Identifying compound efficacy targets in phenotypic drug discovery

Markus Schirle and Jeremy L. Jenkins

Developmental & Molecular Pathways, Novartis Institutes for BioMedical Research, Cambridge, MA 02139, USA

The identification of the efficacy target(s) for hits from phenotypic compound screens remains a key step to progress compounds into drug development. In addition to efficacy targets, the characterization of epistatic proteins influencing compound activity often facilitates the elucidation of the underlying mechanism of action; and, further, early determination of off-targets that cause potentially unwanted secondary phenotypes helps in assessing potential liabilities. This short review discusses the most important technologies currently available for characterizing the direct and indirect target space of bioactive compounds following phenotypic screening. We present a comprehensive strategy employing complementary approaches to balance individual technology strengths and weaknesses.

Introduction

Recently, there has been a resurgence of phenotypic screening in drug discovery as an alternative to the target-centric approaches that have dominated since the molecular biology revolution in the 1980s [1]. Conceptual advantages of phenotypic screens include a pre-selection for compounds with proven cellular activity, the potential to identify novel druggable proteins and mechanisms in a given disease context and the ability to identify compounds that elicit their phenotype via activity on multiple targets, see [2] for an in-depth discussion. Traditionally, however, the follow-up work of target identification and elucidating the mechanism of action (MoA) of hits from phenotypic screens has historically been slow and at times fruitless. Knowledge of the efficacy target, the protein through which the drug elicits the phenotype of interest via direct binding and modulation, is not necessarily an absolute requirement during initial development stages. However, the definition of the on- and off-target space remains a crucial step toward full development for a drug candidate; it facilitates optimization of potency within a lead scaffold and the identification of alternative chemical matter. Moreover, it allows assessment of safety as well as prediction and monitoring of efficacy when moving from a simplified cellular model to *in vivo* models and ultimately the patient. In recent years, the number of available approaches for target deconvolution has increased significantly. These approaches are based on fundamentally different principles and provide complementary information. Thus, a multipronged approach can significantly speed up and increase the success rate of target identification efforts by balancing strengths and weaknesses of the various approaches and enabling prioritization of target hypotheses by intersecting individual hit lists.

The various strategies can be grouped based on the exact question they are addressing (Fig. 1). They provide glimpses of different aspects of the often complex physical and functional interactions of a compound when exerting its biological effects *in vivo*. Importantly, these include not only the efficacy target but also other epistatic proteins relevant for compound effect, off-targets responsible for potentially unwanted secondary phenotypes and proteins involved in transport and modification of the compound (Fig. 2 provides definitions).

Affinity-based approaches focus on the identification of cellular interactors of a compound, which include direct efficacy and off-targets in the context of the phenotype of interest. Genetic and genomic functional approaches focus on identification of genes relevant for the compound effect on the phenotype of interest. These genes will include the efficacy target as well as other components involved in the specific cellular mechanism modulated by the compound. In contrast to the individual protein or gene

Corresponding authors: Schirle, M. (markus.schirle@novartis.com), Jenkins, J.L. (jeremy.jenkins@novartis.com)

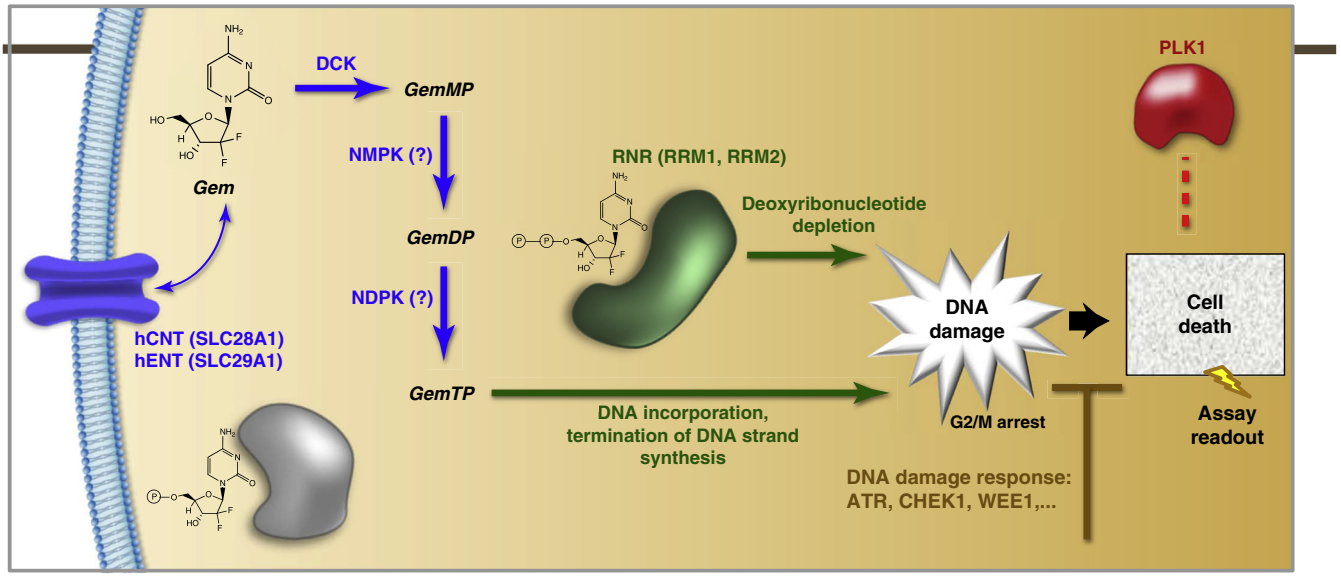
http://dx.doi.org/10.1016/j.drudis.2015.08.001

Focus on: compound properties	*Focus on: target function*
Affinity-based approaches - Small-molecule affinity proteomics - Proteomic target stability profiling - SEC-TID - Protein arrays - Yeast-3-hybrid - Ribosome display	**Functional genetic approaches** - HIP HOP - MutaSeq - Variomics - Compound-sensitized gene screens (si/shRNA, CRISPR/Cas9, cDNA) - Haploid gene trap screens
Knowledge-based approaches - Chemical similarity to references or ensembles - Bioactivity (HTS) fingerprint similarity - Machine learning / classification - MoA-specific compound collections - Computational high-throughput docking - Causal network modeling	**Cellular profiling approaches** - Promoter signature profiling - Cell line cytotoxicity profiling - Gene expression signature analysis - Metabolomics

Drug Discovery Today

FIGURE 1

Overview of available technologies for target deconvolution and elucidation of compound mechanism of action, grouped by underlying principles.



Drug Discovery Today

FIGURE 2

The drug effect pathway, illustrated for gemcitabine (Gem), consists of the efficacy target(s) as well as additional classes of epistatic proteins that influence compound phenotype. Efficacy target - direct target of compound, physical interactor through which the drug elicits the phenotype of interest; here [e.g. ribonucleotide reductase (RNR) DNA] incorporation leads to termination of DNA strand synthesis. Indirect compound effector - not a direct target of the compound but influences drug-dependent effect on phenotype (e.g. complex members modulating target activity, components of cellular mechanisms counteracting compound effect; here, components of DNA damage response). Binding (off)-target - binds drug directly or indirectly, but binding is irrelevant to phenotype of interest; could still mediate different phenotype. Compound-specific effector - influences compound activity but not part of phenotype pathway [e.g. transporters (in/out), bioactivating/metabolizing enzymes; here, deoxycytidine kinase (dKC) and other kinases involved in Gem activation via phosphorylation, nucleoside transporters (hCNT, hENT) required for Gem uptake]. Compound-independent assay effector?modulates assay phenotype when altered, with or without compound treatment; here, PLK1 knockdown alone causes cell toxicity.

resolution provided by these approaches, cellular profiling approaches interrogate the overall cellular response to compound treatment at the level of signaling, gene expression, viability or metabolism. Finally, knowledge-based approaches rely on empirical and computational approaches and a reference collection of compounds with known targets and MoA to make inferences. Ultimately, all approaches generate data linking compounds to targets, edging forward steadily toward the chemical biology aspiration of finding a compound modulator for every amenable target.

Affinity-based approaches

Affinity-based approaches for target deconvolution aspire to identify the full protein-binding spectrum of a compound; this is followed by functional classification of these interactors as efficacy targets, off-targets or compound-specific effectors. Available strategies can be grouped into two broad categories: chemoproteomic approaches interrogate proteins directly in cell lysates or live cells; *in vitro* approaches rely on purified protein or protein domains. The former are generally closer to the naturally occurring proteoform with respect to sequence, post-translational modifications and presence of interacting proteins that can all influence compound binding. However, despite ever-increasing sensitivity of detection methods, the differences in cellular abundance of putative interactors that are being interrogated in parallel still pose a challenge (protein levels span ten orders of magnitude). In addition, most of these methods do not differentiate between direct and indirect binding. A popular version of chemoproteomics combines small-molecule affinity chromatography with mass-spectrometry (MS)-based proteomics [3]. For unbiased target deconvolution, the compound of interest is immobilized on solid support via a functionalized linker introduced at a permissive site so that the resulting affinity probe retains cellular activity. The identification and synthesis of such validated tool compounds can be challenging, in particular for complex natural products. Here, photo-crosslinker matrices can be a viable alternative. Irradiation of the matrix together with the parent compound leads to site-nonselective carbene insertion or addition into the compound. This results in a random compound display on beads with the assumption that at least one of the multiple orientations is compatible with target binding [4]. In all cases, the immobilized compound is then used for enrichment of putative interactors by incubation with a disease-relevant cell lysate. Often, additional steps are taken to further discriminate high-affinity interactors that are more likely to be phenotypically relevant from low-affinity but high-abundance interactors. These include pre-incubation with free compound [5,6] or comparing the enrichment profiles of active and inactive affinity probes from the same scaffold [7]. In these cases, stable isotope-based or label-free approaches to quantitative MS provide the readout for relative quantitation [8,9]. Functional complexes are often preserved during cell lysis and hits will include compound-binding proteins as well as their accessory proteins. This makes further deconvolution by validation of direct binding *in vitro* or by functional approaches necessary. By contrast, this feature has been successfully exploited to demonstrate differential binding of classes of histone deacetylase (HDAC) inhibitors to different and even novel chromatin-remodeling complexes [10]. Noncovalent, lysate-based approaches have been successfully applied to the identification of a wide range of targets, including soluble enzymes such as protein kinases [5], the lipid kinase PIKFyve [11] and many others. However, a number of therapeutically relevant target families, in particular integral membrane proteins such as ion channels and G-protein-coupled receptors, are notoriously incompatible owing to loss of their binding-competent conformation during the experimental workflow [12,13].

By contrast, covalent approaches allow 'freezing' of the interaction between the affinity probe and the target. This also enables application to live cells or crude membrane fractions. For *de novo* target identification, exemplified by the determination of Sec61α as the target of a cyclodepsipeptide inhibitor of translocation [14], the compound of interest is conjugated to a reactive group as well as an affinity handle (e.g. biotin). Reactive groups can be based on a number of chemical scaffolds including protein-reactive natural products [15], general electrophilic groups [16,17] and photoreactive moieties as used in the example above. To improve cell permeability of the affinity probe, the affinity handle is often introduced after covalent bond formation as well as cell lysis, using, for example, click chemistry [18]. Covalent approaches also enable denaturing experimental conditions because proteins do not have to be preserved in their binding-competent conformation during the workflow. As a result, hit lists are biased toward direct interactors reflecting a proximity-driven labeling event. This can be exploited for further target deconvolution in cases where noncovalent approaches identify macromolecular complexes consisting of direct and indirect compound interactors. This concept is exemplified by comparing the data reported for probes based on the class I/IIb HDAC active-site inhibitor suberanilohydroxamic acid (SAHA). As mentioned, a noncovalent SAHA probe yields several, functionally distinct HDAC-containing complexes as specific hits [10]. By contrast, the covalent SAHA–BPyne probe yields only core HDACs and a very small number of additional complex members as specific binders, reflecting HDAC-binding sites located close to the active site for the latter [19]. It should be noted that background as a result of low-affinity interactions and low-efficiency labeling (e.g. for photoreactive groups) remains an issue for covalent approaches. Therefore, additional experimental steps such as competition with free compound or a probe titration are typically required to assess relative affinity and proteome-wide specificity of a probe–target interaction.

In addition to affinity-chromatography-based approaches, strategies have been developed that identify interactions by monitoring changes in target stability upon compound binding. Importantly, these approaches do not require time- and resource-intensive modification of the original compound. The various reported strategies differ in the specific assay format used as readout for protein stability in combination with (quantitative) proteomics [20–23]. The cellular thermal shift assay (CETSA) is analytically straightforward – target stabilization is determined by quantifying the increase of non-heat-denatured protein in solution upon compound treatment. However, the degree of binding-induced stabilization is specific to each individual protein and not every binding event leads to detectable stabilization of the full length protein. This complicates a proteome-wide rank ordering of putative targets. Together with the absence of a functional enrichment step potentially limiting dynamic range, this approach

seems currently best-suited for focused studies on engagement of soluble targets in live cells.

A number of strategies employ a functional readout in living cells for detection of compound–protein interactions. The yeast-3-hybrid approach extends the classical yeast-2-hybrid system for detection of protein–protein interactions by split transcription factor complementation to protein–compound interactions [24]. In a typical setup as used for the recent identification of phosphodiesterase PDE6D as the target of anecortave acetate [25], the compound of interest is conjugated to the bait ligand methotrexate. Interaction with components of a cDNA library leads to a dihydrofolate-reductase-mediated transcriptional activation of a reporter gene. Although the concept has also been extended to mammalian cells [26], applications to compound target ID are still relatively scarce. This might reflect limitations as a result of incomplete coverage of the functional proteome by the mammalian cDNA libraries and the restriction to interactions that can occur in the nucleus.

In addition, a number of biochemical and biophysical approaches are available that use arrayed collections of purified proteins. These are typically based on recombinant expression systems and massive protein production as a first step. Importantly, the normalized protein content can overcome abundance issues inherent to chemoproteomics approaches. Furthermore, these *in vitro* approaches enable a detailed analysis of direct protein–compound interactions, making them attractive for validation of specific target hypotheses generated by other approaches. Beyond such more-focused applications, several platforms have been reported that allow interrogation of a large part of the (human) proteome; protein arrays have been used to assess binding to proteins immobilized on surfaces but require labeled compound and are prone to background issues [27]. By contrast, size-exclusion chromatography for target identification (SEC-TID) assesses binding of an underivatized compound to thousands of individual soluble proteins and domains in a well-by-well SEC format. This allowed the identification of novel interactors of vadimezan and mefruside [28]. Finally, ribosome display-based approaches have been explored in proof-of-concept studies for compound target ID [29]. Notably, as for most affinity-based approaches, these *in vitro* strategies tend to work best for non-transmembrane targets for which binding competence is less dependent on a functional cellular environment.

Functional genomics approaches

A second group of approaches focuses on genetics to identify cellular components that are functionally relevant for the MoA of a compound. Genomic approaches offer powerful opportunities to identify compound efficacy targets in an unbiased manner; however, they are also particularly susceptible to identifying proteins that indirectly influence drug effect (Fig. 2) and can be 'blind' to targets with redundant, compensatory paralogs.

Historically, model organisms have provided useful target identification systems owing to the ability to create gene-deletion libraries that are heterozygous or homozygous (if nonessential). For example, genetically manipulated diploid *Saccharomyces cerevisiae* libraries [30] can be screened in the presence of drug in haploinsufficiency profiling (HIP) or homozygous deletion profiling (HOP) to find genetic deletions that sensitize or suppress drug cytotoxicity [31,32]. Similarly, *Escherichia coli* models that enable systematic drug target titration have helped shed light on drug MoA as well as modalities [33].

Owing to the increasing availability of affordable next-generation sequencing (NGS) technologies, the genome-wide identification of spontaneous or chemically induced resistance-conferring mutations (MutaSeq, DrugTargetSeqR) has gained momentum [34]. MutaSeq is analogous to the clinical scenario of drug resistance through acquired mutations in efficacy targets or compound-specific effectors. Typically, multiple clones can help pinpoint efficacy targets, but also indirect resistance mechanisms emerge (e.g. xenobiotic pumps) [34], necessitating follow-up experiments [35]. Unfortunately, MutaSeq is restricted to compounds with cytotoxic mechanisms. Further, compounds with polypharmacology create a higher hurdle for selecting resistant clones.

An alternative resistance-based approach uses drug treatment of near-haploid human KBM7 cells along with random retroviral introduction of a tagged reporter cassette. Drug-resistant colonies are grown, and the gene containing the inserted reporter cassette is sequenced to reveal targets that suppress drug MoA when disrupted [36]. Like yeast homozygous mutant profiling, haploid genetic screens find efficacy targets only for activating or agonist drug modalities. By contrast, for drugs with inhibitory or antagonist mechanisms, only indirect effectors or compound-specific effectors are revealed.

In addition to genetic knockouts, specific residues responsible for drug binding can be determined systematically by creating protein variants across the genome, an approach dubbed 'variomics' [37]. Functional variomics uses systematic construction of genetic libraries based on error-prone PCR and screening of numerous genetic variants to identify specific point mutations that influence compound efficacy. Although still an emerging approach in human cell lines, variomics holds promise of early prediction of drug-resistance mutations or better understanding target function [38].

Alternatively, similar genetic suppression or potentiation of compound mechanism has been achieved in human cells by combining compound treatment with cDNA overexpression [39] as well as with RNAi, reviewed in [40], in genome-wide screens. siRNA and pooled shRNA screens are amenable to unbiased, genome-wide compound-sensitization experiments. The misconception that RNAi completely knocksdown mRNA has led to confusion as to whether compound–RNAi screens can identify direct drug targets (as opposed to only indirect effectors). In practice, knockdown is variable, and direct and indirect targets are identified because of rampant partial knockdown by siRNA libraries, which creates a synthetic gene insufficiency. As an example, successful target ID was achieved for STF118804, a compound that reduced viability of most B-ALL cell lines, using a genome-wide pooled shRNA screen, where the direct target, nicotinamide phosphoribosyltransferase (NAMPT), was the most statistically significant gene conferring drug sensitivity [41]. In a second example, toxicity of gemicitabine was potentiated by siRNA knockdown of the direct target, Ribonucleotide Reductase (RRM1), as well as by inhibition of Checkpoint Kinase 1 (CHEK1) caused by prevention of DNA damage repair, reinforcing the concept of indirect compound effectors (Fig. 2) [42,43]. Therefore, the 'drug effect pathway' can be well characterized by the total

complement of hits from loss-of-function screens, but the challenge remains to pick the efficacy target among the sensitizing genes. Further, the directionality of the drug effect on its efficacy target – antagonist, agonist or other – presents an additional challenge for interpretation. For agonists, efficacy target knockdown creates a suppressive effect; whereas for antagonists, efficacy target knockdown potentiates compound mechanism. It is ideal therefore to run multiple compound doses in sensitization screens and analyze the data for drug suppression and potentiation.

With the advent of CRISPR–Cas9 genome-editing technologies, additional approaches are emerging to elucidate compound MoA. Transcriptional repression or activation by CRISPRi and CRISPRa, respectively, in combination with ricin toxin or a diphtheria–cholera chimera, revealed pathway members potentiating or suppressing toxin entry and MoA. Moreover, transcription levels were shown to be programmable by doxycycline-inducible Cas9 [41]. Thus, CRISPRi and CRISPRa are a target ID platform analogous to RNAi/cDNA screening, but with greater precision control. Genome editing promises to be a premiere target ID platform in addition to compound target validation [44].

Cellular profiling approaches

Profiling approaches to target ID and MoA elucidation are based on – in the broadest sense – functional readouts of compound action on targets, cells or organisms. In the simplest example, a compound can be tested for direct activity against a panel of targets (e.g. commercially available kinase or receptor panels). By contrast, cellular profiling approaches provide information about the compound MoA by measuring various cellular responses. They can also incorporate correlation with underlying cellular genetic features. In most cases, these approaches are based on correlation of generated profiles to a reference set of compounds with known MoA, using for example pattern matching algorithms. As a result, these approaches are most powerful for binning hits from phenotypic screens into known versus potentially novel MoA. This enables prioritization of the latter for higher resolution affinity- or function-based approaches. The earliest efforts used gene expression profiles generated in yeast [45] or human cell lines: Connectivity Map [46]. Higher throughput has been achieved by using a signature of ~100 genes maximally responsive to compound treatment [47]. Promoter signature profiling (PSP) uses a collection of >40 high-throughput luciferase-based reporter gene assays [48]. Taken together, the resulting profile provides a snapshot of the cellular signaling response to compound treatment and has proven in our hands to be valuable for large-scale correlation to profiles of compounds with known MoA.

In an analogous manner, the quantitative analysis of the cellular metabolome has been successfully applied for target ID of compounds that directly or indirectly affect metabolite levels. An example is the identification of the glucose transporter GLUT1 as the target of glucopiericidin A [49]. Finally, a large collection of cell lines that has been characterized in-depth at the genetic level and beyond – the Cancer Cell Line Encyclopedia (CCLE) [50] – has proven to be a powerful tool for target ID and MoA elucidation of cytotoxic compounds. Here, selective toxicity for a subset of cell lines allows the identification of genetic features that correlate with sensitivity or resistance [50].

Knowledge-based approaches

In general, computational methods for target ID are the fastest shown in Fig. 1, making them the first approaches implemented for deconvoluting phenotypic screening hits. However, like all approaches relying on correlation to reference compounds, they are limited in their breadth of targets that can be predicted, and ultimately depend on experimental validation of efficacy targets. Knowledge-based approaches, at their core, rely on exploitation of integrated compound–target associations gleaned from bioactivity data. Bioactivity data can be assembled from high-throughput screening, dose–response validation or low-throughput experiments found in public, commercial and proprietary sources [51]. The fact that bioactivity data are vast and span many data sources presents a considerable infrastructure burden [52], starting with the problem of target name and chemical structure standardization. Despite these challenges, advances have been made with regard to data in the public domain as a result of efforts like ChEMBL [53] and the open PHACTS project [54]. However, there remains not yet one authoritative source for compound–target associations, which creates an obstacle for most organizations pursuing knowledge-based approaches. Therefore, successful execution of knowledge-based approaches depends on core expertise in cheminformatics and chemical biology to integrate large-scale pharmacology data. In terms of genomic coverage by chemical matter, we estimate that ~4000 human targets have associated chemical modulators from literature, patents or databases available to us, corresponding to roughly 15–20% coverage of the human genome [55]. Even so, direct annotation of hit lists from phenotypic screening, combined with gene set enrichment on annotated targets, is a fast, first-pass approach to guess targets influencing assay biology [55,56].

In lieu of direct target annotations for compounds, a number of knowledge-based approaches exist to predict targets based on the similarity principle, or 'what reference bioactive compound does my compound look like?' A further review is presented in [57]. Ligand-based target prediction requires: (i) a compound descriptor; and (ii) a similarity metric or machine learning technique. Regarding descriptors, many flavors of 2D and 3D chemical descriptors have been evaluated. In recent years, extended connectivity fingerprints have been widely successful. In contrast to chemical descriptors, fingerprints that incorporate compound behavior across aggregated historical assays into a biological fingerprint have proven equally powerful (e.g. HTS fingerprints) [58,59]. For similarity determination, reference database molecules can be searched individually by classical metrics like Tanimoto or Pearson correlation, or as ligand sets, such as the similarity ensemble approach (SEA) [60]. Many machine learning approaches have been explored for training target class models to predict targets of compounds, ranging from Naïve Bayes [61] to Random Forest to Deep Learning [62]. Indeed, machine learning approaches to target prediction have had recent successes, such as application to open access compounds and neglected diseases [63]. Further, many groups are attempting to use network science by projecting compound–target bioactivities for disparate data sources onto graph nodes and edges so that compound–target associations can be ranked [64].

Again, genome coverage is limited for the above computational approaches and, for compounds having unprecedented targets,

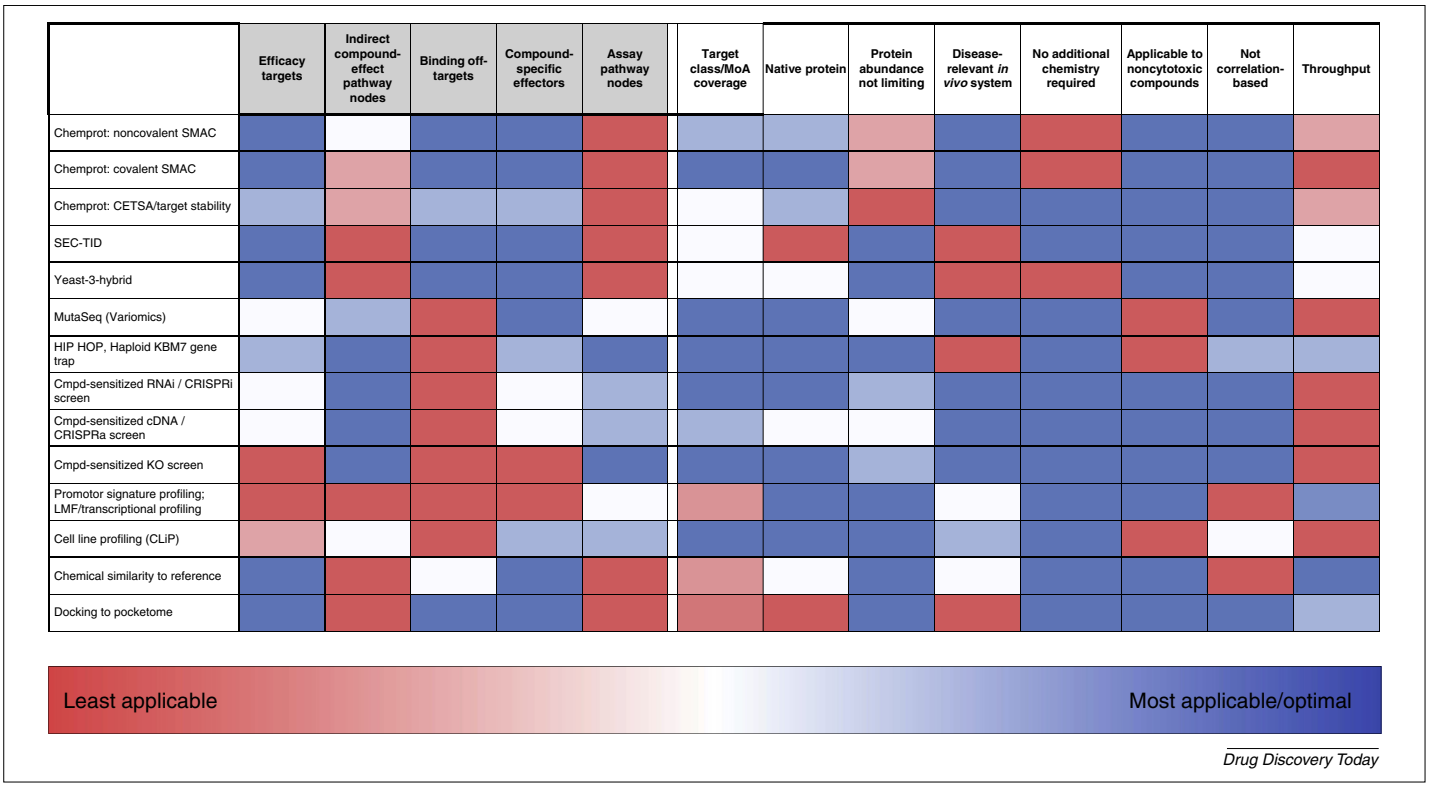


FIGURE 3

Heat map describing applicability of different target deconvolution or mechanism of action (MoA) technologies for identification of various components of the drug effect pathway (gray) as well as performance with regard to relevant parameters and requirements. Note: individual rankings are necessarily subjective. Key abbreviations: chemprot, chemoproteomics; SMAC, small-molecule affinity chromatography; cmpd, compound.

only methods that can extrapolate to undrugged targets will have a chance of success. For example, machine learning approaches that focus on protein domains rather than individual targets can predict novel compound–target pairs [65]. Even more promising, the field of proteochemometrics is attempting to learn on matrices of protein features and chemical features with quantitative activities to allow compound–target prediction for new compounds and/or targets [66]. An alternative approach is to dock compounds to the protein 'pocketome' from X-ray crystal structures to score potential protein interactors [67], a tactic not limited to the liganded genome but rather to the corpus of proteins with crystal structures and ligandable pockets. One final knowledge-based approach to target ID that is not limited to similarity to reference compounds is to apply causal modeling on network graphs seeded with nodes from compound profiling data. For example, differentially expressed genes resulting from compound treatment can be causally linked to the direct compound target 'upstream' in an interaction network (gene–protein and protein–protein edges); proteins in the interactome can thus be predicted to be the compound target if they display the proper connectivity and network directionality commensurate with the magnitude of gene transcriptional effects [68].

Concluding remarks and future outlook

It is for a good reason that a multitude of approaches to target deconvolution exist and are actively being developed. No one method can exquisitely reveal all types of drug efficacy targets regarding the strengths and weaknesses highlighted in Fig. 3. Thorough understanding of these caveats is key to strategic technology deployment in target ID campaigns. We propose here that the distinction between the 'drug–effect pathway' and the pathway driving a phenotype-of-interest (Fig. 2) is useful to guide experiment interpretation. Each technology is susceptible to

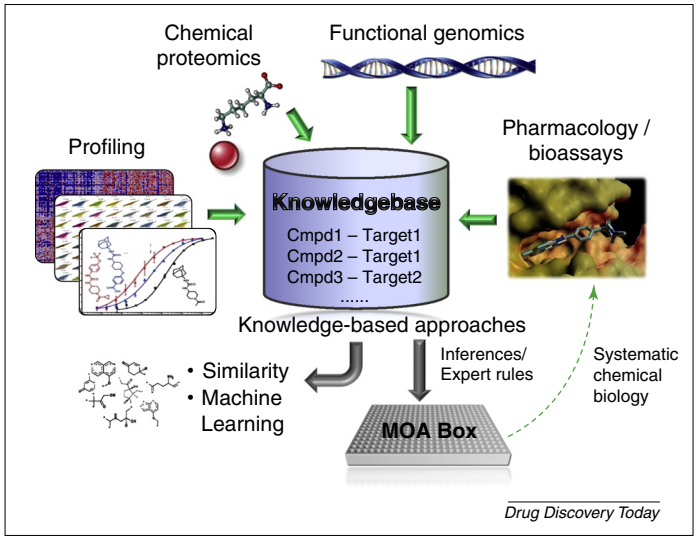


FIGURE 4

Target deconvolution or mechanism of action (MoA) technologies discover new compound–target relationships that, integrated together with bioassay activity databases, form a foundation for chemical biology knowledge. Computational approaches are continuously improved as knowledge grows,

highlighting epistatic proteins that are informative for drug MoA, yet can complicate identification of the efficacy target of the molecule.

In the past few years, target ID technologies have shifted away from approaches that rely solely on correlation to reference compounds in favor of unbiased, genome-wide approaches, enabling more opportunities to discover novel drug targets. New generations of proteomics-based strategies are increasingly compatible with experiments in living cells by employing covalent workflows or target stability-based readouts. Functional genomics, by contrast, has shifted away from synthetic lethal approaches in model organisms toward genome-wide loss-of-function or gain-of-function screens in human cells by genome editing.

The described suite of technologies, together with traditional bioassay data, generates a continuously expanding number of compound–target relationships, which can be integrated into a core knowledge-base feeding back into a phenotypic drug discovery cycle (Fig. 4). This chemical biology knowledge then enables annotations and target prediction for phenotypic hit lists. Further, the integrated pharmacology data yield a virtual framework to construct small compound sets (boxes) with well-annotated mechanisms. For organizations with large compound libraries, dynamic cherry-picking capacity and a database of high-confidence compound–target annotations, we propose that such a 'MoA box' can be a frontline approach to target discovery by phenotypic screening. Ironically, by integrating chemical genetic data, the phenotypic drug discovery pendulum swings back toward 'reverse chemical genetics', enabling systematic chemical biology, and obviating the need for the target ID with the goal of opportunistically linking known targets to new biology through compound-induced phenotypes. Nevertheless, many important compounds in pharmacology defy a simple compound–target modulation relationship – from paclitaxel to thalidomide to metformin. Therefore, phenotypic screening is expected to continue to have a crucial role in drug discovery – complementary but orthogonal to genetic screening and harnessing the true power of chemical biology.

Conflict of interest

Markus Schirle and Jeremy L. Jenkins are employees of Novartis Institutes for BioMedical Research.

Acknowledgment

We thank John Tallarico and Bill Forrester for critical reading of the manuscript and valuable discussions.

References

1 Swinney, D.C. and Anthony, J. (2011) How were new medicines discovered? *Nat. Rev. Drug Discov.* 10, 507–519
2 Lee, J.A. and Berg, E.L. (2013) Neoclassic drug discovery: the case for lead generation using phenotypic and functional approaches. *J. Biomol. Screen.* 18, 1143–1155
3 Rix, U. and Superti-Furga, G. (2009) Target profiling of small molecules by chemical proteomics. *Nat. Chem. Biol.* 5, 616–624
4 Kanoh, N. *et al.* (2005) Photo-cross-linked small-molecule affinity matrix for facilitating forward and reverse chemical genetics. *Angew. Chem. Int. Ed. Engl.* 44, 3559–3562
5 Bantscheff, M. *et al.* (2007) Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat. Biotechnol.* 25, 1035–1044
6 Ong, S.E. *et al.* (2009) Identifying the proteins to which small-molecule probes and drugs bind in cells. *Proc. Natl. Acad. Sci. U. S. A.* 106, 4617–4622
7 Oda, Y. *et al.* (2003) Quantitative chemical proteomics for identifying candidate drug targets. *Anal. Chem.* 75, 2159–2165
8 Bantscheff, M. *et al.* (2007) Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* 389, 1017–1031
9 Bantscheff, M. *et al.* (2012) Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal. Bioanal. Chem.* 404, 939–965
10 Bantscheff, M. *et al.* (2011) Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nat. Biotechnol.* 29, 255–265
11 Cai, X. *et al.* (2013) PIKfyve, a class III PI kinase, is the target of the small molecular IL-12/IL-23 inhibitor apilimod and a player in Toll-like receptor signaling. *Chem. Biol.* 20, 912–921
12 Bassilana, F. *et al.* (2014) Target identification for a Hedgehog pathway inhibitor reveals the receptor GPR39. *Nat. Chem. Biol.* 10, 343–349
13 Akbulut, Y. *et al.* (2015) (−)-Englerin A is a potent and selective activator of TRPC4 and TRPC5 calcium channels. *Angew. Chem. Int. Ed. Engl.* 54, 3787–3791
14 MacKinnon, A.L. *et al.* (2007) Photo-leucine incorporation reveals the target of a cyclodepsipeptide inhibitor of cotranslational translocation. *J. Am. Chem. Soc.* 129, 14560–14561
15 Drahl, C. *et al.* (2005) Protein-reactive natural products. *Angew. Chem. Int. Ed. Engl.* 44, 5788–5809
16 Weerapana, E. *et al.* (2008) Disparate proteome reactivity profiles of carbon electrophiles. *Nat. Chem. Biol.* 4, 405–407
17 Evans, M.J. *et al.* (2005) Target discovery in small-molecule cell-based screens by *in situ* proteome reactivity profiling. *Nat. Biotechnol.* 23, 1303–1307
18 Speers, A.E. *et al.* (2003) Activity-based protein profiling *in vivo* using a copper(i)-catalyzed azide-alkyne [3+2] cycloaddition. *J. Am. Chem. Soc.* 125, 4686–4687
19 Salisbury, C.M. and Cravatt, B.F. (2007) Activity-based probes for proteomic profiling of histone deacetylase complexes. *Proc. Natl. Acad. Sci. U. S. A.* 104, 1171–1176
20 Lomenick, B. *et al.* (2009) Target identification using drug affinity responsive target stability (DARTS). *Proc. Natl. Acad. Sci. U. S. A.* 106, 21984–21989
21 Liu, P.F. *et al.* (2011) Energetics-based discovery of protein–ligand interactions on a proteomic scale. *J. Mol. Biol.* 408, 147–162
22 West, G.M. *et al.* (2010) Quantitative proteomics approach for identifying protein–drug interactions in complex mixtures using protein stability measurements. *Proc. Natl. Acad. Sci. U. S. A.* 107, 9078–9082
23 Savitski, M.M. *et al.* (2014) Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* 346, 1255784
24 Licitra, E.J. and Liu, J.O. (1996) A three-hybrid system for detecting small ligand-protein receptor interactions. *Proc. Natl. Acad. Sci. U. S. A.* 93, 12817–12821
25 Shepard, A.R. *et al.* (2013) Identification of PDE6D as a molecular target of anecortave acetate via a methotrexate-anchored yeast three-hybrid screen. *ACS Chem. Biol.* 8, 549–558
26 Caligiuri, M. *et al.* (2006) MASPIT: three-hybrid trap for quantitative proteome fingerprinting of small molecule-protein interactions in mammalian cells. *Chem. Biol.* 13, 711–722
27 Huang, J. *et al.* (2004) Finding new components of the target of rapamycin (TOR) signaling network through chemical genetics and proteome chips. *Proc. Natl. Acad. Sci. U. S. A.* 101, 16594–16599
28 Salcius, M. *et al.* (2014) SEC-TID: A label-free method for small-molecule target identification. *J. Biomol. Screen.* 19, 917–927
29 Zhu, J. *et al.* (2013) Protein interaction discovery using parallel analysis of translated ORFs (PLATO). *Nat. Biotechnol.* 31, 331–334
30 Giaever, G. and Nislow, C. (2014) The yeast deletion collection: a decade of functional genomics. *Genetics* 197, 451–465
31 Hoepfner, D. *et al.* (2014) High-resolution chemical dissection of a model eukaryote reveals targets, pathways and gene functions. *Microbiol. Res.* 169, 107–120
32 Lee, A.Y. *et al.* (2014) Mapping the cellular response to small molecules using chemogenomic fitness signatures. *Science* 344, 208–211
33 Palmer, A.C. and Kishony, R. (2014) Opposing effects of target overexpression reveal drug mechanisms. *Nat. Commun.* 5, 4296
34 Wacker, S.A. *et al.* (2012) Using transcriptome sequencing to identify mechanisms of drug action and resistance. *Nat. Chem. Biol.* 8, 235–237
35 Kasap, C. *et al.* (2014) DrugTargetSeqR: a genomics- and CRISPR-Cas9-based method to analyze drug targets. *Nat. Chem. Biol.* 10, 626–628

36 Reiling, J.H. *et al.* (2011) A haploid genetic screen identifies the major facilitator domain containing 2A (MFSD2A) transporter as a key mediator in the response to tunicamycin. *Proc. Natl. Acad. Sci. U. S. A.* 108, 11756–11765

37 Huang, Z. *et al.* (2013) A functional variomics tool for discovering drug-resistance genes and drug targets. *Cell Rep.* 3, 577–585

38 Bill, A. *et al.* (2015) Variomics screen identifies the re-entrant loop of the calcium-activated chloride channel ANO1 that facilitates channel activation. *J. Biol. Chem.* 290, 889–903

39 Arnoldo, A. *et al.* (2014) A genome scale overexpression screen to reveal drug activity in human cells. *Genome Med.* 6, 32

40 Palchaudhuri, R. and Hergenrother, P.J. (2011) Transcript profiling and RNA interference as tools to identify small molecule mechanisms and therapeutic potential. *ACS Chem. Biol.* 6, 21–33

41 Matheny, C.J. *et al.* (2013) Next-generation NAMPT inhibitors identified by sequential high-throughput phenotypic chemical and functional genomic screens. *Chem. Biol.* 20, 1352–1363

42 Zhou, J. *et al.* (2013) A kinome screen identifies checkpoint kinase 1 (CHK1) as a sensitizer for RRM1-dependent gemcitabine efficacy. *PLOS ONE* 8, e58091

43 Zhou, J. *et al.* (2010) Modulation of the ribonucleotide reductase-antimetabolite drug interaction in cancer cell lines. *J. Nucleic Acids* 2010, 597098

44 Smurnyy, Y. *et al.* (2014) DNA sequencing and CRISPR-Cas9 gene editing for target validation in mammalian cells. *Nat. Chem. Biol.* 10, 623–625

45 Hughes, T.R. *et al.* (2000) Functional discovery via a compendium of expression profiles. *Cell* 102, 109–126

46 Lamb, J. *et al.* (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929–1935

47 Lai, K. *et al.* (2013) Integrated compound profiling screens identify the mitochondrial electron transport chain as the molecular target of the natural products manassantin, sesquicillin, and arctigenin. *ACS Chem. Biol.* 8, 257–267

48 King, F.J. *et al.* (2009) Pathway reporter assays reveal small molecule mechanisms of action. *J. Assoc. Lab. Autom.* 14, 374–382

49 Kitagawa, M. *et al.* (2010) Metabolomic identification of the target of the filopodia protrusion inhibitor glucopiericidin A. *Chem. Biol.* 17, 989–998

50 Barretina, J. *et al.* (2012) The *Cancer Cell Line Encyclopedia* enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607

51 Jenkins, J.L. *et al.* (2012) Bridging chemical and biological data: implications for pharmaceutical drug discovery. *Comput. Approaches Cheminform. Bioinform.* http://dx.doi.org/10.1002/9781118131411.ch2

52 Lipinski, C.A. *et al.* (2015) Parallel worlds of public and commercial bioactive chemistry data. *J. Med. Chem.* 58, 2068–2076

53 Papadatos, G. and Overington, J.P. (2014) The ChEMBL database: a taster for medicinal chemists. *Future Med. Chem.* 6, 361–364

54 Ratnam, J. *et al.* (2014) The application of the open pharmacological concepts triple store (open PHACTS) to support drug discovery research. *PLOS ONE* 9, e115460

55 Jenkins, J.L. (2012) Large-scale QSAR in target prediction and phenotypic HTS assessment. *Mol. Inform.* 31, 508–514

56 Zhou, Y. *et al.* (2007) Large-scale annotation of small-molecule libraries using public databases. *J. Chem. Inf. Model.* 47, 1386–1394

57 Schenone, M. *et al.* (2013) Target identification and mechanism of action in chemical biology and drug discovery. *Nat. Chem. Biol.* 9, 232–240

58 Wassermann, A.M. *et al.* (2014) A screening pattern recognition method finds new and divergent targets for drugs and natural products. *ACS Chem. Biol.* 9, 1622–1631

59 Plouffe, D. *et al.* (2008) In silico activity profiling reveals the mechanism of action of antimalarials discovered in a high-throughput screen. *Proc. Natl. Acad. Sci. U. S. A.* 105, 9059–9064

60 Keiser, M.J. *et al.* (2007) Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.* 25, 197–206

61 Nidhi, G.M. *et al.* (2006) Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. *J. Chem. Inf. Model.* 46, 1124–1133

62 Ramsundar, B. *et al.* (2015) *Massively multitask networks for drug discovery*. Cornell University Library Available at: http://arxiv.org/abs/1502.02072

63 Martínez-Jiménez, F. *et al.* (2013) Target prediction for an open access set of compounds active against *Mycobacterium tuberculosis*. *PLoS Comput. Biol.* 9, e1003253

64 Chen, B. *et al.* (2012) Assessing drug target association using semantic linked data. *PLoS Comput. Biol.* 8, e1002574

65 Bender, A. *et al.* (2009) Use of ligand based models for protein domains to predict novel molecular targets and applications to triage affinity chromatography data. *J. Proteome Res.* 8, 2575–2585

66 Andersson, C.R. *et al.* (2011) Quantitative chemogenomics: machine-learning models of protein–ligand interaction. *Curr. Top. Med. Chem.* 11, 1978–1993

67 Kufareva, I. *et al.* (2012) Compound activity prediction using models of binding pockets or ligand properties in 3D. *Curr. Top. Med. Chem.* 12, 1869–1882

68 Jaeger, S. *et al.* (2014) Causal network models for predicting compound targets and driving pathways in cancer. *J. Biomol. Screen.* 19, 791–802