Prior to the development of DNA cloning and protein expression technologies, screening for new drugs was based on phenotypic changes induced by candidate drugs in cells, tissues and model organisms. During the 1990s, however, drug discovery was dominated by high-throughput methods applied to specific cloned and expressed molecular targets. Three factors have contributed to the current renaissance in cell-phenotype-based screening for drug candidates: first, the belief that genomics-based target identification and validation needs to be complemented by an unbiased approach based on cell or subcellular function; second, the availability of novel fluorescent cellular probes compatible with the high-throughput instrumentation developed for molecular target screening; and third, the development of technologies for the identification of the drug target and the elucidation of its mechanism of action after a phenotype-based screen [1].

The identification of the molecular target of a drug candidate has several advantages. The most significant advantage is the ability to set up target-based assays and to allow structure-activity relationships (SAR) to guide medicinal chemistry efforts towards lead optimization. Knowing the drug target can also facilitate the identification of potential toxicities or side effects if there is any precedent of toxicities for the identified target.

There are two fundamentally different approaches to target identification following phenotypic screening: first, the direct identification of the target by its binding to the compound of interest, as detected via affinity purification, expression or display cloning, three-hybrid systems or the probing of protein microarrays; and second, indirect approaches in which compound-induced changes in mRNA, protein or metabolite expression profiles are compared with profiles obtained with known target compounds or activity profiles following specific genetic changes. This review will use case studies of target identification efforts and highlight the advantages and disadvantages of the various approaches to target identification after phenotypic screening.
This review will summarize the technologies currently being employed for target identification following phenotype-based screening, and will use recent case studies to highlight both the utility and the limitations of different approaches. Some examples of targets identified after phenotypic screening are listed in Table 1.

**Affinity purification of targets**

The biochemical purification of drug targets from cell or tissue extracts using affinity chromatography was successfully used to identify the targets of drugs including colchicine, cytochalasin B, cyclosporin and FK506 [2,3]. In a phenotype-based screen, Wignall and co-workers tested compounds for their ability to interfere with the *in vitro* assembly of mitotic spindles [4]. One of the most potent compounds was immobilized on agarose beads and used as an affinity resin to isolate specific binding proteins from a cell extract. Two proteins, which were reproducibly enriched, were separated by electrophoresis; the bands were then excised and proteolytically digested, and the resulting peptides were analyzed by MALDI-TOF mass spectrometry. One of the proteins was NQ01, a quinone oxidoreductase that the authors showed to be a component of the microtubule regulatory machinery. In a similar approach, Ki *et al.* [5] used biotinylated radicol to isolate target proteins from a HeLa cell extract. Compound-bound proteins were precipitated with streptavidin beads. After purification, digestion and peptide sequencing, ATP citrate lyase was identified as a radicol target. Shimizu *et al.* [6] have developed an affinity matrix specifically for drug target purification. This latex-based resin features a high surface-to-volume ratio and a hydrophilic spacer arm for compound immobilization, and lacks pores (with the result that target trapping is reduced). They employed this matrix to identify cellular targets of E3330, a novel NFκB pathway compound, and demonstrated that Ref-1, a key effector of redox signaling, is a target [6].

Two fundamental issues must be acknowledged when pursuing a direct identification through affinity chromatography. First, the compound must be immobilized in a manner that maintains its target recognition properties; and second, the extract to be analyzed must contain the target in a form that allows binding of the compound. Three strategies are employed to give confidence that the compound is immobilized in a state that allows target binding: first, SAR data may suggest particular functional groups on the compound that are dispensable for target binding, and can be used for immobilization; second, the compound can be synthesized with the same chemical linker that serves to tether the compound to the affinity matrix and tested for activity; and third, different immobilization sites on the compound can be tried in the hope that at least one is compatible with target binding. To mitigate this problem, a specific linker and functional immobilization group can be included in all the members of a compound library to facilitate immobilization [7].

**Expression library technologies**

The screening of expressed cDNA libraries using the compound as a ‘probe’ to identify targets has been accomplished in a variety of formats. These approaches include library screening of bacterially expressed proteins after transfer to membrane filters, the application of phage or mRNA display technologies, and the intracellular expression of target libraries where a ‘three hybrid’ system is used to detect compound–target interactions. An example of a ‘classic’ filter-based expression library screening approach is the study by Tanaka *et al.* [8] aimed at identifying the target of the anticancer drug HMN-154, which has an unknown mechanism of action. A human cDNA library was induced to express protein and probed with HMN-154 cross-linked to bovine serum albumin (BSA). Protein spots bound by the drug–BSA complex were
detected using anti-BSA antibodies. The advantage of chemically cross-linking the compound tested to BSA is that individual BSA molecules with more than one copy of the drug molecule may then bind polyvalently to the immobilized proteins, which greatly increases the avidity of the probe and allows detection of targets that would escape monovalent probing.

Sche et al. [9] utilized a T7 phage display system to identify the protein target of the small molecule FKS06. Biotinylated FKS06 was used to screen a human cDNA library and phage clones expressing FKBP, a protein target of FKS06, were isolated after only two rounds of affinity enrichment. More recently, Savinov and Austin [10] identified a subunit of ATP synthase as a target for a heterobicyclic compound using phage display, and Jin et al. [11] identified the nucleolar phosphoprotein hNopp140 as a target for the cancer drug doxorubicin.

An alternative protein display is the direct in vitro expression of mRNA followed by the covalent linkage of the expressed protein to the mRNA that encoded it [12]. In a proof-of-concept experiment exploring utility for target identification, McPhearson et al. [13] demonstrated that mRNA display can be used to identify the protein target of FKS06. They successfully isolated FKBP from a library of >10¹¹ recombinants.

One of the key advantages of display systems is that they enable iterative affinity enrichment with multiple rounds of selection interspersed with amplification to identify specific clones. In addition, several strategies have been devised to increase the recovery of higher-affinity, possibly more relevant clones. These affinity enrichment schemes include competition with free ligand, monovalent rather than polyvalent display, and reducing the density of the immobilized compound to reduce avidity resulting from polyvalent interactions.

Three-hybrid systems are conceptually derived from the ‘two hybrid’ technology used to identify and characterize protein–protein interactions. Developed by Licitra and Liu [14], the three-hybrid system is based on the assembly of a multimeric complex that drives the expression of a reporter gene. One of the hybrid proteins consists of a specific DNA binding domain fused to a specific ligand binding domain, the second hybrid protein consists of a specific ligand binding domain fused to a transcriptional activation domain, while the hybrid small molecule, a ‘chemical dimerizer’, consists of two ligands, one for each of the two ligand binding domains, linked by a tether. The successful assembly of a trimeric complex consisting of the two hybrid proteins linked together by the hybrid small molecule drives the expression of a reporter gene. For target identification purposes, the compound of interest is used as one-half of the hybrid small molecule and a cDNA-based library of proteins is expressed for the corresponding collection of hybrid proteins. In a proof-of-concept experiment, Becker et al. [15] employed the three-hybrid system for target discovery using specific kinase inhibitors. In addition to detecting the known kinase targets for the compounds tested, they also identified several kinases not previously recognized as targets of the compounds. Much effort is underway to optimize three-hybrid approaches for target identification, including the development of mammalian cell three hybrid (M3H) systems [16,17]. The technical caveats that apply to other ‘direct’ target identification approaches also hold for the ‘three hybrid’ approach. The compound must be tethered to the hybrid in such a manner that binding to its target is not

---

### TABLE 1

<table>
<thead>
<tr>
<th>Phenotypic screen</th>
<th>Compound</th>
<th>Target</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immunosuppression</td>
<td>FKS06</td>
<td>FKBP (cis-trans peptidyl-prolyl isomerase)</td>
<td>[42]</td>
</tr>
<tr>
<td>Cell cycle arrest</td>
<td>Trichostatin A</td>
<td>Histone deacetylase</td>
<td>[43]</td>
</tr>
<tr>
<td>Reversion of transformed cell phenotype</td>
<td>Geldanamycin</td>
<td>HSP90 (chaperone protein)</td>
<td>[44]</td>
</tr>
<tr>
<td>Angiogenesis inhibition</td>
<td>Fumagillin</td>
<td>MetAP-2 (methionine aminopeptidase)</td>
<td>[45]</td>
</tr>
<tr>
<td>Spicy (hot) pain</td>
<td>Capsaicin</td>
<td>Capsaicin receptor (G-protein-coupled receptor)</td>
<td>[46]</td>
</tr>
<tr>
<td>Tumor growth and angiogenesis inhibition</td>
<td>Dihydroepiandrosterone</td>
<td>LMP2, LMP7 (proteosome subunits)</td>
<td>[47]</td>
</tr>
<tr>
<td>Mitotic arrest</td>
<td>Monastrol</td>
<td>Eg5 (motor kinesin)</td>
<td>[48]</td>
</tr>
<tr>
<td>Cell cycle arrest</td>
<td>Purvalanol B</td>
<td>CDKs, CK1 (kinases)</td>
<td>[49]</td>
</tr>
<tr>
<td>Reversal of transformed cell phenotype</td>
<td>Radicicol</td>
<td>ATP citrate lyase</td>
<td>[50]</td>
</tr>
<tr>
<td>NF-κB inhibition</td>
<td>E3330</td>
<td>Ref-1 (redox-related factor)</td>
<td>[51]</td>
</tr>
<tr>
<td>PIP2-induced actin assembly</td>
<td>187–1</td>
<td>N-WASP (N-Wiskott Aldrich syndrome protein)</td>
<td>[52]</td>
</tr>
<tr>
<td>Anti-tumor activity</td>
<td>Doxorubicin</td>
<td>hNopp140 (nucleolar phosphoprotein)</td>
<td>[53]</td>
</tr>
<tr>
<td>Tumor cell growth inhibition</td>
<td>LAF389</td>
<td>MetAp2 (methionine aminopeptidase)</td>
<td>[54]</td>
</tr>
<tr>
<td>Anti-lipolysis</td>
<td>Niacin (nicotinic acid)</td>
<td>HM74/PUMA-G (G-protein-coupled receptor)</td>
<td>[55]</td>
</tr>
<tr>
<td>Induction of stem cell neurogenesis</td>
<td>TWS119</td>
<td>GSK-3β (glycogen synthase kinase 3β)</td>
<td>[56]</td>
</tr>
<tr>
<td>AP-1 inhibition</td>
<td>PNRI-299</td>
<td>Ref-1 (redox-related factor)</td>
<td>[57]</td>
</tr>
<tr>
<td>In vitro mitotic spindle assembly</td>
<td>NG72 (diminulol)</td>
<td>NQ01 (NADP-dependent oxidoreductase)</td>
<td>[58]</td>
</tr>
</tbody>
</table>
blocked. Likewise the library of target proteins must be expressed in a form capable of binding to the small molecule. The affinity of the compound for its target has to be sufficient to allow the successful assembly of the trimeric complex and to promote enough reporter gene expression to be detectable above background expression levels.

**Protein microarrays**

A different format for the direct identification of protein targets of compounds identified in phenotypic screens is the protein chip or protein microarray. Protein microarrays have been constructed by spotting purified proteins onto chemically derivatized glass or with immobilizing antibodies. Cell-based protein expression arrays have also been assembled [18]. For the purpose of target identification, the labeled compound is incubated with the array and the unbound material washed off, and the array is imaged to identify the locations of specific interactions.

MacBeath and Schreiber [19] demonstrated the successful probing of a protein array with small compounds as a proof of concept for the utility of protein microarrays in target identification. Three known protein–small-molecule pairs (streptavidin–biotin; anti-digoxigen-antibody–digoxigen; and FKBP12–AP1497) were employed in the test. The small molecules were coupled to BSA derivatized with a fluorophore, and the predicted binding pattern was obtained. Again, the use of a BSA conjugate may have facilitated the detection of low-affinity interactions by enabling a polyvalent mode of binding leading to high avidity. Consistent with this, they found that three different ligands for FKBP12 with dissociation constants ranging from nM to µM all yielded equivalent signals. Strategies to differentiate compound–target interactions of different affinities may be important to avoid low-affinity targets not relevant to the compound’s mechanism of action.

A unique cell-based approach to the creation of protein microarrays for target identification was developed by Ziauddin and Sabatini [20]. Plasmid-based expression vectors containing different cDNAs were spotted onto glass slides and overlaid with a transfection reagent and a layer of mammalian cells. Cells on each spot became transfected, leading to an array of cell clusters, each overexpressing a different protein. To demonstrate the utility of this platform for target identification, radiolabeled FK506 was used as a probe of an array. Cells that expressed the target protein FKBP12 were specifically labeled. They also tested their system with a radiolabeled dopamine receptor ligand and, in addition to the labeling of sites expressing the dopamine receptor, phenotypic changes were observed in other cell clusters. These clusters possibly express a protein involved in the activity of the tested compound, illustrating the utility of this approach to identify other proteins involved in the compound’s mechanism of action in addition to the target [20]. The caveats with this approach to target identification after phenotypic screening are similar to those associated with affinity chromatography: the labeling of the compound cannot block its ability to bind to the target, and the target must be present in a form capable of compound binding.

**Systems biology methodologies for target identification**

‘Systems biology’ methodologies are quantitative high-throughput tools for the genome-wide molecular and structural characterization of cells. These technologies include DNA microarrays for measuring specific mRNA levels; electrophoretic or chromatographic approaches for measuring specific protein or metabolite levels; and multi-parameter imaging or microscopy. There are three different ways that systems biology methodologies can be applied to the challenge of target identification following phenotypic screening. First, a detailed description of specific mRNA, protein or metabolite changes that occur after treatment with the compound of interest, coupled with knowledge of pathways and mechanisms, may lead to a set of candidate target molecules. These candidate targets can then be directly tested. Second, the global pattern of measured changes serves as a ‘signature’ or ‘fingerprint’ of the specific compound-induced changes. This particular pattern can then be compared to patterns obtained when analyzing compounds with known targets or with known mechanisms of action. A similar pattern could indicate that they function through the same or a similar target, which can then be tested directly. Third, the signature or fingerprint of changes induced by the compound of interest can be compared with changes obtained through the systematic modulation of specific targets by genetic approaches such as gene deletion or RNA interference. A similar pattern leads to the hypothesis that this protein may be the target for the compound, which can then be tested directly [21,22].

**Differential gene expression signatures**

The advent of high-density DNA microarrays has allowed compound effects on cell phenotype at the level of RNA expression patterns to be characterized in detail. A pioneering study by Scherf et al. [23] used DNA microarrays to analyze the expression of ~8000 genes in NCI’s set of sixty human cancer cell lines (http://dtp.nci.nih.gov/). They then correlated the expression data with the responses of each of these lines to >1400 different compounds; compounds with similar mechanisms of action clustered together and a compound with an unknown mechanism of action could be linked with a particular cluster. The creation of databases of drug signatures of this type is being pursued by several biotechnology companies [24]. Gould-Rothberg et al. [25] characterized the effects of a PPARα agonist on rat liver gene expression and correlated the observed changes with specific metabolic pathways. This data could serve as a reference for compounds acting along similar pathways. Haggerty et al. [26] took compound activity genomic...
profiling one step further by exploring compound-induced changes in the context of different specific genotypes. In addition to analytical and statistical tools of hierarchical clustering and principal component analysis, they describe the use of graph-theory-based descriptors, which they compiled into topological fingerprints. These fingerprints allow one to discern subtle differences in compound-induced activities as a function of the specific genotypic background.

**Proteomics**

Similar to the characterization of patterns of mRNA expression, patterns of protein expression are also being pursued as an approach to target identification. Proteomics-based approaches can be employed where the action of a drug has no impact on RNA levels. Towbin et al. [27] utilized two dimensional gel electrophoresis to identify proteins that showed altered expression levels or altered mobilities after treatment with compounds that cause tumor cell growth arrest. One of the proteins that showed an altered mobility, 14–3-3γ, contained an additional N-terminal methionine. Detailed pharmacologic assessment confirmed the authors’ hypothesis of a direct interaction of the compound with both known methionine aminopeptidases (Type I and II). Proteomics profiling can also be focused on proteins with specific functions and used to compare with the compound of interest’s profile. Plavec et al. [28] measured the levels of proteins involved in inflammation and immune function after treatment of immune cells with both cytokines and after the specific up- or down-regulation of specific immunomodulatory genes. Kunkel et al. [29] extended this work by constructing functional proteomic profiles after a variety of known drug treatments. The resulting profiles can then be compared to those obtained following treatment with the compound of unknown mechanism. Similarities in the responses lead to hypotheses of the specific pathways modulated and the targets that may be involved.

**Metabolite profiling**

An emerging area in systems biology applications and technology development is the broad-based, systematic analysis of the small molecules of the cell [30]. Quantification of the building blocks of catabolism, anabolism and secondary metabolism can be unbiased or alternatively can focus on a particular subclass of metabolites. Watkins et al. [31] described the lipid profile of mouse hepatocytes after treatment with a PPARγ agonist. Lipid profiles or signatures can be obtained after treatment with a broad panel of drugs and compared with the pattern obtained with the compound of interest. As described above for genomics and proteomics technologies, metabolic profiling (metabolomics) can either lead to hypothesis-based testing of specific candidate targets based on the metabolites in flux, or can yield databases of drug or genetic perturbation profiles for comparative analysis.

**Multiparameter cell imaging**

The suite of instrumentation, detection reagents and software that has been developed for multiparameter imaging-based phenotypic screening (also known as high content screening, high throughput microscopy or automated imaging) is proving to have applications for target identification after phenotype screening: characterization of the specific subcellular changes induced by the compounds can lead to specific hypotheses about the targets which may mediate such changes, while a second application is the creation of databases of subcellular morphological changes after known drug or gene knockout treatments. Perlman et al. have demonstrated the utility of high throughput microscopy for target identification in a proof-of-principle study in which they characterized the morphological effects of 100 different compounds on HeLa cells. In general, they found that drugs with common targets yielded similar profiles, whereas broader pharmacological groupings (e.g. protein synthesis inhibitors) yielded diverse profiles. In their analysis they also included three drugs with unknown targets, and were able to functionally match the profile of one of them with that of gene expression inhibitors [32]. Kiger et al. screened a library of gene-specific RNAi constructs for effects on cell morphology using automated fluorescence microscopy. Such genetics-based morphological databases could also be used as a reference for drug target identification, enabling comparison of compound-induced changes to those obtained after specific genes are modulated [33]. One of the key challenges for creating databases of drug ‘fingerprints’ based on cell imaging is the effective handling of data that is not one-dimensional, like a DNA sequence, but rather two-, three- or even four-dimensional (i.e. involving a time variable) in complexity [34].

**Organismal model systems**

If the target of a bioactive compound modulates a highly conserved cellular activity, for example a function associated with the cell cycle, it is possible that the target function will be conserved in a model organism like yeast, in which case the power of yeast genetics can be used for target identification. ‘Forward’ genetics using the yeast system was one of the earliest approaches applied to target identification. ‘Forward’ genetics using the yeast system was one of the earliest approaches applied to target identification. The yeast genome was randomly mutagenised to generate a set of mutant strains, which were then screened for drug resistance. Mutant alleles responsible for drug resistance can subsequently be identified by genetic complementation. The target of rapamycin was identified using this approach [35]. ‘Reverse’ genetics makes use of the systematic production of individual mutant strains corresponding to most or all genes being individually inactivated. The completion of the sequencing of the genomes of model organisms including yeast, Drosophila, C. elegans and zebrafish and the production of extensive mutant collections has enabled a powerful approach to target identification. Using this approach, Parsons et al. [36] performed a proof-of-concept study identifying the targets of twelve...
diverse bioactive compounds. All ~5000 viable haploid yeast deletion strains were tested for drug sensitivity by analyzing the size of the individual colony formed on agar in the presence of the compound. In addition to candidate targets of the compounds being identified, several genes were identified that were associated with sensitivity to several of the compounds, which are candidates for multidrug resistance mechanisms. Giaever et al. used an approach where gene dosage is halved through the creation of heterozygous deletion strains of yeast [37]. Pooled cultures of all strains covering the whole genome were probed for drug resistance or increased drug sensitivity. As each strain has been ‘bar-coded’ with a unique identifier, the relative growth rates of each strain can be quantitated with a custom DNA microarray. Using this method, the authors successfully identified targets of ten diverse bioactive compounds. In addition to isolating known targets, other loci were identified that corresponded to novel targets, as well as to compound uptake mechanisms [38]. Lum et al. [39] recently used this approach to analyze 78 compounds. In addition to target identification, other proteins involved in the compound’s mechanism of action could also be uncovered. For example, Lum et al. discovered that the tricyclic antidepressants they profiled affected the growth of a yeast strain deficient in a P-type ATPase. Thus, in addition to the neurotransmitter re-uptake mode of action already known to underlie the therapeutic action of these compounds, an activity on the human homolog of this ATPase could be the basis of certain aspects of their efficacy, or even their side effects [39]. For both of these yeast systems, the compilation of mRNA gene expression profiles corresponding to the knockout of individual genes can be compared to the mRNA gene expression profiles obtained after compound treatment. Similarities in profiles between compound-treated cells and specific deletion mutants will help identify candidate targets of the compound, which can then be tested directly. Marton et al. [40] demonstrated the feasibility of this approach in the yeast system by finding a high degree of similarity between the mRNA gene expression changes induced by treatment with FK506 and those observed after deletion of the calcineurin gene.

An alternative approach to that of genetic deletions involves knocking down the expression of specific genes through RNA interference or antisense approaches. Gene silencing by RNA interference, mediated either by small interfering RNAs (siRNAs) or short hairpin RNAs (shRNAs) has been applied globally to systematically target many or even all the genes in model systems [41]. Drug resistance or hypersensitivity in particular strains where a protein target is knocked-down implicates that protein as a possible target or suggests that it is involved in signaling or bioavailability.

**Confirmation of the identified target**

Regardless of the methodology used to identify a candidate target, the next question is whether this drug–target interaction is the compound’s mechanism of action. There are several reasons why target confirmation is an important part of the target identification process: first, non-specific interactions can lead to false positives in direct target binding approaches, such as affinity purification or expression cloning; second, the presence of related proteins can result in the isolation of a homolog of the true target; and third, shared activity profiles between the compound and other compounds or specific genetic changes can result from action on a different target.

This ‘target confirmation’ is the converse of ‘target validation’ where the goal is to determine whether a particular target, if pharmacologically modulated, will yield the desired phenotypic (clinical) change. Here you have a compound that is conferring the phenotypic effect desired, and you need to confirm that this change is in fact mediated by the identified target. In addition, binding to multiple targets may underlie the functional activity originally used to identify the compound in the screen. It is therefore critical to identify all molecular interactions that the compound of interest can undertake and then to determine their respective roles and their relevance to both the phenotype they were identified by and their eventual clinical application.

A variety of approaches can be used to support the hypothesis that the identified protein is the *bona fide* target of the bioactive compound, including pharmacological experiments demonstrating potency of the compound on the isolated target comparable with the doses necessary to confer the phenotypic change and the analysis of spatial and temporal expression patterns of the target in the cells relevant to the phenotype. ‘Reverse genetic’ technologies using gene knockouts (e.g. through homologous recombination), gene knock-downs (e.g. through RNA interference) or overexpression-based dominant negatives lend strong support to the confirmation of the target if the phenotypic changes obtained were similar to the changes induced by compound treatment. These latter approaches may be most relevant when the mode of compound action is inhibition (in the case of an enzyme target) or antagonism (in the case of a receptor target). Prudence must be employed in interpreting differences in results, given the differences in how the targets are modulated (pharmacological versus genetic). This is especially true for compounds acting as agonists (whether full, partial or selective) rather than as antagonists or inhibitors. Support for the veracity of the target would also come from the subsequent screening of the target for distinct ligands that yield the same phenotype.

**Conclusions**

We have entered an era when the challenge of finding new medicines with superior therapeutic properties and fewer side effects has galvanized worldwide efforts linking chemical diversity with biological read-outs predictive of therapeutic efficacy. Phenotype-based screens have emerged as a key strategy for the identification of novel efficacious
drug candidates. However, the subsequent preclinical and clinical development of these candidates requires knowledge of the target or targets underlying the mechanisms of the compound’s action. A host of approaches and technological solutions are being developed and optimized for target identification and confirmation after phenotypic screening. Advances in affinity chromatography, expression cloning and protein microarray technologies are allowing identification of targets on the basis of direct binding. Likewise, the creation of databases of gene expression patterns or profiles of specific molecules or substrates of the cell after known drug treatments or specific gene deletions is allowing candidate targets to be identified on the basis of similarity of activity. The advances being made in these technologies and the success of the early proof-of-concept experiments summarized here suggest that we will in future have increasing success in finding the target after screening the phenotype.

References