



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SciVerse ScienceDirect

Current Opinion in  
Biotechnology

# Drug discovery in the age of systems biology: the rise of computational approaches for data integration

Murat Iskar<sup>1,a</sup>, Georg Zeller<sup>1,a</sup>, Xing-Ming Zhao<sup>1</sup>,  
Vera van Noort<sup>1</sup> and Peer Bork<sup>1,2</sup>

The increased availability of large-scale open-access resources on bioactivities of small molecules has a significant impact on pharmacology facilitated mainly by computational approaches that digest the vast amounts of data. We discuss here how computational data integration enables systemic views on a drug's action and allows to tackle complex problems such as the large-scale prediction of drug targets, drug repurposing, the molecular mechanisms, cellular responses or side effects. We particularly focus on computational methods that leverage various cell-based transcriptional, proteomic and phenotypic profiles of drug response in order to gain a systemic view of drug action at the molecular, cellular and whole-organism scale.

## Addresses

<sup>1</sup> Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany

<sup>2</sup> Max-Delbrück-Centre for Molecular Medicine, Berlin, Germany

Corresponding author: Bork, Peer ([bork@embl.de](mailto:bork@embl.de))

<sup>a</sup>These authors contributed equally to this work.

Current Opinion in Biotechnology 2011, 23:1–8

This review comes from a themed issue on  
Systems Biology  
Edited by Sang Yup Lee and Jens Nielsen

0958-1669/\$ – see front matter

© 2011 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.copbio.2011.11.010

## Introduction

Relatively recent efforts led to the creation of large open-access databases of chemicals and associated bioactivity data. Fueled by a change of attitude in the pharmaceutical industry to embrace the academic sector, these databases keep growing and diversifying at a dramatic speed. This trend might revolutionize research in pharmacology, chemogenomics and chemoinformatics, much like the impact DNA sequence archives had on molecular and systems biology. Leveraging these rich and heterogeneous resources requires data integration, mostly through computational approaches. Although challenging, mining the various datasets provides great opportunities for computational research to contribute to a systemic understanding of drug-perturbed molecular and physiological processes, leading to more rational drug discovery.

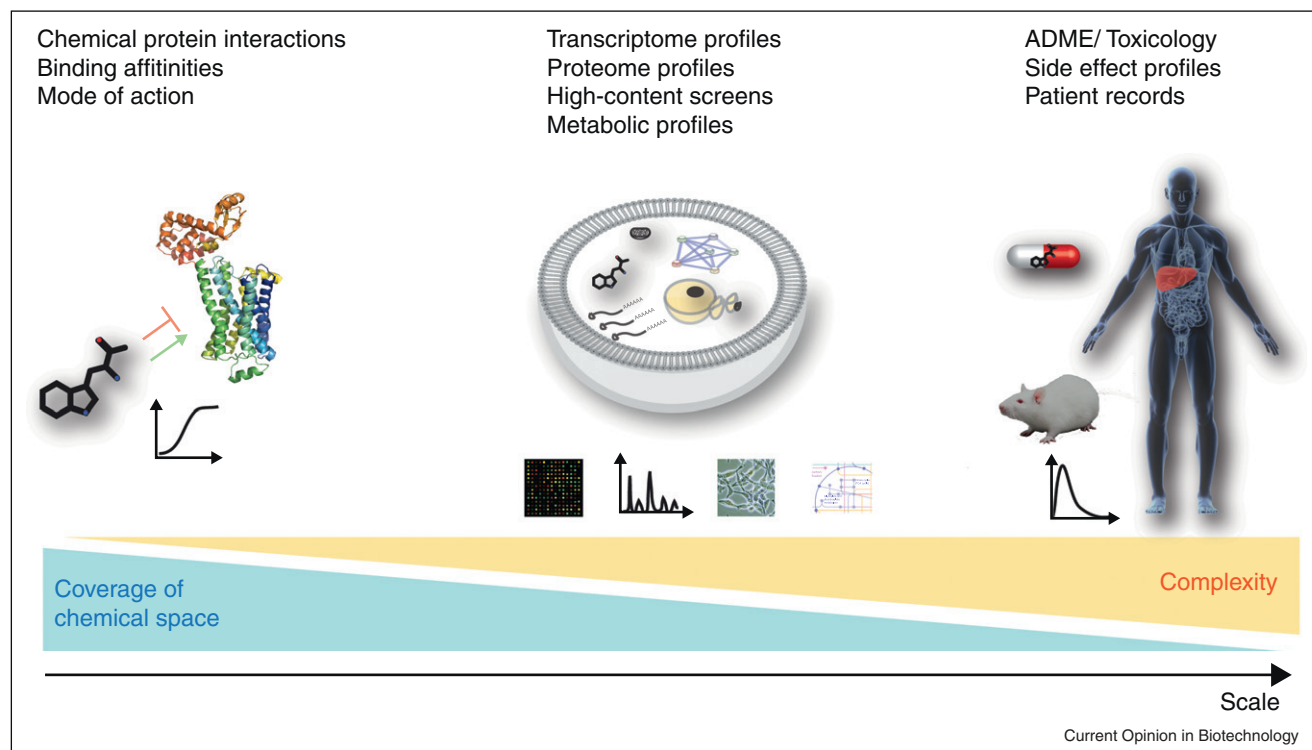
The need for advanced computational methods in drug discovery is further increased by a paradigm shift away from purely drug target-centric approaches [1]. These largely ignore a target's cellular and physiological context, making it difficult to address efficacy and safety problems in early stages of the drug discovery process. Modern concepts try to consider a drug's mechanism of action in a broader sense. These 'systems pharmacology' strategies are enabled by high-throughput screens which profile transcript expression, metabolic states or cellular phenotypes resulting from a wide range of chemical perturbations [2] (see Figure 1). The richness of the resulting multi-parametric readouts provides a strong basis for data-driven computational approaches addressing many important aspects of the drug discovery process, such as a drug's (off-)targets, mechanism of action, metabolism and toxicity. Bioinformatics methods that mine and integrate data from these heterogeneous screens hold great promise for gaining a global understanding of drug action at the molecular, cellular and whole-organism scale. Moreover, such integrative approaches may not only provide systemic insights into mechanisms of drugs and diseases, but also reveal new biological aspects of the 'system' human.

## Data integration utilizing publicly available resources

The availability of data on the bioactivity of chemical compounds in publicly accessible databases is currently increasing dramatically (see Box 1) [3]. For instance, in PubChem, bioassay data have grown from just about 800 to more than 500,000 records in three years [4\*]. Currently, data from *in vitro* target binding assays and chemical perturbation experiments with associated gene expression profiles are routinely deposited in public databases. We expect that in the near future this trend will extend to complex multi-parametric readouts including cell-wide gene expression of large-scale chemical perturbation and other associated proteomics, metabolomics or high-throughput microscopy data deduced from multiple screens [2]. Finally there are data describing drug response of a whole tissue or organism in form of clinical, hematological and histopathological parameters, package labels, including side-effect information, and electronic patient records, although not all are freely accessible (yet). Taken together, these bioactivity data describe drug action at biological scales ranging from molecules via cells to entire organisms.

## 2 Systems Biology

Figure 1



The different scales at which information on bioactivity of chemicals can be profiled. From *in vitro* binding assays to cellular profiling and whole-organism readouts, complexity of the data typically increases. Similarly, assay cost generally increases, whereas experimental throughput and hence availability of the data (coverage of chemical space) decreases along the same axis.

In this context, data integration ideally leads to models that account for several aspects of drug perturbations at multiple scales and may thus provide a systemic view on the biological response processes and networks (see Figure 2). However, these approaches face many challenges. For instance, the lack of controlled and standardized vocabulary describing chemicals, bioactivity assays or pharmacological outcomes of drug treatments poses a major hurdle for data integration from diverse sources which has been recognized by the community [5,6]. Another concern arises from the sheer size of chemical space: Even large-scale screens only achieve partial coverage, so integrative approaches are often hampered by too few chemicals that are shared by different assays (see Figure 2). In this respect, standardization efforts towards establishing priorities for chemical libraries may be a step in the right direction [7]. Despite such obstacles, substantial progress in integrative methodologies has been made during the past years. In the following, we focus on recent computational approaches that address fundamental challenges in drug development and cover different levels of complexity increasing from drug-target prediction over mechanism-of-action inference to drug-safety assessment.

### Integrative approaches to drug-target prediction

Bioactive chemicals exert their function through binding to one or more protein targets. Therefore, identifying a compound's main target(s) is a pivotal step in drug discovery. However, for a thorough systemic understanding of the consequences of drug intake, many other 'readouts' are helpful. For example, identifying binding propensity for off-targets, metabolizing enzymes (such as cytochrome P450s) and transporters (for absorption or excretion) allows safety assessments early in the drug discovery pipeline [19]. Knowledge of the entire profile of drug targets is crucial for assessing both efficacy and safety aspects of drug candidates.

For most pharmacologically relevant protein targets, experimental binding data are incomplete, even though high-throughput screens are more and more employed. This makes computational drug-target prediction attractive for exploring the interaction space between potential targets and bioactive chemicals [20]. Here we will focus on computational approaches that can make inferences for the majority of drug-target interactions. In this respect, many protein structure-based methods are limited by scarce 3D structure data for important target

### Box 1 A selection of publicly available resources on bioactivity of small molecules

#### Pubchem

<http://pubchem.ncbi.nlm.nih.gov/> [4\*]



A comprehensive repository for small molecules and their biological activities. It includes more than 500,000 bioassay results, 130 million experimental bioassay results, 85 million substances representing 30 million chemically unique compounds.

#### chEMBLdb

<https://www.ebi.ac.uk/chembl/db/> [8]



A database that contains chemical properties (e.g. log *P*, Molecular Weight, Lipinski Parameters) and biological activities (e.g. binding constants, pharmacology and ADMET data) for over 700,000 distinct drug-like small molecules including more than 3 million bioactivity results from the scientific literature.

#### DrugBank

<http://drugbank.ca/> [9]



Displays manually annotated fact sheets of drugs (with chemical, pharmacological and pharmaceutical information) and their targets (with sequence, structure, and pathway) of 6796 drug-like chemicals including 1571 FDA-approved small molecule and biotech (protein/peptide) drugs. Additionally, 4285 non-redundant proteins (i.e. drug targets/enzymes/transporters) are associated with drug entries including 230 richly illustrated drug-action pathways.

#### STITCH: Chemical-Protein Interactions

<http://stitch.embl.de/> [10]



A meta-repository to explore known and predicted interactions between more than 300,000 chemicals and their protein targets within the context of protein-protein association networks (2.6 million proteins in 1133 organisms) [11].

#### GEO: Gene expression omnibus

<http://www.ncbi.nlm.nih.gov/geo/> [12]



A public repository that archives and freely distributes 20,000 microarray-based and sequence-based functional genomics studies including disease-associated experiments.

#### Connectivity Map

<http://www.broadinstitute.org/cmap/> [13\*]



A large collection of genome-wide expression profiles generated with microarrays. It contains expression response of four cancer cell lines to over 1300 bioactive, drug-like molecules. With a simple pattern-matching algorithm available on the web server, drugs can be linked to each other or to diseases and genetic perturbations through signature (dis-)similarity search.

#### ACToR

<http://actor.epa.gov/> [14]



An online warehouse that integrates over 500 publicly available chemical toxicity resources including ToxRefDB [15], *in vivo* animal toxicity studies and ToxCastDB, 500 high-throughput assays of a thousand chemicals [16]. It contains data on more than 500,000 chemicals and their potential health and environmental risks.

#### SIDER: Side Effects Resource

<http://sideeffects.embl.de/> [17]



A machine-readable side effect resource that connects 888 approved drugs to 1450 recorded adverse effects extracted from public documents and package inserts. Moreover, it integrates and makes available side-effect frequency, and drug and side effect classifications.

#### NPC: NCGC Pharmaceutical Collection

<http://tripod.nih.gov/npc/> [18]



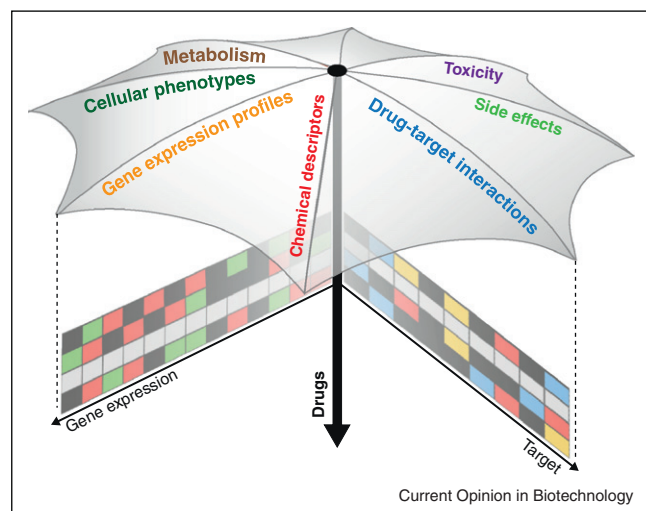
A comprehensive resource of approved and experimental drugs useful for drug repositioning. In addition to being an electronic resource for computational approaches, it also offers a physical compound collection for high-throughput screening experiments.

classes (such as GPCRs), and will be omitted here (but are covered elsewhere, e.g. [21]).

To globally map drug-target associations, Keiser *et al.* characterized drug targets in terms of their known ligands and subsequently inferred novel associations from similarity between ligand sets [22\*,23]. The resulting predictions revealed extensive polypharmacology and ligand promiscuity. They thus challenged the 'one-drug-one-target' dogma and instead support the notion that broad

## 4 Systems Biology

Figure 2



The promise of integrative methods for gaining a systems level understanding of complex biological responses to drug treatment. The growing diversity of publicly accessible bioactivity data has created many possibilities for integrative approaches. For example, gene expression profiles have been combined with drug-target information to infer mechanism of action and feedback regulation of target expression [13<sup>\*</sup>,32,33<sup>\*\*</sup>]. Integrating heterogeneous data from many bioactivity resources under one umbrella (i.e. on the basis of a common set of drugs) will be an important strategy to advance a systems level understanding of drug action.

activity across a family of targets may be required for efficacy against some diseases [22<sup>\*</sup>,24].

Other studies integrated more heterogeneous data types for target identification, yet still relied to some extent on chemical similarity between drugs. One such approach combined structural chemical descriptors of drugs with protein sequence features of the targets [25]. From these a combined distance measure was derived to predict non-trivial drug-target interactions, that is, to propose novel drugs that are not chemically similar to a target's set of known ligands [25].

Systemic concepts that utilized human phenotypic data opened new avenues in drug target prediction and became feasible by using computational approaches. Campillos *et al.* developed ontologies that allowed to quantify side effect similarities derived from package inserts and have shown that from a similarity measure of side-effect profiles, it can be inferred whether two drugs share a target with implications for drug repurposing [26<sup>\*</sup>]. They experimentally validated several novel drug-target relations demonstrating the feasibility of approaches that utilize organism level readouts to predict interactions at the molecular level. Another method by Yamanishi *et al.* [27] integrated chemical similarity and protein-sequence similarity with additional pharmacological data extracted from package labels by text-mining.

From the combination of these heterogeneous data, their machine learning-based method predicted new drug-target interactions [27].

Pursuing a metabolomics approach, Folger *et al.* have introduced a new systemic method for the identification of (potential) anti-cancer drug targets [28<sup>\*\*</sup>]. By integrating a network of human cellular metabolism with cancer expression data, they constructed a generic model of the metabolic fluxes in cancer cells [28<sup>\*\*</sup>]. This approach predicted specific targets whose disruption would severely impact cancer metabolism while exhibiting only mild toxicity according to metabolic models of non-cancerous cells and liver tissue. Importantly, they not only identified many known metabolic anticancer drug-targets and novel candidates, but also provided a mechanistic model of the metabolic consequences of drug interventions. This even allowed the authors to propose combination therapies, where synergistic drug interference appeared most effective in the metabolic model [28<sup>\*\*</sup>].

Recent developments have clearly demonstrated the feasibility and benefits of combining heterogeneous data from systemic readouts for drug-target prediction. In the near future, the increased quantity and diversity of available data from cellular and phenotypic assays will further empower integrative approaches. Growing knowledge of the physiological roles of drug targets in disease etiology and adverse drug reactions will both, profit from and facilitate, accurate computational assessments of a drug's full interaction profile and thus be a key step towards a systemic understanding of drug action.

### Inference of a drug's mechanism of action utilizing cellular readouts

Directly inferring drug target relations from complex cellular readouts may not always be possible owing to convergence of several molecular events onto the same or very similar phenotypes. Nonetheless, such data may be useful for characterizing a drug's general mechanism of action (MoA) on a biological system. MoA refers to the molecular mechanisms by which a drug achieves its pharmacological effect [29]. Elucidating a drug's MoA experimentally is currently still very time and labor intensive, but there is hope that a combination of high-throughput screening technologies and computational analyses can shortcut many aspects by generating hypotheses early on.

There are plenty of computational challenges in the analysis of complex multi-parametric screening data that start with the delineation of (molecular or phenotypic) drug response profiles from several high-throughput screening technologies covering gene expression, proteomics, metabolomics or cellular phenotypes. The hope of large-scale profiling efforts is indeed to capture the molecular or phenotypic signatures of the screened drugs'



MoAs [13<sup>•</sup>]. However, the resulting profiles typically also include general stress or other secondary responses, and effects due to off-target binding. Another challenge is the establishment of adequate similarity measures that reflect the common underlying MoA(s) of chemicals. This is a prerequisite to apply the ‘guilt by association’ rule to infer novel drug indications from profile similarity [30]. Finally, it is an open computational problem how to best integrate heterogeneous data from different screening technologies in order to improve overall prediction accuracy and gain a systemic understanding of a drug’s MoA.

Large-scale profile-based comparisons of drug treatments were pioneered by the Connectivity Map (CMap) project [13<sup>•</sup>]. Generating and comparing genome-wide gene expression readouts of several cell-lines treated with more than 1000 drug-like chemicals illustrated the potential of complex biological readouts for mapping relations between drugs, diseases and genes. Expression-based comparison of drugs has been facilitated by concomitant research into statistical methods for profile similarity search [31,32,33<sup>••</sup>]. Building on the CMap resource, Iorio *et al.* [33<sup>••</sup>] identified network communities of drugs with similar expression profiles. These drug communities were found to generally have similar MoA and higher likelihood of sharing targets or therapeutic effect and community membership was successfully exploited for drug repositioning.

Mechanistic insights into drug action have also been gathered based on their effect on cellular morphology, by exploiting high-content phenotypic screening in conjunction with automated microscopy and image analysis [34<sup>•</sup>,35]. Young *et al.* reduced primary image features to a more compact representation, allowing the computational identification of chemicals with pronounced morphological effects [34<sup>•</sup>]. Clustering these chemicals according to their morphological phenotype revealed relationships to an independent chemical similarity-based clustering. A combination of target prediction and phenotype clustering allowed MoA inference for some chemical classes. Taking high-throughput phenotypic screening to the whole-organism level, behavioral alterations in zebrafish larvae were quantified upon treatment with drug-like chemicals [36<sup>•</sup>,37]. Rihel *et al.* computationally derived profiles from these complex behavioral phenotypes and showed that behavioral profile clustering is strongly correlated with similarity of known bioactivities of drugs in mammals (e.g. target proteins and MoA). This established the screen as a valuable tool for MoA inference of poorly characterized psychotropic compounds which are particularly challenging to study *in vitro* [36<sup>•</sup>].

The concept of profiles capturing drug MoA has been successfully extended to diseases. Diseases, which are accompanied by disruptions of processes and networks at the cellular level, can also be characterized by profiles of

altered gene expression, metabolite concentrations, and so on. Through guilt-by-association, diseases with similar profiles are proposed to be treated with the same drugs [38,39]. Furthermore, drug and disease profiles have been combined based on the idea of ‘reversing disease profiles’: A drug showing an expression profile that is strongly anticorrelated with a disease profile might actually reverse some of the disease effects [13<sup>•</sup>,40,41<sup>•</sup>,42]. Sirota *et al.* systematically developed this concept further into a method, which combines disease expression profiles extracted from publicly available microarray repositories (see Box 1) [12] with drug-induced expression profiles for rational drug repositioning [41<sup>•</sup>]. Extensive experimental follow-up work established that this approach can indeed lead to novel treatments that showed promise in animal models [41<sup>•</sup>,42].

A significant advance towards integration of multiple evidences for drug repositioning was recently made by Gottlieb *et al.* [43<sup>••</sup>]. They combined many different profile-based similarity measures for drugs and diseases to identify novel drug indications using a machine learning-based method. It incorporates drug–drug similarity measures derived from chemical structures, expression profiles, side-effect profiles and properties of the drugs’ targets. Additionally, disease–disease similarity is calculated based on phenotypic as well as molecular profiles (gene expression similarity among others). The integration achieved higher prediction accuracy than each dataset alone and allowed an MoA inference that is robust against weaknesses and errors in individual sources of evidence [43<sup>••</sup>].

**Prediction of higher level systemic features such as drug toxicity and adverse reactions**

Toxicity and adverse reactions of small molecules pose a public health risk, as well as cause high attrition rates in later stages of drug development [44]. This led the U.S. National Research Council (NRC) to propose the development of high-throughput *in vitro* toxicology assays along with controlled vocabularies on toxicity phenotypes and computational methods for assessing a compound’s mechanism-based toxicity [45]. Ultimately, the collected data will allow integrative computational approaches to characterize fundamental biological processes leading to drug-induced organ and organism failures. Such approaches might prove useful to predict toxicity of novel compounds.

Already now, chemical structures and/or physiochemical properties are being used to predict *in vivo* toxicity and adverse drug reactions as has been demonstrated by quantitative structure–activity relationship (QSAR) studies [19,46]. Chemical structure and its predicted protein targets were further associated with adverse drug reactions in order to pinpoint targets and pathways responsible for the undesired drug effect [19]. An integrative

## 6 Systems Biology

model by Sedykh *et al.* [47] combines chemical with biological descriptors mined from PubChem *in vitro* assays (see Box 1) [4<sup>\*</sup>]. It has shown that for assessing *in vivo* toxicity, bioassay data provide unique, valuable information that cannot be derived from chemical structures alone [47].

For the prediction of toxicity-related aspects of drugs, genome-wide expression data derived from drug-perturbed states should be a valuable resource allowing conceptually novel avenues. Natsoulis *et al.* have generated a large-scale dataset of gene expression profiles from liver tissue of rats treated with 344 chemicals [48]. To identify biomarkers predictive of the observed toxicological and pharmaceutical outcomes, they used a supervised classification algorithm, which extracted a small set of ~200 representative genes sufficient to distinguish more than 30 liver toxicity outcomes [48]. To improve the molecular understanding of *in vivo* liver toxicity, Low *et al.* combined drug-induced expression profiles of rat liver with chemical descriptors and provided insights into chemical substructures and regulated genes that are responsible for liver toxicity [49].

Side effects of FDA-approved drugs provide information on human phenotypic response to drug treatment (see Box 1) and their variation in the human population [17]. Fliri *et al.* clustered drugs using either side effect or drug-target profiles [50]. They found seven classes of drugs that are coherent between both clusterings. This implies that for some drugs, target profiles are predictive of side effects. Although these and other results suggest strong associations between side effects and drug targets [26<sup>\*</sup>,50], additional work has shown that targeting distinct proteins that are interaction partners also contributes to similarity between side-effect profiles [51]. Pouliot *et al.* [52] correlated side effect classes summarized at the organ level with heterogeneous bioassay data from Pubchem (see Box 1) [4<sup>\*</sup>] in order to build models that reliably predict safety-related issues [52].

The integrative approaches presented here represent the first steps towards a systemic understanding that is likely to impact toxicology and safety issues in drug development by improving our understanding of mechanism-based toxicity at molecular, cellular and organismal level.

### Summary and outlook

Emphasizing data integration, we have summarized recent computational approaches that might have an impact on rational drug design. Whatever implementation will be accepted by a wider community, we believe that these integrative concepts are promising for the near future, building on the current rapid expansion of publicly accessible resources on the bioactivity of chemical compounds (see Box 1). Integration of diverse types of data describing drug action at multiple levels from

molecular interactions to cellular and organismal aspects will provide a systemic view of drug response processes that is likely to heavily impact drug development. This holistic view will not only enhance drug discovery, but also facilitate a better understanding of the 'system' human. Large-scale drug perturbation experiments represent important new entry points for the reconstruction of human molecular and physiological networks and thus represent important contributions to systems biology [53].

### Acknowledgements

We thank the members of the Bork group for helpful discussions.

### References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Swinney DC, Anthony J: **How were new medicines discovered?** *Nature Reviews. Drug Discovery* 2011, **10**:507-519.
  2. Feng Y, Mitchison TJ, Bender A, Young DW, Tallarico JA: **Multi-parameter phenotypic profiling: using cellular effects to characterize small-molecule compounds.** *Nature Reviews. Drug Discovery* 2009, **8**:567-578.
  3. Gaulton A, Overington JP: **Role of open chemical data in aiding drug discovery and design.** *Future Medicinal Chemistry* 2010, **2**:903-907.
  4. Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Bryant SH:
    - **PubChem: a public information system for analyzing bioactivities of small molecules.** *Nucleic Acids Research* 2009, **37**:W623-W633.
- A publicly available resource for immense amounts of data on chemicals and their bioactivities.
5. Bodenreider O: **The Unified Medical Language System (UMLS): integrating biomedical terminology.** *Nucleic Acids Research* 2004, **32**:D267-D270.
  6. Orchard S, Al-Lazikani B, Bryant S, Clark D, Calder E, Dix I, Engkvist O, Forster M, Gaulton A, Gilson M *et al.*: **Minimum information about a bioactive entity (MIABE).** *Nature Reviews. Drug Discovery* 2011, **10**:661-669.
  7. Hajduk PJ, Galloway WRJD, Spring DR: **Drug discovery: a question of library design.** *Nature* 2011, **470**:42-43.
  8. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B *et al.*: **ChEMBL: a large-scale bioactivity database for drug discovery.** *Nucleic acids research* 2011 doi: 10.1093/nar/gkr777.
  9. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V *et al.*: **DrugBank 3.0: a comprehensive resource for "omics" research on drugs.** *Nucleic Acids Research* 2011, **39**:D1035-D1041.
  10. Kuhn M, Szklarczyk D, Franceschini A, Campillos M, von Mering C, Jensen LJ, Beyer A, Bork P: **STITCH 2 an interaction network database for small molecules and proteins.** *Nucleic Acids Research* 2010, **38**:D552-D556.
  11. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguéz P, Doerks T, Stark M, Müller J, Bork P *et al.*: **The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored.** *Nucleic Acids Research* 2011, **39**:D561-D568.
  12. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM *et al.*: **NCBI GEO: archive for functional genomics data sets – 10 years on.** *Nucleic Acids Research* 2011, **39**:D1005-D1010.
  13. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ,
    - Lerner J, Brunet J-P, Subramanian A, Ross KN *et al.*: **The**

- Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease.** *Science (New York, NY)* 2006, **313**:1929-1935.
- A comprehensive resource on gene expression responses of four cancer cell lines to treatments with ~1300 small molecules. Using a pattern-matching algorithm, drugs, diseases and physiological processes can be linked to each other as exploited by many follow-up studies.
14. Judson R, Richard A, Dix D, Houck K, Elloumi F, Martin M, Cathey T, Transue TR, Spencer R, Wolf M: **ACToR – Aggregated Computational Toxicology Resource.** *Toxicology and Applied Pharmacology* 2008, **233**:7-13.
  15. Martin MT, Judson RS, Reif DM, Kavlock RJ, Dix DJ: **Profiling chemicals based on chronic toxicity results from the U.S. EPA ToxRef Database.** *Environmental Health Perspectives* 2009, **117**:392-399.
  16. Knudsen TB, Houck KA, Sipes NS, Singh AV, Judson RS, Martin MT, Weissman A, Kleinstreuer NC, Mortensen HM, Reif DM et al.: **Activity profiles of 309 ToxCast™ chemicals evaluated across 292 biochemical targets.** *Toxicology* 2011, **282**:1-15.
  17. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P: **A side effect resource to capture phenotypic effects of drugs.** *Molecular Systems Biology* 2010, **6**:343.
  18. Huang R, Southall N, Wang Y, Yasgar A, Shinn P, Jadhav A, Nguyen D-T, Austin CP: **The NCGC pharmaceutical collection: a comprehensive resource of clinically approved drugs enabling repurposing and chemical genomics.** *Science Translational Medicine* 2011, **3**:80ps16.
  19. Bender A, Scheiber J, Glick M, Davies JW, Azzaoui K, Hamon J, Urban L, Whitebread S, Jenkins JL: **Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure.** *ChemMedChem* 2007, **2**:861-873.
  20. Imming P, Sinning C, Meyer A: **Drugs their targets and the nature and number of drug targets.** *Nature Reviews. Drug Discovery* 2006, **5**:821-834.
  21. Haupt VJ, Schroeder M: **Old friends in new guise: repositioning of known drugs with structural bioinformatics.** *Briefings in Bioinformatics* 2011, **12**:312-326.
  22. Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, Hufeisen SJ, Jensen NH, Kujijer MB, Matos RC, Tran TB et al.: **Predicting new molecular targets for known drugs.** *Nature* 2009, **462**:175-181.

This study integrated comprehensive drug-target prediction based on similarity between ligand sets into a global drug-target network. This network revealed a surprising extent of ligand promiscuity and polypharmacology.

  23. Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK: **Relating protein pharmacology by ligand chemistry.** *Nature Biotechnology* 2007, **25**:197-206.
  24. Hopkins AL: **Network pharmacology: the next paradigm in drug discovery.** *Nature Chemical Biology* 2008, **4**:682-690.
  25. Yabuuchi H, Nijijima S, Takematsu H, Ida T, Hirokawa T, Hara T, Ogawa T, Minowa Y, Tsujimoto G, Okuno Y: **Analysis of multiple compound-protein interactions reveals novel bioactive molecules.** *Molecular Systems Biology* 2011, **7**:472.
  26. Campillos M, Kuhn M, Gavin A-C, Jensen LJ, Bork P: **Drug target identification using side-effect similarity.** *Science (New York, NY)* 2008, **321**:263-266.

From similarity of side-effect profiles of approved drugs, the authors inferred whether two drugs share a target. This shows that organismal phenotypic data such as side effects are predictive of binding events at the molecular level supported by experimental validation of novel drug-target relations.

  27. Yamanishi Y, Kotera M, Kanehisa M, Goto S: **Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework.** *Bioinformatics (Oxford, England)* 2010, **26**:i246-i254.
  28. Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T: **Predicting selective drug targets in cancer through metabolic networks.** *Molecular Systems Biology* 2011, **7**:501.

The authors integrated a human metabolic model with cancer expression data to obtain a generic model of cancer metabolism. This correctly predicted many known metabolic anti-cancer drug targets as well as novel candidates. This work is unique in that the mechanistic model offers rich explanations for the predictions making it an ideal tool for rational drug design.

  29. Berger SI, Iyengar R: **Network analyses in systems pharmacology.** *Bioinformatics (Oxford, England)* 2009, **25**:2466-2472.
  30. Dudley JT, Deshpande T, Butte AJ: **Exploiting drug-disease relationships for computational drug repositioning.** *Briefings in Bioinformatics* 2011, **12**:303-311.
  31. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al.: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** In *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**:15545-15550.
  32. Iskar M, Campillos M, Kuhn M, Jensen LJ, van Noort V, Bork P: **Drug-induced regulation of target expression.** *PLoS Computational Biology* 2010, **6**:e1000925.
  33. Iorio F, Bosotti R, Scacheri E, Belcastro V, Mithbaokar P, Ferrero R, Murino L, Tagliaferri R, Brunetti-Pierri N, Isacchi A et al.: **Discovery of drug mode of action and drug repositioning from transcriptional responses.** In *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**:14621-14626.

A drug-drug network based on drug-induced expression profiles was created to reveal drug communities associated with drug MoA and indication. Additional drug-induced expression profiles were inserted into this network to predict MoA of novel drugs based on closest drug neighbors.

  34. Young DW, Bender A, Hoyt J, McWhinnie E, Chirn G-W, Tao CY, Tallarico JA, Labow M, Jenkins JL, Mitchison TJ et al.: **Integrating high-content screening and ligand-target prediction to identify mechanism of action.** *Nature Chemical Biology* 2008, **4**:59-68.

Factor analysis as data reduction tool to handle immense datasets of image-based phenotypic information. Authors compared phenotypic profiles, chemical similarity and predicted drug-target relations to draw mechanism-of-action inferences.

  35. Loo L-H, Wu LF, Altschuler SJ: **Image-based multivariate profiling of drug responses from single cells.** *Nature Methods* 2007, **4**:445-453.
  36. Rihel J, Prober DA, Arvanites A, Lam K, Zimmerman S, Jang S, Haggarty SJ, Kokel D, Rubin LL, Peterson RT et al.: **Zebrafish behavioral profiling links drugs to biological targets and rest/wake regulation.** *Science (New York, NY)* 2010, **327**:348-351.

The value of high-throughput behavioral profiling of zebrafish was used for the discovery and characterization of psychotropic drugs.

  37. Kokel D, Bryan J, Laggner C, White R, Cheung CYJ, Mateus R, Healey D, Kim S, Werdich AA, Haggarty SJ et al.: **Rapid behavior-based identification of neuroactive small molecules in the zebrafish.** *Nature Chemical Biology* 2010, **6**:231-237.
  38. Suthram S, Dudley JT, Chiang AP, Chen R, Hastie TJ, Butte AJ: **Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets.** *PLoS Computational Biology* 2010, **6**:e1000662.
  39. Chiang AP, Butte AJ: **Systematic evaluation of drug-disease relationships to identify leads for novel drug uses.** *Clinical Pharmacology and Therapeutics* 2009, **86**:507-510.
  40. Hu G, Agarwal P: **Human disease-drug network based on genomic expression profiles.** *PLoS ONE* 2009, **4**:e6536.
  41. Sirota M, Dudley JT, Kim J, Chiang AP, Morgan AA, Sweet-Cordero A, Sage J, Butte AJ: **Discovery and preclinical validation of drug indications using compendia of public gene expression data.** *Science Translational Medicine* 2011, **3**:96ra77.

An important pioneering effort to combine disease and drug-induced gene expression profiles for drug repositioning via 'anticorrelated links'.

  42. Dudley JT, Sirota M, Shenoy M, Pai RK, Roedder S, Chiang AP, Morgan AA, Sarwal MM, Pasricha PJ, Butte AJ: **Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease.** *Science Translational Medicine* 2011, **3**:96ra76.

## 8 Systems Biology

43. Gottlieb A, Stein GY, Ruppin E, Sharan R: **PREDICT: a method for inferring novel drug indications with application to personalized medicine.** *Molecular Systems Biology* 2011, **7**:496. A fully integrative approach to infer potential indications for approved or novel drugs. Multiple drug–drug and disease–disease similarity measures were integrated into one platform to predict drug indications with high specificity and sensitivity.
44. Allen J, Jeffrey P: **Approaches to assessing drug safety in the discovery phase.** *Drugs of the Future* 2010, **35**:67-75.
45. Collins FS, Gray GM, Bucher JR: **Toxicology. Transforming environmental health protection.** *Science (New York, NY)* 2008, **319**:906-907.
46. Hughes JD, Blagg J, Price DA, Bailey S, Decrescenzo GA, Devraj RV, Ellsworth E, Fobian YM, Gibbs ME, Gilles RW *et al.*: **Physicochemical drug properties associated with in vivo toxicological outcomes.** *Bioorganic & Medicinal Chemistry Letters* 2008, **18**:4872-4875.
47. Sedykh A, Zhu H, Tang H, Zhang L, Richard A, Rusyn I, Tropsha A: **Use of in vitro HTS-derived concentration-response data as biological descriptors improves the accuracy of QSAR models of in vivo toxicity.** *Environmental Health Perspectives* 2011, **119**:364-370.
48. Natsoulis G, Pearson CI, Gollub JP, Eynon B, Ferng J, Nair R, Idury R, Lee MD, Fielden MR, Brennan RJ *et al.*: **The liver pharmacological and xenobiotic gene response repertoire.** *Molecular Systems Biology* 2008, **4**:175.
49. Low Y, Uehara T, Minowa Y, Yamada H, Ohno Y, Urushidani T, Sedykh A, Muratov E, Kuz'min V, Fourches D *et al.*: **Predicting drug-induced hepatotoxicity using QSAR and toxicogenomics approaches.** *Chemical Research in Toxicology* 2011, **24**: 1251-1262.
50. Fliri AF, Loging WT, Thadejo PF, Volkmann RA: **Analysis of drug-induced effect patterns to link structure and side effects of medicines.** *Nature Chemical Biology* 2005, **1**:389-397.
51. Brouwers L, Iskar M, Zeller G, van Noort V, Bork P: **Network neighbors of drug targets contribute to drug side-effect similarity.** *PLoS ONE* 2011, **6**:e22187.
52. Pouliot Y, Chiang AP, Butte AJ: **Predicting adverse drug reactions using publicly available PubChem BioAssay data.** *Clinical Pharmacology and Therapeutics* 2011, **90**:90-99.
53. Schadt EE, Friend SH, Shaywitz DA: **A network view of disease and compound screening.** *Nature Reviews. Drug Discovery* 2009, **8**:286-295.