Systems biology approaches to a rational drug discovery paradigm

Philip Prathipati* and Kenji Mizuguchi

National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8 Saito-Asagi,
Ibaraki City, Osaka- 567-0085

Abstract: Ligand- and structure-based drug design approaches complement phenotypic and target screens, respectively, and are the two major frameworks for guiding early-stage drug discovery efforts. Since the beginning of this century, the advent of the genomic era has presented researchers with a myriad of high throughput biological data (parts lists and their interaction networks) to address efficacy and toxicity, augmenting the traditional ligand- and structure-based approaches. This data rich era has also presented us with challenges related to integrating and analyzing these multi-platform and multi-dimensional datasets and translating them into viable hypotheses.

Hence in the present paper, we review these existing approaches to drug discovery research and argue the case for a new systems biology based approach. We present the basic principles and the foundational arguments/underlying assumptions of the systems biology based approaches to drug design. Also discussed are systems biology data types (key entities, their attributes and their relationships with each other, and data models/representations), software and tools used for both retrospective- and prospective-analysis, and the hypotheses that can be inferred. In addition, we summarize some of the existing resources for a systems biology based drug discovery paradigm (open TG-GATEs, DrugMatrix, CMap and LINCs) in terms of their strengths and limitations.

Keywords: systems biology, drug discovery, chemoinformatics, chemoproteomics, chemogenomics, genomics, phenomics

INTRODUCTION

Computer-assisted drug design† approaches have traditionally been classified into ligand- and structure-based approaches.[1] [Since this review covers a wide range of research areas, we have prepared a glossary for key technical terms (indicated with a dagger symbol) in Supplementary File 1.] The former approach, based on the idea of similar chemical structures having similar pharmacological effects, led to successful therapeutics even when the mode of action (MOA)†

* Address correspondence to this author at the National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8 Saito-Asagi, Ibaraki City, Osaka, 567-0085, Japan; Tel: +81 72-641-9890; Fax: +81 72-641-9881; E-mail: philip@nibiohn.go.jp

was unknown[2]b,2 The ligand-based approach† mimics medicinal chemistry strategies, such as optimizing physicochemical properties, and is widely used and regarded as the most successful technique to date[3]. However, this approach does not offer a rational design† strategy. Furthermore, most often the structure-activity relationships† (SAR) guided multi-objective optimization† (against both the phenotype† and toxicity†) is fraught with difficulties.[1] This problem has been attributed to key polar functional groups† or substructures [such as basic fragments for G-protein coupled receptors and acid/serine proteases[4] (figure 1), or acidic fragments for phosphatases], which have to retain a particular physicochemical property† (e.g., "basic" in figure 1) for the on-target activity but such a property may be incompatible with favorable absorption, distribution, metabolism, excretion and toxicity (ADMET†) profiles. [1]

The latter, structure-based approach† also presents its own set of problems; methods for docking a compound into a single well-defined protein structure are reasonably mature but it is still a challenge to deal with multiple (functionally distinct) conformational states of the target protein. Furthermore, small molecules can interact with numerous other proteins ("off-targets") and there is also no rational procedure for understanding how the target-interaction profiles ("chemoproteomic profiles"; see section 3.5 for more details) correlate with the phenotype (Fig. 2).[5, 6]

However, since the beginning of this century, the advent of the genomic era has presented researchers with a myriad of high throughput (HT) biological data† (parts lists and their interaction networks), which can assist in the optimization of efficacy and ADMET profiles in the traditional ligand- and structure- based approaches. This data rich era has, on the other hand, presented us with challenges related to integrating and analyzing multi-platform and multi-dimensional datasets and translating them into viable hypotheses.

Hence in the following sections, we elaborate on the existing approaches to drug discovery research and argue the case for a new systems biology based approach (section 1). We present the basic principles and the foundational arguments/underlying assumptions of the systems biology based approaches to drug design (section 2). We then discuss systems biology data types (key entities, their attributes and their relationships with each other), hypotheses that can be inferred, and software and tools used for both retrospective- and prospective-inferences (section 3). We also review data models/representations that capture systems biology data types (section 4), In addition, we summarize some of the existing resources for a systems biology based drug discovery paradigm (open TG-GATEs, DrugMatrix, CMap and LINCs) in terms of their strengths and limitations (section 5). Finally, we review recent machine learning approaches for analyzing systems biology data (section 6).

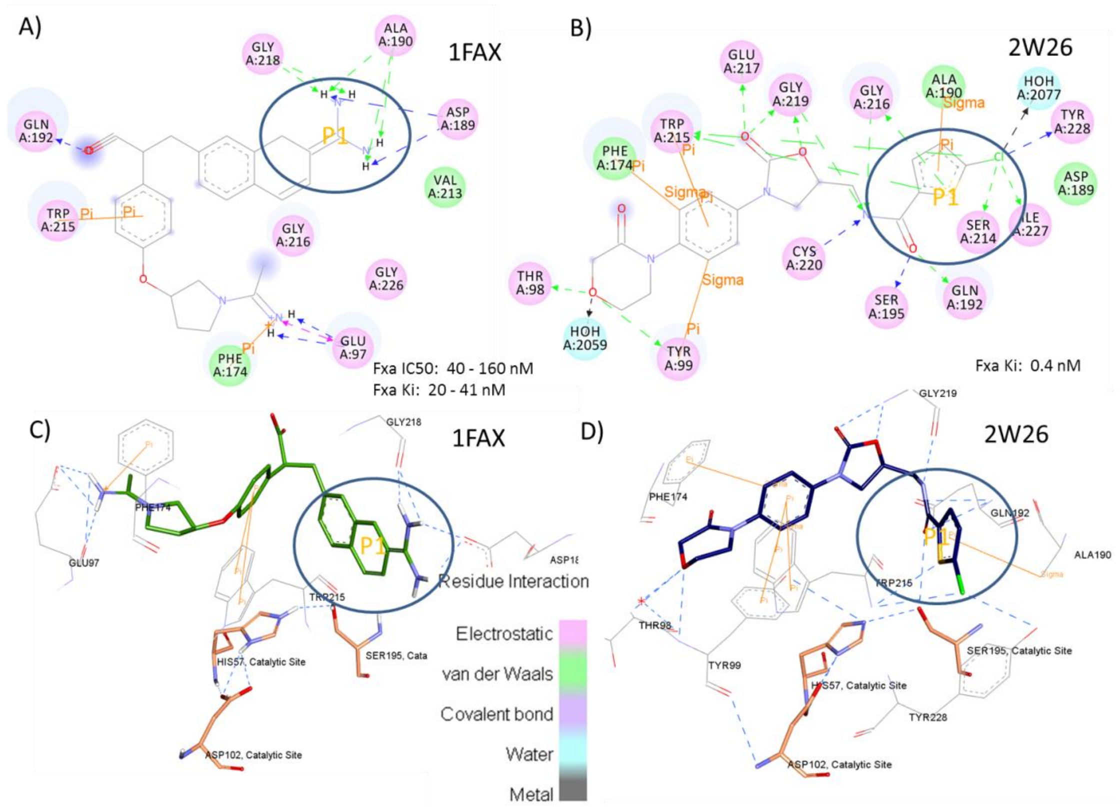

Figure 1: The advantages of structure-based approaches augmenting ligand-based approaches are exemplified using FXa inhibitors Dabigatran (A and C) and Rivaroxaban (B and D). Using a ligand-based approach, a basic fragment at P1 site (circled in A and C) was shown to be essential for FXa inhibition. Hence, the optimization strategy involved exploring other basic fragments at this site, which led to unfavorable ADMET profiles. A structure-based approach was used for replacing the basic amidine fragment of Dabigatran at the P1 pocket with a neutral chlorothiophene fragment by exploiting FXa's binding site information and the full range of intermolecular interactions such pi-pi, cation-pi and anion-pi, in addition to the usual hydrogen bonding, hydrophobic and salt bridge interactions. This strategy has led to the design of the new inhibitor Rivaroxaban (B and D) with favorable ADMET profiles. For more details, the readers can refer to a review by Nar et al.[4]

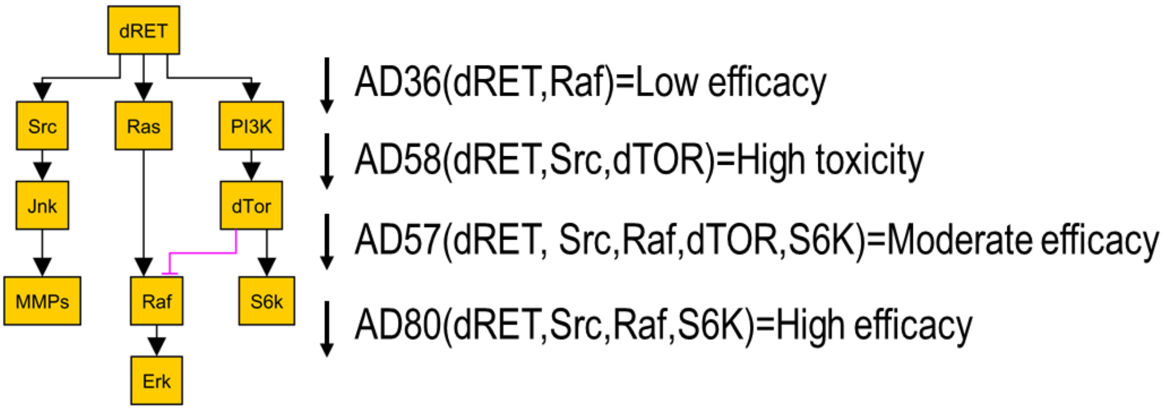
dRET
Src
Ras
PI3K
Jnk
dTor
MMPs
Raf
S6k
Erk
AD36(dRET,Raf)=Low efficacy
AD58(dRET,Src,dTOR)=High toxicity
AD57(dRET, Src,Raf,dTOR,S6K)=Moderate efficacy
AD80(dRET,Src,Raf,S6K)=High efficacy

Figure 2: The cancer model built using the fruit fly Drosophila pathway[†] data was used to rationally arrive at a compound (AD80) that was shown to inhibit four of the 10 selected cancer targets, dRET, Src, Raf and S6K. This chemoproteomic profile (inhibiting these four targets out of the 10) was associated with high efficacy and low toxicity in whole animal screening.. Further SAR analysis also led to the identification of dTOR as an anti-target responsible for toxicity. AD80 proved far more effective and less toxic than standard cancer drugs, which generally focus on a single target. This study by Dar et al.[6] was the first time that whole-animal screening has been used in a rational, step-wise approach to identifying favorable chemoproteomic profiles and laid the case for a rational systems biology approach to drug discovery. For a more general discussion of chemoproteomics, see section 3.5.

1. THE CASE FOR A SYSTEMS BIOLOGY OR NETWORK BASED DRUG DISCOVERY PARADIGM

1.1 From phenotypic and target based screening to systems biology based screening

In the pre-genomic era, two broad types of screens have sequentially dominated early-stage drug development—phenotypic screens and target[†]

-based screens.[7] The former quantifies the effects (phenotypes) that compounds induce in cells[8] and tissues. The correlation between the phenotypes and the compounds' chemical structures is studied and incorporated into ligand-based *in silico* drug design efforts. The latter, target-based screening assesses the effect of compounds on a purified target protein via *in vitro*[†] assays and is supported by structure-based computer-aided drug design techniques. These computational methods have fundamental limitations as described in the previous section, as well as technical limitations discussed in ref1. To address the technical limitations, integrated HT computational protocols have been also proposed, combining ligand- and structure-based approaches.[1,9] Structure-based approaches were integrated into ligand-based approaches with a view to select the model that reflects biological reality,[9a, 10] while ligand-based models were integrated into structure-based models to address issues related to pose prediction and scoring functions.[1,9c, 9d, 11]

One important technical limitation in these computational methods is the accuracy with which molecular recognition events are captured. Thus, molecular dynamics[†] (MD)[12] simulations and quantum chemical calculations[†13] such as Density Functional Theory (DFT) using graphics processing units (GPUs) can complement the traditional HT *in silico* techniques. in drug discovery research. The GPUs have greatly mitigated the computer scalability issues by accelerating the calculations tens of times. While MDs simulations can identify cryptic or allosteric binding sites,[14] enhance traditional virtual-screening methodologies,[15] and aid in the direct prediction of ligand binding energies, quantum chemistry calculations based on *ab initio* and DFT[16] provide estimations of several physicochemical properties with increasing accuracy and establish data-driven sound relationships between structure and observable properties. With constant improvements in both computer power and algorithm design, MD simulations and quantum chemical calculations are likely to play an increasingly important role in the development of novel pharmacological therapeutics.

However, even these integrated and accurate *in silico* approaches have proved insufficient for compound prioritization, because of difficulties in correlating phenotypic and target-based screens.[17] For example, before the advances in molecular biology, phenotypic screens were primarily used with the hope of subsequently identifying the target or targets of intervention.[18] However, target identification and validation[†] proved difficult or impossible in most instances.[7] Hence in the last few decades, phenotypic[†] screens were mostly replaced with target screens, in which the target was validated with genetic studies, in early stages of drug discovery research.[19] However, the over-reliance on target screening manifested as reduced discovery of first in class drugs.[20]

Thus, integrative systems biology based screening methods were proposed as a solution, which combine elements of both phenotypic and target-based screens. This integrated framework hopes to expand and augment target-based screening by providing chemical validation of drug targets, i.e., identifying the target or targets that are modulated by the candidate molecule when the desired phenotype is observed.[21]

1.2 Use of gene†/protein/metabolite profiles as surrogates for phenotypes

For a realistic application of a HT systems biology based integrative approach, phenotypic screens represent a bottleneck and present two major problems. First, phenotypes are difficult to quantify[22] and second, running phenotypic screens (quantification of observable traits) for complex multifactorial diseases such as cancer, diabetes and cardiovascular diseases have logistic difficulties.[23] For instance, according to the current cancer drug discovery paradigm, a rare population of tumor cells with stem cell characteristics (known as cancer stem cells) are considered to be responsible for tumor growth and metastasis and are the focus of much attention. However, monitoring and quantifying cancer stem cell content presents considerable challenges, including a) the requirements of serial biopsies of the tumor or b) counting the number of cells with a particular stem cell marker such as a surface protein or c) injecting the treated cancer stem cells into immune-deficient mice to see if they form tumors. Hence gene expression† signatures that correlate with “stemness” and are sensitive to chemotherapy are often used. [23] In a similar vein, the use of the gene/protein/metabolite expression/concentration signatures as surrogates for many other phenotypic endpoints has steadily been gaining acceptance in drug discovery and toxicological research.[24] Huntington’s disease and Congenital adrenal hyperplasia are two examples where elevated metabolite profiles such as hydroxykynurenine and quinolinate levels and decreased metabolite profiles such as aldosterone and cortisol are used to the respective phenotypic states. Metabolomics is a study that aims to characterize the metabolome† (all endogenous metabolites found in cells and body fluid) under different conditions (for example, in disease states]. [25,26,27,28] Metabolomics can not only help us illustrate the underlying molecular disease-causing mechanisms but also gain broad recognition in discovery of metabolic signatures† [biomarkers] for disease diagnosis. [25,26,27,28]

1.3 Integrative systems biology or network based drug discovery paradigm

Given this background, systems biology approaches analyze how genes†, proteins, metabolites† and other molecular profiles and their interactions are maintained in health and how they become perturbed by genetic and environmental stressors and cause disease (“cause-effect mechanisms”).[29] Thus the *in silico* systems biology or network-based drug discovery paradigm attempts to model all major components and processes involved in early stages of drug discovery and development as described in figure 3.[30] More specifically, the elucidation of cause-effects mechanisms can be realized in three ways: (a) a HT bioinformatics† approach for analyzing retrospectively the cause-effect mechanisms (such as master regulators of phenotypes or cellular states or enriched substructures associated with toxicity or pharmacological activity), (b) advanced machine learning methods for prospective predictions, or (c) both (a) and (b) together in one integrated protocol. We will describe these protocols in more detail in sections 3 and 6.

The analysis of cause-effect mechanisms in chemoinformatics or chemoproteomics (see Table 1 in section 3) may be relatively straightforward for the following reason. These studies deal with a matrix of whether a particular compound interacts with a particular protein (chemoproteomics) or whether a particular chemical possesses a particular substructure (chemocinformatics). Elucidating the cause-effect relationships in chemoproteomics involves identifying statistically significant associations between common structural features (cause) and protein-interaction profiles (effects). This analysis is straightforward, because we model direct binding events in isolated protein (*in vitro*) assays.

However, a proper analysis of cause effect mechanisms in chemogenomics, genomics and phenomics† (see section 3 for the definitions) additionally requires the use of qualitative/quantitative pathway models for understanding the paths between the perturbed elements and the manifested outcomes.[31] For instance, chemogenomics is the study of changes in gene expression profiles induced by chemical compounds. Thus the chemically induced gene expression signature can only be interpreted by modeling protein and gene networks†. [Chemicals binding to proteins transmit the signal to transcription† factors (TFs), which modulate gene expression.][32] Genomics and phonemics studies typically analyze the changes in gene expression profiles in diseased states induced from normal states by non-chemical perturbations. They include: (a) natural substrates (e.g., hormones, neurotransmitters, extra-cellular signals/factors), (b) RNAi† (SiRNA, shRNA) (c) genome† editing (CRISPR) for knockdown or knockout, and (d) cDNA constructs for overexpression of master regulators†. These studies also require the use of gene regulatory networks together with protein and metabolic regulatory networks to interpret meaningfully the path between the signal triggers and output.[33]

Systems biology using qualitative/quantitative models as part of the cause effect elucidation efforts attempts to understand the following elements. First, it concerns the structure of the biological system, i.e., all the elements linking the signal trigger (e.g., ligand binding to its cell-surface receptor) through the diverse pathways and mediators to a specific set of regulator proteins (such as TFs) that are responsible for altering the expression of a large number of genes, which then gives rise to phenotypic changes. (Such a structure is known as the bow-tie architecture[34] because the signal converges to a small number of regulator proteins at the knot region and then diverges.) The second element is the dynamics of the system by constructing predictive qualitative or quantitative models. Third, we need to identify "control elements" based on the predictive model simulations. These control proteins regulate the information flux between the input and the output and are also called bottleneck elements. These bottleneck elements are often prioritized as drug targets. Forth, we also need to understand how the system is constructed by combining known network motifs (i.e., repeating subnetworks such as feed-forward and other regulatory loops).

Thus, to implement a systems biology approach to drug discovery research, we need to integrate chemoinformatic, proteomic, genomic, phenomic, chemoproteomic, chemogenomic data together with qualitative and quantitative cell signaling models.[34] This approach, in addition to elucidating the cause-effect mechanism, can also identify a collection of modifiable drug targets and predict the effect of single- or combinatorial-drug treatments. The modulation of multiple targets may be required, because in most instances, phenotypes have back-up or alternate survival mechanisms, which also need to be perturbed to achieve phenotypic transition.

Although many compound and target prioritization methods have been proposed, most of them are not based on the systems biology framework as described above. For example, compounds and targets that match a given set of proteomic, genomic or metabolomic profiles can be prioritized by searching appropriate databases such as ChEMBL,[35] BindingDB[36], TargetMine[37], and Possum[38] for proteomics,, cMap,[39] CIDD,[40] Toxygates[41], GSEA[42], and QSTAR project[43] for genomics, and MSEA[44] and MetaboAnalyst[45] for metabolomics. The application of integrated approaches such as Galahad[46], Expression2Kinases[47], and CellNOptR[48] which uses both genomic (or proteomic) profiles and protein-protein interaction (PPI) data, is also gaining attention. In its essence, all these prioritization processes involve comparing a molecular profiles (e.g., protein-target interaction or gene expression response) associated with a chemical with a database of disease or pathway signatures such as MSigDB[49], GeneSigDB[50], and EnrichR[51]. The comparisons can be performed using a variety of association measures[39, 52], but have limitations such as ignoring the topology of the regulatory networks and the relative rank of the strength of the association. These limitations can be addressed using the systems biology framework incorporating qualitative and quantitative models, since these models consider the topology of the network and the quantitative measure of molecular activities (such as protein activity, mRNA† expression and metabolite concentration). Thus *in silico* systems biology is regarded as a promising avenue to discover a combination of targets and modulators to produce synergistic effects or avoid antagonist effects.[53,54]

method.[52g]

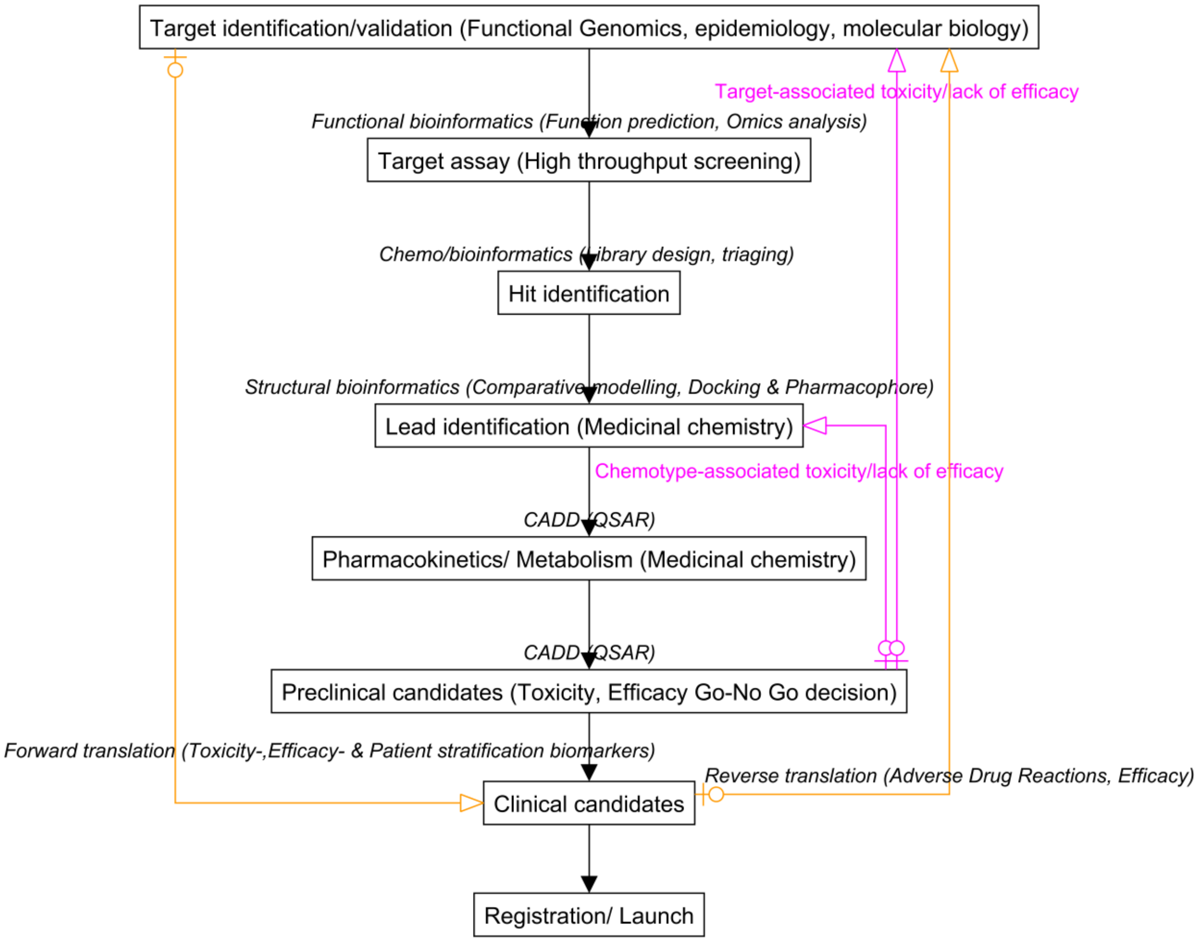
Target identification/validation (Functional Genomics, epidemiology, molecular biology)
Target-associated toxicity/lack of efficacy
Functional bioinformatics (Function prediction, Omics analysis)
Target assay (High throughput screening)
Chemo/bioinformatics (Library design, triaging)
Hit identification
Structural bioinformatics (Comparative modelling, Docking & Pharmacophore)
Lead identification (Medicinal chemistry)
Chemotype-associated toxicity/lack of efficacy
CADD (QSAR)
Pharmacokinetics/ Metabolism (Medicinal chemistry)
CADD (QSAR)
Preclinical candidates (Toxicity, Efficacy Go-No Go decision)
Forward translation (Toxicity-,Efficacy- & Patient stratification biomarkers)
Reverse translation (Adverse Drug Reactions, Efficacy)
Clinical candidates
Registration/ Launch

Figure 3: A schematic flow chart summarizing the process of drug discovery, including major *in silico* contributions (italicized) to chemistry, biology† and ADMET.

2 BASIC PRINCIPLES AND UNDERLYING PREMISES OF SYSTEMS BIOLOGY APPROACHES TO DRUG DESIGN

Since systems biology based approaches attempt to integrate (a) metabolic/chemical concentrations, (b) protein activity, (c) protein expression, (d) gene expression and (e) phenotypic endpoints and use one as a surrogate for the other, it is vitally important to understand some of the underlying assumptions and foundational arguments for application of this strategy for drug discovery. BCL-ABL and epidermal growth factor receptor (EGFR) kinases are prototypical examples, which exemplify correlations between genes, transcripts[†], proteins, and protein activity and led to conceptualization of systems biology approaches[55] (figure 4).

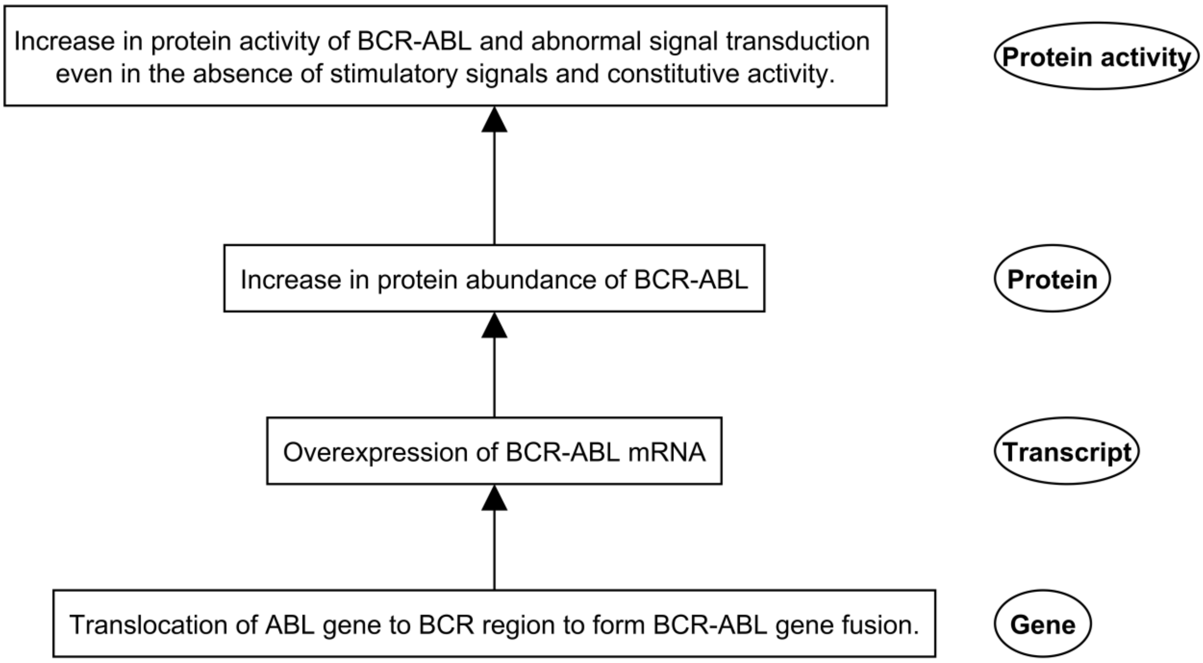
Increase in protein activity of BCR-ABL and abnormal signal transduction even in the absence of stimulatory signals and constitutive activity.
Protein activity
Increase in protein abundance of BCR-ABL
Protein
Overexpression of BCR-ABL mRNA
Transcript
Translocation of ABL gene to BCR region to form BCR-ABL gene fusion.
Gene

Figure 4: BCL-ABL: correlations between gene fusion and protein activity levels. BCR-ABL kinase is a major drug target for a range of blood and solid tumors. Systems biology researchers aim to identify similar such drug targets using HT data and the conceptual workflow described in section 1.3.

2.1 Correlations between mRNA and protein abundance levels

The principle hypothesis of genomics is that steady state mRNA levels correlate with protein concentrations. The reported correlations between mRNA and protein levels by different groups vary from R= 0.4 to R=0.7.[56] The limited correlation was attributed to the importance of post-transcriptional-, translational- and protein degradation-regulation in controlling steady-state protein abundances, and it was suggested that better correlations may be obtained by addressing experimental artefacts.

2.2 Gene co-expression for estimating protein-protein interaction probability

Proteins involved in regulatory interactions are assumed to have more similar gene expression profiles than random pairs. Supporting evidence comes from studies comparing gene expression profile with large-scale PPI data sets.[57] Thus mRNA co-expression profiles may be used to infer cell-signaling networks.

2.3 Existence of upward and downward causation in biological systems

The advances in genetics and epigenetics unequivocally demonstrate the existence of downward causation (e.g., from protein to gene), which is seen as completing a feedback circuit and has formed the founding principle of systems biology[58]. Some of the most important downward causation events include triggers (hormones and neurotransmitters) of cell signaling, control of gene expression by TFs and epigenetic regulators, the protein machinery that reads and repairs making the genome reliable[59] (as described in figure5).

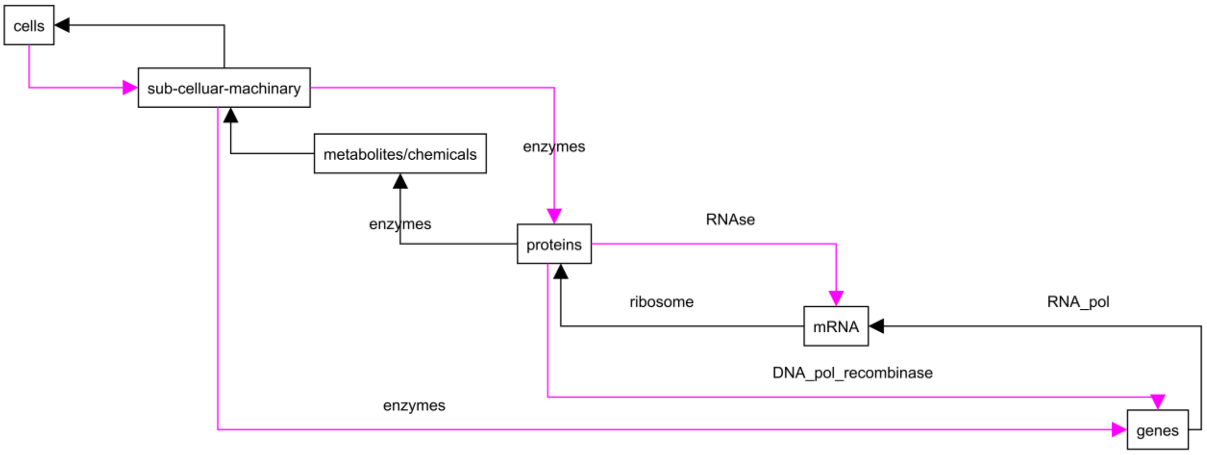

Figure 5: Systems biology view of upward and downward causation in biology as a feedback diagram. While the upward causation is well known (black links), some of major downward causation (blue links) elements include TFs, epigenetic regulators, reverse transcriptases, DNA_polymerases[†], and recombinases.

2.4 Multiple targets for structure-based drug design

Complex multifactorial diseases show perturbations at pathway levels and not necessarily at individual proteins/genes. Hence the systems biology based drug design paradigm focuses on searching for multi-target drugs to perturb disease-associated networks rather than designing selective ligands to target individual proteins. Furthermore, studies on drug promiscuity in proteome-wide binding[60] estimated that an existing drug binds to, on average, 6.3 protein receptors, which include targets (favorable) and anti-targets (unfavorable). Thus given this promiscuity and the systems biology view of diseases, it is important to build actionable models (i.e., models that can predict the effect of perturbations on the relevant physiological process) to guide structure-based drug design (as illustrated in figure 2). [5]

2.5 Correlation between drug-induced protein activity levels and gene expression profiles

The correlation between drug-induced protein activity levels and mRNA abundance is another founding principle of systems biology based drug design. This principle was examined by analyzing the protein activity and mRNA concentrations upon rapamycin treatment.[61] A significant number of proteins with decreased activity caused decreases in their mRNA levels. This result may be explained by common network motifs.[62]

In another study, Iskar et al[63] showed similar trends after analyzing 1,290 drug-target and drug-mRNA profiles, although the sample size was small. A recent study by Koussounadis et al [64] showed that differentially expressed mRNAs correlate significantly better with their protein product than non-differentially expressed mRNAs. These results have increased confidence for the use of differential mRNA expression for biological discovery in various disease systems, as well as providing optimism for the usefulness of inferences from mRNA expression in general.[56d, 65]

2.6 Qualitative and quantitative cell signaling models to simulate the effect of cues on cellular behavior and phenotype

Qualitative and quantitative cell signaling models or network-based computational models are broadly classified as Bayesian, logic based (qualitative) and mass-action (quantitative) models.[54] These models capture the dynamic signaling networks that drive biological decision processes and cellular states in response to cues. In a network based computational model, cellular behavior (e.g., cell migration) or phenotypic states (e.g., metastatic) can be considered a steady state high dimensional vector of gene/protein activities[66],[67]

We can thus expect network-based computational models to be able to simulate diseases such as cancer, which are not just regarded as ones with a genetic basis but as those that are driven by perturbations at the signaling network level. For instance, network based computational models can simulate the events, involving perturbations to some network components, that change normal cells to new malignant states (e.g., EGF induction leading to metastasis), or the events, involving drug mediated perturbations, that might reverse or inhibit certain phenotypic states. [5, 66,67]

3 IMPORTANT ELEMENTS AND RELATIONSHIPS IN SYSTEMS BIOLOGY DATASETS

Systems biology analyzes global profiles of chemicals, proteins and genes to understand and predict biological complexity by using a cross-disciplinary approach.[68,69] Hence it integrates many multi-scale (genes to phenotypes) types of biological information. Systems biology data are best described as a graph, consisting of nodes (elements) and edges (relationships between elements), and analyzed using graph (network) based methods (tables 1 and 2). While Table 1 summarizes some of the datasets that usually are integrated and used in the systems biology based drug design paradigm and hypotheses that can be derived from retrospective analysis, table 2 catalogs open access databases and software packages useful for systems biology approaches to drug discovery.

Table 1: Summary of the datasets used in the systems biology based drug design paradigm and the nature of the hypothesis that can be inferred by analyzing these datasets.

Data type	Parts list	Hypothesis from a retrospective analysis of the interactions
Chemoinformatics	Nodes: Chemicals Node Attributes: *Proteins*, Domains, *Substructures, Enriched fragments*†, *Pharmacophores*, *Toxicophores*, *Physicochemical properties*, *Structural descriptors*, etc	(a) Chemical similarity network analysis that can complement chemoproteomic and chemogenomic analysis.
Proteomics	Nodes: Proteins Node Attributes: *Domain definitions,Ssequence motifs(linear*	(a) Protein similarity and (b) protein interaction networks:

Data type	Parts list	Hypothesis from a retrospective analysis of the interactions
	and non-linear), Superfamily definitions, Sequence descriptors, Cognate ligands, Pathways, other protein interacting partners	
Genomics	Nodes: Genes/ transcripts Node Attributes: *Phenotypes/indications, Perturbagens*† *(small molecule or si/shRNA), motifs, regulators (TFs, Epigenetic factors, Master regulators), Pathways, Literature gene sets.*	(a) Finding and interpreting genes/ transcripts associated with phenotypic changes or perturbations.
Phenomics	Nodes: Diseases/ Indications/ phenotypes Node Attributes: *in vivo*† Biochemical data, Hematology, Organ Weight, Pathology Data, Histology, Pathways, Genes, Proteins (drug targets), Chemicals, Chromatin regulators.	(a) Studying the genotype–phenotype map, (b) Identifying the genetic basis of complex traits.
Chemoproteomics (bipartite networks – edges between chemicals and proteins only)	*Nodes*: Chemicals, Proteins Edge Attributes: Activation, inhibition, degradation.	(a) Analyzing the pharmacological map of the druggable proteome and discovering ligands for undruggable proteome, (b) drug target discovery.
Chemogenomics (bipartite networks – edges between chemicals and genes only)	Nodes: Chemicals, Genes Node Attributes: : *in vivo* Biochemical data, Hematology, Organ Weight, Pathology Data, Histology, Pathways, Genes, Proteins (drug targets), Chemicals, Chromatin regulators. Edge attributes*: activation, repression.*	(a) Determining mode of action, (b) drug repurposing and drug target identification
Qualitative and quantitative network models	Nodes: Chemicals, Genes, Proteins, protein complexes, phenotypes Node Attributes: Activity levels inferred from mRNA or protein expression activity data. Edge Attributes: Regulatory interactions, PTMs,	(a) Represent existing knowledge of biological systems, (b) predict the effect of perturbations on other components of the pathway, (c) identify missing components in a pathway, (d) determine the most critical components of the pathway,

Understanding complex systems often requires a bottom-up analysis which involves investigating a system, not only as individual components but as a whole.[70] Such an investigation can be done by examining the elementary constituents (nodes) individually and then how these are connected. The myriad components of a system and their interactions are best characterized as networks and they are mainly represented as graphs where thousands of nodes are connected with thousands of vertices.

In general, a graph based computations have been successfully applied to the study of biological network topology, from the global perspective of their scale-free, small world, hierarchical nature, to the zoomed-in view of interaction motifs, clusters and modules and the specific interactions between different biomolecules. [70]

In particular, network-based approaches can elucidate the cause-effect mechanisms of existing observations ("retrospective systems biology analysis") by clustering entities and analyzing properties enriched within clusters. For example, genes can be clusters based on expression profiles and common regulators within a co-expressed cluster can be identified by the enrichment of regulator binding sites. Similarly, chemical fragments that preferentially inhibit a given protein family can be identified by clustering chemicals based on their protein interaction profiles.[71].

Weighted gene co-expression network analysis (WGCNA) is one of the best suited data mining methods for retrospectively analyzing the various systems biology data described above. Though it can be applied to most systems biology data sets, it has been most widely used for cause-effect interpretation in genomics. WGCNA allows one to define modules (clusters) and intramodular hubs, correlate the modules with attributes or perform enrichment analysis. Some of the typical hypotheses derived from retrospective analysis of the systems biology databases described in table 1 are discussed below.

Table 2: Open access databases and free software and tools for retrospective analysis. Only tools used and verified by the authors are presented.

Data type	Open access Datasets and tools to import	Open access software
Chemoinformatics	*TargetMine*[37], *Chembl*[35], *Pubchem*[72],	(a) Descriptors*: ChemmineR*[73], *SNG*[74], Rcpi[75], *OpenBabel*, jCMapperCLI.[76]

Data type	Open access Datasets and tools to import	Open access software
	BindingDB[36], *STITCH*[36]	(b) Cluster and enrichment analysis*:WGCNA*[77], *Fabia*[78], *Bicluster, SuperBicuster.*
Proteomics	*TargetMine, STRING*[79], *KEGG*[80], *Uniprot*[81], *etc.*	(a) Descriptors*: Uniprot.WS, Protr*[82], Rcpi[75], *BioMartR.*[83] (b) Cluster and enrichment analysis*: WGCNA*[77], *Fabia*[78], *Bicluster, SuperBicuster.*
Genomics	*NCBI GEO*[84], *Arrayexpress*[84]	(a) Differential gene expression analysis: *Affy*[85], *Limma.*[52f] (b) High Throughput Sequencing Data analysis: *Babraham Bioinformatics*[86] (b) Gene coexpression analysis: *WGCNA*[77], *Fabia*[78], *SuperBicuster.* (c)Databases for gene set enrichment analysis: *MSigDB* [49], *GeneSetDB*[50a], *ConReg, EnrichR*[51], *Hippi.*[87] (d) Tools for gene set enrichment analysis in bioconductor: *SPIA*[88], *gCMAP*[52a], *Piano*[89] together with webservers like *Targetmine*[37] *and DAVID.*[90]
Phenomics	*ICD10/9*[91], *-SIDER*[92], *-OMIM,* etc	(a) Cluster and enrichment analysis*: WGCNA*[77], *Fabia*[78], *Bicluster, SuperBicuster.*
Chemoproteomics	*ChEMBL*[35], *bindingDB*[36], *TargetMine*[37].	(a) Descriptors*: ChemmineR*[73], *SNG*, Rcpi, *OpenBabel, Uniprot.WS, Protr*[82], Rcpi[75], *BioMartR.*[83] (b) Cluster and enrichment analysis*:WGCNA*[77], *Fabia*[78], *Bicluster, SuperBicuster.*
Chemogenomics	*Japanese TGP*[93], *NIBIO immune adjuvant database, SAHA-PIP gene expression profiles*[94], *DrugMatrix*[95], *CMap*[39] *and LINCs*	(a) Differential gene expression analysis: *Affy*[85], *Limma*[52f] Gene coexpression analysis: *WGCNA, Fabia*[78], *SuperBicuster* (b) Databases for gene set enrichment analysis: *MSigDB* [49], *GeneSetDB*[50a], *ConReg, EnrichR*[51], *Hippi.*[87] (c) Tools for gene set enrichment analysis in bioconductor: *SPIA*[88], *gCMAP*[52a], *Piano*[89] together with webservers like *Targetmine*[37] *and DAVID.*[90]
Qualitative and quantitative network models	EBI biomodels[96], KEGG[80].	(a) import and parse systems biology models: *KEGGgraph*[97] for KEGG (b) Network inference: *CellNoptR*[48], *BoolNetR*[98] and *Copasi*[99]. (c)Network simulation: *CellnOptR*[48], *BoolNetR*[98], *Copasi*[99] , *GINsim*[100] and *CellDesigner*[101].

3.1 Chemoinformatics: Chemoinformatics is the field of study of all aspects of the representation and use of chemical and biological information on computers. Since similar chemical structures generally give similar activities, the identification of the links between chemical structures in terms of their attributes can lead to a comprehensive understanding of the nature of the chemical space, inform SAR and polypharmacological profiles, and provide mode of action (MOA) hypotheses for orphan compounds.[102] The chemical similarity network analysis can also complement chemoproteomic and chemogenomic analysis and provide a more complete hypothesis for drug discovery research.

3.2 Proteomics: Proteomics is the large-scale study of proteins, particularly their structures and functions. It is broadly divided into abundance and functional proteomics. While abundance proteomics catalogs protein components, identifies differences between states, finds biomarker and examines post-translational modifications, functional proteomics identifies interactors (PPIs, signal transduction pathways, biochemical machinery), finds enzymatic substrates and studies drug selectivity profiles. The comprehensive characterization of protein similarity and PPI networks can provide useful hypothesis for binding site predictions, and inform on polypharmacology profiles to guide rational optimization of the efficacy and toxicity.[103] The now mature field of PPI network analysis has led to considerable successes in identifying protein modules related to important biological processes and diseases.

3.3 Genomics Genomics is a discipline in genetics that applies recombinant DNA, DNA sequencing and bioinformatics to sequence, assemble and analyze the function and structure of genomes. Genomics research has been used for finding and interpreting genes/transcripts associated with phenotypic changes or perturbations. The identification of events (differential expression of genes; DEG) associated with phenotypic changes is now routinely performed and used for developing diagnostic, prognostic and gene signature assays. However, methods for inferring the causative events and for understanding the cause-effect mechanisms are currently being developed.[104] The understanding of cause-effect mechanisms could provide valuable new points of intervention (drug targets) for restoring the normal phenotypes. It could also provide important regulators for the emerging field of rational cellular reprogramming and phenotypic transitions.[77] As the field of genomics matures the limitations of single gene based DEG approaches and the benefits of gene set based approaches such as gene set enrichment analysis (GSEA)[42] and WGCNA [77] are also being recognized and driving the field.

3.4 Phenomics: Phenomics is a very broad study that deals with measuring how much the physical and biological traits in an organism is affected due to genetic and epigenetic (environmental) effects. Phenomics research contributes to an understanding genetic disorder, cancers and other diseases. It involves mapping genotype to phenotype either directly using single nucleotide polymorphism† (SNP) arrays or via the intermediate gene expression profiles with various phenotypic endpoints. While background normalization† of SNP arrays is major obstacle to SNP array based analysis, the identification of causal elements (master regulators) or interpreting the cause-effect relationships is an area of current phenomics research. [77, 104] Several successful master regulators such as Oct4/Sox2/Nanog for induced pluripotent stem cells were identified retrospectively or prospectively using systems biology tools described in table 2.[105]

3.5 Chemoproteomics: Chemoproteomics is a field of study linking chemicals to molecular targets implicated with therapeutic indications.[38] Chemoproteomic analysis can be used for (a) analyzing known ligand-protein interactions ("druggable protein space") and predicting ligands for proteins with no known small molecules ("extrapolating to the un-druggable protein space"), (b) discovering drug targets and (c) identifying favorable sets of targets responsible for mediating cellular effects. For instance, Crizotinib, which was initially developed as a c-Met inhibitor, was also found to target ALK in NSCLCs. Since ALK mutations were also identified as causative events for NSCLCs, ALK was proposed as a drug target for this indication. [71b, 106]

3.6 Chemogenomics: Chemogenomics is the systematic screening of targeted chemical libraries of small molecules against the global transcriptome space or against individual drug target families. (such as GPCRs and kinases). Chemogenomics analysis of datasets such as cMAP[39] or Japanese Toxicogenomics Project datasets (TGP)[93] have been used to identify well established gene signatures and biomarkers and several successful repurposing efforts were published using the tools presented in table 1.[107] Further analysis of these data sets by using gene set analysis methods such as GSEA[42], Galahad,[46] Expression2Kinases[47] should be able to provide additional hypotheses about potential pathways (and eventually drug targets) or gene sets perturbed by the compound of interest.

3.7 Qualitative and quantitative network models: While all the datasets above are experimentally obtained global profiles, the datasets in this category are theoretical models inferred using the above datasets for understanding specific diseases or pathways. We included this category in table 1 because these models can provide hypotheses and interpretations that are not available from any of the datasets above. These qualitative/quantitative models represent actionable cell signaling pathways and simulations can lead to the identification of drug targets (bottleneck proteins), master regulators (TFs and extracellular cues), missing components of networks, derive novel disease phenotypes (such as heterogeneous mutant cancer cellular states) and can be used for interpreting gene expression analysis.[108]

4 REPRESENTATIONS FOR SYSTEMS BIOLOGY DATA

The practice of systems biology (tables 2) depends upon many software tools, operating on many kinds of data elements from many different sources as described in table 1. These elements may have different attributes (properties described in table 1) and can be connected by different types of edges (links described in table 1), which can be directed or undirected. The edges/links can have physical meaning, denote functional associations or can represent shared characteristics between components.[109] Hence the field of data integration actively researches appropriate frameworks that are applicable to systems biology data.

The resource description framework (RDF) offers a simple mechanism to identify and describe the components and the links between systems biology data. In RDF, the elements are described in terms of their types, attributes and relations to other entities or elements. RDFS (RDF schema) is an extension of RDF and provides additional vocabulary for naming resources (rdfs: labels) and specifying simple type and relational hierarchies (rdfs: subclass of, rdfs: subproperty of). Most RDF resources (such as EBI RDF[109] and bio2rdf) are now implemented as RDFS and can be queried using the SPARQL query language. SPARQL quires may contain triples patterns that can be conjunctively (AND) or disjunctively (OR) combined with mandatory or OPTIONAL triple query patterns.

DrugBank[110] is a prototypical systems biology database, which includes most of the datasets described in table 1. Hence many example protocols to query drugs (as the subject) and various attributes (pka, WaterSolubility, target, DrugClassificationCategory, Indication, Mechanismofaction) as objects are provided as a supplementary file 1. More complex quires can be made integrating the bio2rdf endpoint with other related databases. The readers can refer to EBI’s example SPARQL quires[111] for additional details and bio2rdf example quires[112].

Data warehousing is another method for integrating systems biology datasets. InterMine[113] is an open source data warehousing framework, built specifically for the integration and analysis of complex systems biology data. TargetMine[37] is one of the applications developed using the InterMine framework and was designed specifically for assisting early-stage drug discovery and development. It enables the creation of biological databases accessed by sophisticated web query tools. Parsers are provided for integrating data from many common biological data sources and formats along with a framework for adding your own data, as well as a powerful, scriptable web-service API to allow programmatic access to your data.[114]

5 EXISTING RESOURCES FOR THE SYSTEMS BIOLOGY BASED DRUG DISCOVERY PARADIGM

Table 1 shows multi-disciplinary systems biology datasets on different levels and in the previous section, we described general technologies for data integration. System biology discovers how function arises in dynamical systems (cells) by integrating diverse datasets and infers the missing links between molecules and phenotypes.

Several projects were aimed at integrating multiple datasets and implementing a rational systems biology approach to drug and toxicity research[115], such as Japanese TGP[93], cMAP[39], DrugMatrix[95] and the LINCS[33] project. In table 2, we attempt to discuss the basic structure of the data generated by these projects and their strengths, and highlight the missing links, which limit the use of these data for a systems biology approach.

Japanese TGP[93]is probably one of the richest sources of *in vivo* and *in vitro* chemogneomic data with ~170 compounds-49K transcript profiled in different tissues, at different time points and at different doses (~20K GeneChip assays including histopathological data). Gene signatures and selected gene sensitivity markers were proposed for several toxicological end points but further analysis is required to translate these signatures into anti-targets of interest to drug research. However, a limited number of compounds and a lack of human in vivo or in vitro samples are some of the limitations of the dataset. Since no chemoinformatic or chemoproteomic profiles were considered this protocol has to be further appended by integrating it with chemoinformatic and chemoproteomic datasets to incorporate target identification and structure based drug design aspects.

DrugMatrix[95] was designed along the same lines as the Japanese TGP with 600 compounds, including ~4,000 dose time–tissue combinations, ~2 million dosed tissue samples, ~18,000 microarrays†, ~127,000 histopathology measurements and ~100,000 haematology and chemistry measurements. Furthermore, more than 800 compounds were profiled across 130 in vitro pharmacology assays. Like the Japanese TGP, several gene signatures and selected gene sensitivity markers were proposed for the several toxicological end points but further analysis is required to translate these signatures into anti-targets of interest to drug research. Since DrugMatrix is designed along the same lines as Japanese TGP, it has the same limitations and needs to be integrated with chemoinformatic and chemoproteomic datasets to be useful for drug discovery projects.

The connectivity map[39] is the most cited of the chemogenomics datasets and includes 453 Affymetrix profiles for 164 drugs across multiple cell lines, doses and time points. Several repositioning hypothesis were validated and proposed using positive correlation of transcriptional profiles with other drugs. It has limitations similar to those with the Japanese TGP. In addition, no phenotypic endpoints were measured/reported but can be inferred by linking to ATC or ICD or sider codes.

The Library of Integrated Cellular Signatures (LINCS)[33] is an NIH funded program for the generation of perturbational profiles across multiple cell and perturbation types, as well as read-outs, at a massive scale. To date, LINCS has generated over 1 billion data points of perturbational profiles spanning small-molecules and genetic gain- and loss-of-function across multiple cell types. It currently includes (~5K small molecules+ ~22K CRISPR or cDNA constructs) -induced molecular (1,000 landmark genes+500 kinome) and cellular signatures (876 cell lines). This massive project aims to create a network-based understanding of biology by cataloging changes in gene expression and other cellular phenotypes. The LINCS project addressed most of the limitations of its predecessor the cMAP build 1 and 2 and integrated the chemogenomics dataset with chemoproteomic and chemoinformatic datasets together with several other relevant datasets (e.g., compound-cell perturbations, shRNA-gene perturbations, and kinome scans). However, since only expression profiles of 1000 landmark genes are currently available, LINCS cannot readily be analyzed with conventional approaches such as GSEA. In addition, LINCS does not include cell signaling networks and hence cannot be used to rationally infer the paths between physical target and differentially expressed master regulators.

6 MACHINE-LEARNING METHODS FOR PROSPECTIVE SYSTEMS BIOLOGY DATA ANALYSIS

While the software and tools described in table 2 are mostly used for retrospective analysis (see section 3), network-based approaches can be used for prospective predictions as well, for example, proposing novel biomarkers and validating them by new experiements.[116] In particular, recent advances in machine-learning methods have brought about the possibility of applying these methods to prospective systems biology data analysis. Systems biology data described above have many-to-many relationships (many drugs associate with many targets) and are best analyzed using advanced machine-learning methods.

Drug repurposing is the most sought after predictive hypothesis generation application and the numerous approaches which can be classified as similarity based, 3D structure based, network inference, machine learning are summarized in table 3. As seen from the cited examples in table 3, many methods with unique applicability ranges have already been used for drug repositioning studies. As suggested by Meslamani et al.[117] there is no rationale for considering a single profiling method for drug repositioning. On the basis of a comparative evaluation of several ligand-based and target-based methods in profiling 157 diverse ligands on 2556 different targets, Meslamani et al [118]previously shown that (i) ligand-centric methods should be used whenever possible (which means when enough ligands are known for a particular target), (ii) 2D ligand descriptors are usually preferred to 3D descriptors, with the exception of low molecular-weight apolar ligands, (iii) protein−ligand docking should be reserved to polar and buried active sites of known structure for which few ligands are available, and (iv) receptor−ligand pharmacophore search may then be applied to all other protein structures. [118]

Traditional methods in machine learning and statistics provided data-driven models for predicting a single target or label (Y) either as binary values in classification and real-values in regression. However in recent years, novel application domains such as systems biology datasets have triggered fundamental research on more complicated problems, where multi-target predictions are required.[119] In the realm of systems biology, the targets (Ys) often have diverse relational structures; for instance, biological attributes or entities such as international classification of diseases (ICD) 10/9 annotations, protein domain annotations, Gene Ontology terms, all of which have parent child relationships, while gene co-expression, protein- and chemical- similarity networks are known to be scale-free or follow power law relationships and can be presumed to have a tree shaped hierarchy. Thus a range of machine learning methods have to be considered depending on the data types such as support vector machines (SVMs), neural networks (NN), k-nearest neighbors (kNN), boosting methods for unrelated multi-label datasets and similarity based approaches such as DT-hybrid, kernel regression methods such as lasso or elastic nets or pairwise kernel method (PKM) for related multi-label datasets. [105a-d]

Table 3: Systems biology based prospective methods for drug repositioning can be classified as similarity-, 3D structure-, network inference- and machine learning-based methods. Some of the well-known studies in each category are presented.

Prospective systems biology methods	Similarity based	3D structure based	Network inference	Machine learning
Chemical structure	Gonzalez-Daz et al., (2011)[120], Keiser et al., (2007&2009)[102b, 121].	Meslamani et al.[117] Meslamani et al.[118]	Yildirim et al. (2007)[122]	Ekins et al.(2007 & 2014)[32, 123]

Protein structure (inverse docking)		Li et al., (2006)[124];
		Xie et al., 2011[60b];
		Martínez-Jiménez 2015[125]		

Side effect/phenotype	Campillos et al. (2008)[126] Chen L (2012)[109b]		Yildirim et al. (2007)[122]	Gao YF et al.(2013)[105b] Wu L et al.(2013)[105c] Liu M et al.(2012)[105e]
Transcript expression	Iorio et al. (2010)[127]; Dudley et al. (2011)[107a] Sirota et al. (2011)[107c]			Fernald et al.(2013)[128]
Protein attributes			Yildirim et al. (2007)[122]; Jacoby et al.[129]	Clark et al.(2014)[130]

Chemical and protein attributes			Cheng et al. (2012)[131];
			Zhou et al. (2007)[132];
			van Laarhoven et al. (2011)[133];
				Prathipati et al. (2009)[106a];
				Nidhi et al.(2006)[136];

Prospective systems biology methods	Similarity based	3D structure based	Network inference	Machine learning

			Mei et al. (2013) [134];
			Alaimo et al. (2013) [119a];
			Yamanishi et al (2013).[135]
				Wale et al. (2009) [137]

Pathway			Zhao et al (2014)[53] ;Pan et al (2014)[138]; Han et al. (2008).[139]	

7 CONCLUSION

In this article, we present a case for a systems biology approach to drug discovery research given the issues with either ligand-based or structure-based approaches for mitigating efficacy and toxicity. Some of the foundational principles and presumptions of systems or network approaches were discussed together with representative multi-scale databases used in systems biology research and different tools used to retrospectively or prospectively analyze the data together with the data models used to best capture and retrieve systems biology data. Given the enormous advances in computing technologies, miniaturization and HT experimental technologies, systems biology approaches have enormous potential to change the landscape of healthcare and contribute to drug discovery research.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

THIS WORK WAS IN PART SUPPORTED BY GRANTS-IN-AID FOR SCIENTIFIC RESEARCH FROM THE JAPAN SOCIETY FOR THE PROMOTION OF SCIENCE (GRANT NUMBERS 25430186 AND 25293079) AND FROM JAPAN AGENCY FOR MEDICAL RESEARCH AND DEVELOPMENT ("THE ADJUVANT DATABASE PROJECT") TO K.M. WE ALSO THANK THE REVIEWERS FOR THE NUMBEROUS CONSTRUCTIVE SUGGESTIONS.

SUPPLEMENTARY MATERIAL

Supplementary file 1. The SPARQL R code for querying various drugbank endpoints.

Supplementary file 2. The glossary of various drug discovery technical terms referred in the text.

REFERENCES

1. Prathipati, P.; Dixit, A.; Saxena, A. K., Computer aided drug design: integration of structure based and ligand based approaches. *Current computer aided molecular design* **2007,** *92*, 29-37.

2. Wolber, G.; Seidel, T.; Bendix, F.; Langer, T., Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discov Today* **2008,** *13* (1-2), 23-9.

3. (a) Brustle, M.; Beck, B.; Schindler, T.; King, W.; Mitchell, T.; Clark, T., Descriptors, physical properties, and drug-likeness. *J Med Chem* **2002,** *45* (16), 3345-55; (b) Tarcsay, A.; Keseru, G. M., Contributions of molecular properties to drug promiscuity. *J Med Chem* **2013,** *56* (5), 1789-95; (c) Hopkins, A. L.; Keseru, G. M.; Leeson, P. D.; Rees, D. C.; Reynolds, C. H., The role of ligand efficiency metrics in drug discovery. *Nat Rev Drug Discov* **2014,** *13* (2), 105-21.

4. Nar, H., The role of structural information in the discovery of direct thrombin and factor Xa inhibitors. *Trends Pharmacol Sci* **2012,** *33* (5), 279-88.

5. Pei, J.; Yin, N.; Ma, X.; Lai, L., Systems biology brings new dimensions for structure-based drug design. *J Am Chem Soc* **2014,** *136* (33), 11556-65.

6. Dar, A. C.; Das, T. K.; Shokat, K. M.; Cagan, R. L., Chemical genetic discovery of targets and anti-targets for cancer polypharmacology. *Nature* **2012,** *486* (7401), 80-4.

7. Swinney, D. C.; Anthony, J., How were new medicines discovered? *Nat Rev Drug Discov* **2011,** *10* (7), 507-19.

8. Prathipati, P.; Ma, N. L.; Keller, T. H., Global Bayesian models for the prioritization of antitubercular agents. *J Chem Inf Model* **2008,** *48* (12), 2362-70.

9. (a) Prathipati, P.; Saxena, A. K., Characterization of beta3-adrenergic receptor: determination of pharmacophore and 3D QSAR model for beta3 adrenergic receptor agonism. *J Comput Aided Mol Des* **2005,** *19* (2), 93-110; (b) Prathipati, P.; Pandey, G.; Saxena, A. K., CoMFA and docking studies on glycogen phosphorylase a inhibitors as antidiabetic agents. *J Chem Inf Model* **2005,** *45* (1), 136-45; (c) Narender, T.; Shweta, S.; Tiwari, P.; Papi Reddy, K.; Khaliq, T.; Prathipati, P.; Puri, A.; Srivastava, A. K.; Chander, R.; Agarwal, S. C.; Raj, K., Antihyperglycemic and antidyslipidemic agent from Aegle marmelos. *Bioorg Med Chem Lett* **2007,** *17* (6), 1808-11; (d) Sharma, P.; Singh, S.; Siddiqui, T. I.; Singh, V. S.; Kundu, B.; Prathipati, P.; Saxena, A. K.; Dikshit, D. K.; Rastogi, L.; Dixit, C.; Gupta, M. B.; Patnaik, G. K.; Dikshit, M., alpha-Amino acid derivatives as proton pump inhibitors and potent anti-ulcer agents. *Eur J Med Chem* **2007,** *42* (3), 386-93.

10. Saxena, A. K.; Prathipati, P., Collection and preparation of molecular databases for virtual screening. *SAR QSAR Environ Res* **2006,** *17* (4), 371-92.

11. Prathipati, P.; Saxena, A. K., Evaluation of binary QSAR models derived from LUDI and MOE scoring functions for structure based virtual screening. *J Chem Inf Model* **2006,** *46* (1), 39-51.

12. Durrant, J. D.; McCammon, J. A., Molecular dynamics simulations and drug discovery. *BMC Biol* **2011,** *9*, 71.

13. (a) Merz, K. M., Jr., Using quantum mechanical approaches to study biological systems. *Acc Chem Res* **2014,** *47* (9), 2804-11; (b) Zhou, T.; Huang, D.; Caflisch, A., Quantum mechanical methods for drug design. *Curr Top Med Chem* **2010,** *10* (1), 33-45.

14. (a) Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H.; McCammon, J. A., Discovery of a novel binding trench in HIV integrase. *J Med Chem* **2004,** *47* (8), 1879-81; (b) Ivetac, A.; McCammon, J. A., Mapping the druggable allosteric space of G-protein coupled receptors: a fragment-based molecular dynamics approach. *Chem Biol Drug Des* **2010,** *76* (3), 201-17.

15. Choudhury, C.; Priyakumar, U. D.; Sastry, G. N., Dynamics based pharmacophore models for screening potential inhibitors of mycobacterial cyclopropane synthase. *J Chem Inf Model* **2015,** *55* (4), 848-60.

16. Qu, X.; Latino, D. A.; Aires-de-Sousa, J., A big data approach to the ultra-fast prediction of DFT-calculated bond energies. *J Cheminform* **2013,** *5*, 34.

17. Swamidass, S. J.; Schillebeeckx, C. N.; Matlock, M.; Hurle, M. R.; Agarwal, P., Combined Analysis of Phenotypic and Target-Based Screening in Assay Networks. *J Biomol Screen* **2014,** *19* (5), 782-790.

18. Harrison, C., Phenotypic screening: A more rapid route to target deconvolution. *Nat Rev Drug Discov* **2014,** *13* (2), 102.

19. Choi, H.; Kim, J. Y.; Chang, Y. T.; Nam, H. G., Forward chemical genetic screening. *Methods Mol Biol* **2014,** *1062*, 393-404.

20. Vincent, F.; Loria, P.; Pregel, M.; Stanton, R.; Kitching, L.; Nocka, K.; Doyonnas, R.; Steppan, C.; Gilbert, A.; Schroeter, T.; Peakman, M. C., Developing predictive assays: The phenotypic screening "rule of 3". *Sci Transl Med* **2015,** *7* (293), 293ps15.

21. Kitano, H., A robustness-based approach to systems-oriented drug design. *Nat Rev Drug Discov* **2007,** *6* (3), 202-10.

22. Deans, A. R.; Lewis, S. E.; Huala, E.; Anzaldo, S. S.; Ashburner, M.; Balhoff, J. P.; Blackburn, D. C.; Blake, J. A.; Burleigh, J. G.; Chanet, B.; Cooper, L. D.; Courtot, M.; Csosz, S.; Cui, H.; Dahdul, W.; Das, S.; Dececchi, T. A.; Dettai, A.; Diogo, R.; Druzinsky, R. E.; Dumontier, M.; Franz, N. M.; Friedrich, F.; Gkoutos, G. V.; Haendel, M.; Harmon, L. J.; Hayamizu, T. F.; He, Y.; Hines, H. M.; Ibrahim, N.; Jackson, L. M.; Jaiswal, P.; James-Zorn, C.; Kohler, S.; Lecointre, G.; Lapp, H.; Lawrence, C. J.; Le Novere, N.; Lundberg, J. G.; Macklin, J.; Mast, A. R.; Midford, P. E.; Miko, I.; Mungall, C. J.; Oellrich, A.; Osumi-Sutherland, D.; Parkinson, H.; Ramirez, M. J.; Richter, S.; Robinson, P. N.; Ruttenberg, A.; Schulz, K. S.; Segerdell, E.; Seltmann, K. C.; Sharkey, M. J.; Smith, A. D.; Smith, B.; Specht, C. D.; Squires, R. B.; Thacker, R. W.; Thessen, A.; Fernandez-Triana, J.; Vihinen, M.; Vize, P. D.; Vogt, L.; Wall, C. E.; Walls, R. L.; Westerfeld, M.; Wharton, R. A.; Wirkner, C. S.; Woolley, J. B.; Yoder, M. J.; Zorn, A. M.; Mabee, P., Finding our way through phenotypes. *PLoS Biol* **2015,** *13* (1), e1002033.

23. Gupta, P. B.; Onder, T. T.; Jiang, G.; Tao, K.; Kuperwasser, C.; Weinberg, R. A.; Lander, E. S., Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* **2009,** *138* (4), 645-59.

24. Waters, M. D.; Fostel, J. M., Toxicogenomics and systems toxicology: aims and prospects. *Nat Rev Genet* **2004,** *5* (12), 936-48.

25. Shang, D.; Li, C.; Yao, Q.; Yang, H.; Xu, Y.; Han, J.; Li, J.; Su, F.; Zhang, Y.; Zhang, C.; Li, D.; Li, X., Prioritizing candidate disease metabolites based on global functional relationships between metabolites in the context of metabolic pathways. *PLoS One* **2014,** *9* (8), e104934.

26. Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc* **2007,** *2* (11), 2692-703.

27. Hollywood, K.; Brison, D. R.; Goodacre, R., Metabolomics: current technologies and future trends. *Proteomics* **2006,** *6* (17), 4716-23.

28. Kell, D. B., Metabolomics and systems biology: making sense of the soup. *Curr Opin Microbiol* **2004,** *7* (3), 296-307.

29. Pujol, A.; Mosca, R.; Farres, J.; Aloy, P., Unveiling the role of network and systems biology in drug discovery. *Trends Pharmacol Sci* **2010,** *31* (3), 115-23.

30. Faratian, D.; Clyde, R. G.; Crawford, J. W.; Harrison, D. J., Systems pathology--taking molecular pathology into a new dimension. *Nat Rev Clin Oncol* **2009,** *6* (8), 455-64.

31. Stein, A.; Pache, R. A.; Bernado, P.; Pons, M.; Aloy, P., Dynamic interactions of proteins in complex networks: a more structured view. *FEBS J* **2009,** *276* (19), 5390-405.

32. Ekins, S.; Mestres, J.; Testa, B., In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Br J Pharmacol* **2007,** *152* (1), 9-20.

33. (a) Vempati, U. D.; Chung, C.; Mader, C.; Koleti, A.; Datar, N.; Vidovic, D.; Wrobel, D.; Erickson, S.; Muhlich, J. L.; Berriz, G.; Benes, C. H.; Subramanian, A.; Pillai, A.; Shamu, C. E.; Schurer, S. C., Metadata Standard and Data Exchange Specifications to Describe, Model, and Integrate Complex and Diverse High-Throughput Screening Data from the Library of Integrated Network-based Cellular Signatures (LINCS). *J Biomol Screen* **2014,** *19* (5), 803-816; (b) Vidovic, D.; Koleti, A.; Schurer,

S. C., Large-scale integration of small molecule-induced genome-wide transcriptional responses, Kinome-wide binding affinities and cell-growth inhibition profiles reveal global trends characterizing systems-level drug action. *Front Genet* **2014,** *5* (1664-8021 (Electronic)), 342.

34. (a) Kitano, H., Cancer robustness: tumour tactics. *Nature* **2003,** *426* (6963), 125; (b) Friedlander, T.; Mayo, A. E.; Tlusty, T.; Alon, U., Evolution of bow-tie architectures in biology. *PLoS Comput Biol* **2015,** *11* (3), e1004055.

35. Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Kruger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P., The ChEMBL bioactivity database: an update. *Nucleic Acids Res* **2014,** *42* (Database issue), D1083-90.

36. Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K., BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res* **2007,** *35* (Database issue), D198-201.

37. Chen, Y. A.; Tripathi, L. P.; Mizuguchi, K., TargetMine, an integrated data warehouse for candidate gene prioritisation and target discovery. *PLoS One* **2011,** *6* (3), e17844.

38. Ito, J.; Ikeda, K.; Yamada, K.; Mizuguchi, K.; Tomii, K., PoSSuM v.2.0: data update and a new function for investigating ligand analogs and target proteins of small-molecule drugs. *Nucleic Acids Res* **2015,** *43* (Database issue), D392-8.

39. Lamb, J.; Crawford, E. D.; Peck, D.; Modell, J. W.; Blat, I. C.; Wrobel, M. J.; Lerner, J.; Brunet, J. P.; Subramanian, A.; Ross, K. N.; Reich, M.; Hieronymus, H.; Wei, G.; Armstrong, S. A.; Haggarty, S. J.; Clemons, P. A.; Wei, R.; Carr, S. A.; Lander, E. S.; Golub, T. R., The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **2006,** *313* (5795), 1929-35.

40. San Lucas, F. A.; Fowler, J.; Chang, K.; Kopetz, S.; Vilar, E.; Scheet, P., Cancer in silico drug discovery: a systems biology tool for identifying candidate drugs to target specific molecular tumor subtypes. *Mol Cancer Ther* **2014,** *13* (12), 3230-40.

41. Nystrom-Persson, J.; Igarashi, Y.; Ito, M.; Morita, M.; Nakatsu, N.; Yamada, H.; Mizuguchi, K., Toxygates: interactive toxicity analysis on a hybrid microarray and linked data platform. *Bioinformatics* **2013,** *29* (23), 3080-6.

42. Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.; Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S.; Mesirov, J. P., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **2005,** *102* (43), 15545-50.

43. Verbist, B.; Klambauer, G.; Vervoort, L.; Talloen, W.; Shkedy, Z.; Thas, O.; Bender, A.; Gohlmann, H. W.; Hochreiter, S., Using transcriptomics to guide lead optimization in drug discovery projects: Lessons learned from the QSTAR project. *Drug Discov Today* **2015**.

44. Xia, J.; Wishart, D. S., MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Res* **2010,** *38* (Web Server issue), W71-7.

45. (a) Xia, J.; Mandal, R.; Sinelnikov, I. V.; Broadhurst, D.; Wishart, D. S., MetaboAnalyst 2.0--a comprehensive server for metabolomic data analysis. *Nucleic Acids Res* **2012,** *40* (Web Server issue), W127-33; (b) Xia, J.; Wishart, D. S., Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. *Nat Protoc* **2011,** *6* (6), 743-60.

46. Laenen, G.; Ardeshirdavani, A.; Moreau, Y.; Thorrez, L., Galahad: a web server for drug effect analysis from gene expression. *Nucleic Acids Res* **2015,** *43* (W1), W208-12.

47. Chen, E. Y.; Xu, H.; Gordonov, S.; Lim, M. P.; Perkins, M. H.; Ma'ayan, A., Expression2Kinases: mRNA profiling linked to multiple upstream regulatory layers. *Bioinformatics* **2012,** *28* (1), 105-11.

48. Terfve, C.; Cokelaer, T.; Henriques, D.; MacNamara, A.; Goncalves, E.; Morris, M. K.; van Iersel, M.; Lauffenburger, D. A.; Saez-Rodriguez, J., CellNOptR: a flexible toolkit to train protein signaling networks to data using multiple logic formalisms. *BMC Syst Biol* **2012,** *6*, 133.

49. Liberzon, A.; Subramanian, A.; Pinchback, R.; Thorvaldsdottir, H.; Tamayo, P.; Mesirov, J. P., Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **2011,** *27* (12), 1739-40.

50. (a) Araki, H.; Knapp, C.; Tsai, P.; Print, C., GeneSetDB: A comprehensive meta-database, statistical and visualisation framework for gene set analysis. *FEBS Open Bio* **2012,** *2*, 76-82; (b) Jiang, Z.; Gentleman, R., Extensions to gene set enrichment. *Bioinformatics* **2007,** *23* (3), 306-13.

51. Chen, E. Y.; Tan, C. M.; Kou, Y.; Duan, Q.; Wang, Z.; Meirelles, G. V.; Clark, N. R.; Ma'ayan, A., Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **2013,** *14*, 128.

52. (a) Sandmann, T.; Kummerfeld, S. K.; Gentleman, R.; Bourgon, R., gCMAP: user-friendly connectivity mapping with R. *Bioinformatics* **2014,** *30* (1), 127-8; (b) Fisher, R. A., On the Interpretation of χ2 from Contingency Tables, and the Calculation of P. *Journal of the Royal Statistical Society* **1922,** *85* (1), 87-94; (c) Wilcoxon, F., Individual Comparisons by Ranking Methods. *Biometrics Bulletin* **1945,** *1* (6), 80-83; (d) Wu, D.; Smyth, G. K., Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res* **2012,** *40* (17), e133; (e) Wu, D.; Lim, E.; Vaillant, F.; Asselin-Labat, M. L.; Visvader, J. E.; Smyth, G. K., ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics* **2010,** *26* (17), 2176-82; (f) Ritchie, M. E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C. W.; Shi, W.; Smyth, G. K., limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **2015,** *43* (7), e47; (g) Bauer, S.; Gagneur, J.; Robinson, P. N., GOing Bayesian: model-based gene set analysis of genome-scale data. *Nucleic Acids Res* **2010,** *38* (11), 3523-32.

53. Zhao, B.; Pritchard, J. R.; Lauffenburger, D. A.; Hemann, M. T., Addressing genetic tumor heterogeneity through computationally predictive combination therapy. *Cancer Discov* **2014,** *4* (2), 166-74.

54. Al-Lazikani, B.; Banerji, U.; Workman, P., Combinatorial drug therapy for cancer in the post-genomic era. *Nat Biotechnol* **2012,** *30* (7), 679-92.

55. (a) Jojic, V.; Shay, T.; Sylvia, K.; Zuk, O.; Sun, X.; Kang, J.; Regev, A.; Koller, D.; Best, A. J.; Knell, J.; Goldrath, A.; Joic, V.; Koller, D.; Shay, T.; Regev, A.; Cohen, N.; Brennan, P.; Brenner, M.; Kim, F.; Rao, T. N.; Wagers, A.; Heng, T.; Ericson, J.; Rothamel, K.; Ortiz-Lopez, A.; Mathis, D.; Benoist, C.; Bezman, N. A.; Sun, J. C.; Min-Oo, G.; Kim, C. C.; Lanier, L. L.; Miller, J.; Brown, B.; Merad, M.; Gautier, E. L.; Jakubzick, C.; Randolph, G. J.; Monach, P.; Blair, D. A.; Dustin, M. L.; Shinton, S. A.; Hardy, R. R.; Laidlaw, D.; Collins, J.; Gazit, R.; Rossi, D. J.; Malhotra, N.; Sylvia, K.; Kang, J.; Kreslavsky, T.; Fletcher, A.; Elpek, K.; Bellemarte-Pelletier, A.; Malhotra, D.; Turley, S., Identification of transcriptional regulators in the mouse immune system. *Nat Immunol* **2013,** *14* (6), 633-43; (b) Garnett, M. J.; Edelman, E. J.; Heidorn, S. J.; Greenman, C. D.; Dastur, A.; Lau, K. W.; Greninger, P.; Thompson, I. R.; Luo, X.; Soares, J.; Liu, Q.; Iorio, F.; Surdez, D.; Chen, L.; Milano, R. J.; Bignell, G. R.; Tam, A. T.; Davies, H.; Stevenson, J. A.; Barthorpe, S.; Lutz, S. R.; Kogera, F.; Lawrence, K.; McLaren-Douglas, A.; Mitropoulos, X.; Mironenko, T.; Thi, H.; Richardson, L.; Zhou, W.; Jewitt, F.; Zhang, T.; O'Brien, P.; Boisvert, J. L.; Price, S.; Hur, W.; Yang, W.; Deng, X.; Butler, A.; Choi, H. G.; Chang, J. W.; Baselga, J.; Stamenkovic, I.; Engelman, J. A.; Sharma, S. V.; Delattre, O.; Saez-Rodriguez, J.; Gray, N. S.; Settleman, J.; Futreal, P. A.; Haber, D. A.; Stratton, M. R.; Ramaswamy, S.; McDermott, U.; Benes, C. H., Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **2012,** *483* (7391), 570-5; (c) Barretina, J.; Caponigro, G.; Stransky, N.; Venkatesan, K.; Margolin, A. A.; Kim, S.; Wilson, C. J.; Lehar, J.; Kryukov, G. V.; Sonkin, D.; Reddy, A.; Liu, M.; Murray, L.; Berger, M. F.; Monahan, J. E.; Morais, P.; Meltzer, J.; Korejwa, A.; Jane-Valbuena, J.; Mapa, F. A.; Thibault, J.; Bric-Furlong, E.; Raman, P.; Shipway, A.; Engels, I. H.; Cheng, J.; Yu, G. K.; Yu, J.; Aspesi, P., Jr.; de Silva, M.; Jagtap, K.; Jones, M. D.; Wang, L.; Hatton, C.; Palescandolo, E.; Gupta, S.; Mahan, S.; Sougnez, C.; Onofrio, R. C.; Liefeld, T.; MacConaill, L.; Winckler, W.; Reich, M.; Li, N.; Mesirov, J. P.; Gabriel, S. B.; Getz, G.; Ardlie, K.; Chan, V.; Myer, V. E.; Weber, B. L.; Porter, J.; Warmuth, M.; Finan, P.; Harris, J. L.; Meyerson, M.; Golub, T. R.; Morrissey, M. P.; Sellers, W. R.; Schlegel, R.; Garraway, L. A., The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **2012,** *483* (7391), 603-7.

56. (a) de Sousa Abreu, R.; Penalva, L. O.; Marcotte, E. M.; Vogel, C., Global signatures of protein and mRNA expression levels. *Mol Biosyst* **2009,** *5* (12), 1512-26; (b) Ning, K.; Fermin, D.; Nesvizhskii,

A. I., Comparative analysis of different label-free mass spectrometry based protein abundance estimates and their correlation with RNA-Seq gene expression data. *J Proteome Res* **2012,** *11* (4), 2261-71; (c) Schwanhausser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M., Global quantification of mammalian gene expression control. *Nature* **2011,** *473* (7347), 337-42; (d) Vogel, C.; Marcotte, E. M., Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* **2012,** *13* (4), 227-32.

57. (a) Deane, C. M.; Salwinski, L.; Xenarios, I.; Eisenberg, D., Protein interactions: two methods for assessment of the reliability of high throughput observations. *Mol Cell Proteomics* **2002,** *1* (5), 349-56; (b) Cokus, S.; Mizutani, S.; Pellegrini, M., An improved method for identifying functionally linked proteins using phylogenetic profiles. *BMC Bioinformatics* **2007,** *8 Suppl 4*, S7; (c) Jansen, R.; Yu, H.; Greenbaum, D.; Kluger, Y.; Krogan, N. J.; Chung, S.; Emili, A.; Snyder, M.; Greenblatt, J. F.; Gerstein, M., A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science* **2003,** *302* (5644), 449-53; (d) Wang, J.; Li, C.; Wang, E.; Wang, X., Uncovering the rules for protein-protein interactions from yeast genomic data. *Proc Natl Acad Sci U S A* **2009,** *106* (10), 3752-7; (e) Xing, C.; Dunson, D. B., Bayesian inference for genomic data integration reduces misclassification rate in predicting protein-protein interactions. *PLoS Comput Biol* **2011,** *7* (7), e1002110.

58. Noble, D., Evolution beyond neo-Darwinism: a new conceptual framework. *J Exp Biol* **2015,** *218* (Pt 1), 7-13.

59. Franklin, S.; Vondriska, T. M., Genomes, proteomes, and the central dogma. *Circ Cardiovasc Genet* **2011,** *4* (5), 576.

60. (a) Xie, L.; Ge, X.; Tan, H.; Xie, L.; Zhang, Y.; Hart, T.; Yang, X.; Bourne, P. E., Towards structural systems pharmacology to study complex diseases and personalized medicine. *PLoS Comput Biol* **2014,** *10* (5), e1003554; (b) Xie, L.; Xie, L.; Bourne, P. E., Structure-based systems biology for analyzing off-target binding. *Curr Opin Struct Biol* **2011,** *21* (2), 189-99; (c) Chang, R. L.; Xie, L.; Bourne, P. E.; Palsson, B. O., Antibacterial mechanisms identified through structural systems pharmacology. *BMC Syst Biol* **2013,** *7*, 102; (d) Xie, L.; Evangelidis, T.; Xie, L.; Bourne, P. E., Drug discovery using chemical systems biology: weak inhibition of multiple kinases may contribute to the anti-cancer effect of nelfinavir. *PLoS Comput Biol* **2011,** *7* (4), e1002037.

61. Fournier, M. L.; Paulson, A.; Pavelka, N.; Mosley, A. L.; Gaudenz, K.; Bradford, W. D.; Glynn, E.; Li, H.; Sardiu, M. E.; Fleharty, B.; Seidel, C.; Florens, L.; Washburn, M. P., Delayed correlation of mRNA and protein expression in rapamycin-treated cells and a role for Ggc1 in cellular sensitivity to rapamycin. *Mol Cell Proteomics* **2010,** *9* (2), 271-84.

62. (a) Goentoro, L.; Shoval, O.; Kirschner, M. W.; Alon, U., The incoherent feedforward loop can provide fold-change detection in gene regulation. *Mol Cell* **2009,** *36* (5), 894-9; (b) Shen-Orr, S. S.; Milo, R.; Mangan, S.; Alon, U., Network motifs in the transcriptional regulation network of Escherichia coli. *Nat Genet* **2002,** *31* (1), 64-8; (c) Yeger-Lotem, E.; Sattath, S.; Kashtan, N.; Itzkovitz, S.; Milo, R.; Pinter, R. Y.; Alon, U.; Margalit, H., Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. *Proc Natl Acad Sci U S A* **2004,** *101* (16), 5934-9.

63. Iskar, M.; Campillos, M.; Kuhn, M.; Jensen, L. J.; van Noort, V.; Bork, P., Drug-induced regulation of target expression. *PLoS Comput Biol* **2010,** *6* (9).

64. Koussounadis, A.; Langdon, S. P.; Um, I. H.; Harrison, D. J.; Smith, V. A., Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. *Sci Rep* **2015,** *5*, 10775.

65. Choi, H.; Pavelka, N., When one and one gives more than two: challenges and opportunities of integrative omics. *Front Genet* **2011,** *2*, 105.

66. Creixell, P.; Schoof, E. M.; Erler, J. T.; Linding, R., Navigating cancer network attractors for tumor-specific therapy. *Nat Biotechnol* **2012,** *30* (9), 842-8.

67. Kim, J.; Park, S. M.; Cho, K. H., Discovery of a kernel for controlling biomolecular regulatory networks. *Sci Rep* **2013,** *3* (2045-2322 (Electronic)), 2223.

68. Gomez-Cabrero, D.; Abugessaisa, I.; Maier, D.; Teschendorff, A.; Merkenschlager, M.; Gisel, A.; Ballestar, E.; Bongcam-Rudloff, E.; Conesa, A.; Tegner, J., Data integration in the era of omics: current and future challenges. *BMC Syst Biol* **2014,** *8 Suppl 2*, I1.
69. Hwang, D.; Rust, A. G.; Ramsey, S.; Smith, J. J.; Leslie, D. M.; Weston, A. D.; de Atauri, P.; Aitchison, J. D.; Hood, L.; Siegel, A. F.; Bolouri, H., A data integration methodology for systems biology. *Proc Natl Acad Sci U S A* **2005,** *102* (48), 17296-301.
70. Pavlopoulos, G. A.; Secrier, M.; Moschopoulos, C. N.; Soldatos, T. G.; Kossida, S.; Aerts, J.; Schneider, R.; Bagos, P. G., Using graph theory to analyze biological networks. *BioData Min* **2011,** *4*, 10.
71. (a) Bredel, M.; Jacoby, E., Chemogenomics: an emerging strategy for rapid target and drug discovery. *Nat Rev Genet* **2004,** *5* (4), 262-75; (b) Jacoby, E.; Bouhelal, R.; Gerspacher, M.; Seuwen, K., The 7 TM G-protein-coupled receptor target family. *ChemMedChem* **2006,** *1* (8), 761-82; (c) Glick, M.; Jacoby, E., The role of computational methods in the identification of bioactive compounds. *Curr Opin Chem Biol* **2011,** *15* (4), 540-6; (d) Hartenfeller, M.; Eberle, M.; Meier, P.; Nieto-Oberhuber, C.; Altmann, K. H.; Schneider, G.; Jacoby, E.; Renner, S., Probing the bioactivity-relevant chemical space of robust reactions and common molecular building blocks. *J Chem Inf Model* **2012,** *52* (5), 1167-78.
72. Wang, Y.; Suzek, T.; Zhang, J.; Wang, J.; He, S.; Cheng, T.; Shoemaker, B. A.; Gindulyte, A.; Bryant, S. H., PubChem BioAssay: 2014 update. *Nucleic Acids Res* **2014,** *42* (Database issue), D1075-82.
73. Backman, T. W.; Cao, Y.; Girke, T., ChemMine tools: an online service for analyzing and clustering small molecules. *Nucleic Acids Res* **2011,** *39* (Web Server issue), W486-91.
74. Matlock, M. K.; Zaretzki, J. M.; Swamidass, S. J., Scaffold network generator: a tool for mining molecular structures. *Bioinformatics* **2013,** *29* (20), 2655-6.
75. Cao, D. S.; Xiao, N.; Xu, Q. S.; Chen, A. F., Rcpi: R/Bioconductor package to generate various descriptors of proteins, compounds and their interactions. *Bioinformatics* **2015,** *31* (2), 279-81.
76. Hinselmann, G.; Rosenbaum, L.; Jahn, A.; Fechner, N.; Zell, A., jCompoundMapper: An open source Java library and command-line tool for chemical fingerprints. *J Cheminform* **2011,** *3* (1), 3.
77. Langfelder, P.; Horvath, S., WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **2008,** *9* (1471-2105 (Electronic)), 559.
78. Hochreiter, S.; Bodenhofer, U.; Heusel, M.; Mayr, A.; Mitterecker, A.; Kasim, A.; Khamiakova, T.; Van Sanden, S.; Lin, D.; Talloen, W.; Bijnens, L.; Gohlmann, H. W.; Shkedy, Z.; Clevert, D. A., FABIA: factor analysis for bicluster acquisition. *Bioinformatics* **2010,** *26* (12), 1520-7.
79. Szklarczyk, D.; Franceschini, A.; Wyder, S.; Forslund, K.; Heller, D.; Huerta-Cepas, J.; Simonovic, M.; Roth, A.; Santos, A.; Tsafou, K. P.; Kuhn, M.; Bork, P.; Jensen, L. J.; von Mering, C., STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **2015,** *43* (Database issue), D447-52.
80. 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), D199-205.
81. UniProt, C., Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res* **2014,** *42* (Database issue), D191-8.
82. Xiao, N.; Cao, D. S.; Zhu, M. F.; Xu, Q. S., protr/ProtrWeb: R package and web server for generating various numerical representation schemes of protein sequences. *Bioinformatics* **2015**.
83. Smedley, D.; Haider, S.; Durinck, S.; Pandini, L.; Provero, P.; Allen, J.; Arnaiz, O.; Awedh, M. H.; Baldock, R.; Barbiera, G.; Bardou, P.; Beck, T.; Blake, A.; Bonierbale, M.; Brookes, A. J.; Bucci, G.; Buetti, I.; Burge, S.; Cabau, C.; Carlson, J. W.; Chelala, C.; Chrysostomou, C.; Cittaro, D.; Collin, O.; Cordova, R.; Cutts, R. J.; Dassi, E.; Genova, A. D.; Djari, A.; Esposito, A.; Estrella, H.; Eyras, E.; Fernandez-Banet, J.; Forbes, S.; Free, R. C.; Fujisawa, T.; Gadaleta, E.; Garcia-Manteiga, J. M.; Goodstein, D.; Gray, K.; Guerra-Assuncao, J. A.; Haggarty, B.; Han, D. J.; Han, B. W.; Harris, T.; Harshbarger, J.; Hastings, R. K.; Hayes, R. D.; Hoede, C.; Hu, S.; Hu, Z. L.; Hutchins, L.; Kan, Z.; Kawaji, H.; Keliet, A.; Kerhornou, A.; Kim, S.; Kinsella, R.; Klopp, C.; Kong, L.; Lawson, D.; Lazarevic, D.; Lee, J.

H.; Letellier, T.; Li, C. Y.; Lio, P.; Liu, C. J.; Luo, J.; Maass, A.; Mariette, J.; Maurel, T.; Merella, S.; Mohamed, A. M.; Moreews, F.; Nabihoudine, I.; Ndegwa, N.; Noirot, C.; Perez-Llamas, C.; Primig, M.; Quattrone, A.; Quesneville, H.; Rambaldi, D.; Reecy, J.; Riba, M.; Rosanoff, S.; Saddiq, A. A.; Salas, E.; Sallou, O.; Shepherd, R.; Simon, R.; Sperling, L.; Spooner, W.; Staines, D. M.; Steinbach, D.; Stone, K.; Stupka, E.; Teague, J. W.; Dayem Ullah, A. Z.; Wang, J.; Ware, D.; Wong-Erasmus, M.; Youens-Clark, K.; Zadissa, A.; Zhang, S. J.; Kasprzyk, A., The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res* **2015**.

84. Barrett, T.; Wilhite, S. E.; Ledoux, P.; Evangelista, C.; Kim, I. F.; Tomashevsky, M.; Marshall, K. A.; Phillippy, K. H.; Sherman, P. M.; Holko, M.; Yefanov, A.; Lee, H.; Zhang, N.; Robertson, C. L.; Serova, N.; Davis, S.; Soboleva, A., NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* **2013,** *41* (Database issue), D991-5.

85. Gautier, L.; Cope, L.; Bolstad, B. M.; Irizarry, R. A., affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **2004,** *20* (3), 307-15.

86. Andrews, S., Babraham Bioinformatics - SeqMonk Mapped Sequence Analysis Tool.

87. Schaefer, M. H.; Fontaine, J. F.; Vinayagam, A.; Porras, P.; Wanker, E. E.; Andrade-Navarro, M. A., HIPPIE: Integrating protein interaction networks with experiment based quality scores. *PLoS One* **2012,** *7* (2), e31826.

88. Tarca, A. L.; Draghici, S.; Khatri, P.; Hassan, S. S.; Mittal, P.; Kim, J. S.; Kim, C. J.; Kusanovic, J. P.; Romero, R., A novel signaling pathway impact analysis. *Bioinformatics* **2009,** *25* (1), 75-82.

89. Varemo, L.; Nielsen, J.; Nookaew, I., Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Res* **2013,** *41* (8), 4378-91.

90. Dennis, G., Jr.; Sherman, B. T.; Hosack, D. A.; Yang, J.; Gao, W.; Lane, H. C.; Lempicki, R. A., DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **2003,** *4* (5), P3.

91. Krive, J.; Patel, M.; Gehm, L.; Mackey, M.; Kulstad, E.; Li, J. J.; Lussier, Y. A.; Boyd, A. D., The complexity and challenges of the International Classification of Diseases, Ninth Revision, Clinical Modification to International Classification of Diseases, 10th Revision, Clinical Modification transition in EDs. *Am J Emerg Med* **2015**.

92. Kuhn, M.; Campillos, M.; Letunic, I.; Jensen, L. J.; Bork, P., A side effect resource to capture phenotypic effects of drugs. *Mol Syst Biol* **2010,** *6*, 343.

93. Igarashi, Y.; Nakatsu, N.; Yamashita, T.; Ono, A.; Ohno, Y.; Urushidani, T.; Yamada, H., Open TG-GATEs: a large-scale toxicogenomics database. *Nucleic Acids Res* **2015,** *43* (Database issue), D921-7.

94. Pandian, G. N.; Taniguchi, J.; Junetha, S.; Sato, S.; Han, L.; Saha, A.; AnandhaKumar, C.; Bando, T.; Nagase, H.; Vaijayanthi, T.; Taylor, R. D.; Sugiyama, H., Distinct DNA-based epigenetic switches trigger transcriptional activation of silent genes in human dermal fibroblasts. *Sci Rep* **2014,** *4*, 3843.

95. Ganter, B.; Snyder, R. D.; Halbert, D. N.; Lee, M. D., Toxicogenomics in drug discovery and development: mechanistic analysis of compound/class-dependent effects using the DrugMatrix database. *Pharmacogenomics* **2006,** *7* (7), 1025-44.

96. Chelliah, V.; Juty, N.; Ajmera, I.; Ali, R.; Dumousseau, M.; Glont, M.; Hucka, M.; Jalowicki, G.; Keating, S.; Knight-Schrijver, V.; Lloret-Villas, A.; Natarajan, K. N.; Pettit, J. B.; Rodriguez, N.; Schubert, M.; Wimalaratne, S. M.; Zhao, Y.; Hermjakob, H.; Le Novere, N.; Laibe, C., BioModels: ten-year anniversary. *Nucleic Acids Res* **2015,** *43* (Database issue), D542-8.

97. Zhang, J. D.; Wiemann, S., KEGGgraph: a graph approach to KEGG PATHWAY in R and bioconductor. *Bioinformatics* **2009,** *25* (11), 1470-1.

98. Mussel, C.; Hopfensitz, M.; Kestler, H. A., BoolNet--an R package for generation, reconstruction and analysis of Boolean networks. *Bioinformatics* **2010,** *26* (10), 1378-80.

99. Sutterlin, T.; Kolb, C.; Dickhaus, H.; Jager, D.; Grabe, N., Bridging the scales: semantic integration of quantitative SBML in graphical multi-cellular models and simulations with EPISIM and COPASI. *Bioinformatics* **2013,** *29* (2), 223-9.
100. Chaouiya, C.; Naldi, A.; Thieffry, D., Logical modelling of gene regulatory networks with GINsim. *Methods Mol Biol* **2012,** *804*, 463-79.
101. Matsuoka, Y.; Funahashi, A.; Ghosh, S.; Kitano, H., Modeling and simulation using CellDesigner. *Methods Mol Biol* **2014,** *1164*, 121-45.
102. (a) Noeske, T.; Sasse, B. C.; Stark, H.; Parsons, C. G.; Weil, T.; Schneider, G., Predicting compound selectivity by self-organizing maps: cross-activities of metabotropic glutamate receptor antagonists. *ChemMedChem* **2006,** *1* (10), 1066-8; (b) 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.; Edwards, D. D.; Shoichet, B. K.; Roth, B. L., Predicting new molecular targets for known drugs. *Nature* **2009,** *462* (7270), 175-81; (c) Reymond, J. L.; Awale, M., Exploring chemical space for drug discovery using the chemical universe database. *ACS Chem Neurosci* **2012,** *3* (9), 649-57; (d) Benz, R. W.; Swamidass, S. J.; Baldi, P., Discovery of power-laws in chemical space. *J Chem Inf Model* **2008,** *48* (6), 1138-51.
103. (a) Haupt, V. J.; Schroeder, M., Old friends in new guise: repositioning of known drugs with structural bioinformatics. *Brief Bioinform* **2011,** *12* (4), 312-26; (b) Defranchi, E.; Schalon, C.; Messa, M.; Onofri, F.; Benfenati, F.; Rognan, D., Binding of protein kinase inhibitors to synapsin I inferred from pair-wise binding site similarity measurements. *PLoS One* **2010,** *5* (8), e12214; (c) Zahler, S.; Tietze, S.; Totzke, F.; Kubbutat, M.; Meijer, L.; Vollmar, A. M.; Apostolakis, J., Inverse in silico screening for identification of kinase inhibitor targets. *Chem Biol* **2007,** *14* (11), 1207-14; (d) Kinnings, S. L.; Liu, N.; Buchmeier, N.; Tonge, P. J.; Xie, L.; Bourne, P. E., Drug discovery using chemical systems biology: repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. *PLoS Comput Biol* **2009,** *5* (7), e1000423.
104. (a) Akbani, R.; Ng, P. K.; Werner, H. M.; Shahmoradgoli, M.; Zhang, F.; Ju, Z.; Liu, W.; Yang, J. Y.; Yoshihara, K.; Li, J.; Ling, S.; Seviour, E. G.; Ram, P. T.; Minna, J. D.; Diao, L.; Tong, P.; Heymach, J. V.; Hill, S. M.; Dondelinger, F.; Stadler, N.; Byers, L. A.; Meric-Bernstam, F.; Weinstein, J. N.; Broom, B. M.; Verhaak, R. G.; Liang, H.; Mukherjee, S.; Lu, Y.; Mills, G. B., A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nat Commun* **2014,** *5*, 3887; (b) Weinstein, J. N.; Collisson, E. A.; Mills, G. B.; Shaw, K. R.; Ozenberger, B. A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J. M., The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **2013,** *45* (10), 1113-20.
105. (a) Duran-Frigola, M.; Rossell, D.; Aloy, P., A chemo-centric view of human health and disease. *Nat Commun* **2014,** *5* (2041-1723 (Electronic)), 5676; (b) Gao, Y. F.; Chen, L.; Huang, G. H.; Zhang, T.; Feng, K. Y.; Li, H. P.; Jiang, Y., Prediction of drugs target groups based on ChEBI ontology. *Biomed Res Int* **2013,** *2013* (2314-6141 (Electronic)), 132724; (c) Wu, L.; Ai, N.; Liu, Y.; Wang, Y.; Fan, X., Relating anatomical therapeutic indications by the ensemble similarity of drug sets. *J Chem Inf Model* **2013,** *53* (8), 2154-60; (d) Chen, L.; Zeng, W. M.; Cai, Y. D.; Feng, K. Y.; Chou, K. C., Predicting Anatomical Therapeutic Chemical (ATC) classification of drugs by integrating chemical-chemical interactions and similarities. *PLoS One* **2012,** *7* (4), e35254; (e) Liu, M.; Wu, Y.; Chen, Y.; Sun, J.; Zhao, Z.; Chen, X. W.; Matheny, M. E.; Xu, H., Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. *J Am Med Inform Assoc* **2012,** *19* (e1), e28-35.
106. (a) Prathipati, P.; Ma, N. L.; Manjunatha, U. H.; Bender, A., Fishing the target of antitubercular compounds: in silico target deconvolution model development and validation. *J Proteome Res* **2009,** *8* (6), 2788-98; (b) Nettles, J. H.; Jenkins, J. L.; Bender, A.; Deng, Z.; Davies, J. W.; Glick, M., Bridging chemical and biological space: "target fishing" using 2D and 3D molecular descriptors. *J Med Chem* **2006,** *49* (23), 6802-10; (c) von der Heyde, S.; Bender, C.; Henjes, F.; Sonntag, J.; Korf, U.; Beissbarth, T., Boolean ErbB network reconstructions and perturbation simulations reveal individual drug response in different breast cancer cell lines. *BMC Syst Biol* **2014,**

8, 75; (d) Surgand, J. S.; Rodrigo, J.; Kellenberger, E.; Rognan, D., A chemogenomic analysis of the transmembrane binding cavity of human G-protein-coupled receptors. *Proteins* **2006,** *62* (2), 509-38.
107. (a) Dudley, J. T.; Deshpande, T.; Butte, A. J., Exploiting drug-disease relationships for computational drug repositioning. *Brief Bioinform* **2011,** *12* (4), 303-11; (b) Suthram, S.; Dudley, J. T.; Chiang, A. P.; Chen, R.; Hastie, T. J.; Butte, A. J., Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. *PLoS Comput Biol* **2010,** *6* (2), e1000662; (c) Sirota, M.; Dudley, J. T.; Kim, J.; Chiang, A. P.; Morgan, A. A.; Sweet-Cordero, A.; Sage, J.; Butte, A. J., Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci Transl Med* **2011,** *3* (96), 96ra77.
108. (a) Costello, J. C.; Heiser, L. M.; Georgii, E.; Gonen, M.; Menden, M. P.; Wang, N. J.; Bansal, M.; Ammad-ud-din, M.; Hintsanen, P.; Khan, S. A.; Mpindi, J. P.; Kallioniemi, O.; Honkela, A.; Aittokallio, T.; Wennerberg, K.; Community, N. D.; Collins, J. J.; Gallahan, D.; Singer, D.; Saez-Rodriguez, J.; Kaski, S.; Gray, J. W.; Stolovitzky, G., A community effort to assess and improve drug sensitivity prediction algorithms. *Nat Biotechnol* **2014,** *32* (12), 1202-12; (b) Li, S.; Zhang, B.; Zhang, N., Network target for screening synergistic drug combinations with application to traditional Chinese medicine. *BMC Syst Biol* **2011,** *5 Suppl 1* (1752-0509 (Electronic)), S10; (c) Anchang, B.; Sadeh, M. J.; Jacob, J.; Tresch, A.; Vlad, M. O.; Oefner, P. J.; Spang, R., Modeling the temporal interplay of molecular signaling and gene expression by using dynamic nested effects models. *Proc Natl Acad Sci U S A* **2009,** *106* (16), 6447-52.
109. (a) Jupp, S.; Malone, J.; Bolleman, J.; Brandizi, M.; Davies, M.; Garcia, L.; Gaulton, A.; Gehant, S.; Laibe, C.; Redaschi, N.; Wimalaratne, S. M.; Martin, M.; Le Novere, N.; Parkinson, H.; Birney, E.; Jenkinson, A. M., The EBI RDF platform: linked open data for the life sciences. *Bioinformatics* **2014,** *30* (9), 1338-9; (b) Chen, B.; Ding, Y.; Wild, D. J., Improving integrative searching of systems chemical biology data using semantic annotation. *J Cheminform* **2012,** *4* (1), 6; (c) Wild, D. J.; Ding, Y.; Sheth, A. P.; Harland, L.; Gifford, E. M.; Lajiness, M. S., Systems chemical biology and the Semantic Web: what they mean for the future of drug discovery research. *Drug Discov Today* **2012,** *17* (9-10), 469-74.
110. 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), D1091-7.
111. EBI's Example SPARQL Queries. https://www.ebi.ac.uk/rdf/example-sparql-queries (accessed April, 2015).
112. bio2rdf example SPARQL quires. https://github.com/bio2rdf/bio2rdf-scripts/wiki/Query-repository (accessed April, 2015).
113. Kalderimis, A.; Lyne, R.; Butano, D.; Contrino, S.; Lyne, M.; Heimbach, J.; Hu, F.; Smith, R.; Stepan, R.; Sullivan, J.; Micklem, G., InterMine: extensive web services for modern biology. *Nucleic Acids Res* **2014,** *42* (Web Server issue), W468-72.
114. TargetMine Template quires. http://targetmine.mizuguchilab.org/targetmine/templates.do?filter=Interactions (accessed April,2015).
115. Hendrickx, D. M.; Aerts, H. J.; Caiment, F.; Clark, D.; Ebbels, T. M.; Evelo, C. T.; Gmuender, H.; Hebels, D. G.; Herwig, R.; Hescheler, J.; Jennen, D. G.; Jetten, M. J.; Kanterakis, S.; Keun, H. C.; Matser, V.; Overington, J. P.; Pilicheva, E.; Sarkans, U.; Segura-Lepe, M. P.; Sotiriadou, I.; Wittenberger, T.; Wittwehr, C.; Zanzi, A.; Kleinjans, J. C., diXa: a data infrastructure for chemical safety assessment. *Bioinformatics* **2015,** *31* (9), 1505-7.
116. (a) Carlson, M. R.; Zhang, B.; Fang, Z.; Mischel, P. S.; Horvath, S.; Nelson, S. F., Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. *BMC Genomics* **2006,** *7*, 40; (b) Chen, Y.; Zhu, J.; Lum, P. Y.; Yang, X.; Pinto, S.; MacNeil, D. J.; Zhang, C.; Lamb, J.; Edwards, S.; Sieberts, S. K.; Leonardson, A.; Castellini, L. W.; Wang, S.; Champy, M. F.; Zhang, B.; Emilsson, V.; Doss, S.; Ghazalpour, A.; Horvath, S.; Drake, T. A.; Lusis, A. J.;

Schadt, E. E., Variations in DNA elucidate molecular networks that cause disease. *Nature* **2008,** *452* (7186), 429-35; (c) Dewey, F. E.; Perez, M. V.; Wheeler, M. T.; Watt, C.; Spin, J.; Langfelder, P.; Horvath, S.; Hannenhalli, S.; Cappola, T. P.; Ashley, E. A., Gene coexpression network topology of cardiac development, hypertrophy, and failure. *Circ Cardiovasc Genet* **2011,** *4* (1), 26-35; (d) Gargalovic, P. S.; Imura, M.; Zhang, B.; Gharavi, N. M.; Clark, M. J.; Pagnon, J.; Yang, W. P.; He, A.; Truong, A.; Patel, S.; Nelson, S. F.; Horvath, S.; Berliner, J. A.; Kirchgessner, T. G.; Lusis, A. J., Identification of inflammatory gene modules based on variations of human endothelial cell responses to oxidized lipids. *Proc Natl Acad Sci U S A* **2006,** *103* (34), 12741-6; (e) Ghazalpour, A.; Doss, S.; Zhang, B.; Wang, S.; Plaisier, C.; Castellanos, R.; Brozell, A.; Schadt, E. E.; Drake, T. A.; Lusis, A. J.; Horvath, S., Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet* **2006,** *2* (8), e130; (f) Gong, K. W.; Zhao, W.; Li, N.; Barajas, B.; Kleinman, M.; Sioutas, C.; Horvath, S.; Lusis, A. J.; Nel, A.; Araujo, J. A., Air-pollutant chemicals and oxidized lipids exhibit genome-wide synergistic effects on endothelial cells. *Genome Biol* **2007,** *8* (7), R149; (g) Miller, J. A.; Horvath, S.; Geschwind, D. H., Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *Proc Natl Acad Sci U S A* **2010,** *107* (28), 12698-703; (h) Oldham, M. C.; Konopka, G.; Iwamoto, K.; Langfelder, P.; Kato, T.; Horvath, S.; Geschwind, D. H., Functional organization of the transcriptome in human brain. *Nat Neurosci* **2008,** *11* (11), 1271-82; (i) Oldham, M. C.; Langfelder, P.; Horvath, S., Network methods for describing sample relationships in genomic datasets: application to Huntington's disease. *BMC Syst Biol* **2012,** *6*, 63; (j) Xue, Z.; Huang, K.; Cai, C.; Cai, L.; Jiang, C. Y.; Feng, Y.; Liu, Z.; Zeng, Q.; Cheng, L.; Sun, Y. E.; Liu, J. Y.; Horvath, S.; Fan, G., Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **2013,** *500* (7464), 593-7.

117. Meslamani, J.; Bhajun, R.; Martz, F.; Rognan, D., Computational profiling of bioactive compounds using a target-dependent composite workflow. *J Chem Inf Model* **2013,** *53* (9), 2322-33.

118. Meslamani, J.; Li, J.; Sutter, J.; Stevens, A.; Bertrand, H. O.; Rognan, D., Protein-ligand-based pharmacophores: generation and utility assessment in computational ligand profiling. *J Chem Inf Model* **2012,** *52* (4), 943-55.

119. (a) Alaimo, S.; Pulvirenti, A.; Giugno, R.; Ferro, A., Drug-target interaction prediction through domain-tuned network-based inference. *Bioinformatics* **2013,** *29* (16), 2004-8; (b) Ding, H.; Takigawa, I.; Mamitsuka, H.; Zhu, S., Similarity-based machine learning methods for predicting drug-target interactions: a brief review. *Brief Bioinform* **2014,** *15* (5), 734-47; (c) Pahikkala, T.; Airola, A.; Pietila, S.; Shakyawar, S.; Szwajda, A.; Tang, J.; Aittokallio, T., Toward more realistic drug-target interaction predictions. *Brief Bioinform* **2015,** *16* (2), 325-37.

120. Gonzalez-Diaz, H.; Prado-Prado, F.; Garcia-Mera, X.; Alonso, N.; Abeijon, P.; Caamano, O.; Yanez, M.; Munteanu, C. R.; Pazos, A.; Dea-Ayuela, M. A.; Gomez-Munoz, M. T.; Garijo, M. M.; Sansano, J.; Ubeira, F. M., MIND-BEST: Web server for drugs and target discovery; design, synthesis, and assay of MAO-B inhibitors and theoretical-experimental study of G3PDH protein from Trichomonas gallinae. *J Proteome Res* **2011,** *10* (4), 1698-718.

121. Keiser, M. J.; Roth, B. L.; Armbruster, B. N.; Ernsberger, P.; Irwin, J. J.; Shoichet, B. K., Relating protein pharmacology by ligand chemistry. *Nat Biotechnol* **2007,** *25* (2), 197-206.

122. Yildirim, M. A.; Goh, K. I.; Cusick, M. E.; Barabasi, A. L.; Vidal, M., Drug-target network. *Nat Biotechnol* **2007,** *25* (10), 1119-26.

123. (a) Ekins, S.; Casey, A. C.; Roberts, D.; Parish, T.; Bunin, B. A., Bayesian models for screening and TB Mobile for target inference with Mycobacterium tuberculosis. *Tuberculosis (Edinb)* **2014,** *94* (2), 162-9; (b) Ekins, S.; Freundlich, J. S.; Hobrath, J. V.; Lucile White, E.; Reynolds, R. C., Combining computational methods for hit to lead optimization in Mycobacterium tuberculosis drug discovery. *Pharm Res* **2014,** *31* (2), 414-35.

124. 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), W219-24.

125. Martinez-Jimenez, F.; Marti-Renom, M. A., Ligand-target prediction by structural network biology using nAnnoLyze. *PLoS Comput Biol* **2015,** *11* (3), e1004157.
126. Campillos, M.; Kuhn, M.; Gavin, A. C.; Jensen, L. J.; Bork, P., Drug target identification using side-effect similarity. *Science* **2008,** *321* (5886), 263-6.
127. Iorio, F.; Bosotti, R.; Scacheri, E.; Belcastro, V.; Mithbaokar, P.; Ferriero, R.; Murino, L.; Tagliaferri, R.; Brunetti-Pierri, N.; Isacchi, A.; di Bernardo, D., Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc Natl Acad Sci U S A* **2010,** *107* (33), 14621-6.
128. Fernald, G. H.; Altman, R. B., Using molecular features of xenobiotics to predict hepatic gene expression response. *J Chem Inf Model* **2013,** *53* (10), 2765-73.
129. Jacoby, E.; Tresadern, G.; Bembenek, S.; Wroblowski, B.; Buyck, C.; Neefs, J. M.; Rassokhin, D.; Poncelet, A.; Hunt, J.; van Vlijmen, H., Extending kinome coverage by analysis of kinase inhibitor broad profiling data. *Drug Discov Today* **2015**.
130. Clark, A. M.; Sarker, M.; Ekins, S., New target prediction and visualization tools incorporating open source molecular fingerprints for TB Mobile 2.0. *J Cheminform* **2014,** *6*, 38.
131. 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.
132. Zhou, T.; Ren, J.; Medo, M.; Zhang, Y. C., Bipartite network projection and personal recommendation. *Phys Rev E Stat Nonlin Soft Matter Phys* **2007,** *76* (4 Pt 2), 046115.
133. van Laarhoven, T.; Nabuurs, S. B.; Marchiori, E., Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics* **2011,** *27* (21), 3036-43.
134. Mei, J. P.; Kwoh, C. K.; Yang, P.; Li, X. L.; Zheng, J., Drug-target interaction prediction by learning from local information and neighbors. *Bioinformatics* **2013,** *29* (2), 238-45.
135. 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), i232-40.
136. Nidhi; Glick, M.; Davies, J. W.; Jenkins, J. L., Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. *J Chem Inf Model* **2006,** *46* (3), 1124-33.
137. Wale, N.; Karypis, G., Target fishing for chemical compounds using target-ligand activity data and ranking based methods. *J Chem Inf Model* **2009,** *49* (10), 2190-201.
138. 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), 407-18.
139. Han, S.; Kim, D., Inference of protein complex activities from chemical-genetic profile and its applications: predicting drug-target pathways. *PLoS Comput Biol* **2008,** *4* (8), e1000162.