# Target identification and mechanism of action in chemical biology and drug discovery

**Monica Schenone**[1], **Vlado Dančík**[2], **Bridget K Wagner**[2], and **Paul A Clemons**[2,*]

[1]Proteomics Platform, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA

[2]Chemical Biology Program, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA

## Abstract

Target-identification and mechanism-of-action studies have important roles in small-molecule probe and drug discovery. Biological and technological advances have resulted in the increasing use of cell-based assays to discover new biologically active small molecules. Such studies allow small-molecule action to be tested in a more disease-relevant setting at the outset, but they require follow-up studies to determine the precise protein target or targets responsible for the observed phenotype. Target identification can be approached by direct biochemical methods, genetic interactions or computational inference. In many cases, however, combinations of approaches may be required to fully characterize on-target and off-target effects and to understand mechanisms of small-molecule action.

Modern chemical biology and drug discovery each seek to identify new small molecules that potently and selectively modulate the functions of target proteins. Historically, nature has been an important source for such molecules, with knowledge of toxic or medicinal properties often long predating knowledge of precise target or mechanism. Natural selection provides a slow and steady stream of bioactive small molecules, but each of these molecules must perforce confer reproductive advantage in order for nature to 'invest' in its synthesis. In recent decades, investments in finding new small-molecule probes and drugs have expanded to a paradigm of screening large numbers (typically $10^3$–$10^6$) of compounds for those that elicit a desired biological response[1,2]. In some cases, these studies interrogate natural products[3,4], but more often they involve collections of synthetic small molecules prepared by organic chemistry strategies[5,6] that rapidly yield large collections of relatively pure compounds. Thus, the overall discovery paradigm involves investments both in organic synthesis and in biological testing of large compound collections.

Since the revolution in molecular biology, the biological testing component of screening-based discovery has overwhelmingly involved testing compounds for effects on purified proteins. However, with advances in assay technology, many research programs are

increasingly turning (or returning) to cell- or organism- based phenotypic assays that benefit from preserving the cellular context of protein function[7]. The cost paid for this benefit is that the precise protein targets or mechanisms of action responsible for the observed phenotypes remain to be determined. Even after a relevant target is established, additional functional studies may help to identify unwanted off-target effects or establish new roles for the target protein in biological networks. In this review, we address these important steps in the discovery process, termed target identification or deconvolution[8], illustrating methods available to approach the problem, highlighting recent advances and discussing how findings from multiple approaches are integrated.

## Background

Historically, genetics has provided powerful biological insights, allowing characterization of protein function by manipulation of genetic sequence. A forward genetics (or classical genetics) approach is characterized by identifying, often under experimental selection pressure, a phenotype of interest, followed by identification of the gene (or genes) responsible for the phenotype (see refs. [9, 10]). Modern molecular biological methods, particularly genetic engineering approaches, have given rise to reverse genetics (sometimes equated with molecular genetics), in which a specific gene of interest is targeted for mutation, deletion or functional ablation (for example, with RNAi[11]), followed by a broad search for the resulting phenotype (see refs. [12, 13]).

By analogy to genetics, there are two fundamental approaches to understanding the action of small molecules on biological systems[14,15]. Biochemical screening approaches are analogous to reverse genetics (Fig. 1a). In advance of conducting a high-throughput screen, the protein target is selected and (typically) purified before exposure to small molecules[16,17]. This target validation or credentialing is a time-consuming process that involves demonstrating the relevance of the protein for a particular biological pathway, process or disease of interest[18,19]. Once a target has been validated, it is presumed that binders or inhibitors of this protein will affect the desired process. Often, however, such an impact needs to be characterized more completely in cells or animals by observing compound-induced phenotypes; hence, this approach has been termed reverse chemical genetics.

In contrast, forward chemical genetics refers to the process of testing small molecules directly for their impact on biological processes, often in cells or even in animals (Fig. 1b)[20,21]. Phenotypic screens expose candidate compounds to proteins in more biologically relevant contexts than screens involving purified proteins[7,22]. Because these screens measure cellular function without imposing preconceived notions of the relevant targets and signaling pathways, they offer the possibility of discovering new therapeutic targets. Indeed, several important drug programs have been inspired by phenotypic screening results, including the effects of cyclosporine A and FK506 on T-cell receptor signaling[7,23], leading to the discoveries of FKBP12 (ref. [24]), calcineurin[25] and mTOR[26], and the performance of trapoxin A in differentiation and proliferation assays[27], leading to the discovery of histone deacetylases[28,29]. Importantly, such assays 'prevalidate' the small molecule and its (initially unknown) protein target as an effective means of perturbing the biological process or disease model under study.

However, phenotypic assays require a subsequent effort to discover the molecular targets of bioactive small molecules, which can be a complex endeavor[30]. We often assume that direct interaction with a single target is responsible for phenotypic observations, but this need not be the case. Many drugs show side effects owing to interactions with 'off-target' proteins[31,32], and even small molecule–induced phenotypes observed in cell culture may represent the superposition of effects on multiple targets[33,34]. For drug development, target identification is important to follow-up studies, aiding medicinal chemistry efforts. Furthermore, identifying both the therapeutic target and other targets that might cause unwanted side effects enables optimization of small-molecule selectivity[35]. Conversely, polypharmacology can be considered a new tool, leveraging multiple small-molecule effects to gain maximal effect, once again underscoring the benefit of an unbiased approach to screening and the need for target deconvolution[36].

## Approaches to target identification

In this review, we will cover three distinct and complementary approaches for discovering the protein target of a small molecule: direct biochemical methods, genetic interaction methods and computational inference methods. Direct methods involve labeling the protein or small molecule of interest, incubation of the two populations and direct detection of binding, usually following some type of wash procedure (reviewed in ref. [37]). Genetic manipulation can also be used to identify protein targets by modulating presumed targets in cells, thereby changing small-molecule sensitivity (reviewed in ref. [38]). Target hypotheses, in contrast, can be generated by computational inference, using pattern recognition to compare small-molecule effects to those of known reference molecules or genetic perturbations[39–41]. Mechanistic hypotheses, rather than targets per se, for new compounds emerge from such tests. The target pathway or protein of a new small molecule is inferred but remains to be confirmed[42]. Similarly, hypotheses regarding the mechanism of action of a compound can be generated by gene expression profiling in the presence or absence of compound treatment. Many target-identification projects actually proceed through a combination of these methods, where researchers use both direct measurements and inferences to test increasingly specific target hypotheses. Indeed, we suggest that the problem of target identification will not generally be solved by a single method but rather by analytical integration of multiple, complementary approaches.

## Direct biochemical methods

Biochemical affinity purification provides the most direct approach to finding target proteins that bind small molecules of interest[37]. Because they are based on physical interactions involving mammalian or human proteins, biochemical methods can lead to information about molecular mechanisms of efficacy or toxicity highly relevant to human disease. Similarly, small-molecule optimization efforts can be complemented by biochemical methods when three-dimensional structure information about the target is known. Finally, unbiased protein identification, especially from lysates containing intact protein complexes, potentially allows evaluation of polypharmacology.

Pioneering work in affinity purification involved monitoring chromatographic fractions for enzyme activity after exposure of extracts to compound immobilized on a column, followed by elution[43]. In general, such an approach requires large amounts of extract, possibly prefractionated, and stringent wash conditions. Such approaches have been used with success to identify certain protein targets, including those of both natural[29] and synthetic[44] small molecules, and one such approach has been used to elucidate the mechanism of the teratogenic side effect of thalidomide[35]. However, these methods seem best suited for situations where a high-affinity ligand binds a relatively abundant target protein. High-stringency washes can bias proteins identified to those with the highest-affinity interactions, decreasing the likelihood of finding additional targets that might be important in cellular contexts, as suggested by proteomic profiling studies (for example, those described in refs. [45, 46]). Furthermore, stringent washing will also reduce the ability to identify protein complexes in which the direct target participates and whose members' identities might help illuminate the connection between the direct target and the cellular activity of the small molecule.

Affinity purification experiments also involve the challenge of preparing immobilized affinity reagents that retain cellular activity, so that target proteins will still interact with the small molecule while it is bound to a solid support. A related issue is the identification of appropriate controls and tethers. Different approaches are possible; control beads loaded with an inactive analog[47,48] or capped without compound have been used[49]. These control experiments have limitations, most notably the availability of related inactive compounds. The inactive analog must be sufficiently different from the compound of interest to fail to bind the target, raising the possibility that it will have different physicochemical properties and therefore different nonspecific interactions with proteins. In the case of capped beads, the results are confounded by the background of high-abundance, low-affinity proteins with slight differential binding to the control bead. Elution from the bead-bound small molecule[24] or preincubation of lysate with compound is a viable alternative control[35,49,50] but is limited by compound solubility. The choice of tether, influencing the type of background proteins identified, also becomes a critical parameter[51–55]. These challenges can frustrate individual attempts at affinity purification, particularly if a biochemical approach is used in isolation.

Recent affinity-based methods have attempted to overcome one or more of these challenges. Approaches based on chemical or ultraviolet light–induced cross-linking[56,57] use covalent modification of the protein target to increase the likelihood of capturing low-abundance proteins or those with low affinity for the small molecule. This method requires prior knowledge of the enzyme activity being targeted, making it a slightly biased approach. If there is no bias, cross-linking of the small molecule to proteins is burdened with the possibility of high, nonspecific background owing to the cross-linking event itself. A variation on the use of photoaffinity reagents couples covalent modification to two-dimensional gel electrophoresis in an attempt to deconvolve nonspecific interactions[58]. A nonselective universal coupling method that enables attachment to a solid support by a photoaffinity reaction resulted in the identification of an inhibitor of glyoxalase I[59]. This method assumes that a compound could bind the solid support in multiple ways while some functional relevant sites remain available to interact with the protein target. This approach, however, runs the risk of false-negative results when the functional group is masked in the

coupling reaction. Another method for immobilizing small molecules, that is, coupling them to peptides that allow them to recover the probe-protein complex by immunoaffinity purification[60], was devised to address this issue.

Some small-molecule libraries are prepared with synthetic handles primed for making affinity matrices after an activity is identified[61,62]. Because these methods rely on advance modification of the compound structure, they require additional chemistry expertise and may not be possible for all compound classes. Similarly, if a small molecule can be fluorescently labeled, it can be used to probe proteins separated by microarray (reviewed in ref. [63]). However, such methods are limited to proteins that can be readily manipulated, usually in heterologous expression systems. The use of purified proteins does not necessarily ensure physiological expression levels, giving incorrect information about relative binding to alternative targets in cells and masking effects due to the formation of protein complexes.

Two interesting new target-identification approaches that circumvent the need to immobilize compounds have emerged in the last few years. One of them uses changes in protein susceptibility to proteolytic degradation upon small-molecule binding (reviewed in ref. [64]). The other is based on a characteristic shift in the chromatographic retention-time profile after a compound binds a protein target[65]. Although the generality of these approaches remains to be determined, their combination with quantitative proteomics is quite promising.

Affinity chromatography has been coupled to powerful new techniques in MS, which can possibly provide the most sensitive and unbiased methods of finding target proteins. Quantitative proteomics[66] (reviewed in ref. [67]) has been effective in identifying specific protein-protein interactions by affinity methods[68,69] and has been increasingly applied to protein–small molecule interactions (reviewed in ref. [70]). In the context of target identification, two different quantitative techniques have been used; these can be broadly divided into metabolic and chemical labeling.

Metabolic labeling, more specifically stable-isotope labeling by amino acids in cell culture (SILAC)[71], has been effectively used, in experiments with gentle washing and free soluble competitor pre-incubation of lysates, to provide unbiased assessment of multiple direct and indirect targets (Fig. 2)[49]. It has also been used in combination with serial drug-affinity chromatography to characterize the quantitative changes of the kinome in different phases of the cell cycle[72]. Metabolic labeling has the advantage of allowing sample pooling early in the process, eliminating quantification errors due to sample handling. A disadvantage is that it limits the workflow to immortalized cell lines[73].

Chemical labeling has also been successfully used in the past. Isotope-coded affinity tag (ICAT) technology[74], coupled with beads loaded with active and inactive compound as controls, has been used to identify malate dehydrogenase as a specific target of the anticancer compound E7070 (ref. [47]). Isobaric tags for relative and absolute quantification (iTRAQ)[75], coupled with free soluble competitor for elution as a control, has been used to profile kinases enriched by affinity purification with nonselective kinase-binding small molecules[76]. There are variations on chemical-labeling strategies, such as mass differential tags for relative and absolute quantification (mTRAQ)[75], tandem mass tags (TMT)[77] and

stable isotope dimethyl labeling[78,79]. These chemical labeling strategies provide more versatility regarding the type of samples that can be labeled, but a weakness lies in the fact that they generally rely on labeling at the peptide level, later in the proteomic workflow, and thus are prone to more variation and less accuracy[80].

These unbiased approaches show great promise but will require new software[81–83] and analytical techniques[84] to quantify the relative expression of members of candidate target lists to aid in experimental prioritization. In contrast, biased approaches to target deconvolution based on profiling small-molecule activity against a panel of enzymes are now commercially available and widely used. These methods rely on prior knowledge of the enzyme family (kinases, ubiquitinases, demethylases and so on). For example, assay-performance profiling and compounds with a known mode of action were used to predict kinase inhibitory activity for new compounds emerging from a phenotypic screen[42]. A commercial kinase profiling panel[85] then confirmed activity against a subset of kinases. Having information from genetic studies or computational methods to decide which set of enzymes to investigate provides a valuable tool, again underscoring the importance of integrating all methods at the researcher's disposal.

## Genetic interaction and genomic methods

Target identification based on genetic or genomic methods leverages the relative ease of working with DNA and RNA to perform large-scale modifications and measurements. These methods often use the principle of genetic interaction, relying on the idea of genetic modifiers (enhancers or suppressors) to generate target hypotheses. Gene knockout organisms, RNAi (reviewed in ref. [86]) and small molecules with well-defined mechanisms can each be used to alter the functions of putative targets, uncovering dependencies on activity. For example, if a gene knockdown phenocopies a compound's effects, the evidence that the protein could be the target relevant to that phenotype would be strengthened. In this case, hypersensitivity of individual mutants to sublethal concentrations of compound demonstrates a chemical-genetic interaction (Fig. 3a). Similarly, mating of laboratory and wild yeast strains can reveal patterns of small-molecule sensitivity with specific genetic loci (Fig. 3b)[87].

These concepts have been expanded to include molecularly barcoded libraries of open reading frames[88] and detection of small molecule–resistant clones by microarray (Fig. 3c). Importantly, these studies provide a conceptual framework for understanding complex diseases. Although direct translation to human biology may not be forthcoming, owing to the lack of conservation of some pathways between yeast and human, this framework enables us to envision technical advances that facilitate this type of analysis in mammalian systems. An accompanying perspective in this series[89] addresses these approaches in more detail.

A promising genetics-based technique for target identification involves combining results from small-molecule and RNAi perturbations. This approach enables parallel testing of small-molecule and RNAi libraries for induction of the same cellular phenotype[90,91]. RNAi experiments can be performed on a genome-wide scale to find phenotypes similar to those

induced by small molecules (Fig. 4a)[92]. Alternatively, when some mechanistic clues already exist, a more focused set of RNAi reagents can be used to find pathway members that alter the effects of a small molecule (Fig. 4b)[93]. The main strength of combining small-molecule and RNAi perturbations is the ability to measure phenotypic effects in more physiologically relevant cellular contexts, using mammalian, or even human, cells. Clearly, genetic perturbations cannot always recapitulate or phenocopy the effect of a small molecule[94], for example because of the risk of genetic compensation. Using RNAi libraries that contain entire families of genes may help alleviate this problem. Another way to dissect these effects is to use a suboptimal concentration of small molecule in combination with genetic knockdown. An elegant study provided proof of concept for RNAi-sensitized small-molecule screens[95], which could serve as a follow-up method for the other techniques described in this review.

Genetic target-identification efforts are increasingly focused on mammalian cells. For example, examination of compound-resistant clones of cells, using transcriptome sequencing (RNA-seq), identified intracellular targets of normally cytotoxic compounds[96]. Advantages of this approach include the ability to perform cell type–specific analyses and not having to chemically modify the compound to perform the analysis. Clones of HCT116 colon cancer cells resistant to the polo-like kinase 1 (PLK1) inhibitor BI 2356 were sequenced and compared to the parental line. Although PLK1 was not mutated in every clone, it was the only gene mutated in more than one group; moreover, mutations were present in the known binding site of BI 2356. This proof-of-principle study may pave the way for more rapid target identification in mammalian cells, although this approach is currently limited to cell viability as a phenotype.

Finally, recent efforts to use gene expression signatures to determine compounds' mechanisms of action illustrate the close relationship between genetic techniques and computational methods. Using transcription profiling data from the Connectivity Map[97], a recent study described a weighting scheme to rank order lists of genes across multiple cell lines, resulting in what was termed a prototype ranked list[98]. The authors then used gene-set enrichment analysis[99] to compute pairwise distances between the ranked lists for each compound and constructed a network in which each compound was a node. Clustering revealed communities of nodes connected to each other, two-thirds of which were enriched for similar mechanisms of action, as determined by anatomical therapeutic chemical codes. A limitation of this approach is the reliance on accurate annotation of small-molecule activity, but even so, this type of sophisticated approach will most likely uncover similarities between known bioactive compounds and new screening hits. The line separating genetic and computational approaches increasingly blurs as the technical hurdles to generating massive data sets are surmounted.

## Computational inference methods

On their own, computational methods are used to infer protein targets of small molecules, in addition to providing analytical support for proteomic and genetic techniques. These methods can also be used to find new targets for existing drugs, with the goal of drug repositioning or explaining off-target effects. Profiling methods rely on pattern recognition

to integrate results of parallel or multiplexed experiments, typically from small-molecule phenotypic profiling[100,101]. Ligand-based methods also incorporate chemical structures to predict targets. Structure-based methods rely on three-dimensional protein structures to predict protein–small molecule interactions[102,103]. Owing to their limitation to proteins with known structures, structure-based methods will not be discussed in detail.

In general, bioactivity profiling methods are based on the principle that compounds with the same mechanism of action will have similar behavior across different biological assays. Public databases of, for example, gene expression[97,104] or small-molecule screening[105,106] data sets house measurements that record phenotypic consequences of small-molecule perturbations. Target or mechanism inference frequently relies on the fact that molecules with known mechanisms of action (sometimes called landmark compounds[100]) are also profiled, allowing new compounds to be assessed for similar patterns of performance[107,108]. Small-molecule profiling data can be mined most effectively when there exists a critical mass of data generated using a common set of compounds and cell states (multidimensional screening[109]). A prototypical example of this type of study is the US National Cancer Institute's 'NCI-60' set of 60 cancer cell lines that was exposed to a large collection of small molecules[110]. In a seminal study using the NCI-60 data set, protein targets were connected (via their degree of expression in each cell line) to small-molecule sensitivity patterns in the same cell lines[39] (reviewed in ref. [101]). Similar studies, including those using the COMPARE algorithm[110], have allowed researchers to hypothesize new activities for small molecules[111]. Such approaches represented important conceptual advances in this field because they considered multidimensional profiling as a method to identify small-molecule mechanisms of action.

Gene expression technology has also had an important role in profile-based inference of small-molecule targets and mechanisms. One early study used gene expression profiles of yeast treated with the immunosuppressants FK506 and cyclosporine A. This study not only recapitulated the importance of calcineurin in immunosuppression but also identified a number of other genes that were suggested to be secondary targets[112]. Later, a compendium of yeast gene expression profiles was derived from both genetic mutations and small-molecule perturbations to annotate uncharacterized genes and pathways[40]. Commercial databases of rat tissue–based expression profiles following drug treatment were developed to identify toxicities of new chemical entities[113]. Subsequently, the Connectivity Map was developed as a public database of gene expression profiles derived from small-molecule perturbations of human cell lines (Fig. 5a), which enables the functions of new small molecules to be identified by pattern matching[97].

Several groups have developed profiling methods based on other measurement technologies[114,115], including affinity profiling of biochemical assays to predict ligand binding to proteins[116], and on similarities among side-effect profiles[117]. Affinity profiling was also used to design biologically diverse screening libraries[118]. Several promising studies have involved profiling by high-throughput microscopy (sometimes called high-content screening), allowing phenotypes to be clustered, in a manner analogous to transcriptional profiles, to discover potential small-molecule targets[119,120]. More generally, public databases such as ChemBank[105] or PubChem[106] provide extensible environments for

accumulating rich phenotypic profiles for many compounds exposed to a diverse set of assays and have been exploited to understand both bioinformatic[121] and cheminformatic[101,122] relationships (Fig. 5b).

More focused profiling methods investigate only particular aspects of biological systems. For example, high-throughput screening (HTS) activity profiling and a guilt-by-association approach were used to determine mechanisms of action for hits in an antimalarial screen[123]. Another study[124] examined small molecule–induced cell death and systematically characterized lethal compounds on the basis of modulatory profiles, created experimentally by measuring cell viability. Comparison of modulatory profile clusters with gene expression– and structure-based profiles demonstrated that that modulatory profiles revealed additional relationships. Bioactivity profile similarity search (BASS) is another method that uses profiles of dose-response cell-based assay results to associate targets with small molecules[125]. It is increasingly likely that single measurements will be inadequate to the task of determining mechanism of action.

Recent profiling methods have taken full advantage of high-throughput data available in public and proprietary screening databases, for example by developing HTS fingerprints (HTS-FP) with a goal of facilitating virtual screening and scaffold hopping[126]. Enrichment of gene ontology terms among known protein targets was observed within HTS-FP clusters. Further, HTS-FP can be used to select biologically diverse compounds for screening when testing the full library is not possible.

Using ligand-based predictive modeling to classify compounds is a well-established practice in computational chemistry. Independently, two groups have extended predictive-modeling approaches to explore global relationships among biological targets. In one case[127], protein-ligand interactions were explored with a goal of predicting targets for new compounds. Using Laplacian-modified naive Bayesian modeling, trained on 964 target classes in the WOMBAT database, the researchers were able to predict the top three most likely protein targets and validated their approach by examining therapeutic activities of compounds in the MDL Drug Data Report database. Such modeling can also be used to explain off-target effects or to design target- or family-focused libraries. In a separate but similar approach[128], a large collection of structure-activity relationship data was assembled from various public and proprietary sources, resulting in a set of 836 human genes that are targets of small molecules with binding affinities less than 10 μM. These connections were used to generate a polypharmacology interaction network of proteins in chemical space. This network enables a deeper understanding of compound and target cross-reactivity (promiscuity) and provides rational approaches to lead hopping and target hopping.

Chemical similarity is often used as a metric of success for pattern-recognition approaches. The similarity ensemble approach (SEA)[129] works by quantitatively grouping related proteins on the basis of the chemical similarity of their ligands. For each pair of targets, chemical similarity was computed and properly normalized according to a null model of random similarity. Even though only ligand chemical information was used, biologically related clusters of proteins emerged. Importantly, in the SEA results, there were cases of ligand-based clusters differing from protein sequence–based clusters; for example, ion

channels and G protein–coupled receptors are ligand related, but they have no structure or sequence similarity. In contrast, many neurological receptors with related sequences lack pharmacological similarity. The authors documented the utility of SEA by predicting and confirming off-target activities of methadone, loperamide and emetine. Large-scale prediction of off-target activities using SEA was subsequently described[130].

SEA is well suited to provide insights into the mechanisms of action of small molecules. With that purpose in mind, SEA was used on the MDL Drug Data Report database to predict new ligand-target interactions[131]. The predictions were either confirmed by literature or database search or investigated experimentally. Out of 30 experimentally tested predictions, 23 were confirmed. This work was extended to 'de-orphan' seven US Food and Drug Administration–approved drugs with unknown targets[132]. SEA has also been used to predict targets of compounds that were active in a zebrafish behavioral assay[133]. Out of 20 predictions that were tested experimentally, the authors confirmed 11, with activities ranging from 1 nM to 10 μM. These results suggest that chemical information alone can be sufficient to make predictions using this approach. Additional ligand-based target-identification methods have been reported (reviewed in ref. [103]).

Network-based approaches extend systems biology to drug-target and ligand-target networks and are now known as 'systems chemical biology'[134], 'network pharmacology'[135] or 'systems pharmacology'[136]. Such approaches are necessary, as many phenotypes are caused by effects of compounds on multiple targets. An exploration of the relationship between drug chemical structures, target protein sequences and drug-target network topology resulted in the creation of a unified 'pharmacological space' that could predict ligand-target interactions for new compounds and proteins[137]. Several additional studies report recent advances in network-based approaches to target prediction[138,139].

Inference-based methods arguably have the least bias of any target-identification method as they often rely on experiments done by others. The analyst is distant from the original experimental design and is therefore poised to reveal unanticipated relationships. In contrast, such studies rely on data sets that require substantial time and investment, however distributed, to realize. Fortunately, computational techniques have increasingly emerged to take advantage of the explosion of public data sources. Structured publicly accessible databases are the best hope for the success of such methods; they not only promote the availability of data from multiple sources but also encourage rigor among experimentalists by exposing their data to critical evaluation by the scientific community at large.

## Summary and outlook

Each of the described approaches to target-identification and mechanism-of-action studies has strengths and limitations, and, importantly, different laboratories will have different technical strengths. Laboratories with expertise in chemistry or biochemistry might gravitate toward direct biochemical approaches, whereas genetic or cell biology laboratories might favor genetic interaction approaches, and groups with a high degree of computational experience might first pursue methods that investigate databases for clues. Of course, there

is no 'right answer' about which method is best. Rather, we suggest that a combination of approaches[93,140] is most likely to bear fruit (Fig. 6).

There have been a few examples of successful integration of different methods to help determine the mechanism of action of a small molecule. An integrated approach was used to characterize the ability of a well-known natural product, K252a, to potentiate Nrg1-induced neurite outgrowth[141]. Integrating quantitative proteomic results with a lentivirus-mediated loss-of-function screen to validate candidate target proteins, the authors found that knockdown of AAK-1 reproducibly potentiated Nrg-1–driven neurite outgrowth. Similarly, the mechanism by which another natural product, piperlongumine, selectively kills cancer cells was determined by a combination of direct proteomic affinity-enrichment and short hairpin RNA methods[142].

To determine whether bortezomib neurotoxicity could be attributed to off-target effects, secondary targets of the proteasome inhibitors bortezomib and carfilzomib were explored[143]. The authors used a multifaceted approach, involving screening a panel of purified enzymes, activity-based probe profiling in cells and cell lysates and mining a database of known and predicted proteases, to show that bortezomib inhibits several serine proteases and to hypothesize that its neurotoxicity may be caused by off-target effects.

Computational data integration and network analysis were instrumental in determining Aurora kinase A as a relevant target of dimethylfasudil in acute megakaryoblastic leukemia[144]. Dimethylfasudil is a broad-spectrum kinase inhibitor; integrating quantitative proteomics data, kinase inhibition profiling and RNA-silencing data allowed the researchers to generate a testable hypothesis which not only identified the physiologically relevant target of dimethylfasudil but also a potential therapeutic target for acute megakaryoblastic leukemia. Another elegant application of integrating transcriptional data with proteomic data was used to dissect the synergy between two multikinase inhibitors in chronic myelogenous leukemia cell lines[145]. A combination of proteomic methods to measure drug binding in cell lysates, global phosphoproteomics and genome-wide transcriptomics demonstrated synergy between danusertib and bosutinib in chronic myelogenous leukemia cells harboring the BCR-ABL$^{T315I}$ gatekeeper mutation.

These very recent examples show how powerful the ability to effectively and systematically integrate large sets of disparate data will be in understanding the molecular mechanisms of a small molecule in biological systems. When done in a disciplined and thoughtful manner, such data integration represents a modern instantiation of the scientific method, relying on high-throughput technology, data integration and multidisciplinary approaches to provide clues and avenues to new targets and mechanisms of small-molecule action.

## Acknowledgments

# References

1. Sundberg SA. High-throughput and ultra-high-throughput screening: solution- and cell-based approaches. Curr Opin Biotechnol. 2000; 11:47–53. [PubMed: 10679349]

2. Mayr LM, Bojanic D. Novel trends in high-throughput screening. Curr Opin Pharmacol. 2009; 9:580–588. [PubMed: 19775937]

3. Koehn FE. High impact technologies for natural products screening. Prog Drug Res. 2008; 65(175): 177–210.

4. Lachance H, Wetzel S, Kumar K, Waldmann H. Charting, navigating, and populating natural product chemical space for drug discovery. J Med Chem. 2012; 55:5989–6001. [PubMed: 22537178]

5. Nielsen TE, Schreiber SL. Towards the optimal screening collection: a synthesis strategy. Angew Chem Int Edn Engl. 2008; 47:48–56.

6. O'Connor CJ, Beckmann HS, Spring DR. Diversity-oriented synthesis: producing chemical tools for dissecting biology. Chem Soc Rev. 2012; 41:4444–4456. [PubMed: 22491328]

7. Swinney DC, Anthony J. How were new medicines discovered? Nat Rev Drug Discov. 2011; 10:507–519. [PubMed: 21701501]

8. Terstappen GC, Schlupen C, Raggiaschi R, Gaviraghi G. Target deconvolution strategies in drug discovery. Nat Rev Drug Discov. 2007; 6:891–903. [PubMed: 17917669]

9. García-García MJ, et al. Analysis of mouse embryonic patterning and morphogenesis by forward genetics. Proc Natl Acad Sci USA. 2005; 102:5913–5919. [PubMed: 15755804]

10. Muto A, et al. Forward genetic analysis of visual behavior in zebrafish. PLoS Genet. 2005; 1:e66. [PubMed: 16311625]

11. Fire A, et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature. 1998; 391:806–811. [PubMed: 9486653]

12. Pickart MA, et al. Genome-wide reverse genetics framework to identify novel functions of the vertebrate secretome. PLoS ONE. 2006; 1:e104. [PubMed: 17218990]

13. Ecker A, Bushell ES, Tewari R, Sinden RE. Reverse genetics screen identifies six proteins important for malaria development in the mosquito. Mol Microbiol. 2008; 70:209–220. [PubMed: 18761621]

14. Mitchison TJ. Towards a pharmacological genetics. Chem Biol. 1994; 1:3–6. [PubMed: 9383364]

15. Schreiber SL. Chemical genetics resulting from a passion for synthetic organic chemistry. Bioorg Med Chem. 1998; 6:1127–1152. [PubMed: 9784856]

16. Inglese J, et al. High-throughput screening assays for the identification of chemical probes. Nat Chem Biol. 2007; 3:466–479. [PubMed: 17637779]

17. Macarron R, et al. Impact of high-throughput screening in biomedical research. Nat Rev Drug Discov. 2011; 10:188–195. [PubMed: 21358738]

18. Wyatt PG, Gilbert IH, Read KD, Fairlamb AH. Target validation: linking target and chemical properties to desired product profile. Curr Top Med Chem. 2011; 11:1275–1283. [PubMed: 21401506]

19. Kauselmann G, Dopazo A, Link W. Identification of disease-relevant genes for molecularly-targeted drug discovery. Curr Cancer Drug Targets. 2012; 12:1–13. [PubMed: 22165964]

20. Stockwell BR. Chemical genetics: ligand-based discovery of gene function. Nat Rev Genet. 2000; 1:116–125. [PubMed: 11253651]

21. Stockwell BR. Exploring biology with small organic molecules. Nature. 2004; 432:846–854. [PubMed: 15602550]

22. Clemons PA. Complex phenotypic assays in high-throughput screening. Curr Opin Chem Biol. 2004; 8:334–338. [PubMed: 15183333]

23. Schreiber SL, Crabtree GR. The mechanism of action of cyclosporin A and FK506. Immunol Today. 1992; 13:136–142. [PubMed: 1374612]

24. Harding MW, Galat A, Uehling DE, Schreiber SL. A receptor for the immunosuppressant FK506 is a *cis-trans* peptidylprolyl isomerase. Nature. 1989; 341:758–760. [PubMed: 2477715]

25. Liu J, et al. Inhibition of T cell signaling by immunophilin-ligand complexes correlates with loss of calcineurin phosphatase activity. Biochemistry. 1992; 31:3896–3901. [PubMed: 1373650]

26. Brown EJ, et al. A mammalian protein targeted by G1-arresting rapamycin-receptor complex. Nature. 1994; 369:756–758. [PubMed: 8008069]

27. Yoshida M, Nomura S, Beppu T. Effects of trichostatins on differentiation of murine erythroleukemia cells. Cancer Res. 1987; 47:3688–3691. [PubMed: 2439196]

28. Yoshida M, Kijima M, Akita M, Beppu T. Potent and specific inhibition of mammalian histone deacetylase both *in vivo* and *in vitro* by trichostatin A. J Biol Chem. 1990; 265:17174–17179. [PubMed: 2211619]

29. Taunton J, Hassig CA, Schreiber SL. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science. 1996; 272:408–411. [PubMed: 8602529]

30. McNamara C, Winzeler EA. Target identification and validation of novel antimalarials. Future Microbiol. 2011; 6:693–704. [PubMed: 21707315]

31. Whitebread S, Hamon J, Bojanic D, Urban L. Keynote review: *in vitro* safety pharmacology profiling: an essential tool for successful drug development. Drug Discov Today. 2005; 10:1421–1433. [PubMed: 16243262]

32. Xie L, Bourne PE. Structure-based systems biology for analyzing off-target binding. Curr Opin Struct Biol. 2011; 21:189–199. [PubMed: 21292475]

33. Chen S, et al. Self-renewal of embryonic stem cells by a small molecule. Proc Natl Acad Sci USA. 2006; 103:17266–17271. [PubMed: 17088537]

34. Apsel B, et al. Targeted polypharmacology: discovery of dual inhibitors of tyrosine and phosphoinositide kinases. Nat Chem Biol. 2008; 4:691–699. [PubMed: 18849971]

35. Ito T, et al. Identification of a primary target of thalidomide teratogenicity. Science. 2010; 327:1345–1350. [PubMed: 20223979]

36. Knight ZA, Lin H, Shokat KM. Targeting the cancer kinome through polypharmacology. Nat Rev Cancer. 2010; 10:130–137. [PubMed: 20094047]

37. Burdine L, Kodadek T. Target identification in chemical genetics: the (often) missing link. Chem Biol. 2004; 11:593–597. [PubMed: 15157870]

38. Zheng XS, Chan TF, Zhou HH. Genetic and genomic approaches to identify and study the targets of bioactive small molecules. Chem Biol. 2004; 11:609–618. [PubMed: 15157872]

39. Weinstein JN, et al. An information-intensive approach to the molecular pharmacology of cancer. Science. 1997; 275:343–349. [PubMed: 8994024]

40. Hughes TR, et al. Functional discovery via a compendium of expression profiles. Cell. 2000; 102:109–126. [PubMed: 10929718]

41. Young DW, et al. Integrating high-content screening and ligand-target prediction to identify mechanism of action. Nat Chem Biol. 2008; 4:59–68. [PubMed: 18066055]

42. Fomina-Yadlin D, et al. Small-molecule inducers of insulin expression in pancreatic alpha-cells. Proc Natl Acad Sci USA. 2010; 107:15099–15104. [PubMed: 20696901]

43. Cuatrecasas P, Wilchek M, Anfinsen CB. Selective enzyme purification by affinity chromatography. Proc Natl Acad Sci USA. 1968; 61:636–643. [PubMed: 4971842]

44. Hirota T, et al. Identification of small molecule activators of cryptochrome. Science. 2012; 337:1094–1097. [PubMed: 22798407]

45. Brehmer D, Godl K, Zech B, Wissing J, Daub H. Proteome-wide identification of cellular targets affected by bisindolylmaleimide-type protein kinase C inhibitors. Mol Cell Proteomics. 2004; 3:490–500. [PubMed: 14769951]

46. Wissing J, et al. Chemical proteomic analysis reveals alternative modes of action for pyrido[2,3-*d*]pyrimidine kinase inhibitors. Mol Cell Proteomics. 2004; 3:1181–1193. [PubMed: 15475568]

47. Oda Y, et al. Quantitative chemical proteomics for identifying candidate drug targets. Anal Chem. 2003; 75:2159–2165. [PubMed: 12720356]

48. Wang G, Shang L, Burgett AW, Harran PG, Wang X. Diazonamide toxins reveal an unexpected function for ornithine delta-amino transferase in mitotic cell division. Proc Natl Acad Sci USA. 2007; 104:2068–2073. [PubMed: 17287350]

49. Ong SE, et al. Identifying the proteins to which small-molecule probes and drugs bind in cells. Proc Natl Acad Sci USA. 2009; 106:4617–4622. [PubMed: 19255428]

50. Fleischer TC, et al. Chemical proteomics identifies Nampt as the target of CB30865, an orphan cytotoxic compound. Chem Biol. 2010; 17:659–664. [PubMed: 20609415]

51. Shiyama T, Furuya M, Yamazaki A, Terada T, Tanaka A. Design and synthesis of novel hydrophilic spacers for the reduction of nonspecific binding proteins on affinity resins. Bioorg Med Chem. 2004; 12:2831–2841. [PubMed: 15142543]

52. Speers AE, Cravatt BF. A tandem orthogonal proteolysis strategy for high-content chemical proteomics. J Am Chem Soc. 2005; 127:10018–10019. [PubMed: 16011363]

53. van der Veken P, et al. Development of a novel chemical probe for the selective enrichment of phosphorylated serine- and threonine-containing peptides. Chembiochem. 2005; 6:2271–2280. [PubMed: 16254931]

54. Fonoviis  M, Verhelst SH, Sorum MT, Bogyo M. Proteomics evaluation of chemically cleavable activity-based probes. Mol Cell Proteomics. 2007; 6:1761–1770. [PubMed: 17615255]

55. Verhelst SH, Fonovic M, Bogyo M. A mild chemically cleavable linker system for functional proteomic applications. Angew Chem Int Ed Engl. 2007; 46:1284–1286. [PubMed: 17205587]

56. Evans MJ, Saghatelian A, Sorensen EJ, Cravatt BF. Target discovery in small-molecule cell-based screens by *in situ* proteome reactivity profiling. Nat Biotechnol. 2005; 23:1303–1307. [PubMed: 16200062]

57. Cisar JS, Cravatt BF. Fully functionalized small-molecule probes for integrated phenotypic screening and target identification. J Am Chem Soc. 2012; 134:10385–10388. [PubMed: 22667687]

58. Park J, Oh S, Park SB. Discovery and target identification of an antiproliferative agent in live cells using fluorescence difference in two-dimensional gel electrophoresis. Angew Chem Int Ed Engl. 2012; 51:5447–5451. [PubMed: 22489056]

59. Kawatani M, et al. The identification of an osteoclastogenesis inhibitor through the inhibition of glyoxalase I. Proc Natl Acad Sci USA. 2008; 105:11691–11696. [PubMed: 18695250]

60. Saxena C, et al. Capture of drug targets from live cells using a multipurpose immuno-chemo-proteomics tool. J Proteome Res. 2009; 8:3951–3957. [PubMed: 19537828]

61. Khersonsky SM, et al. Facilitated forward chemical genetics using a tagged triazine library and zebrafish embryo screening. J Am Chem Soc. 2003; 125:11804–11805. [PubMed: 14505387]

62. Kim YK, Chang YT. Tagged library approach facilitates forward chemical genetics. Mol Biosyst. 2007; 3:392–397. [PubMed: 17533452]

63. Tao SC, Chen CS, Zhu H. Applications of protein microarray technology. Comb Chem High Throughput Screen. 2007; 10:706–718. [PubMed: 18045082]

64. Lomenick B, Olsen RW, Huang J. Identification of direct protein targets of small molecules. ACS Chem Biol. 2011; 6:34–46. [PubMed: 21077692]

65. Chan JN, et al. Target identification by chromatographic co-elution: monitoring of drug-protein interactions without immobilization or chemical derivatization. Mol Cell Proteomics. 2012; 11:M111.016642.

66. Aebersold R, Mann M. Mass spectrometry–based proteomics. Nature. 2003; 422:198–207. [PubMed: 12634793]

67. Ong SE, Mann M. Mass spectrometry–based proteomics turns quantitative. Nat Chem Biol. 2005; 1:252–262. [PubMed: 16408053]

68. Blagoev B, et al. A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. Nat Biotechnol. 2003; 21:315–318. [PubMed: 12577067]

69. Ranish JA, et al. The study of macromolecular complexes by quantitative proteomics. Nat Genet. 2003; 33:349–355. [PubMed: 12590263]

70. Rix U, Superti-Furga G. Target profiling of small molecules by chemical proteomics. Nat Chem Biol. 2009; 5:616–624. [PubMed: 19690537]

71. Ong SE, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics. 2002; 1:376–386. [PubMed: 12118079]

72. Daub H, et al. Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. Mol Cell. 2008; 31:438–448. [PubMed: 18691976]

73. Bendall SC, et al. Prevention of amino acid conversion in SILAC experiments with embryonic stem cells. Mol Cell Proteomics. 2008; 7:1587–1597. [PubMed: 18487603]

74. Gygi SP, et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol. 1999; 17:994–999. [PubMed: 10504701]

75. Ross PL, et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol Cell Proteomics. 2004; 3:1154–1169. [PubMed: 15385600]

76. Bantscheff M, et al. Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. Nat Biotechnol. 2007; 25:1035–1044. [PubMed: 17721511]

77. Thompson A, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem. 2003; 75:1895–1904. erratum **75**, 4942 (2003); erratum **78**, 4235 (2006). [PubMed: 12713048]

78. Hsu JL, Huang SY, Chow NH, Chen SH. Stable-isotope dimethyl labeling for quantitative proteomics. Anal Chem. 2003; 75:6843–6852. [PubMed: 14670044]

79. Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJ. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat Protoc. 2009; 4:484–494. [PubMed: 19300442]

80. Mertins P, et al. iTRAQ labeling is superior to mTRAQ for quantitative global proteomics and phosphoproteomics. Mol Cell Proteomics. 2012; 11:M111.014423.

81. Mortensen P, et al. MSQuant, an open source platform for mass spectrometry-based quantitative proteomics. J Proteome Res. 2010; 9:393–403. [PubMed: 19888749]

82. Tsou CC, et al. MaXIC-Q Web: a fully automated web service using statistical and computational methods for protein quantitation based on stable isotope labeling and LC-MS. Nucleic Acids Res. 2009; 37:W661–W669. [PubMed: 19528069]

83. Cox J, et al. Andromeda: A peptide search engine integrated into the MaxQuant environment. J Proteome Res. 2011; 10:1794–1805. [PubMed: 21254760]

84. Margolin AA, et al. Empirical Bayes analysis of quantitative proteomics experiments. PLoS ONE. 2009; 4:e7454. [PubMed: 19829701]

85. Fabian MA, et al. A small molecule–kinase interaction map for clinical kinase inhibitors. Nat Biotechnol. 2005; 23:329–336. [PubMed: 15711537]

86. Boutros M, Ahringer J. The art and design of genetic screens: RNA interference. Nat Rev Genet. 2008; 9:554–566. [PubMed: 18521077]

87. Perlstein EO, Ruderfer DM, Roberts DC, Schreiber SL, Kruglyak L. Genetic basis of individual differences in the response to small-molecule drugs in yeast. Nat Genet. 2007; 39:496–502. [PubMed: 17334364]

88. Pierce SE, et al. A unique and universal molecular barcode array. Nat Methods. 2006; 3:601–603. [PubMed: 16862133]

89. Roemer T, Boone C. Systems-level antimicrobial drug and drug synergy discovery. Nat Chem Biol. 2013; 9:222–231. [PubMed: 23508188]

90. Moffat J, et al. A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell. 2006; 124:1283–1298. [PubMed: 16564017]

91. Wang J, et al. Cellular phenotype recognition for high-content RNA interference genome-wide screening. J Biomol Screen. 2008; 13:29–39. [PubMed: 18227224]

92. Eggert US, et al. Parallel chemical genetic and genome-wide RNAi screens identify cytokinesis inhibitors and targets. PLoS Biol. 2004; 2:e379. [PubMed: 15547975]

93. Guertin DA, Guntur KV, Bell GW, Thoreen CC, Sabatini DM. Functional genomics identifies TOR-regulated genes that control growth and division. Curr Biol. 2006; 16:958–970. [PubMed: 16713952]

94. Knight ZA, Shokat KM. Chemical genetics: where genetics and pharmacology meet. Cell. 2007; 128:425–430. [PubMed: 17289560]

95. Castoreno AB, et al. Small molecules discovered in a pathway screen target the Rho pathway in cytokinesis. Nat Chem Biol. 2010; 6:457–463. [PubMed: 20436488]

96. Wacker SA, Houghtaling BR, Elemento O, Kapoor TM. Using transcriptome sequencing to identify mechanisms of drug action and resistance. Nat Chem Biol. 2012; 8:235–237. [PubMed: 22327403]

97. Lamb J, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006; 313:1929–1935. [PubMed: 17008526]

98. Iorio F, et al. Discovery of drug mode of action and drug repositioning from transcriptional responses. Proc Natl Acad Sci USA. 2010; 107:14621–14626. [PubMed: 20679242]

99. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005; 102:15545–15550. [PubMed: 16199517]

100. Feng Y, Mitchison TJ, Bender A, Young DW, Tallarico JA. Multi-parameter phenotypic profiling: using cellular effects to characterize small-molecule compounds. Nat Rev Drug Discov. 2009; 8:567–578. [PubMed: 19568283]

101. Wagner BK, Clemons PA. Connecting synthetic chemistry decisions to cell and genome biology using small-molecule phenotypic profiling. Curr Opin Chem Biol. 2009; 13:539–548. [PubMed: 19825513]

102. Haupt VJ, Schroeder M. Old friends in new guise: repositioning of known drugs with structural bioinformatics. Brief Bioinform. 2011; 12:312–326. [PubMed: 21441562]

103. Koutsoukas A, et al. From *in silico* target prediction to multi-target drug design: current databases, methods and applications. J Proteomics. 2011; 74:2554–2574. [PubMed: 21621023]

104. Barrett T, et al. NCBI GEO: mining tens of millions of expression profiles—database and tools update. Nucleic Acids Res. 2007; 35:D760–D765. [PubMed: 17099226]

105. Seiler KP, et al. ChemBank: a small-molecule screening and cheminformatics resource database. Nucleic Acids Res. 2008; 36:D351–D359. [PubMed: 17947324]

106. Wang Y, et al. PubChem: a public information system for analyzing bioactivities of small molecules. Nucleic Acids Res. 2009; 37:W623–W633. [PubMed: 19498078]

107. Stegmaier K, et al. Gene expression–based high-throughput screening (GE-HTS) and application to leukemia differentiation. Nat Genet. 2004; 36:257–263. [PubMed: 14770183]

108. Hieronymus H, et al. Gene expression signature–based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. Cancer Cell. 2006; 10:321–330. [PubMed: 17010675]

109. Kim YK, et al. Relationship of stereochemical and skeletal diversity of small molecules to cellular measurement space. J Am Chem Soc. 2004; 126:14740–14745. [PubMed: 15535697]

110. Paull KD, et al. Display and analysis of patterns of differential activity of drugs against human tumor cell lines: development of mean graph and COMPARE algorithm. J Natl Cancer Inst. 1989; 81:1088–1092. [PubMed: 2738938]

111. Zaharevitz DW, et al. Discovery and initial characterization of the paullones, a novel class of small-molecule inhibitors of cyclin-dependent kinases. Cancer Res. 1999; 59:2566–2569. [PubMed: 10363974]

112. Marton MJ, et al. Drug target validation and identification of secondary drug target effects using DNA microarrays. Nat Med. 1998; 4:1293–1301. [PubMed: 9809554]

113. Ganter B, et al. Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. J Biotechnol. 2005; 119:219–244. [PubMed: 16005536]

114. Fliri AF, Loging WT, Thadeio PF, Volkmann RA. Biological spectra analysis: Linking biological activity profiles to molecular structure. Proc Natl Acad Sci USA. 2005; 102:261–266. [PubMed: 15625110]

115. Berg EL, Kunkel EJ, Hytopoulos E, Plavec I. Characterization of compound mechanisms and secondary activities by BioMAP analysis. J Pharmacol Toxicol Methods. 2006; 53:67–74. [PubMed: 16040258]

116. Kauvar LM, et al. Predicting ligand binding to proteins by affinity fingerprinting. Chem Biol. 1995; 2:107–118. [PubMed: 9383411]

117. Campillos M, Kuhn M, Gavin AC, Jensen LJ, Bork P. Drug target identification using side-effect similarity. Science. 2008; 321:263–266. [PubMed: 18621671]

118. Dixon SL, Villar HO. Bioactive diversity and screening library selection via affinity fingerprinting. J Chem Inf Comput Sci. 1998; 38:1192–1203. [PubMed: 9845969]

119. Perlman ZE, et al. Multidimensional drug profiling by automated microscopy. Science. 2004; 306:1194–1198. [PubMed: 15539606]

120. Carpenter AE. Image-based chemical screening. Nat Chem Biol. 2007; 3:461–465. [PubMed: 17637778]

121. Chen B, Wild D, Guha R. PubChem as a source of polypharmacology. J Chem Inf Model. 2009; 49:2044–2055. [PubMed: 19708682]

122. Tanikawa T, et al. Using biological performance similarity to inform disaccharide library design. J Am Chem Soc. 2009; 131:5075–5083. [PubMed: 19298063]

123. Plouffe D, et al. *In silico* activity profiling reveals the mechanism of action of antimalarials discovered in a high-throughput screen. Proc Natl Acad Sci USA. 2008; 105:9059–9064. [PubMed: 18579783]

124. Wolpaw AJ, et al. Modulatory profiling identifies mechanisms of small molecule-induced cell death. Proc Natl Acad Sci USA. 2011; 108:E771–E780. [PubMed: 21896738]

125. Cheng T, Li Q, Wang Y, Bryant SH. Identifying compound-target associations by combining bioactivity profile similarity search and public databases mining. J Chem Inf Model. 2011; 51:2440–2448. [PubMed: 21834535]

126. Petrone PM, et al. Rethinking molecular similarity: comparing compounds on the basis of biological activity. ACS Chem Biol. 2012; 7:1399–1409. [PubMed: 22594495]

127. Nidhi Glick M, Davies JW, Jenkins JL. Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. J Chem Inf Model. 2006; 46:1124–1133. [PubMed: 16711732]

128. Paolini GV, Shapland RH, van Hoorn WP, Mason JS, Hopkins AL. Global mapping of pharmacological space. Nat Biotechnol. 2006; 24:805–815. [PubMed: 16841068]

129. Keiser MJ, et al. Relating protein pharmacology by ligand chemistry. Nat Biotechnol. 2007; 25:197–206. [PubMed: 17287757]

130. Lounkine E, et al. Large-scale prediction and testing of drug activity on side-effect targets. Nature. 2012; 486:361–367. [PubMed: 22722194]

131. Keiser MJ, et al. Predicting new molecular targets for known drugs. Nature. 2009; 462:175–181. [PubMed: 19881490]

132. Gregori-Puigjané E, et al. Identifying mechanism-of-action targets for drugs and probes. Proc Natl Acad Sci USA. 2012; 109:11178–11183. [PubMed: 22711801]

133. Laggner C, et al. Chemical informatics and target identification in a zebrafish phenotypic screen. Nat Chem Biol. 2012; 8:144–146.

134. Oprea TI, Tropsha A, Faulon JL, Rintoul MD. Systems chemical biology. Nat Chem Biol. 2007; 3:447–450. [PubMed: 17637771]

135. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol. 2008; 4:682–690. [PubMed: 18936753]

136. Boran AD, Iyengar R. Systems pharmacology. Mt Sinai J Med. 2010; 77:333–344. [PubMed: 20687178]

137. 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:i232–i240. [PubMed: 18586719]

138. He Z, et al. Predicting drug-target interaction networks based on functional groups and biological features. PLoS ONE. 2010; 5:e9603. [PubMed: 20300175]

139. Mei JP, Kwoh CK, Yang P, Li XL, Zheng J. Drug-target interaction prediction by learning from local information and neighbors. Bioinformatics. 2013; 29:238–245. [PubMed: 23162055]

140. Bach S, et al. Roscovitine targets, protein kinases and pyridoxal kinase. J Biol Chem. 2005; 280:31208–31219. [PubMed: 15975926]

141. Kuai L, et al. AAK1 identified as an inhibitor of neuregulin-1/ErbB4–dependent neurotrophic factor signaling using integrative chemical genomics and proteomics. Chem Biol. 2011; 18:891–906. [PubMed: 21802010]

142. Raj L, et al. Selective killing of cancer cells by a small molecule targeting the stress response to ROS. Nature. 2011; 475:231–234. [PubMed: 21753854]

143. Arastu-Kapur S, et al. Nonproteasomal targets of the proteasome inhibitors bortezomib and carfilzomib: a link to clinical adverse events. Clin Cancer Res. 2011; 17:2734–2743. [PubMed: 21364033]

144. Wen Q, et al. Identification of regulators of polyploidization presents therapeutic targets for treatment of AMKL. Cell. 2012; 150:575–589. [PubMed: 22863010]

145. Winter GE, et al. Systems-pharmacology dissection of a drug synergy in imatinib-resistant CML. Nat Chem Biol. 2012; 8:905–912. [PubMed: 23023260]

146. Parsons AB, et al. Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. Nat Biotechnol. 2004; 22:62–69. [PubMed: 14661025]

147. Perlstein EO, et al. Revealing complex traits with small molecules and naturally recombinant yeast strains. Chem Biol. 2006; 13:319–327. [PubMed: 16638537]
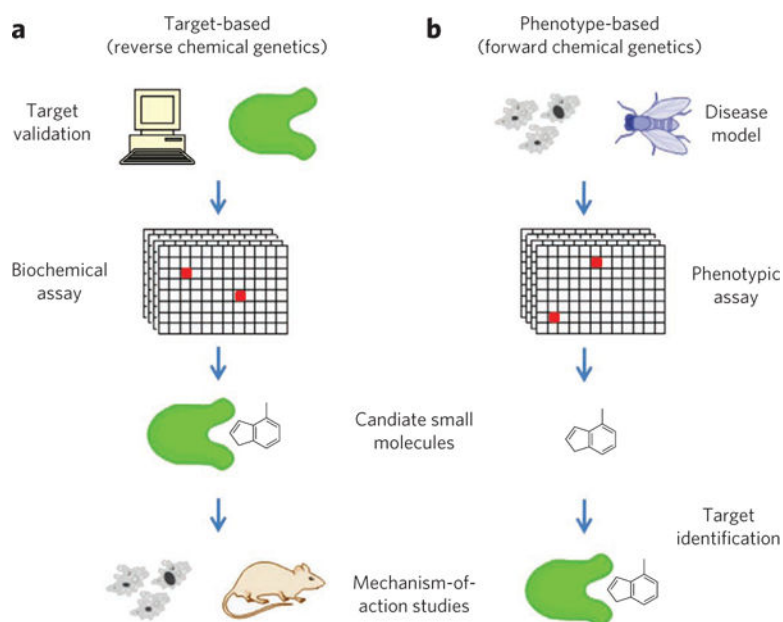
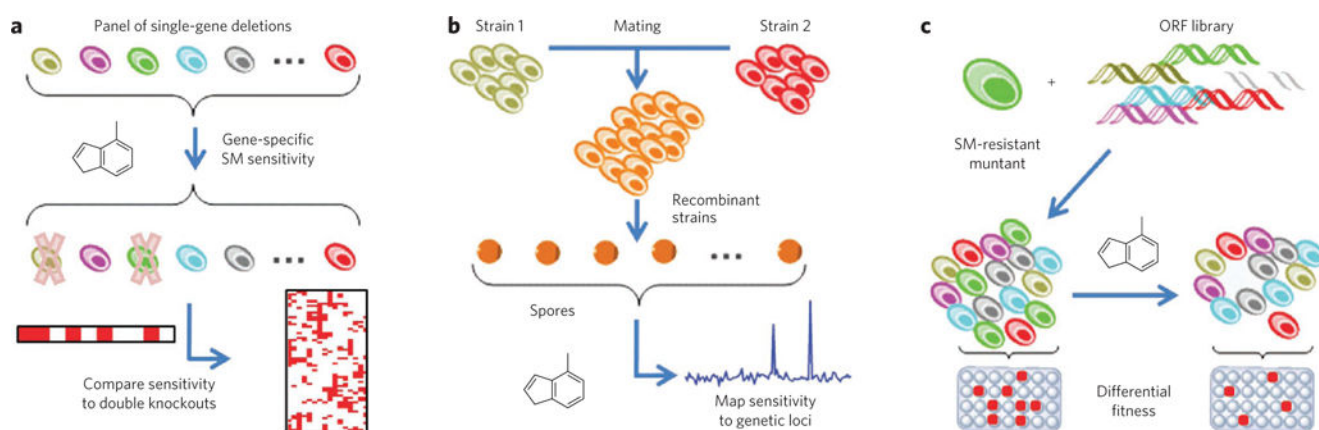**Figure 1. Mechanism-of-action and target identification in chemical genetics**
(**a**) Target-based approaches (reverse chemical genetics) begin with target validation, in which a role is established for a protein in a pathway or disease, followed by a biochemical assay to find candidate small molecules; mechanism-of-action studies are still required to validate cellular activities of candidates and evaluate possible side effects. (**b**) Phenotype-based approaches (forward chemical genetics) begin with a phenotype in a model system and an assay for small molecules that can perturb this phenotype; candidate small molecules must then undergo target-identification and mechanism-of-action studies to determine the protein responsible for phenotypic change.
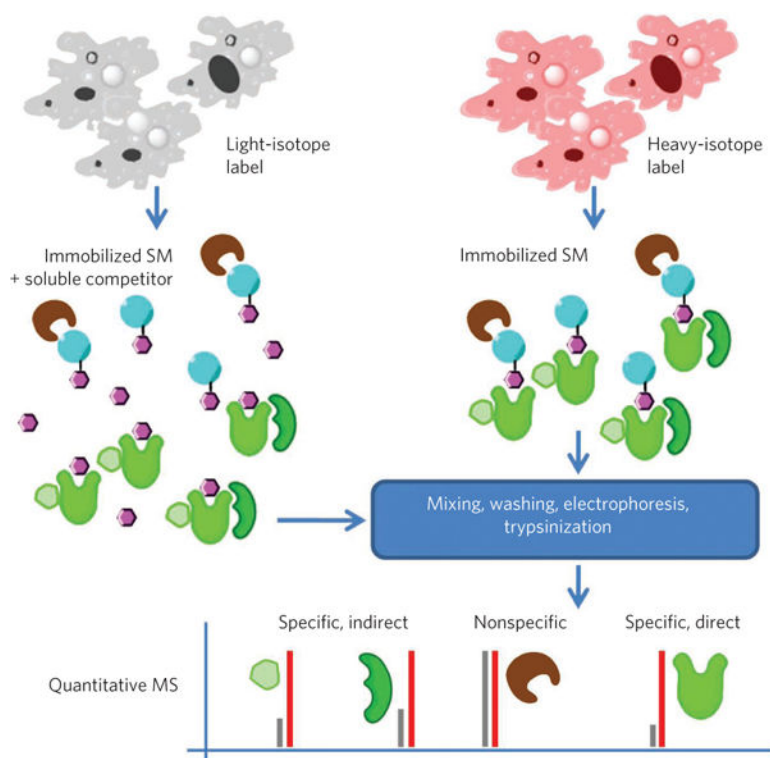
**Figure 2. Illustration of stable isotope labeling and quantitative MS**

Cells are labeled with either heavy- or light-isotope labels. One sample is exposed to bead-immobilized small molecules (SM) in the presence of soluble competitor compound and the other in the absence of competitor. Following mixing, washing and electrophoresis, samples are digested using trypsin and peptide fragments analyzed by quantitative MS[49]. Ratios of heavy- and light-labeled peptides are used to determine specificity of interactions for the small molecule, potentially including both direct and indirect targets (for example, members of complexes including the direct target), but not to differentiate them.

**Figure 3. Illustrations of yeast genomic methods for target-identification and mechanism-of-action studies**

(**a**) A panel of viable single-gene deletions is tested for small-molecule (SM) sensitivity, indicating synthetic-lethal interactions between potential targets and the original deletion; mechanisms are interpreted by comparing interactions to double-knockout strains[146]. (**b**) Different strains of diploid yeast are mated to form F1 recombinants, and meiotic progeny are subjected to small molecules; segregation frequencies allow mapping of small-molecule sensitivity to genetic loci[147]. (**c**) A recessive small molecule–resistant mutant is transformed with a wild-type open reading frame library; transformants obtaining a wild-type copy of the mutant gene are selectively sensitive to small molecules, resulting in their depletion among pooled transformants, as quantified by microarray[89].
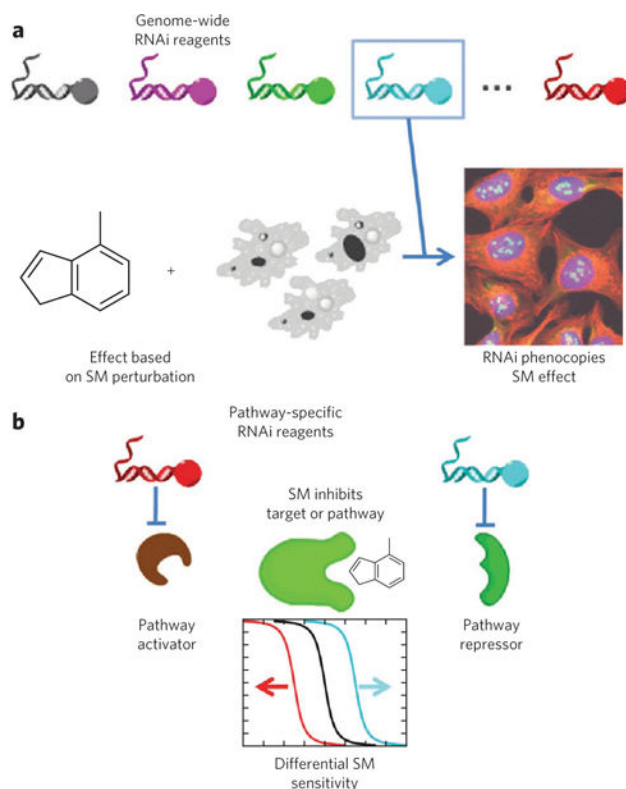
**Figure 4. Illustrations of RNAi-based methods for target-identification and mechanism-of-action studies**

(**a**) In one implementation, phenotypes from genome-wide RNAi are compared to those induced by a small molecule (SM) of interest; full or partial phenocopy of the small-molecule effect by RNAi provides evidence that the gene product is a small-molecule target[92]. (**b**) When prior evidence suggests a particular target pathway, focused sets of RNA reagents can help to generate mechanistic hypotheses. In general, RNAi can enhance or suppress small-molecule effects, as in genetic epistasis analysis; in practice, more complex relationships among proteins than those illustrated may also exist[93].
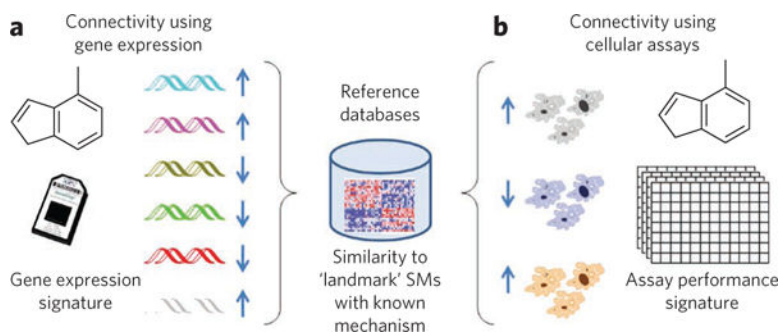
**Figure 5. Illustration of computational inference methods for target-identification and mechanism-of-action studies**

In general, data sets that provide multidimensional readouts of small molecule (SM)-induced phenotypes can be used to provide connections between new small-molecule signatures and reference databases by similarity to 'landmark' compounds with known mechanisms of action.
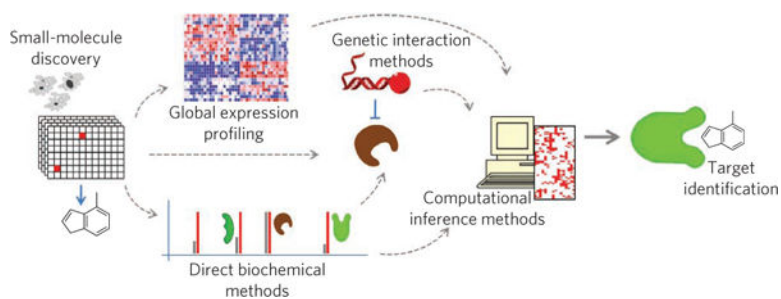
**Figure 6. Illustration of a conceptual workflow for integrated target-identification and mechanism-of-action studies**

Small-molecule discovery often starts with phenotypic screening, and, depending on the expertise available to the researchers, target identification could proceed using any combination of direct biochemical methods, genetic or genomic methods or using computational inference methods. A key component for success is the integration of data from all available methods to produce the most reliable target and mechanistic hypotheses.