Drug Repositioning Through Network Pharmacology

Article *in* Current Topics in Medicinal Chemistry · April 2016

DOI: 10.2174/1568026616666160530181328

5 authors, including:

Hao Ye
CStone Pharmaceuticals Co. Ltd.
48 PUBLICATIONS **868** CITATIONS
SEE PROFILE

Kailin Tang
Tongji University
90 PUBLICATIONS **1,294** CITATIONS
SEE PROFILE

Ritchie J Feuers
U.S. Department of Health and Human Services
115 PUBLICATIONS **2,823** CITATIONS
SEE PROFILE

Huixiao Hong
U.S. Department of Health and Human Services
279 PUBLICATIONS **12,794** CITATIONS
SEE PROFILE

Send Orders for Reprints to reprints@benthamscience.ae

Current Topics in Medicinal Chemistry, 2017, *17*, 1-11

REVIEW ARTICLE

Drug Repositioning Through Network Pharmacology

Hao Ye[1], Jia Wei[2], Kailin Tang[3], Ritchie Feuers[4] and Huixiao Hong[1,*]

[1]*Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR, USA;* [2]*R&D Information, AstraZeneca, Shanghai, China;* [3]*School of Life Sciences and Technology, Tongji University, Shanghai, China;* [4]*Office of Research, National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR, USA*

ARTICLE HISTORY

Received: December 01, 2015
Revised: March 30, 2016
Accepted: April 19, 2016

DOI: 10.2174/1568026616666160530181328

Abstract: Low drug productivity has been a significant problem of the pharmaceutical industry for several decades even though numerous novel technologies were introduced during this period. Currently pharmacologic dogma, "single drug, single target, single disease", is at the root of the lack of drug productivity. From a systems biology viewpoint, network pharmacology has been proposed to complement the established guiding pharmacologic approaches. The rationale for network pharmacology as a major component of drug discovery and development is that a disease can be caused by perturbation of the disease-causing network and a drug may be designed to interact with multiple targets for modulation of such a network from the disease status toward normal status. Therefore, network pharmacology has been applied to guide and assist in drug repositioning. Drugs exerting their therapeutic effects may directly target disease-associated proteins, but they may also modulate the pathways involved in the pathological process. In this review, we discuss the progresses and prospects in network pharmacology, focusing on drug off-targets discovery, disease-associated protein identification, and pathway analysis for elucidating relationships between drug targets and disease-associated proteins.

Keywords: Drug repositioning, network pharmacology, pathway analysis.

DRUG REPOSITIONING

Drug discovery and development is laborious, time consuming, and expensive, making it a very high risk process. The current time frame for bringing a single de novo drug to market averages about fifteen years with an estimated cost of up to or more than one billion dollars ($0.8-1.8 billions) [1,2]. The situation is further exacerbated by the fact that the pharmaceutical industry has faced a low success rate in drug development. Only about 10.45% [3] of candidate drugs that are initiated through clinical testing successfully obtain FDA approval (success rate in each phase: phase I 64.5%, phase II 32.4%, phase III 60.1%, registration 83.2%). Even more problematic, new drug productivity has never kept pace with enormous increases in R&D investment expenses. Thusly, since the mid-1990's the pharmaceutical industry has been faced with the dilemma of high expenditure coupled with low productivity [4]. It has become clear that novel discovery technologies (such as structure-based drug design, combinatorial chemistry, high-throughput screening, and next generation sequencing) were implemented during this period with the specific aim of improving productivity. However, a recent report [5] suggests the current confounding situation is not expected to end in the near future. It has been further suggested that due to the 'patent cliff' that began in 2009 pharmaceutical company sales are suffering the sharpest and most abrupt revenue decline in history. For example, sales at Pfizer have declined from $67.8 billion in 2010 to $51.6 billion in 2013 due to the loss of patent exclusivity of some major drugs, including successful drugs such as Lipitor. Those sales numbers were predicted to decline to $49.65 billion in 2014 [6]. It seems likely that the pharmaceutical industry will not be able to maintain state-of-the-art innovation initiative to replace the loss of revenues from these expired successful drugs. The vicious cycle of decreasing revenues, shrinking R&D expenditures and reduced drug generation coupled with the difficulties of reaching the market places the pharmaceutical industry in a serious dilemma. Therefore, there are urgent needs to propose novel strategies and to utilize new resources in drug discovery.

A potential solution to a portion of these problems would be to simply utilize existing drugs to their most full potential. In that regard, drug repositioning (also known as drug repurposing, drug re-profiling, therapeutic switching and drug re-tasking) is viewed as a promising strategy to improve drug productivity. The aim of such strategies is to identify and develop new indications for existing drugs. Attrition rates of drugs have become significant, and analysis has indicated

*Address correspondence to this author at the Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR 72079, USA;
E-mail: huixiao.hong@fda.hhs.gov

1568-0266/17 $58.00+.00

that a major reason for attrition in clinical drug development is lack of efficacy and clinical safety and/or associated toxicology with each accounting for 30% of the failures [7]. As the safety data of the old drug has already been recorded, drug repositioning can shorten the development cycle by avoiding failure due to safety issues. Moreover, failed drugs, due to low efficacy, may also be revived and effectively utilized. An obvious advantage would be that any newly identified functions can be quickly evaluated in phase II clinical trials [8].

NETWORK PHARMACOLOGY

Recently, Hopkins [9] pointed out that the fundamental problem in drug discovery may not be technological and scientific issues, but guiding strategies. Since 1949 when Pauling first proposed a molecular mechanism underlying classic sickle cell anemia [10], it has been widely accepted that a disease is caused by abnormality of a single gene or protein. Since that time "single gene, single drug, and single disease" strategy has become the predominant paradigm in drug discovery. Based on the 'lock-and-key' model [11], current drug design strategy aims at finding small molecules which can specifically bind the target with high affinity. Any drugs which were found to interact with unintended targets (off-targets) were considered as "dirty" drugs and disregarded for having drug development potential further contributing to drug development attrition.

Recent advances in systems biology have revealed phenotypic robustness in intracellular networks. Most single gene knockouts had little or no effect on phenotype except in the case of a few essential genes, and this was proven by a series of large scale functional genomic studies in model organisms [12-15]. It was further cross validated by the finding of a high proportion of drugs with unsatisfied efficacy in current drug discovery. Human disease, as an intricate phenotype, is rarely a consequence of abnormality in a single gene, but reflects the perturbations of the complex intracellular network in a tissue or an organ system. According to OMIM database [16], more than 14% of diseases have more than one disease-causing gene that can be identified even using a very rigorously statistical qualification. We have listed the top 10 diseases with multiple disease-causing genes in Table **1**. It is obvious that many prevalent diseases such as diabetes, cancer, hypertension and asthma are caused by a set of disease-causing genes. It could be easily inferred that the efficacy of current signal-target drugs is poor for these diseases.

Shifting the drug design paradigm from traditional pharmacology to network pharmacology holds a high expectation

Table 1. Top 10 diseases with multiple disease-causing genes.

MIM ID	Disease	Number of genes	Gene symbols
125853	Diabetes	27	PPARG; GPD2; NEUROD1; IRS1; IGF2BP2; WFS1; CDKAL1; ENPP1; IL6; GCK; PAX4; SLC30A8; TCF7L2; ABCC8; KCNJ11; MAPK8IP1; HNF1A; PDX1; IRS2; LIPC; SLC2A4; HNF1B; GCGR; RETN; AKT2; HNF4A; PTPN1
181500	Schizophrenia	23	MTHFR; CHI3L1; DISC1; DISC2; SYN2; DRD3; SCZD1; SCZD3; DTNBP1; SCZD6; LGR4; SCZD2; DAO; HTR2A; SCZD7; DAOA; AKT1; SCZD10; SCZD8; RTN4R; COMT; APOL2; APOL4
114480	Breast cancer	19	RAD54L; CASP8; BARD1; PIK3CA; HMMR; NQO2; RB1CC1; SLC22A18; ATM; KRAS; BRCA2; AKT1; RAD51; PALB2; TP53; PHB; BRIP1; PPM1D; CHEK2
601626	Acute myeloid leukemia	19	GMPS; MLF1; LPP; CHIC2; KIT; NSD1; NPM1; WHSC1L1; JAK2; NUP214; MLLT10; PICALM; ARHGEF12; ETV6; FLT3; AMLCR2; SH3GL1; CEBPA; RUNX1
114500	Colorectal cancer	18	PLA2G2A; NRAS; ODC1; PIK3CA; TLR2; APC; PDGFRL; TLR4; PTPRJ; CCND1; MLH3; AKT1; BUB1B; TP53; FLCN; AXIN2; AURKA; EP300
209900	Bardet-biedl syndrome 1	16	C2orf86; BBS5; ARL6; BBS7; BBS12; BBS9; TMEM67; TRIM32; BBS1; BBS10; CEP290; TTC8; BBS4; BBS2; MKS1; MKKS
601665	Obesity	16	SDC3; NR0B2; POMC; GHRL; PPARG; UCP1; CARTPT; ADRB2; PPARGC1B; SIM1; ENPP1; ADRB3; UCP3; AGRP; PYY; MC4R
176807	Prostate cancer	15	PCAP; MAD1L1; HIP1; MSR1; KLF6; PTEN; MXI1; CD82; BRCA2; ZFHX3; ELAC2; hpc3; CHEK2; hpc6; AR
211980	Lung cancer	15	CASP8; DLEC1; RASSF1; PIK3CA; IRF1; PARK2; EGFR; BRAF; MAP3K8; ERCC6; SLC22A18; PPP2R1B; KRAS; ERBB2; CYP2A6
145500	Hypertension	13	ECE1; ATP1B1; RGS5; SELE; AGT; AGTR1; ADD1; CYP3A5; NOS3; GNB3; NOS2; PNMT; PTGIS
600807	Asthma	13	HNMT; MUC7; IL13; SCGB3A2; IL12B; ADRB2; HLA-G; TNF; PLA2G7; ALOX5; SCGB1A1; PHF11; CCL11

to increase drug productivity. Instead of searching for a single disease-causing gene, the strategy of network pharmacology focuses on identifying the perturbations of a disease-causing network involving a group of disease-associated proteins/genes. Designing a multiple-target drug that can modulate the network from a disease status to a normal status is very different from developing a single-target drug using traditional pharmacology. Under this revolutionizing strategy many concepts are changing. A "dirty" drug of traditional pharmacology might be an excellent multi-target agent that could efficiently modulate disease status. However, off-targets of "dirty drugs" may not lead only to side effects, but could also generate benefit for other disease therapies. Further, "clean" drugs of traditional strategies are not always "clean". For example, more than 68.70% (1196/1741) of the FDA-approved drugs collected in the latest DrugBank (version 4.1) [17] had been found to bind off-targets after they entered the market.

The modified mechanism of action (MOA) of a drug under network pharmacology strategy can be represented using the drug therapy model as illustrated by Fig. (**1**). In the drug therapy model, drugs exerting a therapeutic effect may not only directly target the disease-related proteins, but can also modulate the pathways that regulate the disease where a pathway can be defined as a manually curated, ascertained and established basic biologically functional unit.

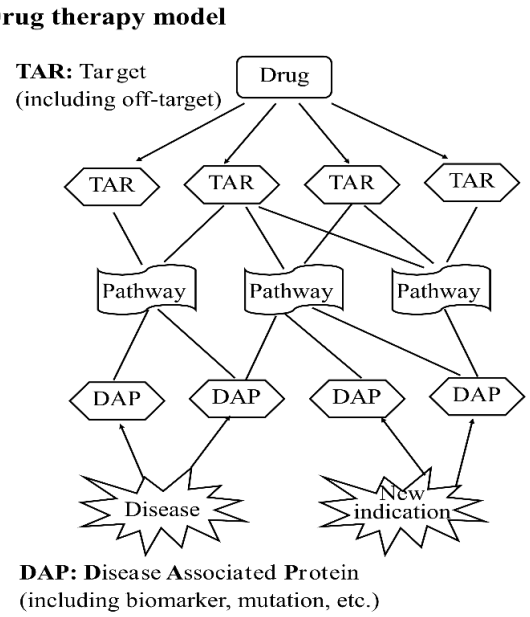

Fig. (1). The modified mechanism of action (MOA) of a drug under network pharmacology. A drug may bind to many targets, including its primary target and off-targets. Disease associated proteins (DAPs) play important roles in the pathological process of a disease. If the targets (especially the off-targets) of a drug and the DAPs of a disease involved in the same biological pathways, the drug may display some therapeutic effects on the disease through those pathways. Therefore, this disease could be a new indication of the drug through regulating those pathways.

OFF-TARGETS IDENTIFICATION

The target (almost always a protein) of a drug and the resultant interaction produces an intended readout or direct effect for the drug's modulation of a disease status. An off-target protein can simply be considered as the protein that was not intentionally targeted yet does interact perhaps yielding potential modulation of a different disease (prospective new target). Whether in traditional pharmacology or network pharmacology, target discovery is the first and most important process for drug repositioning.

Computational methods have been widely used for prediction of drugs' off-targets. For example, molecular docking is a computational technology which simulates the dynamic ligand-receptor recognition process under certain molecular fields, according to the 'lock-and-key' model [11]. It aims to predict the predominant binding modes of a ligand to a receptor that possesses an optimized conformation with a minimized free energy for the overall molecular system. Molecular docking is a tool that has been widely used in structural molecular biology and computer-assisted drug design. According to the statistics from the Swiss institute of bioinformatics (*http://www.click2drug.org/index.html#Docking*), more than 55 programs have been developed to implement molecular docking technologies. Table **2** lists some webservers which are available for free and convenient access.

Unlike virtual chemical screening in which a chemical's conformation library needs to be constructed first, a library of targets and their ligand binding sites is necessary for conducting an off-targets screening. Generally, the ligand binding sites have been selected from the original compound-protein complex after removing or extracting the original ligand. Among the online webservers listed in Table **2**, TarFisDock [18] provides a library of 698 targets of the FDA approved drugs and their ligand binding sites. Without limitation in the targets space, idTarget [28] allows users to build their own targets libraries. It provides a function to automatically establish a library of targets and the ligand binding site from protein structure files (PDB ID list is accepted). These servers are convenient for off-targets discovery.

Drug similarity is also utilized for new target prediction based on the assumption that similar drugs tend to be associated with similar targets. Additionally, chemical structures or side effects have been used to predict new molecular targets for known drugs [33,34]. Wang [35] combined drug-drug similarity and target-target similarity to predict new targets based on heterogeneous graph inference. Fakhraei et al [36] proposed a machine learning method named PSL (probabilistic soft logic) to evaluate the weight of multiple similarities (chemical, ligand, expression, side effect, sequence, protein-protein network, gene ontology) and to predict drug-target interactions (DTIs). A group of traditional machine learning approaches have also been utilized for off-targets predictions based on such features as neural network and support vector machine (SVM) [37-40]. More specifically, Wang [41] used a restricted Boltzmann machine to predict different types of DTIs on a multidimensional network which not only describes binary DTIs, but also encodes their corresponding types of interactions. Numerous network-based approaches have been proposed to exploit latent features of DTI profiles, and those have recently become a popular tool for DTI predictions [37,42-45].

Aside from these computational methods, many drug target databases have been developed by curating/identifying

Table 2. Online docking webservers (comprehensive but not complete)

Web server	URL	Targets library	Support Targets screening	Remark
TarFisDock [18]	http://www.dddc.ac.cn/tarfisdock	698[#]	Yes	Only support structures in mol2 format
SwissDock [19]	http://www.swissdock.ch/docking	User submitted	NO	Support PDB, mol, SDF format. File size <= 5Mb
1-Click Docking[*]	https://mcule.com/apps/1-click-docking/	9871 proteins	Yes	Support a wide of structure format, 2D, 3D, SMILEs, and online-drawing.
DOCK Blaster [20]	http://blaster.docking.org/	User submitted	NO	Support PDB format
ParDOCK [21]	http://www.scfbio-iitd.res.in/dock/pardock.jsp	User submitted	NO	Only PDB format is supported
FlexPepDock [22]	http://flexpepdock.furmanlab.cs.huji.ac.il/	User submitted	NO	Only PDB format is supported
PatchDock [23]	http://bioinfo3d.cs.tau.ac.il/PatchDock/		NO	Only PDB format is supported
MEDock [24]	http://medock.ee.ncku.edu.tw/	User submitted	NO	Support PDB and PQR format
BSP-SLIM [25]	http://zhanglab.ccmb.med.umich.edu/BSP-SLIM	User submitted	NO	SDF and PDB format is acceptable
BioDrugScreen [26]	http://www.biodrugscreen.org/	1589[$]	Yes	Support mol, SDF, PDB
iScreen [27]	http://iscreen.cmu.edu.tw/	User submitted	NO[%]	Support mol, SDF, PDB
idTarget [28]	http://idtarget.rcas.sinica.edu.tw/	User submitted	Yes	Support pdb/mol2/pdbqt/cif as well as online-drawing
Score [29]	http://159.149.85.2/score.htm	User submitted	NO	Support more than 20 formats of stuructures.
Pose & Rank [30]	http://modbase.compbio.ucsf.edu/ligscore/	User submitted	NO	Support PDB and mol2
PLATINUM [31]	http://model.nmr.ru/platinum/	User submitted	NO	Input file size <= 5Mb
LPCCSU [32]	http://bip.weizmann.ac.il/oca-bin/lpccsu	User submitted	NO	Support PDB format

[*]commercial software (free trial is also available); [#]approved drug targets; [$]human targets from human cancer protein interaction network and human drugable proteome; [%]support compound screening on >20000 herbal ingredients.

drug targets from the literature. Further, many DTIs have been discovered after the drugs were approved by the FDA. For example, DrugBank [17] contains not only the primary targets (also called pre-designed target), but also a significant amount of off-target information as ascertained from the literature. Additionally, there are some other resources focusing on the target information of promising compounds with relatively high degrees of success. For example, Ye et al [46] found numerous targets for herbal ingredients form the literature as reported in their HIT publication. We now summarize most, but not a complete list of all these that will now be a publically available resource in Table **3**. Considering the high false positive rate resulting from computational methods, these DTIs that have been identified from the literature require further and thorough experimental validation before they can be positively and effectively categorized as new off-target discoveries.

DISEASE ASSOCIATED PROTEIN DISCOVERY

In order to understand and intervene in molecular mechanisms of diseases, it is essential that causal genes are identified. Those identifications should be the first step in the drug discovery process regardless of whether researchers use a traditional or network pharmacological approach. In order to specifically and clearly identify such genes, genetic linkage analysis and positional cloning methodologies must be employed. Unfortunately, in doing so hundreds of positional candidates that may be disease associated will be identified. Experimentally confirming or evaluating all of these candi-

Table 3. Publically available drug targets resources

Resource	Brief description	URL	Remark
DrugBank [17]	DrugBank v4.1 contains 7740 drug entries including 1584 FDA-approved small molecule drugs, 157 FDA-approved biotech (protein/peptide) drugs, 89 nutraceuticals and over 6000 experimental drugs.	http://www.drugbank.ca	Most popularly used drug target database. All data are manual curated.
TTD [47]	This database currently contains 2,360 targets, including 388 successful, 461 clinical trial, 44 discontinued and 1,331 research targets, 20,667 drugs, including 2,003 approved, 3,147 clinical trial, 14,853 experimental drugs. In addition, the diseases are well matched to ICD-10 and ICD-9 codes.	http://bidd.nus.edu.sg/group/cjttd	Information is manual curated from literature. The targets are well categorized based on corresponding drug status.
PiHelper [48]	It's an open-source framework for drug-target and antibody-target data. It enables integration of ten publicly available drug target resource, including DrugBank, KEGG Drug, et. al	http://bitbucket.org/armish/pihelper	Can instantly acquire the user's interested drug-target interaction from public databases.
KEGG Drug [49]	KEGG DRUG is a comprehensive drug information resource for approved drugs in Japan, USA, and Europe unified based on the chemical structures and/or the chemical components, and associated with target, metabolizing enzyme, and other molecular interaction network information.	http://www.genome.jp/kegg/drug	Contains the targets of crude drugs.
HIT [46]	HIT (Herbal ingredients' Target) is the first comprehensive and fully curated database on herbal ingredients. It currently contains about 1,301 known protein targets derived from more than 3,250 literatures, which covers about 586 active compounds from more than 1,300 reputable Chinese herbs.	http://lifecenter.sgst.cn/hit	The first manual curated herbal ingredient's targets database.
DCDB [50]	The Drug Combination Database (DCDB) is devoted to the research and development of multi-component drugs. DCDB collected 1363 drug combinations (330 approved and 1033 investigational, including 237 unsuccessful usages), involving 904 individual drugs, 805 targets.	http://www.cls.zju.edu.cn/dcdb/index.jsf	The targets information are manual curated from literatures.
MATADOR [51]	MATADOR is a resource for protein-chemical interactions. It differs from other resources such as DrugBank in its inclusion of as many direct and indirect interactions in literatures.	http://matador.embl.de	A complementary source to DrugBank database.
Rask-Andersen [52]	Driving from DrugBank database and extensive manual curation. It contains 435 effect-mediating drug targets in the human genome, which are modulated by 989 unique drugs, through 2,242 drug–target interactions.	http://www.nature.com/nrd/journal/v10/n8/suppinfo/nrd3478.html	Data are derived from DrugBank and literature.
DGIdb [53]	The drug-gene interaction database (DGIdb) drug-gene interactions with expert curation from DrugBank, therapeutic target database (TTD), PharmGKB, a list of targeted agents in lung cancer and ClinicalTrials.gov.	http://dgidb.genome.wustl.edu/	A user friendly interface for collecting interested drug-targets interactions.
CTD [54]	CTD curates specific chemical–gene and –protein interactions in vertebrates and invertebrates from published references, including binding, expression and activity.	http://ctdbase.org/	Contains both curated and predicted chemical-protein interactions.

dates will be time-consuming and expensive. This is clearly demonstrated by the fact that at least 3309 causal genes have been identified and associated with 5369 diseases over the past few decades according to OMIM's statistics (data from December 5, 2014). OMIM as a comprehensive and authoritative resource focusing on the relationships between phenotypes and genotypes can continue to be a valuable resource in this regard.

Several bioinformatic methods have been proposed to facilitate pursuing and then prioritizing disease genes from a long list of candidates that have been generated by positional cloning experiments. Rosario's 2012 review [55] comprehensively covered these bioinformatics methods including those frequently used, such as text mining the literature and measuring similarities with known disease genes.

Differentially expressed genes identified from high throughput techniques such as microarrays and next-generation sequencing (namely, RNA-seq) are frequently used to decipher the molecular basis of diseases. For example, in the arena of cancer research and treatment, the Cancer Genome Atlas (TCGA) [56] was initiated in 2006 with the intent of exploring gene changes in samples from more than 20 cancer types. In addition, many profiling databases are now available that are proving to be useful for exploring the underlying mechanisms of complex diseases. ArrayExpress in EBI [57], Gene Expression Omnibus in NCBI, DDBJ Omics Archive [58] and Stanford Microarray Database [59] are just a few such databases.

MicroRNAs (miRNAs) regulate gene expression. Expressions of some miRNAs are candidate biomarkers of diagnosis of diseases and treatment responses of drugs [60-62]. The emerging technologies such as genotyping microarrays [63,64] and next-generation sequencing technologies [65] facilitated detection of genetic variations and profiling expression of miRNAs to advance the translation of miRNAs as biomarkers clinical practice and safety evaluation [66-68]. Recently, miRNAs were found as important biomarkers involved in disease pathology [69-72]. Some novel bioinformatics methods were proposed to predict the potential disease-related miRNAs. Network and machine learning-based methods are frequently utilized in the area. For example, KATZ was proposed to detect the novel disease-miRNA pairs based on the similarity score from the walks of different lengths between the microRNA and disease nodes and CATAPULT was separately developed as a supervised learning method to predict disease associated miRNAs based on the features from hybrid walks in miRNA-disease network [71].

Table **4** lists some publically available and frequently used databases containing disease genes with experimental validations.

PATHWAY ANALYSIS FOR LINKING DRUG TARGETS AND DISEASE-ASSOCIATED PROTEINS

From a systems biology perspective, proteins do not always perform their functions in isolation, but interact with other cellular components to form complexes or pathways [78]. As basic biologically functioning units particular pathways are frequently used to bridge gaps between biological networks/modules and molecular functions. For example, Ville-Petteri [79] mined the pathway from a gene-regulatory network to gain insights into the causal molecular mechanisms of coronary artery disease. Similarly, Zhao [80] utilized pathway categories to describe function modules in a human liver specific metabolic network. Further, Wang [81] used a similar pathway mining approach to facilitate biomarker discovery in a pancreatic cancer microRNA network. Clearly, these type pathway exploration approaches facilitate the linkage analysis between drug targets and disease-associated proteins in a protein-protein network.

To our knowledge only four studies on this subject have been published, and these focus on drug repositioning through pathway analysis of two groups of proteins (drug targets, disease-associated proteins). Mathur and Dinakarpandian [82] mapped these protein groups into gene ontology (GO) terms of biological processes. As an outcome, protein-protein networks are built by the proteins in each GO term of a guiding biological process. Therein, the GO terms of a biological process overlap with a high degree of centrality for drug targets in the protein-protein network which may hint at the possible effect of the drug upon a disease.

In an effort to further advance this agenda, we proposed a pathway profile association scoring method to evaluate the relationship between drugs and diseases [83]. Our hypothesis was that a drug exerting its therapeutic effect not only directly targets the disease-associated protein but may also modulate the pathways involved in the pathological process. We used pathway enrichment analysis to identify the significantly influenced pathways of drug targets (including primary target and off targets) or disease associated proteins in our curated library. Scores were calculated for the overlapped pathways enriched by drug targets as well as disease-associated proteins to rank the pairs of drug and disease in a disease dataset. The drugs in the top ranked disease-drug pairs have potential to treat all diseases in the pairs, providing candidates of further clinical investigations for drug repositioning.

Pan [84] has more recently developed a pathway analysis approach to implement drug repositioning through drug targets' pathway enrichment analysis and the manual curation of diseases' causal pathways. Li [85] had proposed a pathway-based method for drug repositioning using the causal chain in a layered drug-target-pathway-gene-disease network. In summary, the overlapped pathways have been first identified through mapping the drug targets and disease-associated proteins to a pathway database. Therefore, a set of causal chains are generated using the topological location of drug targets and disease-associated proteins in the overlapped pathways. Finally, the novel drug indications can be predicted through the maximum likelihood estimation of these causal chains between the drug and diseases.

Because associations between drug targets and disease-associated proteins in pathways have no directions, network pharmacology analysis can only provide possible new indications of a drug for the diseases in the associated pathways. Further investigations on the candidate new indications are needed to confirm the therapeutic effects that would be potential for drug repositioning study. The side

Table 4. The publically available and frequently used databases

Database	Brief description	URL	Remark
OMIM [16]	OMIM is a comprehensive, authoritative compendium of human genes and genetic phenotypes that is freely available and updated daily. It contains 5369 disease phenotypes, 3309 genes (Data on 8th, Dec., 2014).	http://omim.org	Frequently used disease gene database with high credit.
CTD [54]	CTD contains curated and inferred gene–disease associations. The curated information contains 4927 diseases, 7705 genes, 30382 disease-gene records	http://ctdbase.org	Gene-disease data are updated from literature weekly.
DisGeNET [73]	DisGeNET is a discovery platform integrating information on gene-disease associations from several public data sources and the literature. The current version contains 381056 associations, between 16666 genes and 13172 diseases. The curation records contain 7108 genes, 5466 disease, 22678 disease associations.	http://www.disgenet.org/web/DisGeNET/v2.1	An integrative disease-gene database.
GAD [74]	The Genetic Association Database is a database of genetic association data from complex diseases and disorders. Although, the data was frozen by 09/01/2014, it is still available.	http://geneticassociationdb.nih.gov	The data was frozen on Sept 1, 2014.
COSMIC [75]	Catalogue of somatic mutations in cancer (COSMIC) contains cancer related somatic mutation information and associated details extracted from literatures.	http://cancer.sanger.ac.uk/cancergenome/projects/cosmic	Focus on gene mutations in cancer.
DISEASES [76]	DISEASES is a frequently updated web resource that integrates evidence on disease-gene associations from automatic text mining, manually curated literature, cancer mutation data, and genome-wide association studies. In addition, it provides different channels to download, text mining, knowledge and experiments.	http://diseases.jensenlab.org	Both predicted and manually curated disease-associations are included.
SwissVar [77]	SwissVar is a portal to search variants in Swiss-Prot entries of the UniProt Knowledgebase (UniProtKB). It summarized all the information related to a particular gene variant, including manual annotation on the genotype-phenotype relationship of each specific variant based on literature. Currently, it contains 3371 disease related proteins, 4472 diseases. (Data from 12/10/2014)	http://swissvar.expasy.org	The details of the evidences for each disease-gene association are well summarized.

effects of a drug that interact with many off target could also be useful information for safety evaluation of the drug in drug repositioning process.

FUTURE PROSPECTS

We can now define network pharmacologically based multi-target modulation as leading where a drug modulates many genes in a disease associated pathway to change the pathway from disease status to normal status. Newly proposed strategies promise to dramatically reduce time to development, failure rates, and developmental costs, while ultimately improving quality of new drugs with specific disease treatment capabilities. These type advances in network pharmacology are changing the traditional concept that new candidate drugs with off-targets (only leading to side effects) should be discarded. More importantly, there is an abundance of so called dirty drugs that may now be revived by investigating the drug targets' biological effects in each disease causing protein-protein network. In this regard, pathway analysis serves as an effective and simplified strategy to decipher the biological linkage between drug targets and disease-associated proteins. Although, major advances have been made in this field, we identify three major future challenges yet to be overcome.

Affinity Difference Between Primary Target and Off-Targets

In a majority of cases a drug will display much higher affinity to its primary target than off-targets. This results from the fact that as a principle of selectivity drug design, affinity optimization is a primary objective resulting in efficient and effective blockage of the target. Therefore, in the majority of cases discarded drugs that may to be considered for repositioning would have a high affinity to their originally intended primary (predesigned) target and low affinities to its off-targets. Indeed, differences in extent of inhibition could influence the extent of a drug's regulatory effects in disease relevant pathways. In fact, Sorin [86] proposed a method to

evaluate such effects using pathway enrichment analysis with consideration of expression levels of the genes (in fold change) as well as topological signal flow in a pathway. The existence of these affinity differences provides sufficient motivation for moving toward identification of highly regulated pathways affected by a drug and to precisely evaluate the influence on the drug targets and disease-associated proteins.

Protein Expression Levels in Different Tissues

It is well known that protein expression levels are tissue dependent, and tissue differences could directly influence drug perturbation of pathways directly involved in a disease. This would especially be the case for tissue or organ specific diseases. Some mRNA tissue distribution databases have been developed based on microarray technology, e.g. TissueDistributionDBs [87], TiGER [88], PaGenBase [89], and TiSGeD [90]. However, the translations of mRNA molecules (genes) into proteins (the molecules displaying actual biological functions) are different; thus, the expression level of a gene is not necessarily the same as its protein products. Fortunately, using antibody technology, a human proteome project called HUMAN PROTEIN ATLAS [91] has just updated its latest version (November 2014) covering 16,975 proteins and their corresponding express levels in 213 tissues. Such tissue-specified protein expression data holds significant promise as that information could greatly facilitate future pathway analysis.

Genes Not Able to be Mapped to Pathways Due to the Incomplete Resource

Within our current context a pathway can be described as a series of actions among various molecules or substrates that result in a particular biological effect. Many pathway databases have been independently developed, and range from the KEGG pathway [51], MetaBase pathway (also called GeneGO, https://portal.genego.com), IPA (http://www.ingenuity.com/), Pathway studio [92], BioCarta [93], PANTHER [94], Reactome [95], PID [96], to the Pathway Commons [97]. Yet none of them can lay claim to being a complete or most comprehensive database. In fact, different databases have their advantages and limitations. For example, KEGG contains more than 250 pathways covering more than 6000 genes. Different from KEGG, MetaBase collected a large number of cellular signaling pathways (MetaBase v6.5 contains >800 pathways, <5400 genes). Due to inherent shortcomings there is an urgent need to merge these resources in order to comprehensively investigate drug targets and disease-associated proteins using network pharmacology. Without these advances it is likely that some drug targets or disease-associated proteins will fail to be mapped to pathways.

A major challenge within this field will be to reduce redundancy effects among these varied resources. Donato et al [98] recently proposed a method to modify crosstalk (could be considered as the genes participating in multiple pathways) effects in pathway analysis, suggesting this would provide clues useful for reduction in the redundancy across different pathway databases. However, overlapped genes may continue to imply redundancy among different pathway resources to a certain extent.

In network pharmacology, pathway analysis has become a promising method for drug repositioning. It could be used to extract biologically meaningful MOAs from complex biological network and to discover new indications of old drugs. Evaluation of the relationships between targets of drugs and the disease-associated proteins is the key for the success of drug repurposing. New progress in pathway analysis is required and then expected to facilitate and accelerate drug repurposing.

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflict of interest.

ACKNOWLEDGMENTS

This research was supported in part by an appointment to the Research Participation Program at the National Center for Toxicological Research (Hao Ye) administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Food and Drug Administration. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Food and Drugs Administration.

REFERENCES

[1] Adams, C. P.; Brantner, V. V. Estimating the cost of new drug development: is it really 802 million dollars? *Health Aff (Millwood),* **2006**, *25*(2), 420-428.

[2] Khanna, I. Drug discovery in pharmaceutical industry: productivity challenges and trends. *Drug Discov. Today,* **2012**, *17*(19), 1088-1102.

]3] Hay, M.; Thomas, D. W.; Craighead, J. L.; Economides, C.; Rosenthal, J. Clinical development success rates for investigational drugs. *Nat. Biotechnol.,* **2014**, *32*(1), 40-51.

[4] Ashburn, T. T.; Thor, K. B. Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug. Discov.,* **2004**, *3*(8), 673-683.

[5] Lendrem, D.; Senn, S. J.; Lendrem, B. C.; Isaacs, J. D. R&D productivity rides again? *Pharm Stat.,* **2014**, *14*(1), 1-3.

[6] Pfizer's Acquisition Strategy: Will Pfizer Still Go For A Big Acquisition? Available online: http://seekingalpha.com/article/2762625-pfizers-acquisition-strategy-will-pfizer-still-go-for-a-big-acquisition (Accessed on Feb 22, 2015).

[7] Kola, I. The State of Innovation in Drug Development. *Clin. Pharmacol. Ther.,* **2008**, *83*(2), 227-230.

[8] Chong, C. R.; Sullivan, D. J., Jr. New uses for old drugs. *Nature,* **2007**, *448*(7154), 645-646.

[9] Hopkins, A. L. Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.,* **2008**, *4*(11), 682-690.

[10] Pauling, L.; Itano, H. A.; *et al.* Sickle cell anemia a molecular disease. *Science,* **1949**, *110*(2865), 543-548.

[11] Morris, G.M.; Lim-Wilby, M. Molecular docking. *Methods Mol. Biol.*, **2008**, *443*[443]: 365-382.

[12] Zambrowicz, B. P.; Sands, A. T. Modeling drug action in the mouse with knockouts and RNA interference. *Drug Discov. Today,* **2004**, *3*(5), 198-207.

[13] Winzeler, E. A.; Shoemaker, D. D.; Astromoff, A.; Liang, H.; Anderson, K.; Andre, B.; Bangham, R.; Benito, R.; Boeke, J. D.; Bussey, H.; Chu, A. M.; Connelly, C.; Davis, K.; Dietrich, F.; Dow, S. W.; El Bakkoury, M.; Foury, F.; Friend, S. H.; Gentalen, E.; Giaever, G.; Hegemann, J. H.; Jones, T.; Laub, M.; Liao, H.; Liebundguth, N.; Lockhart, D. J.; Lucau-Danila, A.; Lussier, M.;

M'Rabet, N.; Menard, P.; Mittmann, M.; Pai, C.; Rebischung, C.; Revuelta, J. L.; Riles, L.; Roberts, C. J.; Ross-MacDonald, P.; Scherens, B.; Snyder, M.; Sookhai-Mahadeo, S.; Storms, R. K.; Veronneau, S.; Voet, M.; Volckaert, G.; Ward, T. R.; Wysocki, R.; Yen, G. S.; Yu, K.; Zimmermann, K.; Philippsen, P.; Johnston, M.; Davis, R. W. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science,* **1999**, *285*(5429), 901-906.

[14] Giaever, G.; Chu, A. M.; Ni, L.; Connelly, C.; Riles, L.; Veronneau, S.; Dow, S.; Lucau-Danila, A.; Anderson, K.; Andre, B.; Arkin, A. P.; Astromoff, A.; El-Bakkoury, M.; Bangham, R.; Benito, R.; Brachat, S.; Campanaro, S.; Curtiss, M.; Davis, K.; Deutschbauer, A.; Entian, K. D.; Flaherty, P.; Foury, F.; Garfinkel, D. J.; Gerstein, M.; Gotte, D.; Guldener, U.; Hegemann, J. H.; Hempel, S.; Herman, Z.; Jaramillo, D. F.; Kelly, D. E.; Kelly, S. L.; Kotter, P.; LaBonte, D.; Lamb, D. C.; Lan, N.; Liang, H.; Liao, H.; Liu, L.; Luo, C.; Lussier, M.; Mao, R.; Menard, P.; Ooi, S. L.; Revuelta, J. L.; Roberts, C. J.; Rose, M.; Ross-Macdonald, P.; Scherens, B.; Schimmack, G.; Shafer, B.; Shoemaker, D. D.; Sookhai-Mahadeo, S.; Storms, R. K.; Strathern, J. N.; Valle, G.; Voet, M.; Volckaert, G.; Wang, C. Y.; Ward, T. R.; Wilhelmy, J.; Winzeler, E. A.; Yang, Y.; Yen, G.; Youngman, E.; Yu, K.; Bussey, H.; Boeke, J. D.; Snyder, M.; Philippsen, P.; Davis, R. W.; Johnston, M. Functional profiling of the Saccharomyces cerevisiae genome. *Nature,* **2002**, *418*(6896), 387-391.

[15] Deutschbauer, A. M.; Jaramillo, D. F.; Proctor, M.; Kumm, J.; Hillenmeyer, M. E.; Davis, R. W.; Nislow, C.; Giaever, G. Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics,* **2005**, *169*(4), 1915-1925.

[16] Hamosh, A.; Scott, A. F.; Amberger, J. S.; Bocchini, C. A.; McKusick, V. A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.,* **2005**, *33*(Database issue), D514-517.

[17] Law, V.; Knox, C.; Djoumbou, Y.; Jewison, T.; Guo, A. C.; Liu, Y.; Maciejewski, A.; Arndt, D.; Wilson, M.; Neveu, V.; Tang, A.; Gabriel, G.; Ly, C.; Adamjee, S.; Dame, Z. T.; Han, B.; Zhou, Y.; Wishart, D. S. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.,* **2014**, *42*(Database issue), D1091-D1097.

[18] Li, H.; Gao, Z.; Kang, L.; Zhang, H.; Yang, K.; Yu, K.; Luo, X.; Zhu, W.; Chen, K.; Shen, J.; Wang, X.; Jiang, H. TarFisDock: a web server for identifying drug targets with docking approach. *Nucleic Acids Res.,* **2006**, *34*(Web Server issue), W219-W224.

[19] Grosdidier, A.; Zoete, V.; Michielin, O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res.,* **2011**, *39*(Web Server issue), W270-W277.

[20] Irwin, J. J.; Shoichet, B. K.; Mysinger, M. M.; Huang, N.; Colizzi, F.; Wassam, P.; Cao, Y. Automated docking screens: a feasibility study. *J Med Chem.,* **2009**, *52*(18), 5712-5720.

[21] Gupta, A.; Gandhimathi, A.; Sharma, P.; Jayaram, B. ParDOCK: an all atom energy based Monte Carlo docking protocol for protein-ligand complexes. *Protein Pept. Lett.,* **2007**, *14*(7), 632-646.

[22] London, N.; Raveh, B.; Cohen, E.; Fathi, G.; Schueler-Furman, O. Rosetta FlexPepDock web server--high resolution modeling of peptide-protein interactions. *Nucleic Acids Res.,* **2011**, *39*(Web Server issue), W249-W253.

[23] Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.,* **2005**, *33*(Web Server issue), W363-W367.

[24] Chang, D. T.; Oyang, Y. J.; Lin, J. H. MEDock: a web server for efficient prediction of ligand binding sites based on a novel optimization algorithm. *Nucleic Acids Res.,* **2005**, *33*(Web Server issue), W233-W238.

[25] Lee, H. S.; Zhang, Y. BSP-SLIM: a blind low-resolution ligand-protein docking approach using predicted protein structures. *Proteins,* **2012**, *80*(1), 93-110.

[26] Li, L.; Bum-Erdene, K.; Baenziger, P. H.; Rosen, J. J.; Hemmert, J. R.; Nellis, J. A.; Pierce, M. E.; Meroueh, S. O. BioDrugScreen: a computational drug design resource for ranking molecules docked to the human proteome. *Nucleic Acids Res.,* **2010**, *38*(Database issue), D765-D773.

[27] Tsai, T.-Y.; Chang, K.-W.; Chen, C.-C. iScreen: world's first cloud-computing web server for virtual screening and de novo drug design based on TCM database@Taiwan. *Journal of Computer-Aided Mol. Design,* **2011**, *25*(6), 525-531.

[28] Wang, J. C.; Chu, P. Y.; Chen, C. M.; Lin, J. H. idTarget: a web server for identifying protein targets of small chemical molecules with robust scoring functions and a divide-and-conquer docking approach. *Nucleic Acids Res.,* **2012**, *40*(Web Server issue), W393-W399.

[29] Vistoli, G.; Pedretti, A.; Mazzolari, A.; Testa, B. In silico prediction of human carboxylesterase-1 (hCES1) metabolism combining docking analyses and MD simulations. *Bioorg. Med. Chem.,* **2010**, *18*(1), 320-329.

[30] Fan, H.; Schneidman-Duhovny, D.; Irwin, J. J.; Dong, G.; Shoichet, B. K.; Sali, A. Statistical potential for modeling and ranking of protein-ligand interactions. *J. Chem. Inf. Model.,* **2011**, *51*(12), 3078-3092.

[31] Pyrkov, T. V.; Chugunov, A. O.; Krylov, N. A.; Nolde, D. E.; Efremov, R. G. PLATINUM: a web tool for analysis of hydrophobic/hydrophilic organization of biomolecular complexes. *Bioinformatics,* **2009**, *25*(9), 1201-1202.

[32] Sobolev, V.; Sorokine, A.; Prilusky, J.; Abola, E. E.; Edelman, M. Automated analysis of interatomic contacts in proteins. *Bioinformatics,* **1999**, *15*(4), 327-332.

[33] Keiser, M. J.; Setola, V.; Irwin, J. J.; Laggner, C.; Abbas, A. I.; Hufeisen, S. J.; Jensen, N. H.; Kuijer, M. B.; Matos, R. C.; Tran, T. B.; Whaley, R.; Glennon, R. A.; Hert, J.; Thomas, K. L. H.; Edwards, D. D.; Shoichet, B. K.; Roth, B. L. Predicting new molecular targets for known drugs. *Nature,* **2009**, *462*(7270), 175-181.

[34] Lounkine, E.; Keiser, M. J.; Whitebread, S.; Mikhailov, D.; Hamon, J.; Jenkins, J. L.; Lavan, P.; Weber, E.; Doak, A. K.; Cote, S.; Shoichet, B. K.; Urban, L. Large-scale prediction and testing of drug activity on side-effect targets. *Nature,* **2012**, *486*(7403), 361-367.

[35] Wang, W.; Yang, S.; Li, J. Drug target predictions based on heterogeneous graph inference. *Pac. Symp. Biocomput.,* **2013**, 53-64.

[36] Fakhraei, S.; Huang, B.; Raschid, L.; Getoor, L. Network-Based Drug-Target Interaction Prediction with Probabilistic Soft Logic. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **2014**, *11*(5), 775-787.

[37] Bleakley, K.; Yamanishi, Y. Supervised prediction of drug–target interactions using bipartite local models. *Bioinformatics,* **2009**, *25*(18), 2397-2403.

[38] Nagamine, N.; Sakakibara, Y. Statistical prediction of protein–chemical interactions based on chemical structure and mass spectrometry data. *Bioinformatics,* **2007**, *23*(15), 2004-2012.

[39] Faulon, J.-L.; Misra, M.; Martin, S.; Sale, K.; Sapra, R. Genome scale enzyme–metabolite and drug–target interaction predictions using the signature molecular descriptor. *Bioinformatics,* **2008**, *24*(2), 225-233.

[40] Jacob, L.; Vert, J. P. Protein-ligand interaction prediction: an improved chemogenomics approach. *Bioinformatics,* **2008**, *24*(19), 2149-2156.

[41] Wang, Y.; Zeng, J. Predicting drug-target interactions using restricted Boltzmann machines. *Bioinformatics,* **2013**, *29*(13), i126-i134.

[42] Chen, X.; Liu, M.-X.; Yan, G.-Y. Drug–target interaction prediction by random walk on the heterogeneous network. *Mol. BioSyst.,* **2012**, *8*(7), 1970-1978.

[43] Cheng, F.; Liu, C.; Jiang, J.; Lu, W.; Li, W.; Liu, G.; Zhou, W.; Huang, J.; Tang, Y. Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput. Biol.,* **2012**, *8*(5), e1002503.

[44] Yamanishi, Y.; Araki, M.; Gutteridge, A.; Honda, W.; Kanehisa, M. Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics,* **2008**, *24*(13), i232-i240.

[45] Zhao, S.; Li, S. Network-based relating pharmacological and genomic spaces for drug target identification. *PloS One,* **2010**, *5*(7), e11764.

[46] Ye, H.; Ye, L.; Kang, H.; Zhang, D.; Tao, L.; Tang, K.; Liu, X.; Zhu, R.; Liu, Q.; Chen, Y. Z.; Li, Y.; Cao, Z. HIT: linking herbal active ingredients to targets. *Nucleic Acids Res.,* **2011**, *39*(Database issue), D1055-D1059.

[47] Zhu, F.; Shi, Z.; Qin, C.; Tao, L.; Liu, X.; Xu, F.; Zhang, L.; Song, Y.; Liu, X.; Zhang, J.; Han, B.; Zhang, P.; Chen, Y. Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. *Nucleic Acids Res.,* **2012**, *40*(Database issue), D1128-D1136.

[48] Aksoy, B. A.; Gao, J.; Dresdner, G.; Wang, W.; Root, A.; Jing, X.; Cerami, E.; Sander, C. PiHelper: an open source framework for drug-target and antibody-target data. *Bioinformatics,* **2013**, *29*(16), 2071-2072.

[49] Kanehisa, M.; Goto, S.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.,* **2014**, *42*(Database issue), D199-D205.

[50] Liu, Y.; Hu, B.; Fu, C.; Chen, X. DCDB: drug combination database. *Bioinformatics,* **2010**, *26*(4), 587-588.

[51] Gunther, S.; Kuhn, M.; Dunkel, M.; Campillos, M.; Senger, C.; Petsalaki, E.; Ahmed, J.; Urdiales, E. G.; Gewiess, A.; Jensen, L. J.; Schneider, R.; Skoblo, R.; Russell, R. B.; Bourne, P. E.; Bork, P.; Preissner, R. SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Res.,* **2008**, *36*(Database issue), D919-D922.

[52] Rask-Andersen, M.; Almen, M. S.; Schioth, H. B. Trends in the exploitation of novel drug targets. *Nat. Rev. Drug Discov.,* **2011**, *10*(8), 579-590.

[53] Griffith, M.; Griffith, O. L.; Coffman, A. C.; Weible, J. V.; McMichael, J. F.; Spies, N. C.; Koval, J.; Das, I.; Callaway, M. B.; Eldred, J. M.; Miller, C. A.; Subramanian, J.; Govindan, R.; Kumar, R. D.; Bose, R.; Ding, L.; Walker, J. R.; Larson, D. E.; Dooling, D. J.; Smith, S. M.; Ley, T. J.; Mardis, E. R.; Wilson, R. K. DGIdb: mining the druggable genome. *Nat. Methods,* **2013**, *10*(12), 1209-1210.

[54] Davis, A. P.; Grondin, C. J.; Lennon-Hopkins, K.; Saraceni-Richards, C.; Sciaky, D.; King, B. L.; Wiegers, T. C.; Mattingly, C. J. The Comparative Toxicogenomics Database's 10th year anniversary: update 2015. *Nucleic Acids Res.* **2014**, *43*(Database issue), D914-D920.

[55] Piro, R. M.; Di Cunto, F. Computational approaches to disease-gene prediction: rationale, classification and successes. *FEBS J.,* **2012**, *279*(5), 678-696.

[56] International Cancer Genome, C.; Hudson, T. J.; Anderson, W.; Artez, A.; Barker, A. D.; Bell, C.; Bernabe, R. R.; Bhan, M. K.; Calvo, F.; Eerola, I.; Gerhard, D. S.; Guttmacher, A.; Guyer, M.; Hemsley, F. M.; Jennings, J. L.; Kerr, D.; Klatt, P.; Kolar, P.; Kusada, J.; Lane, D. P.; Laplace, F.; Youyong, L.; Nettekoven, G.; Ozenberger, B.; Peterson, J.; Rao, T. S.; Remacle, J.; Schafer, A. J.; Shibata, T.; Stratton, M. R.; Vockley, J. G.; Watanabe, K.; Yang, H.; Yuen, M. M.; Knoppers, B. M.; Bobrow, M.; Cambon-Thomsen, A.; Dressler, L. G.; Dyke, S. O.; Joly, Y.; Kato, K.; Kennedy, K. L.; Nicolas, P.; Parker, M. J.; Rial-Sebbag, E.; Romeo-Casabona, C. M.; Shaw, K. M.; Wallace, S.; Wiesner, G. L.; Zeps, N.; Lichter, P.; Biankin, A. V.; Chabannon, C.; Chin, L.; Clement, B.; de Alava, E.; Degos, F.; Ferguson, M. L.; Geary, P.; Hayes, D. N.; Hudson, T. J.; Johns, A. L.; Kasprzyk, A.; Nakagawa, H.; Penny, R.; Piris, M. A.; Sarin, R.; Scarpa, A.; Shibata, T.; van de Vijver, M.; Futreal, P. A.; Aburatani, H.; Bayes, M.; Botwell, D. D.; Campbell, P. J.; Estivill, X.; Gerhard, D. S.; Grimmond, S. M.; Gut, I.; Hirst, M.; Lopez-Otin, C.; Majumder, P.; Marra, M.; McPherson, J. D.; Nakagawa, H.; Ning, Z.; Puente, X. S.; Ruan, Y.; Shibata, T.; Stratton, M. R.; Stunnenberg, H. G.; Swerdlow, H.; Velculescu, V. E.; Wilson, R. K.; Xue, H. H.; Yang, L.; Spellman, P. T.; Bader, G. D.; Boutros, P. C.; Campbell, P. J.; Flicek, P.; Getz, G.; Guigo, R.; Guo, G.; Haussler, D.; Heath, S.; Hubbard, T. J.; Jiang, T.; Jones, S. M.; Li, Q.; Lopez-Bigas, N.; Luo, R.; Muthuswamy, L.; Ouellette, B. F.; Pearson, J. V.; Puente, X. S.; Quesada, V.; Raphael, B. J.; Sander, C.; Shibata, T.; Speed, T. P.; Stein, L. D.; Stuart, J. M.; Teague, J. W.; Totoki, Y.; Tsunoda, T.; Valencia, A.; Wheeler, D. A.; Wu, H.; Zhao, S.; Zhou, G.; Stein, L. D.; Guigo, R.; Hubbard, T. J.; Joly, Y.; Jones, S. M.; Kasprzyk, A.; Lathrop, M.; Lopez-Bigas, N.; Ouellette, B. F.; Spellman, P. T.; Teague, J. W.; Thomas, G.; Valencia, A.; Yoshida, T.; Kennedy, K. L.; Axton, M.; Dyke, S. O.; Futreal, P. A.; Gerhard, D. S.; Gunter, C.; Guyer, M.; Hudson, T. J.; McPherson, J. D.; Miller, L. J.; Ozenberger, B.; Shaw, K. M.; Kasprzyk, A.; Stein, L. D.; Zhang, J.; Haider, S. A.; Wang, J.; Yung, C. K.; Cros, A.; Liang, Y.; Gnaneshan, S.; Guberman, J.; Hsu, J.; Bobrow, M.; Chalmers, D. R.; Hasel, K. W.; Joly, Y.; Kaan, T. S.; Kennedy, K. L.; Knoppers, B. M.; Lowrance, W. W.; Masui, T.; Nicolas, P.; Rial-Sebbag, E.; Rodriguez, L. L.; Vergely, C.; Yoshida, T.; Grimmond, S. M.; Biankin, A. V.; Bowtell, D. D.; Cloonan, N.; deFazio, A.; Eshleman, J. R.; Etemadmoghadam, D.; Gardiner, B. B.; Kench, J. G.; Scarpa, A.; Sutherland, R. L.; Tempero, M. A.; Waddell, N. J.; Wilson, P. J.; McPherson, J. D.; Gallinger, S.; Tsao, M. S.; Shaw, P. A.; Petersen, G. M.; Mukhopadhyay, D.; Chin, L.; DePinho, R. A.; Thayer, S.; Muthuswamy, L.; Shazand, K.; Beck, T.; Sam, M.; Timms, L.; Ballin, V.; Lu, Y.; Ji, J.; Zhang, X.; Chen, F.; Hu, X.; Zhou, G.; Yang, Q.; Tian, G.; Zhang, L.; Xing, X.; Li, X.; Zhu, Z.; Yu, Y.; Yu, J.; Yang, H.; Lathrop, M.; Tost, J.; Brennan, P.; Holcatova, I.; Zaridze, D.; Brazma, A.; Egevard, L.; Prokhortchouk, E.; Banks, R. E.; Uhlen, M.; Cambon-Thomsen, A.; Viksna, J.; Ponten, F.; Skryabin, K.; Stratton, M. R.; Futreal, P. A.; Birney, E.; Borg, A.; Borresen-Dale, A. L.; Caldas, C.; Foekens, J. A.; Martin, S.; Reis-Filho, J. S.; Richardson, A. L.; Sotiriou, C.; Stunnenberg, H. G.; Thoms, G.; van de Vijver, M.; van't Veer, L.; Calvo, F.; Birnbaum, D.; Blanche, H.; Boucher, P.; Boyault, S.; Chabannon, C.; Gut, I.; Masson-Jacquemier, J. D.; Lathrop, M.; Pauporte, I.; Pivot, X.; Vincent-Salomon, A.; Tabone, E.; Theillet, C.; Thomas, G.; Tost, J.; Treilleux, I.; Calvo, F.; Bioulac-Sage, P.; Clement, B.; Decaens, T.; Degos, F.; Franco, D.; Gut, I.; Gut, M.; Heath, S.; Lathrop, M.; Samuel, D.; Thomas, G.; Zucman-Rossi, J.; Lichter, P.; Eils, R.; Brors, B.; Korbel, J. O.; Korshunov, A.; Landgraf, P.; Lehrach, H.; Pfister, S.; Radlwimmer, B.; Reifenberger, G.; Taylor, M. D.; von Kalle, C.; Majumder, P. P.; Sarin, R.; Rao, T. S.; Bhan, M. K.; Scarpa, A.; Pederzoli, P.; Lawlor, R. A.; Delledonne, M.; Bardelli, A.; Biankin, A. V.; Grimmond, S. M.; Gress, T.; Klimstra, D.; Zamboni, G.; Shibata, T.; Nakamura, Y.; Nakagawa, H.; Kusada, J.; Tsunoda, T.; Miyano, S.; Aburatani, H.; Kato, K.; Fujimoto, A.; Yoshida, T.; Campo, E.; Lopez-Otin, C.; Estivill, X.; Guigo, R.; de Sanjose, S.; Piris, M. A.; Montserrat, E.; Gonzalez-Diaz, M.; Puente, X. S.; Jares, P.; Valencia, A.; Himmelbauer, H.; Quesada, V.; Bea, S.; Stratton, M. R.; Futreal, P. A.; Campbell, P. J.; Vincent-Salomon, A.; Richardson, A. L.; Reis-Filho, J. S.; van de Vijver, M.; Thomas, G.; Masson-Jacquemier, J. D.; Aparicio, S.; Borg, A.; Borresen-Dale, A. L.; Caldas, C.; Foekens, J. A.; Stunnenberg, H. G.; van't Veer, L.; Easton, D. F.; Spellman, P. T.; Martin, S.; Barker, A. D.; Chin, L.; Collins, F. S.; Compton, C. C.; Ferguson, M. L.; Gerhard, D. S.; Getz, G.; Gunter, C.; Guttmacher, A.; Guyer, M.; Hayes, D. N.; Lander, E. S.; Ozenberger, B.; Penny, R.; Peterson, J.; Sander, C.; Shaw, K. M.; Speed, T. P.; Spellman, P. T.; Vockley, J. G.; Wheeler, D. A.; Wilson, R. K.; Hudson, T. J.; Chin, L.; Knoppers, B. M.; Lander, E. S.; Lichter, P.; Stein, L. D.; Stratton, M. R.; Anderson, W.; Barker, A. D.; Bell, C.; Bobrow, M.; Burke, W.; Collins, F. S.; Compton, C. C.; DePinho, R. A.; Easton, D. F.; Futreal, P. A.; Gerhard, D. S.; Green, A. R.; Guyer, M.; Hamilton, S. R.; Hubbard, T. J.; Kallioniemi, O. P.; Kennedy, K. L.; Ley, T. J.; Liu, E. T.; Lu, Y.; Majumder, P.; Marra, M.; Ozenberger, B.; Peterson, J.; Schafer, A. J.; Spellman, P. T.; Stunnenberg, H. G.; Wainwright, B. J.; Wilson, R. K.; Yang, H. International network of cancer genome projects. *Nature,* **2010**, *464*(7291), 993-998.

[57] Rustici, G.; Kolesnikov, N.; Brandizi, M.; Burdett, T.; Dylag, M.; Emam, I.; Farne, A.; Hastings, E.; Ison, J.; Keays, M.; Kurbatova, N.; Malone, J.; Mani, R.; Mupo, A.; Pedro Pereira, R.; Pilicheva, E.; Rung, J.; Sharma, A.; Tang, Y. A.; Ternent, T.; Tikhonov, A.; Welter, D.; Williams, E.; Brazma, A.; Parkinson, H.; Sarkans, U. ArrayExpress update--trends in database growth and links to data analysis tools. *Nucleic Acids Res.,* **2013**, *41*(Database issue), D987-D990.

[58] Kosuge, T.; Mashima, J.; Kodama, Y.; Fujisawa, T.; Kaminuma, E.; Ogasawara, O.; Okubo, K.; Takagi, T.; Nakamura, Y. DDBJ progress report: a new submission system for leading to a correct annotation. *Nucleic Acids Res.,* **2014**, *42*(Database issue), D44-D49.

[59] Hubble, J.; Demeter, J.; Jin, H.; Mao, M.; Nitzberg, M.; Reddy, T. B.; Wymore, F.; Zachariah, Z. K.; Sherlock, G.; Ball, C. A. Implementation of GenePattern within the Stanford Microarray Database. *Nucleic Acids Res.,* **2009**, *37*(Database issue), D898-D901.

[60] Koturbash, I.; Tolleson, W.H.; Guo, L.; Yu, D.; Chen, S.; Hong, H.; Matters, W.; Ning, B. MicroRNAs as pharmacogenomic biomarkers for drug efficacy and drug safety assessment. *Biomark. Med.*, **2015**, *9*(11), 1153-1176.

[61] Wang, Y.; Liu, Z.; Zou, W.; Hong, H.; Fang, H.; Tong, W. Molecular regulation of miRNAs and potential biomarkers in the progression of hepatic steatosis to NASH. *Biomark. Med.*, **2015**, *9*(11), 1189-1200.

[62] Liu, J.; Jennings, S.; Tong, W.; Hong, H. Next-generation sequencing for profiling expression of miRNAs: technical progress and applications in drug development. *J. Biomed. Sci. Eng.*, **2011**, *4*(10), 666–676.

[63] Hong, H.;, Xu, L.; Liu, J.; Jones, W.D.; Su, Z.; Ning, B.; Perkins, R.; Ge, W.; Miclaus, K.; Zhang, L.; Park, K.; Green, B.; Han, T.; Fang, H.; Lambert, C.G.; Vega, S.C.; Lin, S.M. Jafari, N.; Czika,

W.; Wolfinger, R.D.; Goodsaid, F.; Tong, W.; Shi, L. Technical reproducibility of genotyping SNP arrays used in genome-wide association studies. *PLoS One,* **2012**, *7*(9), e44483.

[64] Hong, H.; Su, Z.; Ge, W.; Shi, L.; Perkins, R.; Fang, H.; Xu, J.; Chen, J.J.; Han, T.; Kaput, J.; Fuscoe, J.C.; Tong, W. Assessing Batch Effects of Genotype Calling Algorithm BRLMM for the Affymetrix GeneChip Human Mapping 500K Array Set Using 270 HapMap Samples. *BMC Bioinformatics*, **2008**, *9*(Suppl 9), S17.

[65] Hong, H.; Zhang, W.; Shen, J.; Su, Z.; Ning, B.; Han, T.; Perkins, R.; Shi, L.; Tong, W. Critical role of bioinformatics in translating huge amounts of next-generation sequencing data into personalized medicine. *Sci. China Life Sci.*, **2013**, *56*(2), 110-8.

[66] Hong, H.; Slikker, W. Jr. Advancing translation of biomarkers into regulatory decision making. *Biomark. Med.*; **2015**, *9*(11), 1043-1046.

[67] Hong, H.; Tong, W. Emerging efforts for discovering new biomarkers of liver disease and hepatotoxicity. *Biomark. Med.*; **2014**, *8*, 143–146.

[68] Hong, H.; Goodsaid, F.; Shi, L.; Tong, W. Molecular biomarkers: a US FDA effort. *Biomark. Med.*; **2010**, 4, 215–225.

[69] Zou, Q.; Li, J.; Song, L.; Zeng, X.; Wang, G. Similarity computation strategies in the microRNA-disease network: a survey. *Brief. Funct. Genomics*, **2016**, *15*(1), 55-64.

[70] Zeng, X.; Zhang, X.; Zou, Q. Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks. *Brief. bioinform.*, **2016**, *17*(2), 193-203.

[71] Zou, Q.; Li, J.; Hong, Q.; Lin, Z.; Wu, Y.; Shi, H.; Ju, Y. Prediction of MicroRNA-Disease Associations Based on Social Network Analysis Methods. *Biomed. Res. Int.*, **2015**, *2015*, 810514.

[72] Zeng, X.; Liao, Y.; Liu, Y.; Zou, Q. Prediction and validation of disease genes using HeteSim Scores. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **2016**, doi: 10.1109/TCBB.2016.2520947.

[73] Bauer-Mehren, A.; Rautschka, M.; Sanz, F.; Furlong, L. I. DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks. *Bioinformatics,* **2010**, *26*(22), 2924-2926.

[74] Becker, K. G.; Barnes, K. C.; Bright, T. J.; Wang, S. A. The genetic association database. *Nat. Genet.,* **2004**, *36*(5), 431-432.

[75] Forbes, S. A.; Beare, D.; Gunasekaran, P.; Leung, K.; Bindal, N.; Boutselakis, H.; Ding, M.; Bamford, S.; Cole, C.; Ward, S.; Kok, C. Y.; Jia, M.; De, T.; Teague, J. W.; Stratton, M. R.; McDermott, U.; Campbell, P. J. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.,* **2015**, *43*(Database issue), D805-D811.

[76] Pletscher-Frankild, S.; Pallejà, A.; Tsafou, K.; Binder, J. X.; Jensen, L. J. DISEASES: Text mining and data integration of disease–gene associations. Methods, **2015**, *74*, 83-89.

[77] Mottaz, A.; David, F. P.; Veuthey, A. L.; Yip, Y. L. Easy retrieval of single amino-acid polymorphisms and phenotype information using SwissVar. *Bioinformatics,* **2010**, *26*(6), 851-852.

[78] Barabasi, A. L.; Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.,* **2004**, *5*(2), 101-113.

[79] Makinen, V. P.; Civelek, M.; Meng, Q.; Zhang, B.; Zhu, J.; Levian, C.; Huan, T.; Segre, A. V.; Ghosh, S.; Vivar, J.; Nikpay, M.; Stewart, A. F.; Nelson, C. P.; Willenborg, C.; Erdmann, J.; Blakenberg, S.; O'Donnell, C. J.; Marz, W.; Laaksonen, R.; Epstein, S. E.; Kathiresan, S.; Shah, S. H.; Hazen, S. L.; Reilly, M. P.; Coronary, A. D. G.-W. R.; Meta-Analysis, C.; Lusis, A. J.; Samani, N. J.; Schunkert, H.; Quertermous, T.; McPherson, R.; Yang, X.; Assimes, T. L. Integrative genomics reveals novel molecular pathways and gene networks for coronary artery disease. *PLoS Genet.,* **2014**, *10*(7), e1004502.

[80] Zhao, J.; Geng, C.; Tao, L.; Zhang, D.; Jiang, Y.; Tang, K.; Zhu, R.; Yu, H.; Zhang, W.; He, F.; Li, Y.; Cao, Z. Reconstruction and analysis of human liver-specific metabolic network based on CNHLPP data. *J. Proteome Res.,* **2010**, *9*(4), 1648-1658.

[81] Wang, J.; Sen, S. MicroRNA functional network in pancreatic cancer: From biology to biomarkers of disease. *J. Biosci.,* **2011**, *36*(3), 481-491.

[82] Mathur, S.; Dinakarpandian, D. Drug repositioning using disease associated biological processes and network analysis of drug targets. *AMIA Annu. Symp. Proc.,* **2011**, *2011*, 305-311.

[83] Ye, H.; Yang, L.; Cao, Z.; Tang, K.; Li, Y. A pathway profile-based method for drug repositioning. *Chinese Sci. Bulletin,* **2012**, *57*(17), 2106-2112.

[84] Pan, Y.; Cheng, T.; Wang, Y.; Bryant, S. H. Pathway analysis for drug repositioning based on public database mining. *J. Chem. Inf. Model.,* **2014**, *54*(2), 407-418.

[85] Li, J.; Lu, Z. Pathway-based drug repositioning using causal inference. *BMC Bioinformatics,* **2013**, *14*(16), 1-10.

[86] Draghici, S.; Khatri, P.; Tarca, A. L.; Amin, K.; Done, A.; Voichita, C.; Georgescu, C.; Romero, R. A systems biology approach for pathway level analysis. *Genome Res.,* **2007**, *17*(10), 1537-1545.

[87] Kogenaru, S.; del Val, C.; Hotz-Wagenblatt, A.; Glatting, K.-H. TissueDistributionDBs: a repository of organism-specific tissue-distribution profiles. *Theor. Chem. Acc.,* **2010**, *125*(3-6), 651-658.

[88] Liu, X.; Yu, X.; Zack, D. J.; Zhu, H.; Qian, J. TiGER: a database for tissue-specific gene expression and regulation. *BMC Bioinformatics,* **2008**, *9*, 271.

[89] Pan, J. B.; Hu, S. C.; Shi, D.; Cai, M. C.; Li, Y. B.; Zou, Q.; Ji, Z. L. PaGenBase: a pattern gene database for the global and dynamic understanding of gene function. *PLoS One,* **2013**, *8*(12), e80747.

[90] Xiao, S. J.; Zhang, C.; Zou, Q.; Ji, Z. L. TiSGeD: a database for tissue-specific genes. *Bioinformatics,* **2010**, *26*(9), 1273-1275.

[91] Uhlen, M.; Oksvold, P.; Fagerberg, L.; Lundberg, E.; Jonasson, K.; Forsberg, M.; Zwahlen, M.; Kampf, C.; Wester, K.; Hober, S.; Wernerus, H.; Bjorling, L.; Ponten, F. Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.,* **2010**, *28*(12), 1248-1250.

[92] Nikitin, A.; Egorov, S.; Daraselia, N.; Mazo, I. Pathway studio—the analysis and navigation of molecular networks. *Bioinformatics,* **2003**, *19*(16), 2155-2157.

[93] Pathway information provided by BioCarta. Available online: http://cgap.nci.nih.gov/Pathways/BioCarta_Pathways (Accessed on Feb25, 2015).

[94] Mi, H.; Muruganujan, A.; Thomas, P. D. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.,* **2013**, *41*(Database issue), D377-D386.

[95] Croft, D.; Mundo, A. F.; Haw, R.; Milacic, M.; Weiser, J.; Wu, G.; Caudy, M.; Garapati, P.; Gillespie, M.; Kamdar, M. R.; Jassal, B.; Jupe, S.; Matthews, L.; May, B.; Palatnik, S.; Rothfels, K.; Shamovsky, V.; Song, H.; Williams, M.; Birney, E.; Hermjakob, H.; Stein, L.; D'Eustachio, P. The Reactome pathway knowledgebase. *Nucleic Acids Res.,* **2014**, *42*(Database issue), D472-D477.

[96] Schaefer, C. F.; Anthony, K.; Krupa, S.; Buchoff, J.; Day, M.; Hannay, T.; Buetow, K. H. PID: the Pathway Interaction Database. *Nucleic Acids Res.,* **2009**, *37*(Database issue), D674-D679.

[97] Cerami, E. G.; Gross, B. E.; Demir, E.; Rodchenkov, I.; Babur, O.; Anwar, N.; Schultz, N.; Bader, G. D.; Sander, C. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.,* **2011**, *39*(Database issue), D685-D690.

[98] Donato, M.; Xu, Z.; Tomoiaga, A.; Granneman, J. G.; Mackenzie, R. G.; Bao, R.; Than, N. G.; Westfall, P. H.; Romero, R.; Draghici, S. Analysis and correction of crosstalk effects in pathway analysis. *Genome Res.,* **2013**, *23*(11), 1885-1893.