In Silico Systems Pharmacology to Assess Drug's Therapeutic and Toxic Effects

Orozco, Alejandro Aguayo; Audouze, Karine; Brunak, Soren; Taboureau, Olivier

Published in:
Current Pharmaceutical Design

DOI:
10.2174/1381612822666160907093215

Publication date:
2016

Document version
Peer reviewed version

Send Orders for Reprints to reprints@benthamscience.ae

In silico systems pharmacology to assess drug's therapeutic and toxic effects

Alejandro Aguayo Orozco[a], Karine Audouze[b], Søren Brunak[a], Olivier Taboureau[*a,b].
[a]*Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark;* [b]*UMRS 973 Inserm, Université Paris Diderot, Sorbonne Paris Cité, Paris 75013, France.*

Abstract: For many years, the "one target, one drug" paradigm has been the driving force behind developments in pharmaceutical research. With the recent advances in molecular biology and genomics technologies, the focus is shifting toward "drug-holistic" systems based approaches (i.e. systems pharmacology). The integration of large and diverse amount of data from chemistry and biology coupled with the development and the application of network-based approaches to cope with these data is the next paradigm of drug discovery. Systems pharmacology offers a novel way of approaching drug discovery by developing models that consider the global physiological environment of protein targets and their modification by drugs. Studying drug action across multiple scales of complexity from molecular and cellular to tissue and organism levels may help identify new druggable disease genes and to design new drugs with a better efficacy and clinical safety.

The abstract should not exceed 250 words for review papers summarizing the essential features of the article.

Keywords: Systems Pharmacology, biological network, drug, protein-protein interactions, pathways, gene expression, pharmacogenomics, toxicity.

1. INTRODUCTION

Although the number of new drug candidates reaching the market has been stable since 2010 (with 40 new molecular entities appearing each year), the attrition rate remains high in late stage clinical development phases (II and III) leading to a key challenge in drug discovery [1]. The main reasons for this attrition in drug development are essentially due to a lack of efficacy and the clinical safety (toxicology) [2]. Furthermore, the conventional assumption that selective ligands act on a single target has slowed down the new drug discovery and development process in the past years, especially for complex diseases like cancers, neurological disorders and diabetes [3-5].

Recent advances in chemical biology and systems biology have shown that most drugs interact with multiple targets and although activities against several targets might be beneficial, the drawback is that it can also lead to dramatic toxicity effects [6,7]. Within the age of big data and with the advances in genomic technologies, Next-Generation Sequencing (NGS) and Genome-Wide Association Studies (GWAS), together with the improvements in mass spectroscopy, large amounts of data on genome-wide gene expression profiles, proteins and their interactions with other biomolecules are being provided [8]. Therefore, the mechanisms of action and the safety of drugs can be explored not only at the molecular level but also at the level of the whole biological system, i.e. systems pharmacology.
Systems pharmacology considers drug actions and side effects in a regulatory network context, drug target and disease gene product related, that is understanding the mechanism underlying the multiple actions of drugs. In fact, multi-target drugs exponentially increase the number of pharmacologically relevant target molecules. Low-affinity binding of multi-target drugs eases the restrictions of druggability, and increases the size of the drugable proteome [9-12].
For example, Imatinib, a drug used in the treatment of chronic myelogenous leukemia (CML) is reported to block the activity of several nonspecific tyrosine kinases [13]. Rosiglitazone, which has been used for the treatment of type II diabetes mellitus, not only stimulates the peroxisome proliferator-activated receptor gamma, but also blocks interferon gamma-induced chemokine expression in Graves' disease or ophthalmopathy [14-15].

Computational approaches and network biology are essential components of systems pharmacology, which in turn can generate testable hypotheses from large and diverse amounts of data. Network biology helps to identify the biological mechanism associated with a disease or a treatment at several layers of complexity (from molecular and cellular to tissue, organs and systems) based on the concept that the functions of molecular components in a human cell are closely connected and thus a disease is not only a consequence of genetic variation but also a result of perturbations of intracellular and extracellular networks linking tissue and organ systems [16]. Targeting a small number of nodes by partial inhibition may be more efficient than the total inhibition of a single node. So, network biology analysis predictions, point to the capacity to perturb robust phenotypes by modulating multiple proteins, instead

XXX-XXX/14 $58.00+.00

of deleting individual nodes (gene or protein) within disease networks.

Systems pharmacology is also applied to assess the toxicity of a compound (i.e. systems toxicology). In systems toxicology, molecular changes in the context of an exposure are measured quantitatively, thereby deciphering the casual chain of molecular events across different levels of organization that link exposure to adverse outcomes. Interestingly, such an approach has been gaining attention over the past few years through a program launched by the OECD in 2012 for the development of Adverse Outcome Pathways (AOP) [17]. With the opportunity to integrate and to combine all this information in systems pharmacology-based drug safety predictions, it is expected that some of the current challenges in drug discovery can be tackled.

Here, we will review the status of systems pharmacology and showcase data currently exploited in different studies as well as the network analysis considered to decipher the mechanisms of drug action. All the data sources cited are accessible through the hyperlinks in listed table 1.

2. Towards systems level data integration

2.1 Drug-Target

The role of small molecules in biological systems can only be understood in relation to their targets' functions that can be defined as molecular events. Therefore data and knowledge on the interaction between proteins and small molecules are necessary in order to understand molecular and cellular functions [18]. However, chemical-protein interaction information is widely spread across various data sources. Over many years, the interaction of chemicals with their targets have been studied in both biochemistry and pharmacology, but much of the existing data is dispersed in the vast amounts of literature, locked up in commercial databases or sequestered in private datasets. This information is now compiled, regularly updated and available in large databases such as PubChem [19], ChEMBL [20], ChemProt [21], DrugBank [22] or OpenPHACTS [23] giving a better overview of the available knowledge that drives system level approaches. Based on this data, *in silico* chemogenomic methods have been developed for predicting the polypharmacological profiles of bioactive compounds and for the identification of potential new targets (target fishing approach) [24-25].

Another crucial step in target assessment is the quantification of the likelihood of discovering a therapeutic molecule that is at the same time safe, and efficacious. For this reason it is important to gather information of this kind in databases bbased on the toxicity of the compounds, together with the interaction of the small molecules with the possible targets of a relevant disease-pathway.

Various research programs in U.S.A., Europe and Japan have been for several years focusing their efforts on high-throughput screening technologies to address the current lack of toxicity evaluation of thousands of chemicals, in order to improving drug-target information. The U.S. Environmental Protection Agency (EPA) ToxCast research program and Tox21, involve coordinated effort of governmental regulatory and research entities to use high-throughput bioassays aiming to characterize key elements of toxicity pathways, and key biological events that may suppose potential targets for chemicals whose interactions may lead to disease [26]. Its goal is to acquire enough information on a range of chemicals to evaluate their bioactivity profiles and predict possible patterns of toxic effects and phenotypes that correlate with observed *in vivo* toxicity [27-28].

From the European side, SEURAT-1 (Safety Evaluation Ultimately Replacing Animal Testing-1) is a research program whose underlying assumption is the possibility to identify mechanisms of action (MOAs) that are relevant to human toxicity, based on existing knowledge such as adverse drug reactions (ADRs) of marketed drugs. The SEURAT-1 project's goal was to establish *in vitro* assays to characterize and represent these MOAs. The experimental results supporting SEURAT-1 are stored in a web-accessible shared repository called the ToxBank Data Warehouse [29]. Recently, an H2020-supported collaborative project called EU-ToxRisk has been established to translate molecular mechanistic understanding of toxicity into safety testing strategies (http://www.eu-toxrisk.eu/).

2.2 Chemogenomics and Toxicogenomics

Chemogenomics and Toxicogenomics have suffered from a lack of large publicly available standardized datasets. This problem has been addressed in the past years by the release of two highly important datasets: the TGP (TG-GATEs) [30] dataset and DrugMatrix [31]. The first is a toxicogenomics database that stores gene expression profiles together with traditional toxicological data obtained from *in vivo* (rat) and *in vitro* (primary rat hepatocytes, primary human hepatocytes) exposure to different compounds at multiple doses and time points. DrugMatrix covers drug-dose-time-tissue combination profiles for approximately 600 different compounds administered to rats. They demonstrate that the use of both traditional toxicity measurements together with gene expression analysis enriches the understanding of individual compound effects.

While several toxicogenomics projects made their data available via public databases, such as ArrayExpress [32], Gene Expression Omnibus (GEO) [33] and Expression Atlas [34], data from other projects are more difficult to access. Moreover, different experimental designs make it difficult to compare and to analyze the results. diXa [35] is a warehouse that aims to overcome these drawbacks by

defining standard workflows for data preprocessing and standard formats for metadata annotation. Besides, diXa integrates information from toxicology, chemistry and human disease databases along with the original data, enhancing and easing the interpretation of data analysis results.

Recently, the National Institutes of Health (NIH) provided a Library of Integrated Network-based Cellular Signatures (LINCS) containing gene expression profiles for more than 11,000 compounds measured on several cell lines [36]. Such data reporting how a drug induces changes in gene expression in different cell-lines coupled with network biology and phenotypes can lead to a better understanding of drug action and toxicities.

2.3 Protein-Protein Interactions & Drug-Pathways

One major challenge in the post-genomic era is integration of experimental and computational data into specific biological pathways, to achieve the understanding of higher-level complexities of molecular mechanisms in cellular processes including the final phenotype. Towards this end some databases have been developed that contain wired gene-gene and protein-protein information in interaction pathways. How is a protein or a gene that has been related to an experimental stress or cell disturbance connected to a physiological mechanism and a disease phenotype? Pathway databases help make this connection by enabling the connection between drugs, genes and proteins in a model. KEGG (Kyoto Encyclopedia of Genes and Genomes) contains protein networks stored as a collection of pathway maps that represent wired diagrams of proteins and other biomolecules responsible for the cellular functions [37]. Other databases contain annotated interactions in the form of pathways from the curation of peer-reviewed journals, such as Reactome [38], WikiPathways [39] or PID (The Pathway Interaction Database [40]. Furthermore, since proteins infrequently operate in isolation but rather function in highly interconnected cellular pathways, integration of protein-protein interactions (PPI), derived essentially from high-throughput approaches, including yeast two-hybrid screens, immunoprecipitation studies followed by mass spectrometry analysis or small scale experiments have become a valuable resource of information to implemented in network biology and systems pharmacology. Databases such as HPRD [41], Intact [42] or STRING [43] have been extensively used to identify new proteins associated to a phenotype [44-45]. They enable the creation of large PPI networks that can help to gain insight into disease and phenotype mechanisms.

2.4 Tissue/Organ specificity

High-throughput sequence methods have made it possible to explore gene expression patterns genome-widely, thereby offering the opportunity to describe the expression of gene within specific human organs and tissues. Such data can facilitate a better comprehension of the differences in the expression at the specific point where a disease takes place. Organ-specific patterns of gene expression may be of importance due to the promiscuity of genes, as they are expressed in many different tissues in the organism and at different levels. For example, Ponten F, et al. demonstrated a different protein expression profile specific to tissue based on an analysis of over 2 million immunohistochemistry images annotated by pathologists [46]. This study demonstrated as well that proteins expressed in a well-defined set of cell types are rather important for the function of those cells [47].

Databases such as HOMER (Human Organ-specific Molecular Electronic Repository) [48], HPA (Human Protein Atlas) [49], GNF tissue atlas [50] or Pagenbase [51] aim to cover gene expression information in an organ-specific manner. The information contained is filtered from organ-specific gene/proteins and disease as a result of the cross-linking of several available data sources.

2.5 Phenotypic outcomes

When trying to understand the pathological phenotype it is also necessary to make the connection drug-disease and drug-target-disease. Hence databases connecting pathways, targets and molecules to unhealthy phenotypes are necessary. These phenotypes may be disease or toxicological side effects produced by the effect of a small molecule perturbation. There are many online sources that contain clinical information and allow the collection of data from clinicians, patients and pharmaceutical companies.

Among others, a public computer-readable side effect resource (SIDER) that connects 888 drugs to 1,450 side effect terms was developed [52]. The Comparative Toxicogenomics Database (CTD) [53] is another example of a derived resource that includes curated information and data analysis tools about gene-chemical and chemical-disease associations that promote the understanding of the effects of chemicals on human health. AERSs (created by the US FDA, World Health Organization, and Health Canada [54]), EudraVigilance (a system designed for collecting reports of suspected side effects created by the European medicines agency [55]), and JAPIC (which manages all package insert information of pharmaceutical products in Japan, under the approval of Health and Welfare Minister of Japan [56]) are other valuable data sources. However, there is no common terminology for therapeutic effects and side effects in each organization. Thus, to facilitate the coding of "regulatory data" in biopharmaceutical development and clinical trials, and the reporting of therapeutic and side effects, several dictionaries (terminologies) have been developed. The most widely used are MeSH (Medical Subject Headings [57]), MedDRA (Medical Dictionary for Regulatory Activities [58]), ICD-10 (International classification of diseases [59]), SNOMED CT (Systematized Nomenclature of Medicine- Clinical Terms [60]), the ATC

Classification System (Anatomical Therapeutic Chemical Classification System [61]), UMLS (Unified Medical Language System [62]), and J-ART (Japanese Adverse Reaction Terminology).

3. Systems pharmacology

3.1 From Target to Protein-Protein Interactions and Pathways Using Network Biology Approaches

A disease can not only be a consequence of a genetic variation but might be also be a consequence of perturbations of complex intracellular and extracellular networks linking tissue and organ systems [16]. In fact, multiple changes in the same or different protein complexes and pathways can contribute to a disease. So the perception that human diseases are associated with a few dominant factors (reductionist view) is now replaced by a view of diseases as the outcome of many weak contributors (holistic view).

The systems pharmacology method emerged as an all-inclusive approach that analyzes events from all the interactions point of view, helping to understand the interaction mechanism between drugs and complex diseases (Figure 1) [63]. In systems pharmacology, the data are centered into network-based associations with an entity (drug, protein, gene, cell, phenotype, …) defined by a node and the connection (interaction, association, deregulation, …) between two entities representing by an edge. These edges can be weighted and of different types to specify the degree and the direction of the connection. For example, drugs-target proteins can be compiled in a network based on protein-protein interactions data. Nodes define drugs and proteins, and edges define drug-protein and protein-protein interactions. Then, including protein expression in cells and/or tissues, or phenotypes associated for each protein can complicate such network, but add to their usefulness. Similarly, differentiation of genes by a set of drugs can be embedded within a network. Such data can then be enriched with pathway information and connected to some physiological dysfunction observed for these drugs, allowing identifying the drug's action. Each physiological process is known to be regulated by signaling networks of chemicals and biomolecules. These reactions have different kinetics and time points, and complex networks control each phenotype or physiological function. For example, it is known that a drug is characterized by an Adverse Drugs Reaction (ADR) profile, which can be described as a toxicity network. This network may consist of one or more sites of actions that may also be dependent or independent of one another. Looking for example to hepatotoxicity, there is large panel of physiological features that can be generated by a particular drug and lead to drug-induced liver injury (DILI). Among them we notice, hepatitis, cholestasis, steatosis, hepatic granuloma or hepatic necrosis. Interestingly, although causing the same ADR, the drugs will not necessary interact with the same set of proteins which will not perturb the same pathways and consequently might lead to different physiological effects associated to DILI. With systems pharmacology, it is possible to represent the drug-target-pathway-physiological interactions in a network and to analyze the potential link and overlap between all the molecular events leading to DILI. By extension of this concept, a systematic association between drug, targets, clinical outcomes and side effects can be integrated in a network leading to the identification of novel drug actions [64].

The release of large set of gene expression data contributes also in Systems pharmacology. Based on the differential expression of more than 1,300 drugs to four human cell lines, the Connectivity Map (CMap) project produced gene expression signatures allowing for comparison of small molecules sharing similar physiological processes and diseases [65]. Calvert et al. used systems pharmacology/toxicology to prove the capabilities of this approach in drug repositioning. They used CMap to identify drugs with overlapping gene expression profiles with caloric restriction (CR), shown to induce retard aging in several animal models. Based on their longevity network analysis, they suggested eleven different compounds of which five were tested in *C. elegans* and four produced a lifespan effect [66]. Segura-Cabrera et al. developed a network framework that combines biomolecular interactions and known drug-target interactions for the prioritization of drugs, genes and pathways according to the biological context. Thereby demonstrating that various gene expression data or drug-target screening can prioritize drugs and pathways, respectively [67]. Finally, Chen et al. assessed if structurally similar compounds have similar cellular responses using the chemical structures and gene expression profiles of 11,000 compounds from LINCS. They concluded that two compounds tend to share similar gene expression profiles in cell lines for ~20% of the data [68].

3.2 Biological enrichment- diseases- toxicities classification limitation

The more the data is integrated in biological network, the more complex is the analysis. To overcome this challenge, unsupervised and supervised graph theory approaches have been developed. Subnetworks (hubs) on highly connected nodes can be identified using for example MCODE [69] or GIANT [70] with the help of Cytoscape. Biological enrichment is another possibility to organize the entities. Typically, when a set of genes is identified as interesting in relation to a disease, these genes are analyzed in the context of their PPI networks. Further analysis is usually carried out to enrich these networks with known pathways and disease-associations. There are quite a lot of biological human pathway resources, but most of them do not contain disease, drug or tissue specific information. Therefore computer programming, data curation and biostatistics are necessary for this data enrichment and data

characterization to lead to impact. Zhang et al. developed IPAD (Integrated Pathway Analysis Database), a resource for systematic enrichment analysis, by analyzing, identifying and validating drug, pathway, organ, disease and their associations [9]. More recently Handen et al. developed a tool called Lens for Enrichment and Network Studies of human proteins (LENS that executes systematic network, pathway and disease enrichment analysis on genes of interest [71].

With the integration of tissue information, biological enrichment studies have been reported. For example, a combination of gene expression in human protein complexes revealed tissue specificity and pathology [72]. Similarly, McCall et al. [73] leveraged data from the GEO and ArrayExpress public repositories to build statistical models for the most annotated genes for 131 human and 89 mouse tissue types to address which genes are expressed in a given cell type. Petrovskiy et al. presented an approach to assess tissue-specific gene knockout effects through target-centric gene network. The model connects the expression of the group of target genes with their expression with machine learning models trained on expression data [74]. Such information is useful in understanding the chemical mechanism of action linked to the drug tolerance, drug efficacy, side effects, and risk assessment.

Systems pharmacology studies have also been reported including Genome Wide Association Studies (GWAS). Sun et al. [75] compared network properties of the genes causing a disease and drug targets for five major disease categories (cancer, cardiovascular diseases, metabolic diseases, immune system disease, and nervous system disease). They collected the disease genes from genome wide association studies (GWAS) and their corresponding drugs based on drugs' Anatomical Therapeutic Chemical (ATC) classification. With a network approach they found that disease genes were significantly enriched in targets, especially for cancers. In another study a human disease network involving disease-disease links among 108 diseases was built based on mRNA expression data and differential co-expression analysis. This enriched network shared known disease genes and drugs more significantly than those based on different expression analysis. Some new disease relationships were discovered, for instance obesity and psoriasis, which have recently been found to share similar molecular mechanisms. Additionally, it was also found that both the type of disease and the tissue affected influenced the degree of disease similarity. This led to a global perspective of the human diseasome from the viewpoint of regulation mechanisms [76]. Finally, Zickenrott et al. introduced a novel network-based approach for predicting target genes and other bioactive compounds that could revert disease phenotypes, throughout the reconstruction of gene regulatory networks (GNRs) corresponding to both disease and healthy phenotypes [77]. To validate it they predicted drug candidates for Rheumatoid arthritis. The enriched genes in the core of the disease network were targeted by copper sulfate and cyclosporine. Theses results confirmed a previous study [78].

3.3 Future research in the field of Systems pharmacology

3.3.1) Dose-response

In the context of a drug it is relevant not only to investigate the action of the drug at the target site, but also understand how the drug reached the target and the effective concentration. Taking a mechanistic view of the involved processes, pharmacokinetics/ pharmacodynamics (PK/PD) models might be useful to model the signaling network. It is important to study the dose together with time-response and gene expression, as it may explain why some diseases, such as cancer may respond to a treatment until at a certain point the treatment fails. Robinson et al. using a systems-based toxicogenomics approach assessed quantitatively dose- and time- dependent effects on gene expression, in enriched GO biological processes perturbed by MeHg in mouse embryos during cranial neural tube closure. Altered expression was observed for 883 genes, including several previously characterized as crucial for the neural tube development. These genes were associated with specific GO biological processes that may underlie MeHg-induced teratogenic and neurodevelopmental toxicity outcome [79].

With the recent development of experimental tools that can generate sufficiently quantitative and thorough data, mathematical models such as exposure models, physiologically-based pharmacokinetic (PBPK) and biologically-based dose response (BBDR) can be used to approximate ADME (absorption, distribution, metabolism and excretion) processes, thus understanding the dose-time response effects of a drug in a more effective manner [80]. For example, Kirouac et al. developed a mathematical model incorporating cellular networks into PK-PD models to capture systems-level architectural features of oncogenic signaling networks [81].

3.3.2 Patient response variability

With the concept that everyone responds to drug differently due to genetic predisposition, pharmacogenetics and pharmacogenomics began to be carried out. The objective is to explain why an individual responses differently to a drug therapy from being beneficial to an almost complete lack of therapeutic efficacy [82-83]. As knowledge of human genomics and therapeutics grows, it has become clear that drug response phenotypes are complex pathways, involving genes in pharmacokinetics and pharmacodynamics pathways and targets in downstream signaling parts of the pathways. Schizophrenia is an interesting example. One-third of the patients fail to respond favorably to a drug treatment. In addition, adverse drug reactions and side effects are often associated to the antipsychotics treatment leading to a lack of tolerability. Based on GWAS studies, a large panel of genetic associations and variations implicated in this pathology has

been suggested. In combination to pathway and protein-protein interactions, patterns of genes belonging to the same biological pathways or on genes that show evidence of coregulation through gene expression analysis can be selected [84]. Two of them, Catechol-o-methyl transferase (COMT) and serotonin receptor 5-HT1A are drug-target considered for the treatment. However, it has been shown that the specific polymorphism of these two proteins had a negative impact on patient treated with clozapine, the gold standard drug treatment for schizophrenia [85]. Not only, multiscale network modeling approaches can be used to understand the etiology of a disease but they can also suggest the risk of non-efficacy due to drug target polymorphism.

4. CONCLUSION

With the emergence of big data and the advances in pharmacology that have taken place in the past years, there is a pressure to maximize drug efficacy, reduce toxicity and select responsive patients. Massive amount of data are generated and accumulated by new experimental technologies such as transcriptomics and genetics. Furthermore, centralized systems that facilitate the integration and standardization of diverse federated resources are in development. So, drug action can be explored across multiple scale of complexity, from molecular and cellular to tissue and organism levels over time and dose. With the development of new mathematical models and network-based strategies, systems pharmacology allows the assessment of the chemical effect at the biological systems (i.e., systems pharmacology and systems toxicology) more comprehensively.

At the population level, understanding the individual differences in drug response in the context of biological networks is the new challenge in pharmacogenomics and personalized medicine. To do so, methods that integrate drug-target, clinical-outcome, and genetic factors using network biology have started to be reported [86-88]. Such analysis would definitively contribute to a better understanding of the variability in drug response and a more personalized approach to therapy.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest

ACKNOWLEDGEMENTS

We would like to acknowledge EU-ToxRisk, a project funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 681002.

REFERENCES

[1] Waring MJ, Arrowsmith J, Leach AR, Leeson PD, Mandrell S, Owen RM, *et al.* An analysis of the attrition of drug candidates from four major pharmaceutical companies. Nat. Rev. Drug. Discov 2015; 14(7): 475-486.

[2] Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nat Biotechnol 2014; 32: 40-51.

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

[4] Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov 2004; 3(8): 711-715.

[5] Sams-Dodd F. Target-based drug discovery: is something wrong? Drug Discov Today 2005; 10(2):139-147.

[6] Crema F, Modini C, Croci T, Langlois M, de Ponti F. Intestinal prokinesia by two esters of 4-amino-5-chloro-2- methoxybenzoic acid: involvement of 5-hydroxytryptamine-4 receptors and dissociation from cardiac effects in vivo. J Pharmacol Exp Ther 1999; 288:1045-1052.

[7] Walker BD, Singleton CB, Bursill JA, Wyse KR, Valenzuela SM, Qiu MR, *et al*. Inhibition of the human ether-a-go-go-related gene (HERG) potassium channel by cisapride: affinity for open and inactivated states. Br J Pharmacol 1999;128: 444-450.

[8] Chandra N, Padiadpu J. Network approaches to drug discovery. Expert Opin Drug Discov 2013; 8(1): 7-20.

[9] Zhang F, Drabier R. IPAD: the Integrated Pathway Analysis Database for systematic enrichment analysis. BMC Bioinformatics 2012; 13 Suppl 15: S7.

[10] Bai JP, Abernethy DR. Systems pharmacology to predict drug toxicity: integration across levels of biological organization. Annu Rev Pharmacol Toxicol 2013; 53: 451-473.

[11] Berger SI, Iyengar R. Network analyses in systems pharmacology. Bioinformatics 2009; 25(19): 2466-2472.

[12] Zhou W, Wang Y, Lu A, Zhang G. Systems pharmacology in small molecular drug discovery. Int J Mol Sci 2016: 17(2).

[13] Kerkela R, Grazette L, Yacobi R, Iliescu C, Patten R, Beahm C, *et al*. Cardiotoxicity of the cancer therapeutic agent imatinib mesylate. Nat Med 2006; 12: 908-916.

[14] Antonelli A, Ferrari SM, Fallahi P, Piaggi S, Paolicchi A, Franceschini SS, *et al*. Cytokines (interferon-gamma and tumor necrosis factoralpha)-induced nuclear factor-kappaB activation and chemokine (C-X-C motif) ligand 10 release in Graves' disease and ophthalmopathy are modulated by pioglitazone. Metabolism 2011; 60: 277-283.

[15] Antonelli A, Ferrari SM, Frascerra S, Ruffilli I, Pupilli C, Bernini G, *et al*. β (CCL2) and α (CXCL10) chemokine modulations by cytokines and peroxisome proliferator-activated receptor-α agonists in Graves' ophthalmopathy. J Endocrinol 2012; 213:183-191.

[16] Barabasi AL, Oltvai Z. Network Biology: understanding the cell's functional organization. Nat Rev Genet 2004; 5: 101-113.

[17] Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, *et al*. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environmental Toxicology and Chemistry 2010; 29(3): 730-741.

[18] Sharan R, Ulitsky I, Shamir R. Network-based prediction of protein function. Mol Syst Biol 2007; 3: 88.

[19] Chen B, Wild D, Guha R. PubChem as a source of polypharmacology.

Formatted: English (US)

J Chem Inf Model 2009; 49: 2044-2055.

[20] Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, *et al*. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 2011; 40(Database issue): D1100-D1107.

[21] Kringelum J, Kjaerulff SK, Brunak S, Lund O, Oprea TI, Taboureau O. ChemProt-3.0: a global chemical biology diseases mapping. Database 2016; bav123

[22] Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, *et al.* Drugbank 4.0: shedding new light on drug metabolism. Nucleic Acids Res 2014; 42(1): D1091-D1107.

[23] Azzaoui K, Jacoby E, Senger S, Rodriguez EC, Loza M, Zdrazil B, et al. Scientific competency questions as the basis for semantically enriched open pharmacological space development. Drug Discov Today 2013; 18(17-18): 843-852.

[24] Lavecchia A, Cerchia C. In silico methods to address polypharmacology: current status, applications and future perspectives. Drug Discov Today 2016; 21(2): 288-298.

[25] Masoudi-Nejad A, Mousavian Z, Bozorgmehr JH, Drug-target and disease networks: polypharmacology in the post-genomic era. In Silico Pharmacol 2013; 1: 17.

[26] Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ. The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci 2007; 95(1): 5-12.

[27] Kavlock R, Chandler K, Houck K, Hunter S, Judson R, Kleinstreuer N, *et al*. Update on EPA's ToxCast program: providing high throughput decision support tools for chemical risk management. Chem Res Toxicol 2012; 25(7): 1287-1302.

[28] Mahadevan B, Snyder RD, Waters MD, Benz RD, Kemper RA Tice RR, Richard AM. Genetic toxicology in the 21st century: reflections and future directions. Environ Mol Mutagen 2011; 52(5): 339-354.

[29] Kohonen P, Benfenati E, Bower D, Ceder R, Crump M, Cross K, *et al*. The ToxBank Data Warehouse: Supporting the Replacement of In Vivo Repeated Dose Systemic Toxicity Testing. Molecular Informatics 2013; 32(1): 47-63.

[30] Igarashi Y, Nakatsu N, Yamashita T, Ono A, Urushidani T, Yamada H. Open TG-GATEs: a large-scale toxicogenomics database. Nucleic Acids Res 2015; 43(Database issue): D921-D927.

[31] Ganter B, Tugendreich S, Pearson CI, Ayanoglu E, Baumhueter S, Bostian KA, *et al.* Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. J Biotechnol 2005; 119(3): 219-244.

[32] Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, *et al*. ArrayExpress--a public repository for microarray gene expression data at the EBI. Nucleic Acids Res 2003; 31(1): 68-71.

[33] Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, *et al*. NCBI GEO: archive for functional genomics data sets-update. Nucleic Acids Res 2013; 41: D991-995.

[34] Petryszak R, Keays M, Tang YA, Fonseca NA, Barrera E, Burdett T, *et al.* Expression Atlas update-an integrated database of gene and protein expression in humans, animals and plants. Nucleic Acids Res 2016; 44: D746-52.

[35] Hendrickx DM, Aerts HJ, Caiment F, Clark D, Ebbels TM, Evelo CT, *et al*. diXa: a data infrastructure for chemical safety assessment. Bioinformatics 2015; 31(9): 1505-7.

[36] Liu C, Su J, Yang F, Wei K, Ma J, Zhou X. Compound signature detection on LINCS 1000 big data. Mol. Biosys 2015; 11, 714-722.

[37] Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic Acids Res 2004; 32: D277-D280.

[38] Haw R, Hermjakob H, D'Eustachio P, Stein L. Reactome pathway analysis to enrich biological discovery in proteomics data sets. Proteomics 2011; 11(18): 3598-613.

[39] Kelder T, van Iersel MP, Hanspers K, Kutmon M, Conklin BR, Evelo CT, Pico AR. WikiPathways: building research communities on biological pathways. Nucleic Acids Res 2012; 40: D1301-1307.

[40] Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, *et al*. PID: the Pathway Interaction Database. Nucleic Acids Res 2009; 37: D674-D679.

[41] Mishra G, Suresh M, Kumaran K, Kannabiran N, Suresh S, Bala P, et al. Human protein reference database – 2006 update. Nucleic Acids Res 2006; 34: D411-D414.

[42] Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, Orchard S, *et al*. IntAct: an open source molecular interaction database. Nucleic Acids Res 2004; 32: D452-D455.

[43] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015; 43: D447-D452.

[44] Lage K, Karlberg EO, Størling ZM, Olason OI, Pedersen AG, Rigina O, et al. A human phenome-interactome network of protein complexes implicated in genetic disorders. Nat. Biotechnol 2007; 25: 309-316.

[45] Kim YA, Cho DY, Przytycka TM. Understanding Genotype-Phenotype effects in Cancer via network approaches. PLoS Comput. Biol 2016; 12(3): e1004747.

[46] Ponten F, Gry M, Fagerberg L, Lundberg E, Asplund A, Berglund L, *et al*. A global view of protein expression in human cells, tissues and organs. Mol Syst Biol 2009; 5: 337.

[47] Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, et al. Proteomics. Tissue-based map of the human proteome. Science 2015; 347 (6220):1260419.

[48] Zhang F, Chen JY. HOMER: a human organ-specific molecular electronic repository. BMC Bioinformatics 2011; 12: S4.

[49] Ponten F, Jirström K, Uhlen M. The human protein atlas – a tool for pathology. J. Pathol. 2008; 216: 387-393.

[50] Su AI, Wiltshire T, Batalov S, Lapp H, Ching, *et al*. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc. Natl. Acad. Sci. USA 2004; 101: 6062-6067.

[51] Pan JB, Hu SC, Shi D, Cai MC, Li YB, et al. PaGenBase: a pattern gene database for the global and dynamic understanding of gene function. PLoS One 2013; 8(12): e80747.

[52] Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. Nucleic Acids Res. 2016; 44(D1):D1075-1079.

[53] Davis AP, et al. The Comparative Toxicogenomics Database's 10th year anniversary: update 2015. Nucleic Acids Res 2015; 43: D914-920.

[54] Szarfman A, Tonning JM, Doraiswamy PM. Pharmacovigilance in the 21st century: new systematic tools for an old problem. Pharmacotherapy 2004; 24:1099-1104.

[55] Vermeer NS, Straus SM, Mantel-Teeuwisse AK, Domergue F, Egberts TC, Leufkens HG, et al. Traceability of biopharmaceuticals in spontaneous reporting systems: a cross-sectional study in the FDA Adverse Event Reporting System (FAERS) and EudraVigilance. Drug Saf 2013; 36:617-625.

Formatted: English (US)

Formatted: Spanish

Formatted: Swedish

[56] Shiokawa T. Background, introduction, and activity of the Japan Primary Registries Network. J Evid Based Med 2009; 2: 41-43.

[57] US National Library of Medicine, National Institutes of Health. Medical subject headings, http://www.nlm.nih.gov/mesh/.

[58] Wang L, Jiang G, Li D, Liu H. Standardizing drug adverse event reporting data. Stud Health Technol Inform 2013; 192: 1101.

[59] Hohl CM, Karpov A, Reddekopp L, Stausberg J. ICD-10 codes used to identify adverse drug events in administrative data: a systematic review. J Am Med Inform Assoc 2014; 21: 547-557.

[60] Wasserman H, Wang J. An applied evaluation of SNOMED CT as a clinical vocabulary for the computerized diagnosis and problem list. AMIA Annu Symp Proc 2003; 2003: 699-703.

[61] WHO Collaborating Centre for Drug Statistic Methodology. ATC/DDD Index, ,http://www.whocc.no/atc_ddd_index/. 2015.

[62] Gabetta M, Larizza C, Bellazzi R. A unified medical language system (UMLS) based system for literature-based discovery in medicine. Stud Health Technol Inform 2013; 192: 412-416.

[63] Butcher EC, Berg EL, Kunkel EJ. Systems biology in drug discovery. Nat Biotechnol 2004; 22(10): 1253-1259.

[64] Oprea TI, Nielsen SK, Ursu O, Yang JJ, Taboureau O, Mathias SL, *et al*. Associating drugs, targets, and clinical outcomes into an integrated network affords a new platform for computer-aided drug repurposing. Mol Inform 2011; 30:100-101.

[65] Lamb J, Crawfoed ED, Peck D, Modell JW Blat IC, Wrobel MJ, *et al*. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006; 313(5795): 1929-1935.

[66] Calvert S, Tacutu R, Sharifi S, Teixeira R, Ghosh P, de Magalhaes JP. A network pharmacology approach reveals new candidate caloric restriction mimetics in C. elegans. Aging Cell 2016; 15(2): 256-66.

[67] Segura-Cabrera A., Singh N, Komurov K. An integrated network platform for contextual prioritization of drugs and pathways. Mol Biosyst 2015; 11(11): 2850-2859.

[68] Chen B, Greenside P, Paik H, Sirota M, Hadley D, Butte AJ. Relating chemical structure to cellular response: An integrative analysis of gene expression, bioactivity and structural data across 11,000 compounds. CPT Pharmacometrics Syst Pharmacol 2015; 4: 576-584.

[69] Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 2003; 4: 2.

[70] Cumbo F, Paci P, Santoni D, Di Paola L, Giulani A. GIANT: A cytoscape plugin for modular networks. PLoS One 2014; 9: e:105001.

[71] Handen A, Ganapathiraju MK. LENS: web-based lens for enrichment and network studies of human proteins. BMC Med Genomics 2015; 8 Suppl 4: S2.

[72] Bornigen D, Pers TH, Thorrez L, Huttenhower C, Moreau Y, Brunak S. Concordance of gene expression in human protein complexes reveals tissue specificity and pathology. Nucleic Acids Res 2013; 41: e171.

[73] McCall MN, Uppal K, Jaffee HA, Zilliox MJ, Irizarry RA. The gene expression barcode: leveraging public data repositories to begin cataloging the human and murine transcriptomes. Nucleic Acids Res 2011; 39: D1011-1015.

[74] Petrovskiy ED, Saik OV, Tiys ES, Lavrik IN, Kolchanov NA, Ivanisenko VA. Prediction of tissue-specific effects of gene knockout on apoptosis in different anatomical structures of human brain. BMC Genomics 2015; 16 Suppl 13: S3.

[75] Sun J, Zhu K, Zheng W, Xu H. A comparative study of disease genes and drug targets in the human protein interactome. BMC Bioinformatics 2015; 16 Suppl 5: S1.

[76] Yang J, Wu SJ, Dai WT, Li YX, Li YY. The human disease network in terms of dysfunctional regulatory mechanisms. Biol Direct 2015; 10: 60.

[77] Zickenrott S, Angarica VE, Upadhyaya BB, del Sol A. Prediction of disease-gene-drug relationships following a differential network analysis. Cell Death Dis 2016; 7: e2040.

[78] Wells G, Haguenauer D, Shea B, Suarez-Almazor ME, Welch VA, Tugwell P. Cyclosporine for rheumatoid arthritis. Cochrane Database Syst Rev 2000; 2: CD001083.

[79] Robinson JF, Guerrette X, Yu X, Hong S, Faustman EM. A systems-based approach to investigate dose- and time-dependent methylmercury-induced gene expression response in C57BL/6 mouse embryos undergoing neurulation. Birth Defects Res B Dev Reprod Toxicol 2010; 89(3): 188-200.

[80] Waters MD, Fostel JM. Toxicogenomics and systems toxicology: aims and prospects. Nat Rev Genet 2004; 5(12): 936-948.

[81] Kirouac DC, Onsum MD. Using network biology to bridge pharmacokinetics and pharmacodynamics in Oncology. CPT Pharmacometrics Syst Pharmacol 2013; 2: e71.

[82] Guttmacher AE, and Collins FS. Realizing the promise of genomics in biomedical research. JAMA 2005; 294(11): 1399-402.

[83] Wang L. Pharmacogenomics: a systems approach. Wiley Interdiscip Rev Syst Biol Med 2010; 2(1): 3-22.

[84] Collier DA, Eastwod BJ, Malki K, Mokrab Y. Advances in the genetics of schizophrenia: toward a network and pathway view for drug discovery. Ann N Y Acad Sci 2016; 1366: 61-75.

[85] Bosia M, Lorenzi C, Pirovano A, Guglielmino C, Cocchi F, Spangaro M, Bramanti P, Smeraldi E, Cavallero R. COMT Val158Met and 5-HT1A-R -1019 C/G polymorphisms: effect on the negative symptom response to clozapine. Pharmacogenomics 2015; 16: 35-44.

[86] Chen L, Zhang L, Zhao Y, Xu L, Shang Y, Wang Q, *et al*. Prioritizing risk pathways: a novel association approach to searching for disease pathways fusing SNPs and pathways. Bioinformatics 2009; 25: 237-242.

[87] O'Dushlaine C, Kenny E, Heron EA, Segurado R, Gill M, Morris DW, *et al*. The SNP ratio test: pathway analysis of genome-wide association datasets. Bioinformatics 2009; 25: 2762-2763.

[88] Schlosberg CE, Schwantes-An TH, Duan W, Saccone NL. Application of Bayesian network structure learning to identify causal variant SNPs from resequencing data. BMC Proc 2011; 5: S109.

FIGURES

Figure 1: Representation of the in silico systems pharmacology strategy

TABLES

Table 1: Systems pharmacology data sources

Resrouce Type	Name	URL	Description
Chemical-Protein Databases	PubChem	https://pubchem.ncbi.nlm.nih.gov	Information of the interaction between chemicals and their targets
	ChEMBL	https://www.ebi.ac.uk/chembl	
	ChemProt	http://potentia.cbs.dtu.dk/ChemProt	
	DrugBank	http://www.drugbank.ca	
	OpenPHACTS	https://www.openphacts.org	
	Toxcast & Tox21	https://ncats.nih.gov/tox21	
Chemogenomic and Toxicogenomic databases	TG-GATEs	http://toxico.nibiohn.go.jp	Databases storing the gene expression profiles from *in vivo* and *in vitro* exposure to different compounds, at different concentrations, times or both
	ToxBank	http://toxbank.net	
	DrugMatrix	https://ntp.niehs.nih.gov/drugmatrix	
	ArrayExpress	https://www.ebi.ac.uk/arrayexpress	
	GEO	http://www.ncbi.nlm.nih.gov/geo	
	Expression Atlas	https://www.ebi.ac.uk/gxa/home	
	diXa	http://www.dixa-fp7.eu	
	LINCS	https://lincs.ed.gov	
PPIs and Drug-pathways databases	KEGG	http://www.genome.jp/kegg	These databases have been developed to contain the wired gene-gene and protein-protein interaction information in the form of interaction pathways
	Reactome	http://www.reactome.org	
	Wikipathways	http://www.wikipathways.org	
	PID	https://wiki.nci.nih.gov	
	HPRD	http://www.hprd.org	
	Intact	http://www.ebi.ac.uk/intact	
	STRING	http://string-db.org	
Tissue/Organ expression specificity databases	HOMER	http://discovery.informatics.iupui.edu/HOMER	Tissue/organ specific gene expression profiles are stored in theses databases
	HPA	http://www.proteinatlas.org	
	GNF tissues atlas	https://cgwb.nci.nih.gov/cgi-bin/	
	Pagenbase	http://bioinf.xmu.edu.cn/PaGenBase	
Phenotypic outcomes databases	SIDER	http://sideeffects.embl.de	Databases connecting pathways, targets and molecules to unhealthy phenotypes, such as disease or toxicological side effects
	CTD	http://ctdbase.org	
	AERSs	https://open.fda.gov/data/faers	
	EudraVigilance	https://eudravigilance.ema.europa.eu	
	JAPIC	http://database.japic.or.jp	
Terminology dictionaries	MeSH	https://www.nlm.nih.gov/mesh/	Therapeutic and side effect terminologies dictionaries
	MedDRA	http://www.meddra.org	
	ICD-10	http://www.who.int/classifications/icd/	
	SNOMED CT	http://www.ihtsdo.org/snomed-ct	
	ATC	http://www.who.int/classifications/atcddd/	
	UMLS	https://www.nlm.nih.gov/research/umls/	
	J-ART		
Network analysis and enrichment tools	CMap	http://www.broadinstitute.org/cmap/	Collection of GWAS data and pattern-matching algorithms that enable the discovery of functional connection between drugs, genes and diseases
	Cytoscape	http://www.cytoscape.org	Software for the visualization of complex interaction networks
	IPAD	http://bioinfo.hsc.unt.edu/ipad/	Resource for systematic enrichment analysis (associations between drug, pathway, organ and disease)
	LENS	http://severus.dbmi.pitt.edu/LENS/	Software for the systematic network, pathway and disease enrichment analysis on genes of interest