovery. *Methods*: Databases and networks are integrated into a more complete network to better present the relationships among drugs, targets, genes, phenotypes and diseases. After discussing the limitations and obstacles of the recent research, we suggest several strategies to build a successful and practical drug–target network. *Results/conclusion*: A useful, integrated network can be built from various databases and networks by resolving several issues, such as limited coverage and inconsistency. This integrated network can be completed by the prediction of missing links, biological network comparison and drug target identification. Possible applications are multi-target drug development, drug repurposing, estimation of drug effect on target perturbations in the whole system and extraction of the suitable purpose of the drug–target sub-network.

Keywords: drug–disease network, drug–target network, network analysis, network integration, target identification

Expert Opin. Drug Discov. (2009) 4(11):1-13

1. Introduction

Due to the exponential increase in biological, chemical and interaction data by advanced high-throughput experimental techniques, there is an increasing demand on integrating these data for developing a new drug in a more efficient manner. One such example is a drug–target network [1-5] that helps to find both new drug targets and effective new drugs at a lower cost. Along with the construction of a drug–target network, other types of networks, such as drug–disease networks, drug–side effect networks and drug–molecular fragment networks, have been developed by various researchers [6,7]. Similar to these types of graphs are drug–drug networks. For instance, the use of quantitative structure–activity relationship models to construct drug–drug networks of antiviral drugs [8] and anti-fungal compounds [9]. The links in these networks connect drug–drug pairs, and the information contained in the link is related to the activity of the drug for different targets. In this sense, the networks mentioned above are drug–target networks as well.

These networks can be used for the following: i) to construct the whole network by inferring missing links from the information of known links; ii) to find new drug targets; iii) to infer potential side effects from the network; iv) to reposition existing drugs; v) to design multi-target drugs that interact with only effective targets; vi) to design an effective drug combination that can maximize the efficacy of targeting disease; and vii) to find new relationships among disease, treatment, patients, targets, drugs and genes.

informa
healthcare

We start our review by introducing the 'central dogma of drug action', shown in **Figure 1A**. It states that a single, or multiple drugs, interact with appropriate single or multiple target protein(s); thus, as a result disease can be modulated. Based upon these steps of drug action, it is clear that not only information on individual biological components (such as drugs, genes, proteins, cell types and diseases), but also information on the component relationships is essential to understanding the whole process.

Undoubtedly, the most important aspect of the entire drug-action process is the drug-target–disease relationship. However, it is also critical to have an appropriate understanding of phenotypic context during this process. There are four distinctive types of biological states that determine a specific phenotypic context: genomic, proteomic, metabolomic and cellular states. **Figure 1B** shows the relationships between drugs and disease in terms of phenotypic context. Genomic context specifies a particular genomic state, including epigenetic state and genetic variations that govern the gene expression level of all genes. Proteomic context is determined by the expression level of all proteins, including different isoforms with specific post-translational modification. Metabolomics is the study of small-molecule metabolite profiles and the collection of all metabolites in a biological organism. Cellular context is a defined cellular state of a specific tissue type at a particular disease state. The perturbation of disease state by a drug can be detected by genetic profiling data; this perturbation changes the disease state through genetic pathways. Most drugs also affect disease state by interacting with proteins. The primary effect of this interaction is to change a specific disease state through protein pathways by modulating disease-related proteins (target proteins). However, modulation of adverse drug reaction (ADR)-related proteins can cause unexpected drug–protein interactions. Information regarding drug action in a cell is useful for testing drug efficacy and toxicity. The real drug action in the human body can be predicted from this *in vitro* data using *in vivo–in vitro* correlation techniques.

In this article, we aim to review useful biological and chemical databases for building complete and effective drug–target networks, in addition to recent work pioneering the modeling of drug–target networks and other drug-related networks. Based on the drug–disease relationship diagram (**Figure 1**), we introduce databases for each component and each relationship. Drug–target networks and other drug-related networks can be built from the interaction data. After introducing current research, we discuss limitations of this work and the future direction of drug–target network research.

2. Databases to build drug–target and other drug-related networks

A network consists of nodes and links. To create drug-related networks, it is necessary to collect information on drugs, genes, protein targets and diseases as nodes and their interactions as links. **Table 1** summarizes several representative databases.

2.1 Databases for drugs, proteins, phenotypes and diseases

Information for chemicals can be obtained from various sources. The representative databases are PubChem, DrugBank, ChemBank and KEGG LIGAND. PubChem [10] provides information on chemical structures and their biological activities. Using PubChem, one can find information on chemical probes discovered by high-throughput screening of small molecules that modulate the activity of gene products. DrugBank [11] is a knowledge-based database that combines detailed drug data with comprehensive drug action and drug target information. It contains nearly 4800 drug entries including > 1350 FDA-approved small molecule drugs, 123 FDA-approved biotech drugs, 71 nutraceuticals and > 3243 experimental drugs. ChemBank [12] and KEGG [13] LIGAND also contain public information about chemical substances, reactions and other chemoinformatics resources.

Drug targets are typically proteins whose activities are modulated by specific chemicals. The Protein Data Bank (PDB) and the SWISS-PROT are the most widely used databases for the structures and sequences of proteins, respectively. PDB [14] was established in 1971 and it contains an archive of information about experimentally determined 3D structures of proteins, nucleic acids and complex assemblies. SWISS-PROT [15] is a manually curated protein sequence database that has been valued for its high quality annotation and the use of standardized nomenclature.

In addition to the primary sequence and structure databases, there are several databases that provide information on specific proteins that interact with drugs, target proteins and diseases. For example, TTD [16] provides information about the known and explored therapeutic protein and nucleic acid targets, the targeted diseases, pathway information and the corresponding drugs/ligands. TargetDB [17] provides the status information on target sequences and tracks their progress through the various stages of protein production and structure determination. PDTD [18] is a database containing current and potential drug targets with known 3D structures; it contains 1207 entries covering 841 drug targets. SuperTarget [19] integrates drug-related information about medical indication, adverse drug effects, drug metabolism, pathways and Gene Ontology terms of the target proteins. TPDB [20] is a comprehensive, curated, searchable, documented compilation of publicly available information on the protein targets of reactive metabolites of 18 well-studied chemicals and drugs of known toxicity. Although it seems relatively small, it is a good example of how other databases should be designed in the future.

Other databases cover information on specific proteins that induce ADRs. DITOP [21] includes information on drug-induced toxicity related proteins (DITRPs) that mediate toxicity through their interactions with drugs or reactive metabolites. Currently, it contains 618 literature-reported DITRPs, 529 drugs and 418 toxicities. DART [22] provides comprehensive information about adverse effects of drugs

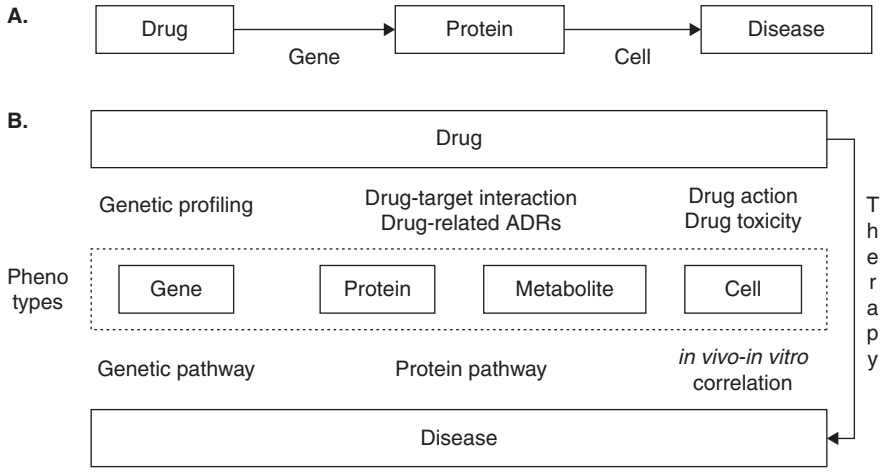

Figure 1. Relationships among drugs, disease and phenotypic contexts. A. The central dogma of drug action. **B.** The drug–disease relationship with phenotypic context that includes genomic, proteomic and cellular states.
ADR: Adverse drug reaction.

described in the literature and potential targets involved in ADR that are not yet confirmed. It gives the physiological function of each target and corresponding ADRs induced by drug–target binding. MDL toxicity database [23] is serviced by MDL Information Systems, Inc. and it offers chemical structure-based access to data on > 150,000 known toxic substances. Although this system provides the potential toxicological profile of drug candidates under investigation in a very large scale, it is not public.

Genetic profiling and cellular experimental results are useful phenotypes for building networks, such as drug–gene networks and disease-related gene networks. The Connectivity Map project by the Broad institute created a reference collection of gene-expression profiles on human cell lines treated with bioactive small molecules [24]. It contains > 7000 expression profiles representing 1309 compounds. In addition, Gene Expression Omnibus (GEO) [25], ArrayExpress [26] and NCI60 [27,28] provide genetic profiles of microarray experiments. GEO and ArrayExpress contain abundant gene expression data under various biological conditions, while NCI60 is a specified archive for screens of tumors. The NCI60 database offers both gene expression profiles and drug activity patterns against 60 human cancer cell lines, which contain both genetic and cellular profiles. It can be used to build both individual networks (i.e., drug–gene networks and drug–cell networks) and an integrated network from various individual networks.

Online Mendelian Inheritance in Man (OMIM) [29] is a comprehensive and authoritative compendium of human genes and disease. It contains information on all known mendelian disorders for > 12,000 genes.

2.2 Databases for the interactions among drugs, proteins and diseases

In a drug–target network, biological components (i.e., drugs, proteins, genes and diseases) are represented by nodes; it is necessary to link nodes by edges based on the relationships between them. Several databases offer information on these interactions on a large-scale. DrugBank and TTD provide a list of drugs targeting proteins with information on related diseases. SuperTarget provides a list of target proteins found by searching the database with the information of a query drug (i.e., adverse drug effect and structural similarity). STITCH [30] is a resource to explore known and predicted interactions of chemicals and proteins (interaction networks). It contains interactions for > 68,000 chemicals and > 1.5 million proteins in 373 species. MATADOR [19] is a resource for protein–chemical interactions. It differs from other resources, such as DrugBank and TTD (which usually contain only the main mode of interactions), by its inclusion of many direct and indirect interactions.

Other databases of drug–target interactions offer interaction strength information quantitatively based on the experimental binding data. The PubChem BioAssay database contains bioactivity screens (i.e., percentage of activity inhibition) of chemical substances described in PubChem. It currently contains > 1400 bioassay depositions and 45 million biological activity outcomes for > 700,000 compounds. The Psychoactive Drug Screening Program Ki database [31] provides screening of novel psychoactive compounds for pharmacological and functional activity at cloned human or rodent CNS receptors, channels and transporters. BindingDB [32] is a database of experimentally determined protein–ligand binding affinities that provides results of various binding assays. In addition, the Biological General Repository for Interaction Datasets database provides 167,660 non-redundant protein and genetic interactions from 22 model organisms, along with many drug–gene interactions [33].

The interaction among off-target, non-therapeutic proteins and drugs may cause ADRs. Previously mentioned databases on ADR-related proteins include the DITOP, DART and

Table 1. Representative databases related with drug–target network.

Information type	Database name	Content	Website	Coverage
Drug	PubChem	Structures and activities of chemicals	http://pubchem.ncbi.nlm.nih.gov	> 10 million chemicals
	DrugBank	Drug and target information	http://www.drugank.ca	4800 drugs
	ChemBank	Structures and activities of chemicals	http://chembank.broadinstitute.org	> 1.2 million chemicals
	KEGG LIGAND	Structures and reactions of chemicals	http://www.genome.jp/ligand	15,790 chemicals
Protein	PDB	Structures of proteins	http://www.rcsb.org/pdb	54,428 proteins
	SWISS-PROT	Sequences of proteins	http://www.ebi.ac.uk/swissprot	470,369 entries
Target protein	TTD	Target, disease, pathway and corresponding drugs	http://xin.cz3.nus.edu.sg/group/ttd	1535 targets, 2107 drug/ligands
	TargetDB	Target registration (experimental progress)	http://targetdb.pdb.org	14,200 targets (human)
	PDTD	Structures of potential drug targets	http://www.dddc.ac.cn/pdtd	841 potential targets
	SuperTarget	Drug targets, metabolization and pathway	http://insilico.charite.de/supertarget	> 2500 proteins, 1500 drugs
	TPDB	Reactive metabolite target proteins	http://tpdb.medchem.ku.edu/tpdb.html	13 human targets, 32 chemicals
ADR protein	DITOP	Drug-induced toxicity related proteins	http://bioinf.xmu.edu.cn/databases/ADR	618 proteins, 529 drugs/ligands
	DART	Adverse effect targets of drugs	http://xin.cz3.nus.edu.sg/group/drt	236 proteins, 1327 drugs/ligands
	MDL toxicity	*In vitro* and *in vivo* data of toxicity (commercial)	http://www.symyx.com	150,000 chemicals
Drug–target	TTD	Drugs and related diseases of a target	http://xin.cz3.nus.edu.sg/group/ttd	1535 targets, 2107 drug/ligands
	SuperTarget	Targets of a drug	http://insilico.charite.de/supertarget	> 2500 proteins, 1500 drugs
	STITCH	Known and predicted interactions of drug–target	http://stitch.embl.de	> 1.5 million proteins, 68,000 chemicals
	MATADOR	Manually annotated target–drug interactions	http://matador.embl.de	2901 proteins, 801 drugs
	PubChem BioAssay	Bioactivity screens of chemical substances	http://pubchem.ncbi.nlm.nih.gov	> 45 million activities, 700,000 chemicals
	PDSP Ki DB	Pyschoactive drug screens	http://pdsp.med.unc.edu	47,310 K_i values
	BindingDB	Binding affinities of protein-chemicals	http://www.bindingdb.org	28,112 chemicals
	BioGRID	Protein, drug and genetic interactions	http://www.thebiogrid.org	167,660 interactions
Genetic profile	Connectivity Map	Gene expression, diseases and bioactive small molecules	http://www.broadinstitute.org/cmap	> 7000 expression profiles, 1309 compounds
	GEO	Gene expression/molecular abundance	http://www.ncbi.nlm.nih.gov/geo	404,071 samples (RNA, genome, protein)
	ArrayExpress	Gene expression under biological conditions	http://www.ebi.ac.uk/microarray-as/ae	8521 experiments, 246,071 assays
	NCI60		http://genome-www.stanford.edu/nci60	> 70,000 chemicals, > 8000 genes, 60 cells

ADR: Adverse drug reaction.

Table 1. Representative databases related with drug–target network (continued).

Information type	Database name	Content	Website	Coverage
		Gene expression and drug activity against 60 cancer cells		
	OMIM	Relationships of genes and diseases	http://www.ncbi.nlm.nih.gov/omim	2239 genes, 3770 diseases
Cellular profile	NCI60	Gene expression and drug activity against 60 cancer cells	http://genome-www.stanford.edu/nci60	> 70,000 chemicals, > 8000 genes, 60 cells
Pathway	KEGG	Pathways of biological substances and xenobiotics	http://www.genome.jp/kegg	93,318 pathways
	MetaCyc	Non-redundant experimental metabolic pathways	http://metacyc.org	12,000 pathways

ADR: Adverse drug reaction.

MDL toxicity databases; these databases can be used to infer the relationship of a drug and related ADRs.

The KEGG PATHWAY is a collection of manually drawn pathway maps of the molecular interaction and reaction networks for metabolism of various biological substances and xenobiotics, genetic information processing, environmental information processing, cellular processes and human diseases. MetaCyc [34] is a database of non-redundant, experimentally elucidated metabolic pathways. It contains > 1200 pathways from > 1600 different organisms involved in both primary and secondary metabolism.

3. Drug–target networks and other drug-related networks

3.1 Drug–target networks

Recently, there have been several pioneering studies to build drug–target networks. Paolini *et al.* [35] presented global mapping of pharmacological space by integrating several sources of medicinal chemistry structure–activity relationship data. They built a human polypharmacology interaction network representing relationships between proteins in chemical space by using binding data from Pfizer. In addition, they defined three types of indexes to measure human protein promiscuity. In the following year, Yildirim *et al.* [2] built a bipartite graph of drug–protein interactions from FDA-approved drugs and their target proteins. In this graph, a drug and protein are connected to each other if the protein is a known target of the drug (drug–target network). In addition, they built a human disease network generated from OMIM-based disorder–disease gene associations. Yamanishi *et al.* [3] built four classes of drug–target interaction networks in humans involving enzymes, ion channels, GPCRs and nuclear receptors. It is notable that they linked unknown interactions between chemical/genomic and pharmacological space from known information by developing a computational algorithm. In an attempt to build complete networks, they used known data in combination with a prediction algorithm. Yamanishi *et al.* used the predictions to fill missing links and they used predicted pairs to build a complete network. Keiser *et al.* [36] developed a method that quantitatively groups and relates proteins based on the chemical similarity of their ligands. They defined a similarity score between ligand sets, and thus built a similarity network for 246 enzymes and receptors by using the scores between ligand sets of enzymes and receptors. In 2009, Cases and Mestres [37] defined the complete cardiovascular target space that contains both therapeutic targets and off-targets by integrating previous knowledge on cardiovascular targets and 44,032 small molecules. They built a ligand–protein interaction map between the 14,734 scaffolds from all molecules annotated to cardiovascular targets and the 581 cardiovascular-relevant proteins. Although their map covers a small region of the whole pharmacological space, their work is useful because of the strong potential application for drug discovery. Recently, Vina *et al.* [4] built a drug–target network by using predicted drug–target pairs based on drug connectivity and protein receptor sequences. While most previous work for predicting target proteins of a drug found only a single target, their algorithm predicted more than one target.

Nagamine *et al.* [38,39] made predictions for protein–chemical interactions by using protein data (i.e., amino-acid sequences) in conjunction with chemical structures and mass spectrometry data. Their work [38,39] in 2009 improved prediction accuracy by proposing a strategy to repeatedly feed back experimental results into computational predictions with consideration of biological effects of interest. Recently, Xie *et al.* [40] introduced a novel computational strategy to identify protein–ligand binding profiles on a genome-wide scale. Furthermore, they applied this technique to elucidate the

molecular mechanisms of adverse drug effects. They built a Cholesteryl Ester Transfer Protein (CETP) off-target binding network and inferred the mechanisms of adverse drug effects of CETP inhibitors. SuperPred [41], developed by Dunkel *et al*. in 2008, predicts potential targets of a query chemical structure based on the chemical similarities of predicted Anatomic Therapeutic Chemical (ATC)-codes using drug–target interactions and pathway information. Different from Paolini's and Campillos's target identifiers, Li and Lai [42] predicted potential drug targets based on simple sequence properties and not on the known drug–target interaction information alone. Their identifier can be used to find potential targets, and thus to complete the missing nodes of networks instead of the missing links.

3.2 Drug–ADR target networks

The information on side effects can be used to build a drug–target network and an off-target network. Under the assumption that drugs with similar *in vitro* protein binding profiles tend to cause similar side effects, Campillos *et al*. [7] developed an identifier of drug targets using side effect similarity. They used phenotypic side effect similarities to infer whether two drugs share a target, and thus built a network of drugs predicted to have common protein targets to identify target proteins. Xie *et al*. [40] applied their computational tool of identification of protein–ligand binding profiles to build an off-target binding network of CETP as mentioned before.

3.3 Disease and treatment networks

Instead of focusing on target and off-target proteins, many researchers have attempted to build networks related to drugs and treatment. One such example, the human disease network [6], was built by Goh *et al*. They classified each disorder into one of 22 classes based on the physiological system. From the diseasome bipartite graph, they also generated two biologically relevant network projections: human disease network on the genome space and disease gene network on the disease space. Another similar example is the therapy network [43] developed by Nacher and Schwartz. They investigated the human network corresponding to interactions between FDA-approved drugs and human therapies. Nacher and Schwartz defined five bipartite graphs whose nodes can be classified into two disjoint sets of drugs and therapies at five individual hierarchical levels of ATC classification. Yildirim *et al*. also built a human disease network generated from OMIM-based disorder–disease gene associations [2].

3.4 Problems of network integration

In order to be useful, drug–target networks should be combined with many diverse biological information sources and then properly analyzed. To understand more realistic *in vivo* dynamics of numerous components in cells, many different kinds of interactions (such as protein–protein interactions, protein–DNA interactions, epigenetic changes, metabolite effects and functional pathway information) should be considered simultaneously. These interactions should be integrated because the whole system together determines the cellular activity or phenotypic result. However, there are many problems and obstacles with network integration.

First, there are inconsistent chemical names in many databases and networks; thus, consistency in chemical naming is necessary to solve this problem. The consistent process of naming has to also be performed for other biological components, such as targets and diseases. Another problem is different coverage and depth of various databases and networks. Each database contains large-scale data individually, but the number of common components for various databases is too small, such that the integrated database and network cannot cover whole biological systems. Integrating the same types of databases with different coverage increases the number of data; thus, integrating databases could compensate for this coverage problem. Another related problem is incomplete network information. More complete networks can be built by using the prediction results of unknown biological components and missing links between them. Predicting missing biological components and links is also a good solution for the coverage problem mentioned above.

A more severe problem is that there exist inconsistencies between databases. A great portion of interaction data was derived from error-prone high-throughput experiments [44]. As a result, those interaction data inevitably contain many errors, which cause serious problems when integrating networks. However, it should be realized that network integration can provide an effective way to detect the inconsistent interaction data and clean up the noise that exists in many interaction networks. Finally, there is a problem in heterogeneity of data types. For example, some types of data (such as K_i values) are real-valued, while data regarding protein–protein interaction are Boolean (true or false). Some types of interaction data have physical origin, while some other types of data describe mere statistical co-dependency. Data heterogeneity poses a big challenge in developing a unified framework for interpreting the network, especially once an integrated network has been made.

4. Biological network comparison

Recently, an exponential increase in information on molecular interactions has enabled us to compare networks of different model species, tissues and cell-types under varying conditions. Because it is believed that the accumulated biological knowledge of different model species could be transferred to humans, it is of great interest to know the systematic differences between different model species or any other different conditions for the purpose of drug discovery. Accordingly, comparative network analysis gives an opportunity to predict new functional interactions that are poorly understood in one species and to judge what kind of network information can be directly transferred to elucidate drug effects in human body [45-56].

Previously, people have tried to develop methods for comparing biological networks that were similar to the sequence comparison methods, even though such methods were less stable than the sequence comparison [46,48,51,52,54,56]. For example, Matthews *et al.* [51] and Yu *et al.* [56] devised a basic pair-wise network alignment method to compare protein–protein interaction networks and regulatory networks across different species using a sequence-based search. In addition to the simple alignment of single interactions, many heuristic approaches have been devised for solving computationally hard network alignment problems by considering a whole array of network structures [52]. More generally beyond the pair-wise network alignment, multiple network alignments evaluating more than two networks have also been devised and tested to handle general relationships across multiple networks [54,55]. Similar to the sequence-based construction of a phylogenetic tree, the relationship between different networks (e.g., different species, tissues, cell-types or environmental conditions) can be represented by a tree that defines global topology of various networks (**Figure 2**). Furthermore, integrating the flood of biological contents into the network comparison framework would be needed. This would include a procedure to define appropriate objective function to score the matched network with different biological information. For a general review about network alignments, see the review paper by Sharan and Ideker [53].

Based on the network comparison framework, an integrated network that consists of different, but necessary information together has been analyzed. The related studies focused on identifying unknown protein interactions, conserved sub-networks (modules), regulatory networks, functional orthology, enzyme clusters along the genomic location and so on [54,57,58]. These kinds of analyses will become more meaningful as there is more knowledge about molecular interactions.

5. Systematic drug target identification and multi-target strategy

Drug target identification has been a major challenge in drug discovery. Until now, several methods have been applied to identify a suitable target. Ideally, a good target would regulate the pathway of interest and blocking the target would result in effective medical treatment. Because this kind of target-based rational drug discovery was proposed, the most important issue has been finding and selecting a specific target-binding ligand. The rationale for this strategy is that the specificity to the selected target leads to reduced side effects that may be caused by undesirable, non-therapeutic off-target binding [59,60]. Recently, however, multi-targeted drug strategy has come into the spotlight. Multi-targeted drug strategy is based upon several observations that many effective drugs (including Gleevec, anticancer drugs and NSAIDs) have turned out to act unexpectedly and cause secondary off-target effects [59,60]. Therefore, it is now believed that several highly efficient drugs might affect multiple targets simultaneously, and, therefore, synergistically influence pathways related to the disease of interest [59-68].

Based on this notion, a systematic way to identify multiple drug targets is necessary to more clearly reveal the effects of drug candidates. The most advanced, high-throughput screening system will be greatly helpful for multiple target identification [69]; however, usually such systems typically have limitations, such as insufficient coverage or expensive costs. Thus, at this stage, a reliable computational method to predict multiple drug targets can be a valuable tool when complementarily used with experimental validations [1,3,18,30,38,42,70]. A common feature of these studies is performing proteome-wide scale prediction between all possible protein–chemical interactions. Systematic predictions of protein–protein interactions have also been studied by various ways, but might be more difficult to predict due to intrinsic complexity [71].

Accordingly, computational approaches and accompanying high-throughput experiments are indispensible for detecting unrecognized or weak binding drug targets. This combination is needed because such targets are difficult to detect by experimental technique alone, making it more difficult to uncover the complex mode of drug actions [72,73]. Once potential drug targets are discovered, then the identified single/multiple drug targets can be analyzed in terms of a biological network [74-77], as discussed in the next section.

6. Estimating drug effects (target perturbations) in the network

Assuming reliable network information is given, an important question for drug discovery would be which target subset is optimal to achieve favorable therapeutic effects while simultaneously reducing non-therapeutic side effects. In other words, how can we estimate the perturbation effects of drugs from our network descriptions and topology to infer what combination of drugs and targets is the best for treating a disease [75,78,79]. Can we link the analysis results in the multi-layered molecular information to the complex disease at the phenotypic level?

Based on previous research, successful promiscuous drugs seem to bind to multiple distinctive targets which may be related to functional pathways [60]. Thus, methods to predict drug targets and estimate target-perturbation effects in a global integrated network containing varying biological information would be a highly useful tool. However, such technology is still in its infancy [80-83]. As an example, Csermely *et al.* tried to model global effect related to various network perturbations, but the model is too simplified to be directly applied to the integrated network [61].

Another associated problem is drug repurposing or drug repositioning. Because approved drugs in current markets already have safe drug profiles, converting the drug purpose can be a very efficient strategy to discover new drugs [84]. However, we also need to know which drug targets are relevant to the new therapeutic effect and their likelihood

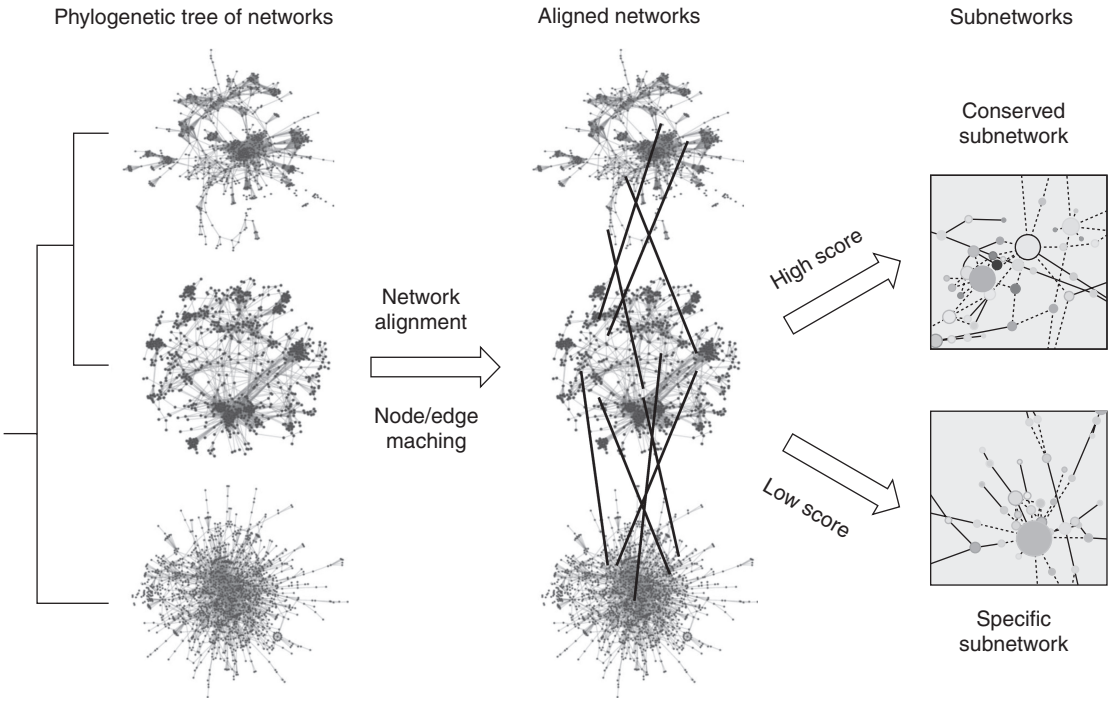

Figure 2. Schematic representation of network alignment and important sub-networks.

of becoming effective drugs in the new positions. Similar problems such as multi-drug resistance and drug side effects can be similarly approached.

To estimate drug effects, the integrated network should be analyzed in terms of the perturbations of multiple drug targets. Next, the perturbation effects should be transformed to some sort of score to estimate total treatment effect for a certain disease. Relating the systematic analysis of drug network perturbations to diseases would be of great interest. Finally, the treatment effect for various diseases by various perturbations would be predicted quantitatively as shown in the heat-map of **Figure 3**.

7. Purpose-suitable drug–target networks

In the previous sections, we examined several main issues of building an integrated drug–target network. In this section, we discuss a practical strategy to apply the network to drug design. As the number of available network components grows, more computational power will be needed. Practically speaking, all components, their interactions and all kinds of descriptions could not be simultaneously considered due to the limitation in computational power. Thus, alternatively, we can focus on a particular part of the whole network or group individual abstract components into higher-order organization to shrink network size and computational demand. Network modularization may be a feasible way for such organization [82,85-88]. For example, in protein–protein interaction networks, protein complexes often acts as a functional unit and can be regarded as one module [89]. Thus, higher-level representation of the whole network minimizes information loss while faster and more efficient algorithms are necessary for analyzing the integrated large network.

Figure 4 displays a strategy for integrated network application for a specific purpose. From the complete integrated network built by integration of different types of biological interaction data and predictions of missing data, a purpose-suitable sub-network can be extracted for a specific application. Sub-networks could be used for finding a new drug target, developing a promiscuous drug design for depression, or inferring side effects of a new developed drug. The integrated network consists of multiple layers (five layers in the example). A sub-network of a specific layer (disease, gene and drug layer in the example) or a specific node (protein2 in the example) can be extracted depending on the specific purpose. These sub-networks display only the network part that is relevant for the purpose of the specific application. This makes the analysis of these networks easier because of their smaller sizes.

8. Conclusion

A drug–target network approach of integrating information on a drug, gene, protein, cell and a disease can be a good solution for more efficient drug discovery, especially as more biological, chemical data and interaction data are produced. There have been several pioneer studies on building and investigating drug–target networks, disease networks and

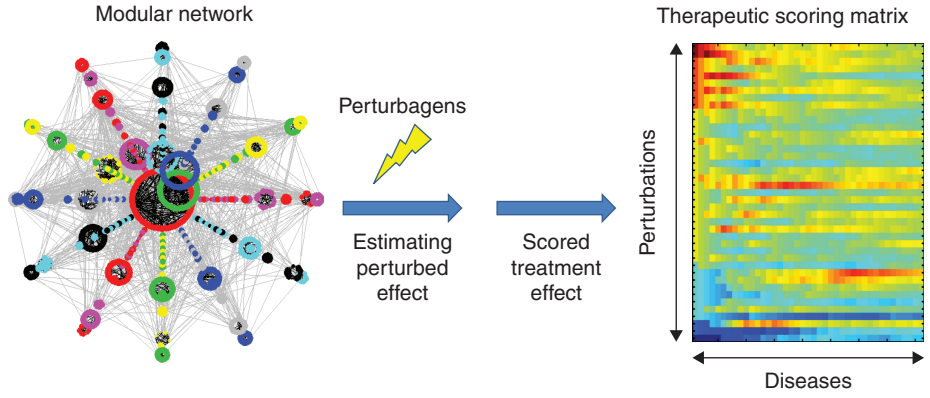

Figure 3. Schematic diagram for estimating perturbation effects.

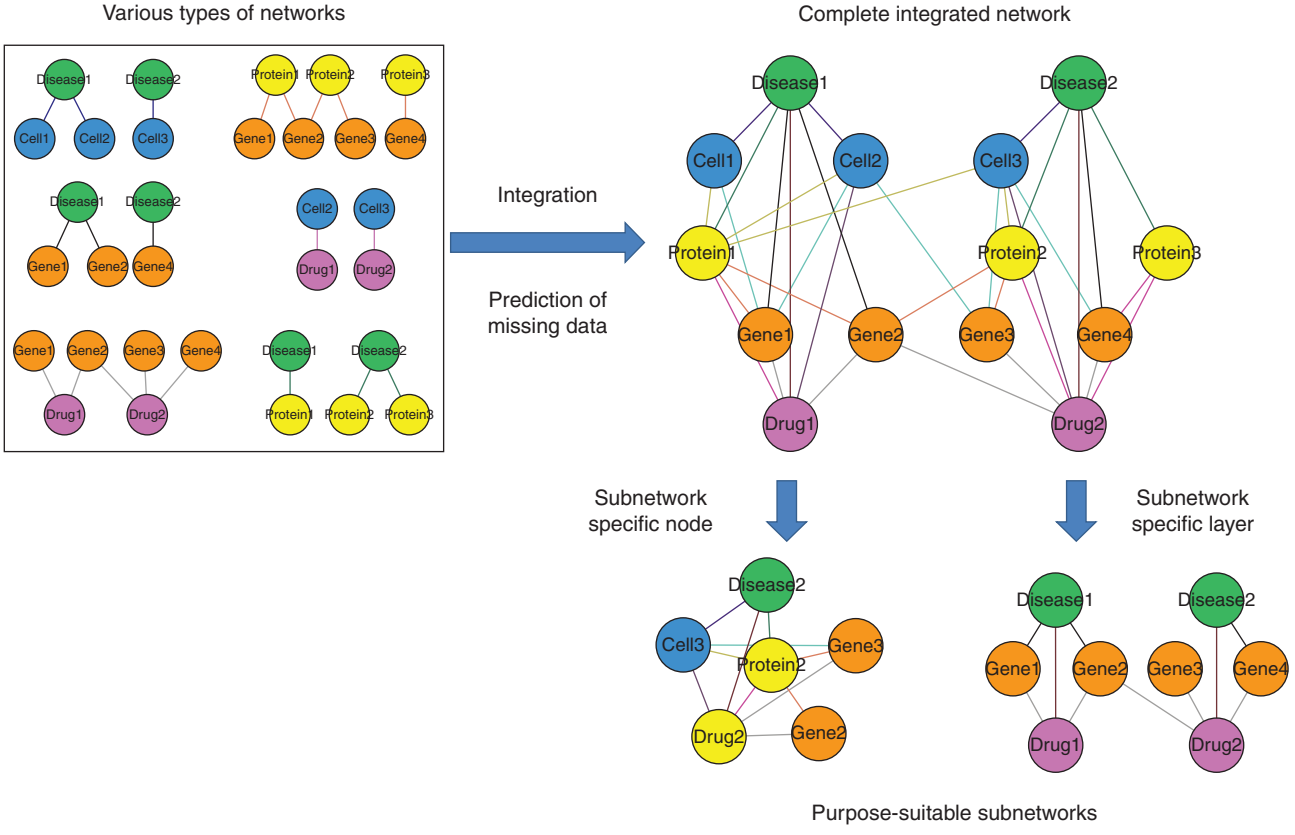

Figure 4. A strategy for building a complete integrated network and its application.

other drug-related networks (**Table 1**). However, such networks typically cover just a small part of the whole biological and chemical network space and include many missing and erroneous nodes and links. Thus, a more complete network should be built by integrating various types of networks. However, the process of integrating networks has many problems, such as inconsistency of chemical names and data heterogeneity. Moreover, additional problems are associated with different coverage of various databases, filling in missing data through predictions and removing inconsistency between different data sets. After building a more complete integrated network, the network should be properly analyzed to render itself useful for practical applications, such as drug target identification and drug repurposing. As a solution, we suggest several strategies, such as biological network comparison, estimation of drug effects (target perturbations) in the systematic view and extraction of a purpose-suitable drug–target network. Drug–target networks integrated with many other

biological networks could become an immensely useful source for future drug development.

9. Expert opinion

Throughout the manuscript so far, we have integrated our 'expert opinion' with the contents to assert the belief that a network and biological data should be integrated to make them most useful. Therefore, most of our opinions in this section are simply the restatement of our discussions in the previous sections from a slightly different perspective.

As stated earlier, the most difficult issue associated with an integrated network is that all of its components are heterogeneous. For example, drugs and proteins are molecules whose nature is obviously physical; while the nature of genes is not so obvious, it can have different meaning depending on the definition of genes and the context [90]. In addition, the nature of disease is clearly different from that of drugs and molecules. Moreover, the relationships between these components are also heterogeneous. The interactions between drugs and molecules, such as drug–protein interactions and protein–protein interactions, have a physical origin that can be expressed in terms of free energy and molecular interactions. On the other hand, disease–protein or disease–gene interactions typically imply that a certain protein or gene is associated with a particular disease, regardless of a specific mechanism. Other related problems are that the network is often incomplete and unevenly sampled and as a result some regions of the network are densely populated while others are not. Such problems of heterogeneity and an incomplete network pose a difficulty in developing a rigorous computational framework for interpreting the network. As a way to assess the importance of nodes, people frequently calculate the network centrality. If the links are heterogeneous (functional or physical, Boolean or real-valued), it is unclear how to handle heterogeneity in a rigorous way. Additionally, if there exists a severe sampling problem, any conclusion drawn from some type of global network analysis may be misleading. Problems could be alleviated if a proper normalization scheme is developed.

The network approach is useful because it can give a global view by allowing system-level analysis, which proves useful for drug discovery. However, in our opinion, the ultimate usefulness of the network approach is yet to be proven. It is true that an integrated network can provide a global view and provide a valuable insight on the general relationship among drugs, target proteins and diseases. However, if one becomes interested in a specific drug discovery project (for example, the repositioning of Gleevec) and then eventually makes a firm decision to launch the project, it is unclear how global analysis on drug–target networks would benefit the project. Under this circumstance, it is likely that the investigative team has already collected a greater amount of relevant information than would be needed for the repositioning of Gleevec. Therefore, it is unlikely that a team could gain additional useful information on Gleevec repositioning by analyzing the drug–target–disease network, if the network is a mere collection of known information. The situation would be different only if the network is not a mere collection of known information. Therefore, it is critical to create an information-rich network by augmenting the network with as many diverse sources of information as possible, by predicting missing nodes and links, and by cleaning up the inconsistency in original data sources. Perhaps more critical is to develop a rigorous computational framework for analyzing the network to generate non-trivial information that can ultimately benefit drug discovery.

Acknowledgement

We thank Byung-Chul Lee and the members of Protein Bio-Informatics Laboratory (PBIL) for helpful discussions. S Lee and K Park contributed equally to this work.

Declaration of interest

The authors received funding from the National Research Foundation of Korea.

Bibliography

1. Kuhn M, Campillos M, Gonzalez P, et al. Bork P. Large-scale prediction of drug-target relationships. FEBS Lett 2008;582(8):1283-90
2. Yildirim MA, Goh KI, Cusick ME, et al. Drug-target network. Nat Biotechnol 2007;25(10):1119-26
3. Yamanishi Y, Araki M, Gutteridge A, et al. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. Bioinformatics 2008;24(13):I232-I40
4. Vina D, Uriarte E, Orallo F, et al. Alignment-free prediction of a drug-target complex network based on parameters of drug connectivity and protein sequence of receptors. Molecular Pharmaceutics 2009;6(3):825-35
5. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol 2008;4(11):682-90
6. Goh KI, Cusick ME, Valle D, et al. The human disease network. Proc Natl Acad Sci USA 2007;104(21):8685-90
7. Campillos M, Kuhn M, Gavin AC, et al. Drug target identification using side-effect similarity. Science 2008;321(5886):263-6
8. Prado-Prado FJ, de la Vega OM, Uriarte E, et al. Unified QSAR approach to antimicrobials. 4. Multi-target QSAR modeling and comparative multi-distance study of the giant components of antiviral drug-drug complex networks. Bioorgan Med Chem 2009;17(2):569-75
9. Gonzalez-Diaz H, Prado-Prado FJ. Unified QSAR and network-based computational chemistry approach to antimicrobials, part 1: Multispecies activity models for antifungals. J Comput Chem 2008;29(4):656-67
10. Wang Y, Xiao J, Suzek TO, et al. PubChem: a public information system for analyzing bioactivities of small molecules. Nucleic Acids Res 2009;37(Web Server issue):W623-33
11. Wishart DS, Knox C, Guo AC, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 2008;36(Database issue):D901-6
12. Seiler KP, George GA, Happ MP, et al. ChemBank: a small-molecule screening and cheminformatics resource database. Nucleic Acids Res 2008;36(Database issue):D351-9
13. Kanehisa M, Goto S, Hattori M, et al. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res 2006;34(Database issue):D354-7
14. Berman HM, Westbrook J, Feng Z, et al. The protein data bank. Nucleic Acids Res 2000;28(1):235-42
15. Boeckmann B, Bairoch A, Apweiler R, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res 2003;31(1):365-70
16. Chen X, Ji ZL, Chen YZ. TTD: therapeutic target database. Nucleic Acids Res 2002;30(1):412-5
17. Chen L, Oughtred R, Berman HM, Westbrook J. TargetDB: a target registration database for structural genomics projects. Bioinformatics 2004;20(16):2860-2
18. Gao Z, Li H, Zhang H, et al. PDTD: a web-accessible protein database for drug target identification. BMC Bioinformatics 2008;9:104
19. Gunther S, Kuhn M, Dunkel M, et al. SuperTarget and matador: resources for exploring drug-target relationships. Nucleic Acids Res 2008;36(Database issue):D919-22
20. Hanzlik RP, Koen YM, Theertham B, et al. The reactive metabolite target protein database (TPDB)–a web-accessible resource. BMC Bioinformatics 2007;8:95
21. Zhang JX, Huang WJ, Zeng JH, et al. DITOP: drug-induced toxicity related protein database. Bioinformatics 2007;23(13):1710-2
22. Ji ZL, Han LY, Yap CW, et al. Drug Adverse Reaction Target Database (DART): proteins related to adverse drug reactions. Drug Saf 2003;26(10):685-90
23. Database Mt. http://www.symyx.com. MDL information Systems, Inc. 2003
24. Lamb J, Crawford ED, Peck D, et al. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. Science 2006;313(5795):1929-35
25. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30(1):207-10
26. Parkinson H, Kapushesky M, Shojatalab M, et al. ArrayExpress–a public database of microarray experiments and gene expression profiles. Nucleic Acids Res 2007;35(Database issue):D747-50
27. Scherf U, Ross DT, Waltham M, et al. A gene expression database for the molecular pharmacology of cancer. Nat Genet 2000;24(3):236-44
28. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. Nat Rev Cancer 2006;6(10):813-23
29. Amberger J, Bocchini CA, Scott AF, Hamosh A. McKusick's Online Mendelian Inheritance in Man (OMIM). Nucleic Acids Res 2009;37(Database issue):D793-6
30. Kuhn M, Mering C, Campillos M, et al. STITCH: interaction networks of chemicals and proteins. Nucleic Acids Res 2008;36(Database issue):D684-8
31. NIMH. http://pdsp.med.unc.edu. NIMH
32. Liu T, Lin Y, Wen X, et al. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res 2007;35(Database issue):D198-201
33. Stark C, Breitkreutz BJ, Reguly T, et al. BioGRID: a general repository for interaction datasets. Nucleic Acids Research 2006;34:D535-D9
34. Caspi R, Foerster H, Fulcher CA, et al. MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res 2006;34(Database issue):D511-6
35. Paolini GV, Shapland RH, van Hoorn WP, et al. Global mapping of pharmacological space. Nat Biotechnol 2006;24(7):805-15
36. Keiser MJ, Roth BL, Armbruster BN, et al. Relating protein pharmacology by ligand chemistry. Nat Biotechnol 2007;25(2):197-206
37. Cases M, Mestres J. A chemogenomic approach to drug discovery: focus on cardiovascular diseases. Drug Discov Today 2009;14(9-10):479-85
38. Nagamine N, Sakakibara Y. Statistical prediction of protein chemical interactions based on chemical structure and mass spectrometry data. Bioinformatics 2007;23(15):2004-12

39. Nagamine N, Shirakawa T, Minato Y, et al. Integrating statistical predictions and experimental verifications for enhancing protein-chemical interaction predictions in virtual screening. PLoS Comput Biol 2009;5(6):e1000397

40. Xie L, Li J, Xie L, Bourne PE. Drug Discovery using chemical systems biology: identification of the protein-ligand binding network to explain the side effects of CETP Inhibitors. PLoS Comput Biol 2009;5(5):e1000387

41. Dunkel M, Gunther S, Ahmed J, et al. SuperPred: drug classification and target prediction. Nucleic Acids Res 2008;36(Web Server issue):W55-9

42. Li Q, Lai L. Prediction of potential drug targets based on simple sequence properties. BMC Bioinformatics 2007;8:353

43. Nacher JC, Schwartz JM. A global view of drug-therapy interactions. BMC Pharmacol 2008;8:5

44. Braun P, Tasan M, Dreze M, et al. An experimentally derived confidence score for binary protein-protein interactions. Nat Methods 2009;6(1):91-8

45. Bandyopadhyay S, Sharan R, Ideker T. Systematic identification of functional orthologs based on protein network comparison. Genome Res 2006;16(3):428-35

46. Berg J, Lassig M. Local graph alignment and motif search in biological networks. Proc Natl Acad Sci USA 2004;101(41):14689-94

47. Dutkowski J, Tiuryn J. Identification of functional modules from conserved ancestral protein-protein interactions. Bioinformatics 2007;23(13):I149-I58

48. Guo X, Hartemink AJ. Domain-oriented edge-based alignment of protein interaction networks. Bioinformatics 2009;25(12):I240-I6

49. Kelley BP, Sharan R, Karp RM, et al. Conserved pathways within bacteria and yeast as revealed by global protein network alignment. Proc Natl Acad Sci USA 2003;100(20):11394-9

50. Li YL, de Ridder D, de Groot MJL, Reinders MJT. Metabolic pathway alignment between species using a comprehensive and flexible similarity measure. Bmc Systems Biology 2008;2:111

51. Matthews LR, Vaglio P, Reboul J, et al. Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs". Genome Res 2001;11(12):2120-6

52. Ogata H, Fujibuchi W, Goto S, Kanehisa M. A heuristic graph comparison algorithm and its application to detect functionally related enzyme clusters. Nucleic Acids Res 2000;28(20):4021-8

53. Sharan R, Ideker T. Modeling cellular machinery through biological network comparison. Nat Biotechnol 2006;24(4):427-33

54. Sharan R, Suthram S, Kelley RM, et al. Conserved patterns of protein interaction in multiple species. Proc Natl Acad Sci USA 2005;102(6):1974-9

55. Stuart JM, Segal E, Koller D, Kim SK. A gene-coexpression network for global discovery of conserved genetic modules. Science 2003;302(5643):249-55

56. Yu HY, Luscombe NM, Lu HX, et al. Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. Genome Res 2004;14(6):1107-18

57. Bandyopadhyay S, Sharan R, Ideker T. Systematic identification of functional orthologs based on protein network comparison. Genome Res 2006;16(3):428-35

58. Koyuturk M, Grama A, Szpankowski W. An efficient algorithm for detecting frequent subgraphs in biological networks. Bioinformatics 2004;20(Suppl 1):i200-7

59. Frantz S. Playing dirty. Nature 2005;437(7061):942-3

60. Roth BL, Sheffler DJ, Kroeze WK. Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. Nat Rev Drug Discov 2004;3(4):353-9

61. Csermely P, Agoston V, Pongor S. The efficiency of multi-target drugs: the network approach might help drug design. Trends Pharmacol Sci 2005;26(4):178-82

62. Espinoza-Fonseca LM. The benefits of the multi-target approach in drug design and discovery. Bioorg Med Chem 2006;14(4):896-7

63. Hampton T. "Promiscuous" anticancer drugs that hit multiple targets may thwart resistance. Jama-J Am Med Assoc 2004;292(4):419

64. Hopkins AL, Mason JS, Overington JP. Can we rationally design promiscuous drugs? Curr Opin Struct Biol 2006 Feb;16(1):127-36

65. Mencher SK, Wang LG. Promiscuous drugs compared to selective drugs (promiscuity can be a virtue). BMC Clin Pharmacol 2005;5(1):3

66. Morphy R, Kay C, Rankovic Z. From magic bullets to designed multiple ligands. Drug Discov Today 2004;9(15):641-51

67. Park K, Lee S, Ahn HS, Kim D. Predicting the multi-modal binding propensity of small molecules: towards an understanding of drug promiscuity. Mol Biosyst 2009;5(8):844-53

68. Wermuth CG. Multitargeted drugs: the end of the "one-target-one-disease" philosophy? Drug Discov Today 2004;9(19):826-7

69. Shoemaker BA, Panchenko AR. Deciphering protein-protein interactions. Part I. Experimental techniques and databases. PLoS Comput Biol 2007;3(3):e42

70. Engelen S, Trojan LA, Sacquin-Mora S, et al. Joint evolutionary trees: a large-scale method to predict protein interfaces based on sequence sampling. PLoS Comput Biol 2009;5(1)

71. Shoemaker BA, Panchenko AR. Deciphering protein-protein interactions. Part II. Computational methods to predict protein and domain interaction partners. PLoS Comput Biol 2007;3(4):e43

72. Zimmermann GR, Lehar J, Keith CT. Multi-target therapeutics: when the whole is greater than the sum of the parts. Drug Discov Today 2007;12(1-2):34-42

73. Keith CT, Borisy AA, Stockwell BR. Multicomponent therapeutics for networked systems. Nat Rev Drug Discov 2005;4(1):71-U10

74. Yeh P, Tschumi AI, Kishony R. Functional classification of drugs by properties of their pairwise interactions. Nat Genet 2006;38(4):489-94

75. Lehar J, Zimmermann GR, Krueger AS, et al. Chemical combination effects predict connectivity in biological systems. Mol Sys Biol 2007;3:80

76. Lehar J, Krueger AS, Avery W, et al. Synergistic drug combinations tend to improve therapeutically relevant selectivity

(vol 7, pg 659, 2009). Nat Biotechnol 2009;27(7):1

77. Borisy AA, Elliott PJ, Hurst NW, et al. Systematic discovery of multicomponent therapeutics. Proc Natl Acad Sci USA 2003;100(13):7977-82
78. Segre D, DeLuna A, Church GM, Kishony R. Modular epistasis in yeast metabolism. Nat Genet 2005;37(1):77-83
79. Araujo RP, Petricoin EF, Liotta LA. A mathematical model of combination therapy using the EGFR signaling network. Biosystems 2005;80(1):57-69
80. Young DW, Bender A, Hoyt J, et al. Integrating high-content screening and ligand-target prediction to identify mechanism of action. Nat Chem Biol 2008;4(1):59-68
81. Calvano SE, Xiao W, Richards DR, et al. A network-based analysis of systemic inflammation in humans. Nature 2005;437(7061):1032-7
82. Tanay A, Sharan R, Kupiec M, Shamir R. Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data. Proc Natl Acad Sci USA 2004;101(9):2981-6
83. Fitzgerald JB, Schoeberl B, Nielsen UB, Sorger PK. Systems biology and combination therapy in the quest for clinical efficacy. Nat Chem Biol 2006;2(9):458-66
84. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov 2004;3(8):673-83
85. Banks E, Nabieva E, Chazelle B, Singh M. Organization of physical interactomes as uncovered by network schemas. PLoS Comput Biol 2008;4(10)
86. Kashtan N, Alon U. Spontaneous evolution of modularity and network motifs. Proc Natl Acad Sci USA 2005;102(39):13773-8
87. Ravasz E, Somera AL, Mongru DA, et al. Hierarchical organization of modularity in metabolic networks. Science 2002;297(5586):1551-5
88. Royer L, Reimann M, Andreopoulos B, Schroeder M. Unraveling protein networks with power graph analysis. PLoS Comput Biol 2008;4(7)
89. Han S, Kim D. Inference of protein complex activities from chemical-genetic profile and its applications: predicting drug-target pathways. PLoS Comput Biol 2008;4(8):e1000162
90. Gerstein MB, Bruce C, Rozowsky JS, et al. What is a gene, post-ENCODE? History and updated definition. Genome Res 2007;17(6):669-81

Affiliation

Soyoung Lee, Keunwan Park & Dongsup Kim[†]
[†]Author for correspondence
KAIST,
Department of Bio and Brain Engineering,
335 Gwahak-ro,
Yuseong-gu, Daejeon,
305-701 Korea, Republic of Korea
Tel: +82 42 350 4317; Fax: +82 42 350 4310;
E-mail: kds@kaist.ac.kr