Review

For reprint orders, please contact reprints@future-science.com

Future Medicinal Chemistry

Next-generation phenotypic screening

Phenotypic drug discovery (PDD) strategies are defined by screening and selection of hit or lead compounds based on quantifiable phenotypic endpoints without prior knowledge of the drug target. We outline the challenges associated with traditional phenotypic screening strategies and propose solutions and new opportunities to be gained by adopting modern PDD technologies. We highlight both historical and recent examples of approved drugs and new drug candidates discovered by modern phenotypic screening. Finally, we offer a prospective view of a new era of PDD underpinned by a wealth of technology advances in the areas of *in vitro* model development, high-content imaging and image informatics, mechanism-of-action profiling and target deconvolution.

Scott J Warchal[1], Asier Unciti-Broceta[1] & Neil O Carragher*,[1]

[1]Cancer Research UK Edinburgh Centre, Institute of Genetics & Molecular Medicine, University of Edinburgh, Crewe Road South, Edinburgh EH4 2XR, UK

*Author for correspondence:
Tel.: +44 0131 651 8671
n.carragher@ed.ac.uk

First draft submitted: 29 January 2016; Accepted for publication: 4 April 2016; Published online: 30 June 2016

Keywords: • high content • image informatics • mechanism of action • phenotypic • screening • target deconvolution • transcriptomics • pathway profiling

Background

During the past 5 years, the drug discovery field has witnessed a re-emerging interest in phenotypic drug discovery (PDD) strategies and increased research activity in phenotypic assay development and screening. PDD describes the screening and selection of hit or lead compounds based on quantifiable phenotypic endpoints from cell-based assays or model organisms without prior knowledge of the drug target. The renewed interest in phenotypic screening may be attributed to several factors including: the demand to identify high-value novel drug targets to feed contemporary target-directed drug discovery (TDD) capabilities and commercial drug discovery pipelines; high attrition rates in late stage clinical development and an overall decrease in pharmaceutical R&D productivity, while not directly attributed to limitations of TDD, nevertheless, correlate with the widespread adoption of the TDD operating model in favor of PDD strategies [1–5]; significant duplication of effort and focus upon a relatively small number of well-characterized targets across industrial and academic drug discovery groups; urgent unmet medical need in complex human conditions such as heterogeneous solid cancers and neurodegeneration, where target biology is poorly understood; recent retrospective analysis of all drugs approved by the US FDA since 1999 indicating significant success rates in development of novel, first-in-class, small-molecule drugs by PDD approaches [6–8].

While the three recent retrospective studies of drug approval rates present discrepancies in the number of drug approvals attributed to PDD and TDD strategies, primarily because of differences in terminology, disease area focus and period of analysis, all three studies demonstrate that PDD approaches are providing a significant contribution to clinical approval rates of first-in-class drugs [6–8]. This recent clinical success of PDD is con-

FUTURE SCIENCE part of fsg

sidered by many as remarkable given the relatively low investments in PDD in comparison to TDD by translational funding bodies in academia and industry over the past three decades. PDD, however, does not represent a new drug discovery strategy and was indeed the preferred drug discovery approach prior to increased understanding of human disease at the genetic level and the emergence of molecular biology techniques, which advanced elegant molecular pharmacology studies and high-throughput screening of specific targets [2]. Traditional PDD approaches have utilized a variety of biological model systems such as *in vivo* physiological and behavioral models, *ex vivo* tissue-based assays and basic *in vitro* cellular assays to guide drug development. While many drugs successfully used in the clinic today were discovered using such early PDD approaches, traditional PDD methods were laborious and did not provide ample mechanistic information and thus tend to favor the discovery of less selective agents including cytotoxics rather than novel classes of targeted therapies. Given the duration of time between early-stage drug discovery and clinical approval, many of the example first-in-class medicines attributed to PDD described in the three recent retrospective articles [6–8] are somewhat historical. They utilize, by modern standards, rudimentary phenotypic assays, and thus such retrospective analysis may indeed underestimate the true value of modern phenotypic screening strategies, with regard to identifying novel targets and translation into clinical success.

A significant driving force behind the resurgence of PDD may be attributed to substantial technology developments across several inter-related areas, which advance the PDD paradigm. Such advances include more sophisticated cell-based and small model organism-based automated screening platforms [9–11]. Advances in the development of more complex and more disease-relevant phenotypic assays incorporating: multicellular co-cultures, 3D models, patient-derived primary and Induced Pluripotent Stem Cell (IPSc) models, including gene-edited and isogenic controls, which recapitulate key disease driver mutations, are all well-placed to advance the molecular and pathophysiological relevance of phenotypic screening assays. Improvements in cell-based assay technologies are further complemented by advances in target deconvolution strategies including affinity mass spectrometry, cellular thermal shift assays and cDNA expression microarray technologies among others (Box 1) [12–14]. Also the development of new methodologies, which enable profiling drug mechanism-of-action (MOA) in complex biological samples at genomic, proteomic and phenotypic levels at scale [15–17], supports informed mechanistic classification and triaging of phenotypic hits to assist further target deconvolution or progress preclinical development of phenotypic hits in the absence of target knowledge.

In this article, we attempt to address some common misconceptions and challenges associated with phenotypic screening. We highlight both historical and recent success stories of approved drugs and new drug candidates discovered by PDD. We describe some of the challenges and pitfalls of poorly designed phenotypic screening and target deconvolution strategies and how these may be resolved by the application of new technologies. We place specific emphasis upon the evolution of new gene transcription, pathway profiling and multiparametric high-content screening technologies, which support more advanced phenotypic screening and MOA studies. We provide specific examples and discuss the advantages and limitations of each new approach. Finally, we conclude by discussing how the

Box 1. Target deconvolution methods

Chemical proteomics
- Affinity chromatography and mass spectrometry [12,18–20]
- Quantitative proteomic and silac labeling [21]
- Thermostability shift assays: *in vitro* and in cells [13,22]

Expression cloning
- Phage display [23]
- Yeast three-hybrid assays [24,25]
- cDNA cell microarray [26]

Genetic-based screens
- Yeast deletion collections [27]
- Haploinsufficiency profiling [28]
- Resistance screens combined with Next-Generation Sequencing (NGS) profiling of resistant clones [29]
- Small model organism knock-out (KO) and genetic mutant collections (under development in Zebrafish, *Caenorhabditis elegans* and Drosophila)
- Modifier screens: si/shRNA or CRISPR-Cas9 library screens to identify modulators of small-molecule activity [30,31]
- Activity-based protein profiling [32]

This is a nonexhaustive list of target deconvolution methods and reflects some of the most common approaches selected by the authors.

combination of new technology developments such as more advanced primary and induced pluripotent stem cell (IPSc) culture techniques, gene editing, high-throughput gene transcription, pathway profiling and multiparametric high-content screening technologies are well-placed to advance phenotypic screening toward increased success across multiple disease areas.

Historical examples of drugs discovered by phenotypic screening

Comprehensive discussion of approved drugs originating from PDD strategies have been reviewed previously [6–8,33]. In this article, we highlight specific examples of PDD drugs currently used in the clinics, which challenge conservative views on the necessity for target deconvolution to progress candidate drugs through clinical development. We further describe new lead compounds and candidate drugs discovered by more modern phenotypic screening strategies, which guide chemical design toward specific MOA and which can integrate with ligand-based drug design and TDD strategies to develop highly potent and selective lead compounds and drug candidates.

Metformin (Figure 1, structure 1) belongs to the biguanide class of compounds and represents the first-line standard-of-care therapy for Type 2 diabetes by virtue of its confirmed physiological effects upon decreased glucose production by the liver. Approved in Europe in 1957, metformin has been used for decades as a safe and efficacious medicine to manage the morbidity and mortality associated with Type 2 diabetes and represents a core component of new drug combination therapies for diabetes [34,35]. It is, however, only recently that the MOA by which metformin regulates glucose levels has been revealed. In mouse hepatocytes, metformin leads to the accumulation of AMP and related nucleotides, which inhibit adenylate cyclase, reduces levels of cyclic AMP and PKA activity, abrogates phosphorylation of downstream protein targets of PKA and blocks glucagon-dependent glucose output from hepatocytes [36]. These new insights into metformin MOA will most likely pave the way to development of novel antidiabetic drugs.

Further examples of approved drugs derived from compound library screening in phenotypic models were the molecular target of the drug is not known include, daptomycin (Figure 1, structure 2), a naturally occurring antibiotic targeting cell membrane function of Gram-positive bacteria to treat systemic and life-threatening infections [37]. Pemirolast (Figure 1, structure 3) is an antiallergic drug therapy that is proposed to work through suppression of mast cell degranulation, histamine release and eosinophil activation, although precise target mechanism remains to be confirmed [38]. Rufinamide (Figure 1, structure 4) is a triazole derivative used as an anticonvulsant/antiepileptic medication to treat several seizure disorders including Lennox–Gastaut syndrome [39]. The specific molecular target or targets of rufinamide remain to be established.

These examples, described in Figure 1, serve to highlight that if target deconvolution was always a prerequisite for drug development, valuable treatments for such serious human disorders would not have been developed and would not progress the next generation of therapies for many of the most serious and life-threatening conditions.

Sirolimus also known as rapamycin, (Figure 1, structure 5) is a macrolide derived from the bacterium, *Streptomyces hygroscopicus* and discovered through PDD to possess immunosuppressive, antifungal and anticancer properties [40–42]. The National Cancer Institute (NCI) Developmental Therapeutics Program demonstrated that rapamycin inhibited cell growth in panels of tumor cell lines [43]. Subsequent mechanistic studies indicated that the MOA was mediated through inhibition of a serine/threonine protein kinase critical to cell growth, proliferation and survival, the subsequently named mammalian target of rapamycin (mTOR) [44]. Inhibition is mediated through forming a complex between rapamycin bound FKBP12 with mTORC1 [45]. It is worth noting that while rapamycin has been clinically approved, it violates the Lipinski rule of 5 defining optimal lead and drug like properties and thus may never have been developed through a conventional small molecule drug discovery program. The development of rapamycin through PDD and subsequent understanding of mTOR signaling and the target of rapamycin within preclinical and clinical settings provided important target validation data to support the development of several rapamycin analogs known as rapalogs and second generation ATP-competitive mTOR kinase inhibitors targeting mTOR catalytic activity associated with both mTORC1 and mTORC2 complexes [46,47].

Recent examples of phenotypic screening outcomes

We recently reported application of an iterative process consisting of ligand-based design and phenotypic screening of focused chemical libraries to develop novel antiproliferative inhibitors. The strategy employs promiscuous kinase inhibitors as templates to design high-quality small-molecule collections to facilitate the concurrent search for enhanced physicochemical properties and novel pharmacological features. Using this method, target deconvolution of identified hits and leads is largely simplified (for example, focused kinome screening), thereby assisting subsequent lead optimization campaigns [47]. The application of this strategy resulted in the discovery of the first subnanomolar SRC

inhibitor with 1000-fold selectivity over ABL [48] and highly potent dual mTORC1 and mTORC2 inhibitors (eCF309 – Figure 2, structure 6) with high selectivity over other family kinases [47]. A further example of a highly selective kinase inhibitor derived from a phenotypic screen is the allosteric inhibitor of MEK, Trametinib (Figure 2, structure 7), which was initially identified by screening for increased mRNA expression of the cyclin-dependent kinase inhibitor p15 and cell proliferation [49].

Modern high-content phenotypic screening assays, which quantify specific functional endpoints, can also be used to identify compounds with precise target MOA such as the identification of novel Eg5 kinesin inhibitors which induce the monopolar and monaster phenotype [50]. Similar approaches were used to discover and confirm the MOA of second generation Eg5 kinesin inhibitors (AZD4877), which have progressed into clinical development (Figure 2, structure 8) [51,52]. Olesoxime (Figure 2, structure 9) was originally discovered by performing a screen of 40,000 small molecules in an *in vitro* cell-based assay to identify compounds capable of preventing motor neuron cell death in the absence of trophic support [53].

The historical exemplars of approved drugs discovered by PDD chosen (metformin, daptomycin, pemirolast and rufinamide, Figure 1, structures 1–4) highlight that clinically useful and safe medicines can be developed without precise knowledge of the target mechanism. Examples of approved drugs discovered by PDD also serve to highlight that compound structures, which lie out with conventional characteristics of drug likeness can be proved effective in patients (rapamycin, Figure 1, structure 5). Further recent examples of approved drugs discovered by phenotypic screening demonstrate that highly selective targeted therapies can be discovered by phenotypic screening. These examples include drugs targeting ubiquitously expressed regulators of critical cellular functions such as protein synthesis (mTOR inhibitors, eCF309, Figure 2, structure 6), MAPK/ERK signaling (trametinib, Figure 2, structure 7), Eg5 kinesin and mitotic spindle assembly

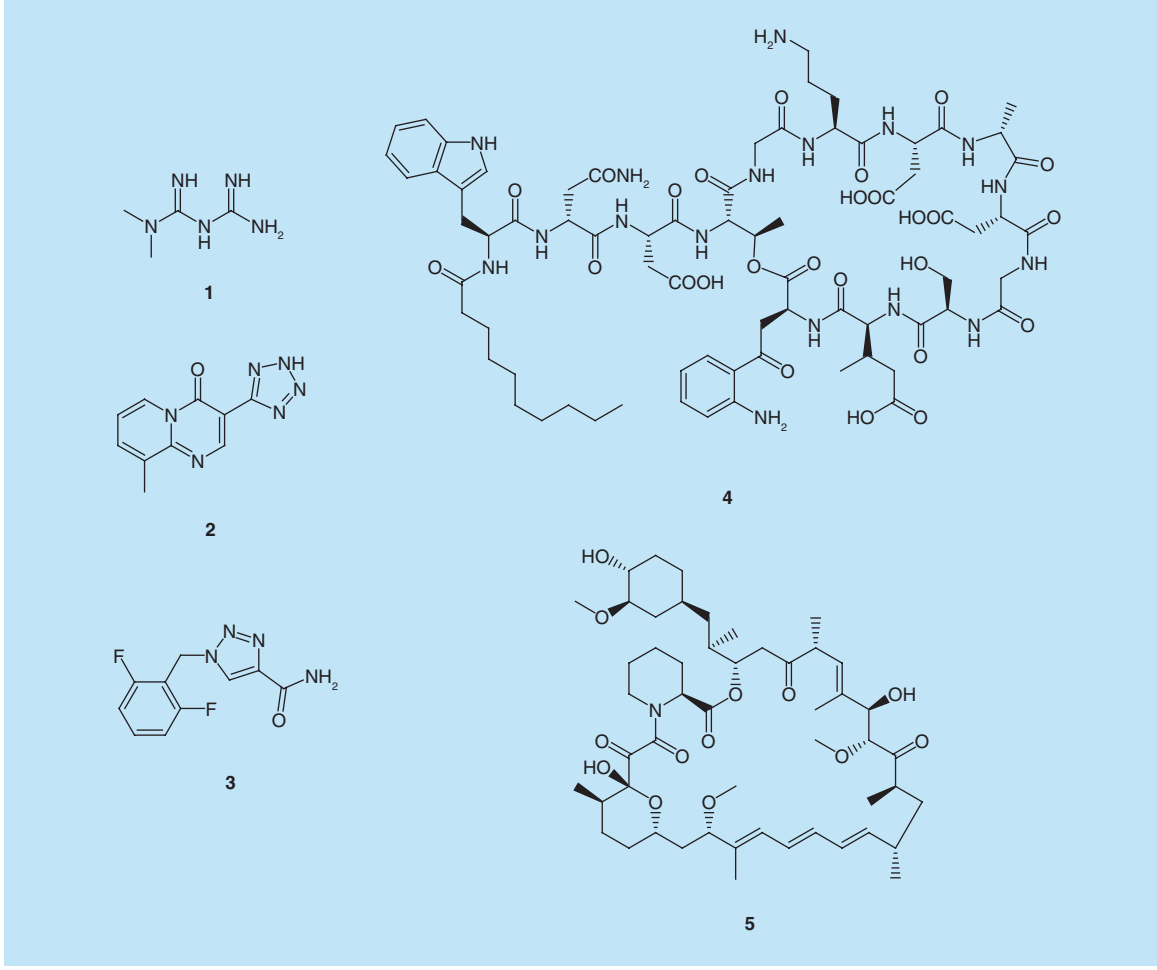

Figure 1. Compound structures of historical examples of drugs discovered by phenotypic drug discovery.
1: Metformin; 2: Daptomycin; 3: Pemirolast; 4: Rufinamide; 5: Rapamycin/sirolimus.

(AZD4877, Figure 2, structure 8) and mitochondrial function (olesoxime, Figure 2, structure 9). Such target classes are unlikely to be prioritized by current drug target review or translational funding committees using conventional target selection criteria to support investment in novel therapeutic targets. Thus, PDD approaches have resulted in the development of many clinically valuable drugs, which would not be developed by TDD programs. The recent examples, which we highlight in Figures 1 and 2 represent a small number of drugs and drug candidates discovered by PDD. For more comprehensive listings of drugs approved by PDD, we direct readers to three recent review articles providing in depth description on the origins of drugs discovered by PDD [6,8,33].

While the overwhelming development of modern targeted therapies has been derived from TDD approaches, these recent examples highlight how advanced phenotypic screening can efficiently direct structure–activity relationships (SAR) and identify novel chemotypes with high potency and selectivity. The above examples further highlight how PDD and TDD approaches complement one another and how new opportunities for combining PDD and TDD strategies are supported by more advanced phenotypic screening, MOA profiling and target deconvolution technologies.

Pitfalls of poorly designed phenotypic screens & black box assays

The phrase 'phenotypic screening' is a broadly used term to describe the extraction of quantifiable readouts of biological relevance from any cell-, tissue- or organism-based system suitable for medium- to high-throughput chemical or functional genomic screening in a target agnostic fashion. Phenotypic screens can range from simplistic 2D cell line viability or reporter assays/pathway screens to more complex multicellular, 3D and multiparametric assays. Phenotypic screening is applied in both the industrial and academic research settings to support functional genomic studies, discover novel candidate drugs and/or useful chemical probes and pharmacological tools for further exploring biology. Thus, analysis and debate on success and challenges of phenotypic screening and target deconvolution strategies must be placed into appropriate context of the value and information provided by the primary phenotypic screen. Traditional single endpoint cell viability and reporter-based cellular assays provide limited information of drug MOA and thus limited opportunity to triage and precisely direct further development of phenotypic hits prior to target deconvolution. Such traditional phenotypic assays, which provide limited mechanistic data, so called 'black box' assays may amplify phenotypic screening challenges and common pitfalls such as, preferential selection of cytotoxic compounds, pan-assay interference compounds or PAINS and sharp activity cliffs, which confound SAR studies. Such pitfalls can largely be avoided by development of information-rich phenotypic screening assays such as multicellular co-culture assays to discriminate phenotypic effect between distinct cell types or multiparametric high-content phenotypic profiling assays, which provide more informative insights into cellular pharmacology. Such high-content assays can classify MOA based upon specific cell targeting or by phenotypic fingerprint similarity with compounds of known MOA and target binding [54]. High-content screening in co-culture assays incorporating target and nontarget cell types may help guide hit selection and chemical design away from toxicity toward enhanced efficacy and novel target space within a single primary high-throughput phenotypic screen [55]. While many 'black box' phenotypic assays represented the state-of-the-art at the time of their development and have had many notable successes in supporting the development of novel drugs, including many of the examples described in Figures 1 and 2. In contrast to modern high-content phenotypic assays, 'black box' assays provide limited opportunity to design screens, which guide selection of hits and leads toward increased therapeutic index and novel phenotypic and target space.

Several review articles, editorials and commentators also suggest that phenotypic screening may help reduce high attrition rates observed during late-stage clinical development specifically the high failure rate observed during Phase II clinical trials resulting from lack of efficacy [4,56,57]. However, the ability of a phenotypic screen to reduce attrition from poor efficacy is directly related to the ability of the primary phenotypic screening assays and any secondary phenotypic assays used for hit selection to predict clinical outcomes. For many complex diseases it is unlikely that the primary screen will recapitulate the full complexity of human disease. Thus, phenotypic screening assays must be developed that ask specific clinical questions or recapitulate key segments of disease pathophysiology to inform subsequent decision-making and effectively guide the next stages of preclinical development and validation. This approach is supported by recent advances in cell-based assay methodology and technologies. Examples in the oncology area include techniques for culturing glioma progenitor cells representing the cancer stem cell niche [58], 3D tumor and fibroblast co-culture organotypic assays, which recapitulate the dense fibrosis and poor drug penetration of poorly vascularized tumors [59] and 3D tumor spheroid cultures, which recapitulate the hypoxic and host cell stromal microenvironment

of many tumors [60]. Screening phenotypic hits across suites of such assays raises the bar with regard to early assessment of the clinical relevance of hit and lead compounds, and also informs subsequent preclinical and clinical development strategies. Ongoing advances in IPSc, gene editing and microfluidic technologies support the development of more physiologically relevant assays across disease areas further advancing more robust approaches to prioritizing phenotypic hits.

Challenges in target deconvolution

An emerging simplistic view of phenotypic screening is that it is an effective strategy for identification of new therapeutic targets from physiological-based models to feed TDD. However, as discussed above, it is unlikely that a primary phenotypic screening assay by itself will predict clinical efficacy and it is also unlikely that initial chemical hits from a large chemical library phenotypic screen will have sufficient potency or selectivity to support rapid and robust target deconvolution. Caution should, therefore, be taken to ensure that poorly designed phenotypic screening and target deconvolution strategies do not create expensive new drug discovery bottlenecks in target deconvolution and further investment of significant chemistry resources on poorly validated targets.

As discussed below, target deconvolution is a challenging and expensive endeavor with limited success rates, we therefore propose that the pathway from phenotypic screening to target deconvolution should not directly follow one another. Rather phenotypic hits should be carefully triaged through increasingly more complex and disease relevant secondary phenotypic assays to build further confidence in the translational potential of the phenotypic hit. Further panel screening across *in vitro* toxicity assays, physiologically relevant assay formats and phenotypic profiling against reference compound libraries will help select the most novel and desirable compounds for target deconvolution. Further experimental medicine studies, including transcriptomic and proteomic analysis, across genetically defined cell assays to prioritize biomarker and drug combination strategies support subsequent chemical optimization using specific pathway reporter assays, target deconvolution and preclinical development. This more in-depth biological investigation will then shift the PDD bottleneck from target deconvolution toward increased disease relevance, novelty, safety and hopefully improved efficacy and drug discovery productivity.

Target deconvolution

Following careful triaging and selection of high-value lead compounds identified by PDD, a number of dis-

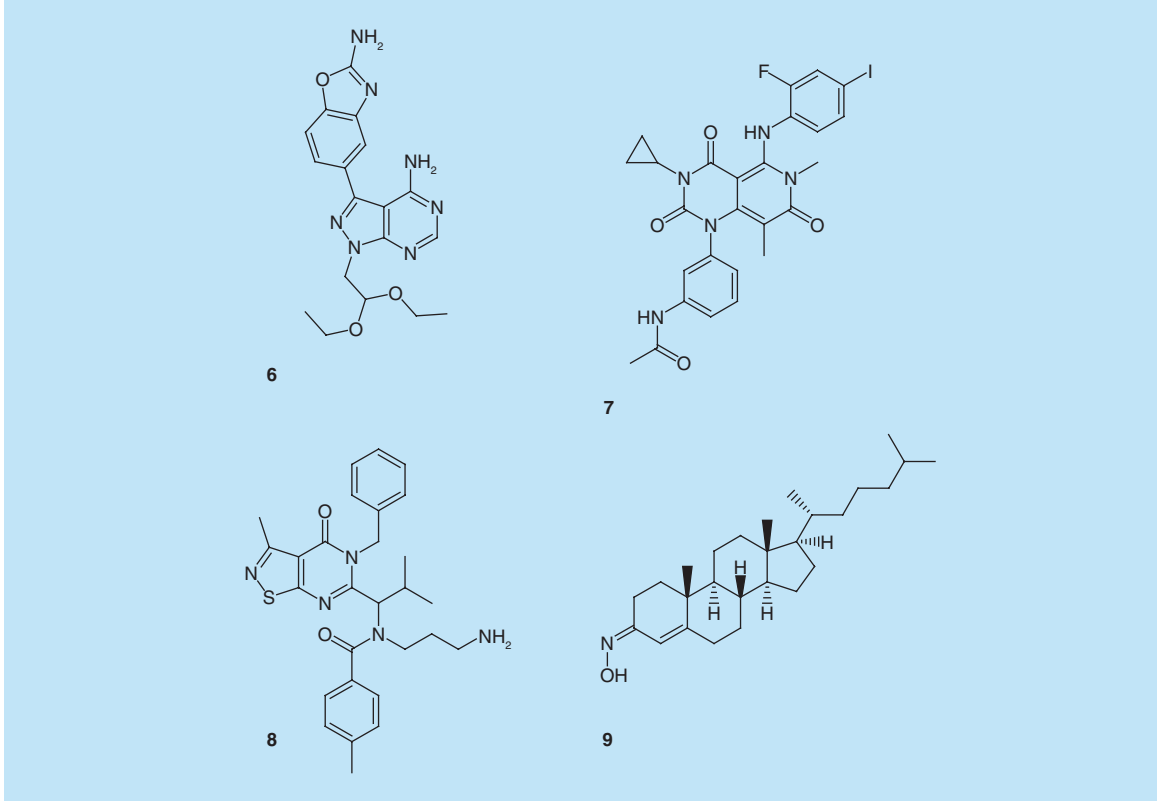

Figure 2. Compound structures from recent examples of phenotypic drug discovery and modern phenotypic screening. 6: eCF309; 7: Trametinib; 8: AZD4877; 9: Olesoxime.

tinct and complementary target deconvolution strategies can be employed (Box 1). Many early target deconvolution studies relied heavily upon affinity-based chemical proteomics approaches which have demonstrated success in identifying targets for a variety of novel inhibitors including hedgehog pathway, bromodomain and N-acetyltransferase inhibitors derived from phenotypic screens [18–20]. However, despite such examples of clear target deconvolution success stories, affinity-based chemical proteomics are often hampered by nonspecific binding of proteins [61–63]. The recent development of publicly available databases characterizing nonspecific protein binding contaminants associated with affinity-based proteomics methods attempts to address the background noise associated with affinity-based proteomics [62]. Competition assays were parent compound is preincubated with cell lysates prior to adding the conjugated affinity capture compound have been developed to determine nonspecific binding to affinity capture reagents and can be combined with databases describing common background contaminant profiles to increase confidence in identifying specific drug-target interactions [61,62]. Such approaches, however, do not completely resolve the issue of nonspecific binding, and affinity-based chemical proteomics is limited to providing lists of potential target binders rather than conclusive evidence of which target is responsible for the phenotypic response, thus further target confirmation studies are required. Several new and complementary target deconvolution strategies are rapidly emerging (Box 1), although no target identification method provides conclusive evidence of which target is responsible for the complete pharmacological profile of a compound. The application of distinct target deconvolution methods (Box 1) combined with other MOA profiling tools may provide strong corroborative evidence to prioritize target hypothesis, which may be responsible for phenotypic response. However, validation of target hypothesis will only be confirmed through further biochemical and cell pharmacology studies. Established and emerging target deconvolution strategies have been reviewed in depth elsewhere [64,65] and so will not be covered in further detail here; however, we do highlight the latest trends in target deconvolution strategies in Box 1.

Mechanism-of-action profiling

A critical success factor in any drug discovery project is the understanding of candidate drug MOA within complex and physiologically relevant biological settings. Several new technologies enable rapid MOA profiling in complex cell models at genetic, proteomic and phenotypic levels at scale. Such higher throughput MOA profiling can facilitate the selection of appropriate phenotypic hits to take forward the further preclinical development, identify new assay endpoints and biomarkers to support early hit-to-lead chemical optimization, provide corroborative evidence for target deconvolution studies and support further preclinical development and translation toward clinical studies with or without conclusive target identification. Recent advances in MOA profiling technologies include: high-throughput gene transcription profiling, pathway profiling at the post-translational level and high-throughput phenotypic imaging and image informatics [15–17,66]. The latest developments and application of these approaches in PDD are described in further detail in the following sections.

High-throughput gene transcription profiling

Gene transcription-based profiling approaches using whole genome expression arrays provide a comprehensive overview of gene activity in biological samples. Common applications of gene expression arrays include genome-wide differential expression studies, disease classification and drug MOA analysis. The concept of using gene transcription profiling to elucidate drug MOA and deconvolve therapeutic targets was first applied by Hughes *et al.* who created a compendium of 300 yeast deletion strains and associated transcription profiles [27]. By correlating similarity of transcription profiles from drug-treated cells with those derived from each individual yeast deletion they identified the C-8 sterol isomerase, ERG2 as the target for the anesthetic Dyclonine [27]. To progress a more systematic comparative bioinformatics analysis of gene expression profiles, the Connectivity Map concept and public repository of transcription profiles was developed [14]. Connectivity Map combines a catalog of gene expression profiles from large panels of compound perturbed samples with computational and statistical methods to support similarity profiling of gene expression patterns to infer compound MOA [14,67]. As a proof-of-concept study, connectivity map gene expression profiling was applied to identify the MOA of the compound, Gedunin, identified as a hit from a screen for androgen receptor inhibitors. Gene transcription profiles of LNCaP prostate cancer cells treated for 6 h with Gedunin were used to query the Connectivity Map database, which identified high similarity with multiple HSP90 inhibitors; subsequent studies further support Gedunin as an inhibitor of HSP90 function [14]. The Connectivity Map approach has proven particularly useful for discovering the MOA of natural products from traditional remedies. A recent example used Connectivity Map to identify the MOA of Berberine, an isoquinoline alkaloid used in traditional Chinese

herbal medicine and which has demonstrated anticancer properties in phenotypic assays [68]. Transcription profiles of HepG2 cells treated with Berberine for 4 h demonstrated similarity with gene expression profiles of the protein synthesis inhibitor cycloheximide as well as several mTOR and HSP90 inhibitors. Subsequent cellular pharmacology studies demonstrated that Berberine inhibits protein synthesis, Akt activity but not mTOR activity and induces AMPK-mediated endoplasmic reticulum stress and autophagy [68]. Therefore, in this case, the initial application of Connectivity Map and gene transcription similarity profiling identified mechanistically similar compounds with known target activities to guide subsequent studies to further elucidate the MOA of Berberine.

Recent technical advances in gene expression profiling include the development of higher throughput and low-cost gene-expression methods such as the L1000™ platform. L1000™ Expression Profiling is based upon the rapid quantification of a reduced number of landmark transcripts in 384-well plate format and a computational model to infer expression across the genome [15]. The L1000™ technology underpins the Library of Integrated Cellular Signatures (LINCS) NIH program, which funds the generation of perturbed gene expression profiles across multiple cell and perturbation types supporting drug MOA profiling at scale [15,69,70]. While gene transcription profiling has proven effective in elucidation of compound MOA, success is dependent upon the use of appropriate biological assays where the relevant target pathway for any given compound is activated. A further dependency is the cross referencing to a comprehensive and well-annotated reference set of compound signatures also generated under appropriate biological context. An alternative approach to inferring MOA from gene expression signatures is the comparison of drug sensitivity/phenotypic response across large panel of cells with their basal gene expression profiles. A recent study used correlation-based analyses to associate the sensitivity of 481 compounds tested across 860 human cancer cell lines with the basal gene expression profile of each cell line [69]. The study included 115 small molecules of unknown mechanism with the aim of identifying novel targets for these compounds; correlation analysis was focused on single-transcript correlation outliers to prioritize potential target hypothesis [69]. Cancer cell sensitivity to the compounds BRD5468 and ML239 correlated with high expression of the monoglyceride lipase MGLL and the fatty acid desaturase FADS2, respectively [69]. Treatment with the MGLL inhibitor, JZL184 or shRNA knockdown of MGLL attenuated the cytotoxicity of BRD5468 and FADS2 knockdown and cotreatment with the selective FADS2 inhibitor SC-26196 reduced ML239 cytotoxicity [69]. These studies demonstrate that correlation of drug sensitivity profiles with basal gene expression patterns across large cell panels can reveal specific target hypothesis. An advantage of correlating transcription profiles across large panels of cells is the ability to distinguish between distinct transcript correlations with drug sensitivity from coregulated transcripts thereby prioritizing the most likely targets. However, limitations of this approach include the prerequisite for compounds that display distinct sensitivity across cell panels, which also display differential gene expression patterns and confounding correlation with mechanisms of metabolism or indirect regulators of compound sensitivity. Indeed the analysis by Rees *et al.* demonstrated that for 57% of the compounds tested, no significant correlation with any target could be detected [69]. Despite the recent advances in transcription-based profiling technologies, the costs associated with such analysis limit high-throughput application to larger compound sets and dose–response and temporal studies. Transcription-based profiling may also only reveal the downstream effects of compound exposure rather than the direct therapeutic targets.

Pathway profiling across panels of primary cell-based assays

Profiling compound response at the post-translational pathway level across panels of primary cells and pathway targets has also demonstrated success in determining drug MOA, confirming selectivity, identifying toxicity liabilities and guiding SAR [71]. For example the BioMAP® – Human Primary Cell Phenotypic Profiling Services provided by DiscoverRX consists of panels of primary human cell-based assay systems, a database of reference compound profiles, and computational data mining and analysis tools to support drug MOA analysis [72]. The comparison of BioMAP profiles from testing of two p38MAPK inhibitors, PD169316 and SB203580 revealed activity features unique to SB203580 including inhibition of VCAM-1, E-selectin, IL-8 and P-selectin expression [71]. To further explore the structural determinants of the unique activities of SB203580, the BioMAP profiles of several well-studied p38MAPK inhibitors and SB203580 analogs were generated for comparison. These studies reveal that many of the unique activities of SB203580 represent secondary off-target activities independent of catalytic activity [71]. The BioMAP approach is applicable to large numbers of compounds tested across dose–response and time–series studies supporting precise SAR studies upon pathway responses. While the assays and core pathways tested represent highly sensitive readouts for multiple biological mechanisms,

the biological space covered will not be appropriate for elucidating the MOA of all molecules.

Reverse phase protein microarray

Reverse phase protein microarray (RPPA) represents a highly sensitive and quantitative high-throughput antibody-based proteomics methodology for measuring abundance of multiple proteins and phospho-proteins across large sample sets [73]. Key applications of RPPA include, dynamic pathway profiling at the post-translational network level following chemical or genetic perturbation, screening modulators of key pathway markers and protein biomarker discovery in clinical and preclinical studies [73–75]. Recent advances in RPPA technology include more sophisticated sample handling, quality control, better quality affinity reagents and optical detection, including planar waveguide detection systems providing femtomole to zeptomole sensitivity in protein analyte detection in formats suitable for medium-throughput applications [76]. The development of ultra-sensitive RPPA facilitates large-scale multiplex analysis of multiple post-translational markers across small samples from *in vitro*, preclinical or clinical biopsies. Thus, RPPA technology is particularly suited to proteomic analysis of miniaturized assay formats of a few thousand cells from an individual well of a microtiter plate and microfluidic devices. Similar to the BioMAP and gene-expression approaches, multiple pathways can be monitored across large sets of assay panels and RPPA profiles compared with reference compound can help predict MOA and triage common/nonnovel pathway inhibitors or highly promiscuous pathway inhibitors with toxic liabilities. Retrospective analysis of esophageal adenocarcinoma patients who were also under treatment with metformin (Figure 1, structure 1) for diabetes demonstrated a better response to chemoradiation therapy compared with patients who were not receiving metformin [77]. However, the MOA of metformin in esophageal cancer was unknown. RPPA analysis applied to esophageal cancer cells treated with metformin revealed inhibition of PI3K/mTOR signaling pathway, which correlated with reduced cell growth and increased apoptosis [78]. In a similar approach, a Danish study comparing recurrence rates for breast cancer between Simvastatin users and nonusers demonstrated a significant reduction in recurrence rates in the statin users [79]. RPPA analysis of triple-negative breast cancer cell lines following Simvastatin treatment demonstrated decreased phosphorylation of FOXO3a. Subsequent knockdown of FOXO3a attenuated the effect of Simvastatin on mammosphere formation and migration [80]. Corilagin has recently been identified as a major active component in a well-known herbal medicine (*Phyllanthus niruri L.*) with antitumor activity although the antitumor mechanism has not been clearly defined. RPPA analysis of a panel of ovarian cancer cell lines treated with Corilagin demonstrated inhibited activation of canonical Smad and non-canonical ERK/AKT pathways, which correlated with inhibition of TGF-β secretion and TFG-β pathway activation [81]. Similar to correlation of drug sensitivity across cell panels with basal gene expression profiles, drug sensitivity across cell panels have also been correlated with basal protein levels and pathway activation states by RPPA to identify both MOA and mechanism-of-resistance [82,83]. Correlation of sensitivity of a panel of small-cell lung cancer lines treated with the PARP inhibitor BMN 673 with RPPA analysis indicated the compound sensitivity is associated with elevated baseline expression levels of several DNA repair proteins [83]. This study identified a novel 'DNA repair score' consisting of a group of 17 DNA repair proteins, which predict sensitivity to BMN 673 [83]. Small-cell lung cancer insensitivity or resistance to BMN 673 correlated with baseline activation of the PI3K/mTOR pathway identifying a potential drug combination hypothesis [83]. While the majority of exemplar studies describing RPPA applications in drug MOA analysis have been applied to late-stage drug candidates or approved drugs, many of which have come from target-directed drug discovery, the success of this approach indicates that it will also be a useful method for uncovering the MOA of hits and lead compounds derived from phenotypic screens. A significant advantage of antibody-based proteomic profiling approaches is that they can help identify translatable pharmacodynamic or predictive biomarker reagents to guide appropriate preclinical proof-of-concept studies and clinical development strategies of drug candidates with or without conclusive target deconvolution.

High-content image-based multiparametric phenotypic profiling

Advances in automated microscopic image acquisition and image analysis tools enable the extraction of functional phenotypic endpoints from complex assay formats including 3D and co-culture models. Integration of high-throughput imaging assays with new image informatics resources enable high-throughput phenotypic profiling and classification of MOA across multiple assays, dose–response and time–series studies. We outline below the development in high-content imaging and image informatics methods and the new opportunities they present to phenotypic screening.

Evolution of high-content imaging & image informatics methods applicable to phenotypic screening

The rapid development of automated microscope platforms has enabled the ability to generate tens of thou-

sands of images a day on a single platform supporting medium- and high-throughput image-based phenotypic screening. With image analysis software capable of extracting several hundred measurements per cell from these images, researchers can detect and quantify subtle phenotypic changes that would otherwise be missed with the naked eye or from a single endpoint assay. These developments have stimulated a new field of biological profiling in cell-based assay systems called high-content analysis [9,84]. However, due to the high-dimensional nature of the high-content datasets, tried-and-tested methods to determine hits and guide SAR developed in TDD are no longer applicable. This means new methods for hit selection and triaging are required, and with the parallel developments in machine learning and other quantitative fields there are many options open to researchers.

Image-based phenotypic measurements can be recorded on two levels: an average of whole-well/-cell population measurements or individual cell measurements. Whole-well measurements are less computationally intensive and easier to obtain and can prove useful when individual cell segmentation is not feasible. Measurements taken from individual cells can be much more detailed, such as individual cell areas or number of organelles per cell. However, individual cell measurements generate large datasets that can become unwieldy and difficult to analyze without significant computing power and data handling pipelines. Therefore, many image-based phenotypic assays use well or population averages of data obtained from individual cell measurements, describing the mean or median cell within each image. While this reduces the amount of data, and allows for more simple analyses, calculating a population average removes any information about heterogeneity or possible phenotypic subpopulations. In instances of two equal sized subpopulations, a well average phenotypic measurement may be a representative of few cells within that image and thus does not accurately record phenotypic response across the cell population. A method to quantify cellular heterogeneity within cell populations has recently been suggested based on three simple statistic procedures: percentage of outliers; the Kolmogorov–Smirnov (KS) test of normality; and quadratic entropy. This method can then be used to classify a cellular population according to the type of heterogeneity observed [85].

The development of such cellular subpopulation analysis methods is important as the origins behind heterogeneity within clonal populations are not well-understood and the diverse response to therapeutics can be a driver underlying clonal selection, a well-known contributor to the evasion of anticancer therapeutics observed in many tumors. New methods calculating heterogeneity and the impact of pharmacological intervention upon heterogeneous cell populations are thus especially relevant to anticipating therapeutic response and monitoring evolution of the disease in response to treatment within complex tumor microenvironments. Several studies have also reported that the expression of specific transcription factors associated with stem cell pluripotency are expressed in a heterogeneous fashion in embryonic stem cell cultures. For example, approximately 80% of embryonic stem cells express Nanog, while 10–20% do not [86]. Stem cell heterogeneity and conversion between distinct pluripotent or differentiated stem cell fates also impact upon therapeutic areas dependent upon endogenous stem cell differentiation and reprogramming such as tissue regeneration and repair. The evolution of image-based methods monitoring cell heterogeneity and classification of subpopulation responses at the single-cell level support the development of more complex and clinically relevant heterogeneous and multicellular models for automated cell-based screening. However, the challenge remains in how to distill such complex multiparametric data to enable key decision-making. Advances in the fields of multivariate statistics and machine learning offer potential solutions.

Development of multivariate high-content methods to predict compound MOA

In 2004, Perlman *et al.* published a landmark paper describing the use of compound 'fingerprints' derived from phenotypic measurements. It was shown that compounds with known similar MOA exhibited similar phenotypic fingerprints [17] and this could be used to predict the MOA of unknown compounds by their similarity to that of known compounds. In order to create the compound fingerprints a modified KS test was developed to compare the distribution of every measurement against the same measurement for the negative control, producing a list of numbers for each compound [87]. These vectors were aligned to other compound vectors in order to maximize correlation to account for differences in potency across ranges of concentrations. The pairwise Euclidean distance was calculated to create a similarity matrix between all the tested compounds; following hierarchical clustering, compounds with similar MOA were found closely aligned to one another. This was the first published demonstration that image-based phenotypic information proved descriptive enough to discern compounds from one another [17]. Further development on multiparametric phenotypic assays combined with different compound profiling methods utilizing multivariate statistics, machine learning and artificial neural networks have steadily evolved [88–92]. In a recent study, 2725

compounds were profiled in a multiparametric high-content assay measuring phenotypic effects upon the nucleus, cytoplasm, endoplasmic reticulum, golgi and cytoskeleton of the U2OS osteosarcoma cell line [54]. The high-content phenotypic fingerprints subsequently generated were used to cluster mechanistically similar compounds using the Markov Clustering Algorithm and then each compound cluster was analyzed for enrichment of individual targets and gene sets to facilitate MOA analysis [54]. Individual target annotations for compounds were obtained from public and commercial drug target databases such as, ChEMBL, Drugbank, GVK (GOSTAR), Integrity and Metabase. Gene set enrichments were obtained from the following databases: Biosystems, Metabase, Integrity, Metabase pathway-derived gene sets (Metabase noodles) and Gene Go Ontologies [54]. Two compounds, 6-[6-(diethylaminopyridin-3-yl]-*N*-[4-(4-morpholinyl)phenyl]-9*H*-purin-2-amine and Silmitasertib clustered with each other and a collection of other compounds inducing similar phenotypic response. In contrast to the majority of compounds in this cluster, which were associated with gene sets enriched in PI3K/Akt/mTOR, the previously described Jak3 inhibitor, 6-[6-(diethylaminopyridin-3-yl]-*N*-[4-(4-morpholinyl)phenyl]-9*H*-purin-2-amine and the Casein kinase II inhibitor, Silmitasertib had not previously been associated with direct inhibition of PI3K/AKT/mTOR pathway targets. Subsequent biochemical analysis revealed 6-[6-(diethylaminopyridin-3-yl]-*N*-[4-(4-morpholinyl)phenyl]-9*H*-purin-2-amine inhibited 3-phophoinositide-dependent protein kinase 1 (PDPK1), a component of PI3K/AKT/mTOR signaling and Silmitasertib inhibited mTOR and PI3K-α with IC50 of 390 nM and 461 nM, respectively [54]. These studies demonstrate that novel compound–target associations can be identified from image-based multiparametric high-content profiling. In contrast to transcription or post-translational pathway profiling methods (BioMAP and RPPA) previously discussed, multiparametric high-content profiling assays can run in high-throughput across arrayed whole genome screens, large chemical libraries and compound profiling studies incorporating dose response and time series if necessary.

Integrating phenotypes & SAR to predict MOA

The study of SAR by the generation and screening of compounds with similar chemical structures, is one of the fundamental methods used by medicinal chemists to determine which structural motifs are required for inducing a biological effect on a particular protein, cell or organism. In principle, compounds with analogous chemical structures often bind to the same or similar protein targets, a principle that is used to develop derivatives with improved drug metabolism and pharmacokinetic (DMPK) properties, and as Perlman *et al.* demonstrated, compounds with similar MOA produce similar phenotypes. Young *et al.* then filled the gap in this reasoning by investigating if compounds with comparable chemical structures produce similar phenotypes [87]. They screened HeLa cells with a small molecule compound library and performed factor analysis on 36 features to produce a fingerprint for each compound, with which a pair-wise similarity matrix was created by the cosine distance between phenotypic fingerprints. In order to determine the similarity between chemical structures, they defined the molecular structure through radial atom neighbors and a structure similarity matrix was constructed through Tanimoto distances between the compounds. The two similarity matrices, one for phenotype and one for chemical structure, were clustered by phenotypic similarity, which revealed distinct phenotypic clusters that matched up to distinct groups of structurally similar compounds [87]. Performing SAR studies in cell-based phenotypic assays is significantly challenged by the fact that effects of compound modulation upon phenotypic activity are multifactorial, influenced not only by target engagement, but also cellular permeability (cLogP/D-mediated), subcellular distribution, cell transport mechanisms, membrane interactions and off-target activities. These issues increase the likelihood of obtaining sharp activity cliffs, which hinder directional SAR studies contributing to more complex and lengthy medicinal chemistry programs. In the study by Young *et al.*, it was found that small changes in chemical structure were associated with large phenotypic differences indicating that sharp chemical activity cliffs are retained in information-rich high-content screening data [87]. Further biological investigation into the distinct multiparametric profiles obtained between chemically similar analogs may reveal the underlying causes of activity cliffs. For example, multiparametric high-content analysis can help diagnose if loss of activity is a consequence of reduced potency on a specific target mechanism, impaired distribution and influence within specific subcellular compartments or completely distinct MOA indicative of new off-target activities. Thus, high-content analysis supports a more in-depth cellular pharmacology approach to guiding subsequent chemical library design from initial phenotypic hits.

High-content imaging quality control/assay standards

To gain meaningful results from any screen, including phenotypic-guided SAR, assay quality control is critical. For many, the lack of reliable metrics to determine assay robustness, such as the z-factor in high-throughput

biochemical screens [93] is a deterrent for widespread adoption of multiparametric high-content methods in industrial drug discovery. Despite criticism of the inappropriate use of the z-factor in high-content studies by many groups, there is still not a universally accepted replacement. Any metric suggested has to address three primary concerns: ability to work with multivariate data; assay independence; and ease of implementation and interpretation. Many attempts to develop such a method have used the z-factor as their basis [94,95], although none have addressed all the issues or gained widespread adoption. The same principles apply for identifying phenotypic endpoints to guide SAR. Feature extraction and selection methods can reduce the data to a single value analogous to an IC_{50} to guide chemical design toward specific areas of phenotypic space. Successful implement of such phenotypic-guided SAR is, however, critically dependent upon reproducible assay formats, appropriate feature extraction and selection methods and deep biological insight to ensure phenotypic features guide chemical design toward desired outcomes. The use of open datasets, such as the Broad Bioimage Benchmark Collection [96,97] enables researchers to compare image analysis and informatics methods on a common collection of annotated images, allowing iterative improvement of methods through collaboration and the replication of results. Further collaborative initiatives to develop common standards for HCS and image analysis methods will promote further adoption and stimulate cross-collaboration between both academic and industrial groups to advance the field of high-content image-based phenotypic screening and profiling.

Conclusion

To date, the majority of phenotypic screening assays that have been implemented in chemical or si/shRNA library screening campaigns and examples of compounds and drugs developed through PDD strategies have used simple biological models and assay readouts. Historical success of PDD combined with acceptance of a significant contribution to recent drug approvals has stimulated renewed interest in PDD strategies. In this article, we describe limitations of traditional PDD approaches and highlight solutions and new opportunities for PDD presented by recent advances in assay development and image-based screening technology. With new advances in precise gene editing technologies such as CRISPR-cas9, primary patient-derived cell culture, IPSc differentiation combined with multiparametric high-content phenotypic profiling, all advance the applications of phenotypic screening under more relevant and well-defined biological contexts. We propose that further development and adoption of new phenotypic assay technologies are well-placed to advance a new era of next-generation phenotypic screening contributing to both PDD and TDD success rates.

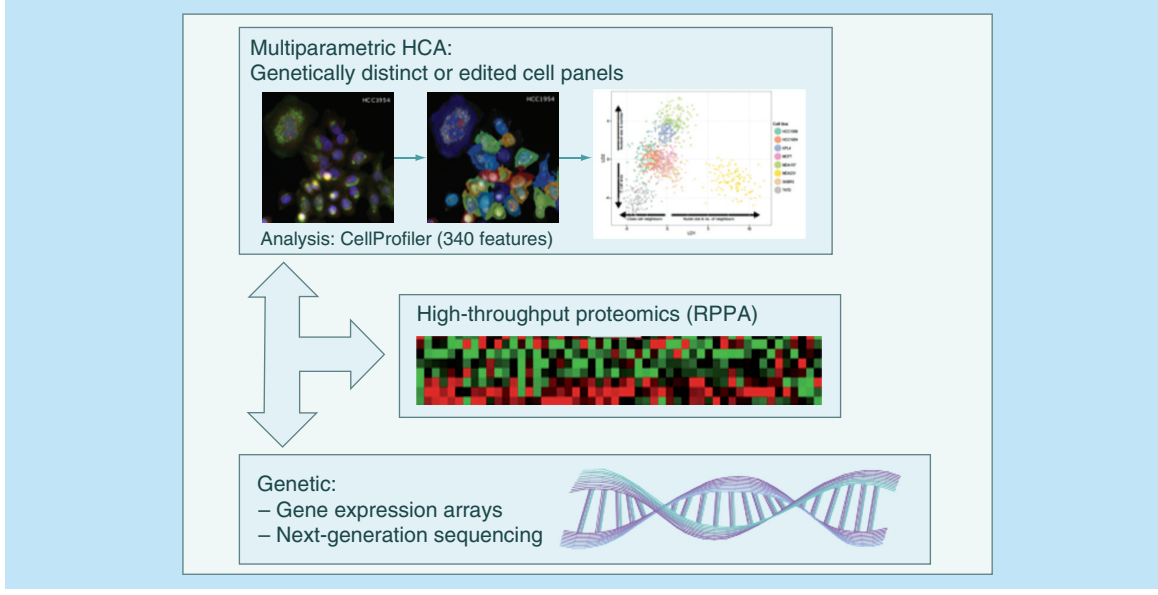

Figure 3. Integration of high-content multiparametric phenotypic profiles with proteomic and genetic datasets.
Representative displays of multiparametric cell morphometry analysis using optimized cell staining and CellProfiler image-analysis protocols followed by correlation of distinct drug-induced phenotypes between cells with molecular data at post-translational and genetic levels. Application of these methods support understanding of drug mechanism of action; identification of resistance mechanisms/pathways to guide biomarker discovery; novel drug combination hypotheses and high-throughput pharmacogenomics incorporating more complex phenotypes across disease areas and across advanced multicellular or 3D models.
RPPA: Reverse phase protein microarray.

Future perspective

Despite over 10 years of research carried out with high-content phenotypic screening, the majority of studies have focused their efforts on a small selection of established cell lines, picked primarily due to amenable cell culture propagation and imaging properties rather than relevance to human disease. The reasons behind this are understandable, as cost and speed represent important criteria in medium- to high-throughput screening projects. Thus, selection of cell lines, which can be rapidly bulked up and accurately and reliably segmented into 2D cell culture assays using readily available image analysis methods, are attractive. An important advantage of image-based high-content screening over other screening platforms is the ability to extract functional endpoints from more complex *in vitro* assays, which extend beyond simple 2D cultures and may include 3D multicellular tissue models and small model organism screens, which exploit more complex biology. The development and adoption of more complex *in vitro* assays may benefit PDD in several ways:

- Application of assays which more accurately represent disease pathophysiology thus contributing to improved translation and clinical success rates;
- Identify novel target space including unbiased identification of novel target classes, which are not currently being pursued by drug discovery groups;
- Identify targets with more relevant functional validation, increasing confidence in target hypothesis to justify subsequent TDD investments;
- Recapitulate intact autocrine, paracrine and juxtacrine pathway signaling networks supporting discovery and development of novel multitargeted therapies and combination approaches.

The primary goal of PDD is to identify small molecules that beneficially modify a disease-associated phenotype, selecting a single cell line to model the disease can, however, prove risky. As demonstrated in cystic fibrosis disease models, there is little overlap between compounds that show efficacy in correcting the CFTR trafficking defect when the mutant CFTR protein is expressed across multiple cell lines [98]. This should lead us to question how well we place our trust in conclusions drawn from an experiment modeled in a single cell line. Application of high-content screening across genetically distinct primary cells or precise CRISPR-cas9 gene-edited cell panels can help elucidation of drug MOA by linking phenotype to genotype and also stimulate biomarker and drug combinations studies (Figure 3) [69,99].

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Executive summary

Background

Historical examples of drugs discovered by phenotypic screening

- Approved drugs with unknown target mechanism.
- Recent examples of modern phenotypic screening outcomes, including development of highly potent and selective targeted agents.
- Pitfalls of poorly designed phenotypic screens/black-box assays.
- Challenges in target deconvolution.
- New approaches in target deconvolution.
- New approaches in mechanism-of-action determination (genomic profiling, proteomics and high-content analysis).

Evolution of high-content imaging & image informatics methods applicable to phenotypic screening

- Early multiparametric high-content methods for compound classification.
- More advanced image analysis and image informatics, including integration of multiparametric phenotypic fingerprints with chemical similarity.
- Current limitations in high-content analysis and high-throughput image informatics.

Future perspective

- New disease models, incorporating 3D assays, induced pluripotent stem cell and gene editing technologies.
- New opportunities for application of phenotypic screening across genetically distinct/gene-edited cell panels linking phenotype to genotype to support high-throughput genomics and personalized healthcare strategies in new disease areas.

References

Papers of special note have been highlighted as:
• of interest; •• of considerable interest

1 Sams-Dodd F. Target-based drug discovery: is something wrong? *Drug Discov. Today* 10(2), 139–147 (2005).

2 Lee JA, Berg EL. Neoclassic drug discovery: the case for lead generation using phenotypic and functional approaches. *J. Biomol. Screen* 18(10), 1143–1155 (2013).

3 Scannell JW, Blanckley A, Boldon H, Warrington B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov.* 11(3), 191–200 (2012).

4 Carragher NO, Brunton VG, Frame MC. Combining imaging and pathway profiling: an alternative approach to cancer drug discovery. *Drug Discov. Today* 17(5–6), 203–214 (2012).

5 Carragher NO. Advancing high content analysis towards improving clinical efficacy. *European Pharmaceutical Review* (2011). www.europeanpharmaceuticalreview.com/5648

6 Eder J, Sedrani R, Wiesmann C. The discovery of first-in-class drugs: origins and evolution. *Nat. Rev. Drug Discov.* 13(8), 577–587 (2014).

7 Swinney DC, Anthony J. How were new medicines discovered? *Nat. Rev. Drug Discov.* 10(7), 507–519 (2011).

8 Moffat JG, Rudolph J, Bailey D. Phenotypic screening in cancer drug discovery – past, present and future. *Nat. Rev. Drug Discov.* 13(8), 588–602 (2014).

9 Bickle M. The beautiful cell: high-content screening in drug discovery. *Anal. Bioanal. Chem.* 398(1), 219–226 (2010).

10 Chang TY, Pardo-Martin C, Allalou A, Wahlby C, Yanik MF. Fully automated cellular-resolution vertebrate screening platform with parallel animal processing. *Lab Chip* 12(4), 711–716 (2012).

11 Isherwood B, Timpson P, Mcghee EJ *et al.* Live cell *in vitro* and *in vivo* imaging applications: accelerating drug discovery. *Pharmaceutics* 3(2), 141–170 (2011).

12 Rix U, Superti-Furga G. Target profiling of small molecules by chemical proteomics. *Nat. Chem. Biol.* 5(9), 616–624 (2009).

13 Martinez Molina D, Nordlund P. The cellular thermal shift assay: a novel biophysical assay for *in situ* drug target engagement and mechanistic biomarker studies. *Annu. Rev. Pharmacol. Toxicol.* 56 141–161 (2016).

14 Lamb J, Crawford ED, Peck D *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313(5795), 1929–1935 (2006).

• **Development of the Connectivity Map database to correlate transcription profiles from compound perturbed samples with transcription profiles from other compounds and disease to support mechanism-of-action analysis and disease positioning.**

15 Liu C, Su J, Yang F, Wei K, Ma J, Zhou X. Compound signature detection on LINCS L1000 big data. *Mol. Biosyst.* 11(3), 714–722 (2015).

16 Pawlak M, Schick E, Bopp MA, Schneider MJ, Oroszlan P, Ehrat M. Zeptosens' protein microarrays: a novel high performance microarray platform for low abundance protein analysis. *Proteomics* 2(4), 383–393 (2002).

17 Perlman ZE, Slack MD, Feng Y, Mitchison TJ, Wu LF, Altschuler SJ. Multidimensional drug profiling by automated microscopy. *Science* 306(5699), 1194–1198 (2004).

•• **First demonstration that multiparametric high-content profiling can discriminate and predict compound mechanism-of-action.**

18 Lee J, Wu X, Pasca Di Magliano M *et al.* A small-molecule antagonist of the hedgehog signaling pathway. *Chembiochem* 8(16), 1916–1919 (2007).

19 Filippakopoulos P, Qi J, Picaud S *et al.* Selective inhibition of BET bromodomains. *Nature* 468(7327), 1067–1073 (2010).

20 Larrieu D, Britton S, Demir M, Rodriguez R, Jackson SP. Chemical inhibition of NAT10 corrects defects of laminopathic cells. *Science* 344(6183), 527–532 (2014).

21 Ong SE, Schenone M, Margolin AA *et al.* Identifying the proteins to which small-molecule probes and drugs bind in cells. *Proc. Natl Acad. Sci. USA* 106(12), 4617–4622 (2009).

22 Jafari R, Almqvist H, Axelsson H *et al.* The cellular thermal shift assay for evaluating drug target interactions in cells. *Nat. Protoc.* 9(9), 2100–2122 (2014).

23 Geuijen CA, Bijl N, Smit RC *et al.* A proteomic approach to tumour target identification using phage display, affinity purification and mass spectrometry. *Eur. J. Cancer* 41(1), 178–187 (2005).

24 Moser S, Johnsson K. Yeast three-hybrid screening for identifying anti-tuberculosis drug targets. *Chembiochem* 14(17), 2239–2242 (2013).

25 Chidley C, Haruki H, Pedersen MG, Muller E, Johnsson K. A yeast-based screen reveals that sulfasalazine inhibits tetrahydrobiopterin biosynthesis. *Nat. Chem. Biol.* 7(6), 375–383 (2011).

26 Sandercock AM, Rust S, Guillard S *et al.* Identification of anti-tumour biologics using primary tumour models, 3-D phenotypic screening and image-based multi-parametric profiling. *Mol. Cancer* 14, 147 (2015).

27 Hughes TR, Marton MJ, Jones AR *et al.* Functional discovery via a compendium of expression profiles. *Cell* 102(1), 109–126 (2000).

28 Giaever G, Shoemaker DD, Jones TW *et al.* Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nat. Genet.* 21(3), 278–283 (1999).

29 Wacker SA, Houghtaling BR, Elemento O, Kapoor TM. Using transcriptome sequencing to identify mechanisms of drug action and resistance. *Nat. Chem. Biol.* 8(3), 235–237 (2012).

30 Burgess DJ, Doles J, Zender L *et al.* Topoisomerase levels determine chemotherapy response *in vitro* and *in vivo*. *Proc. Natl Acad. Sci. USA* 105(26), 9053–9058 (2008).

31 Heynen-Genel S, Pache L, Chanda SK, Rosen J. Functional genomic and high-content screening for target discovery and deconvolution. *Expert Opin. Drug Discov.* 7(10), 955–968 (2012).

32 Dominguez E, Galmozzi A, Chang JW *et al.* Integrated phenotypic and activity-based profiling links Ces3 to obesity and diabetes. *Nat. Chem. Biol.* 10(2), 113–121 (2014).

33 Swinney DC. The contribution of mechanistic understanding to phenotypic screening for first-in-class medicines. *J. Biomol. Screen.* 18(10), 1186–1192 (2013).

34 Nathan DM, Buse JB, Davidson MB *et al.* Medical management of hyperglycemia in Type 2 diabetes: a consensus algorithm for the initiation and adjustment of therapy: a consensus statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care* 32(1), 193–203 (2009).

35 Ross S, Thamer C, Cescutti J, Meinicke T, Woerle HJ, Broedl UC. Efficacy and safety of empagliflozin twice daily versus once daily in patients with Type 2 diabetes inadequately controlled on metformin: a 16-week, randomized, placebo-controlled trial. *Diabetes Obes. Metab.* 17(7), 699–702 (2015).

36 Miller RA, Chu Q, Xie J, Foretz M, Viollet B, Birnbaum MJ. Biguanides suppress hepatic glucagon signalling by decreasing production of cyclic AMP. *Nature* 494(7436), 256–260 (2013).

37 Pogliano J, Pogliano N, Silverman JA. Daptomycin-mediated reorganization of membrane architecture causes mislocalization of essential cell division proteins. *J. Bacteriol.* 194(17), 4494–4504 (2012).

38 Kawashima T, Iwamoto I, Nakagawa N, Tomioka H, Yoshida S. Inhibitory effect of pemirolast, a novel antiallergic drug, on leukotriene C4 and granule protein release from human eosinophils. *Int. Arch. Allergy. Immunol.* 103(4), 405–409 (1994).

39 Hakimian S, Cheng-Hakimian A, Anderson GD, Miller JW. Rufinamide: a new anti-epileptic medication. *Expert Opin. Pharmacother.* 8(12), 1931–1940 (2007).

40 Vezina C, Kudelski A, Sehgal SN. Rapamycin (AY-22, 989), a new antifungal antibiotic. I. Taxonomy of the producing streptomycete and isolation of the active principle. *J. Antibiot. (Tokyo)* 28(10), 721–726 (1975).

41 Martel RR, Klicius J, Galet S. Inhibition of the immune response by rapamycin, a new antifungal antibiotic. *Can. J. Physiol. Pharmacol.* 55(1), 48–51 (1977).

42 Houchens DP, Ovejera AA, Riblet SM, Slagel DE. Human brain tumor xenografts in nude mice as a chemotherapy model. *Eur. J. Cancer Clin. Oncol.* 19(6), 799–805 (1983).

43 Seto B. Rapamycin and mTOR: a serendipitous discovery and implications for breast cancer. *Clin. Transl. Med.* 1(1), 29 (2012).

44 Brown EJ, Albers MW, Shin TB *et al.* A mammalian protein targeted by G1-arresting rapamycin-receptor complex. *Nature* 369(6483), 756–758 (1994).

45 Lipinski CA. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol.* 1(4), 337–341 (2004).

46 Pike KG, Malagu K, Hummersone MG *et al.* Optimization of potent and selective dual mTORC1 and mTORC2 inhibitors: the discovery of AZD8055 and AZD2014. *Bioorg. Med. Chem. Lett.* 23(5), 1212–1216 (2013).

47 Fraser C, Carragher NO, Unciti-Broceta A. eCF309: a potent, selective, cell-permeable mTOR inhibitor. *Med. Chem. Comm.* 7, 471–477 (2016).

48 Fraser C, Dawson JC, Dowling R,*et al.* Rapid discovery and structure–activity relationships of pyrazolopyrimidines that potently suppress breast cancer cell growth via SRC kinase inhibition with exceptional selectivity over ABL kinase. *J. Med. Chem.* 59(10), 4697–4710 (2016).

49 Yoshida T, Kakegawa J, Yamaguchi T *et al.* Identification and characterization of a novel chemotype MEK inhibitor able to alter the phosphorylation state of MEK1/2. *Oncotarget* 3(12), 1533–1545 (2012).

50 Mayer TU, Kapoor TM, Haggarty SJ, King RW, Schreiber SL, Mitchison TJ. Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. *Science* 286(5441), 971–974 (1999).

51 Theoclitou ME, Aquila B, Block MH *et al.* Discovery of (+)-N-(3-aminopropyl)-N-[1-(5-benzyl-3-methyl-4-oxo-[1,2]thiazolo[5,4-d]pyrimidin -6-yl)-2-methylpropyl]-4-methylbenzamide (AZD4877), a kinesin spindle protein inhibitor and potential anticancer agent. *J. Med. Chem.* 54(19), 6734–6750 (2011).

52 Gerecitano JF, Stephenson JJ, Lewis NL *et al.* A Phase I trial of the kinesin spindle protein (Eg5) inhibitor AZD4877 in patients with solid and lymphoid malignancies. *Invest. New Drugs* 31(2), 355–362 (2013).

53 Bordet T, Buisson B, Michaud M *et al.* Identification and characterization of cholest-4-en-3-one, oxime (TRO19622), a novel drug candidate for amyotrophic lateral sclerosis. *J. Pharmacol. Exp. Ther.* 322(2), 709–720 (2007).

54 Reisen F, Sauty De Chalon A, Pfeifer M, Zhang X, Gabriel D, Selzer P. Linking phenotypes and modes of action through high-content screen fingerprints. *Assay Drug Dev. Technol.* 13(7), 415–427 (2015).

- **Identification of target hypothesis by correlating multiparametric phenotypic profiling with gene enrichement datasets of phenotypically similar compounds.**

55 Isherwood BJ, Walls RE, Roberts ME *et al.* High-content analysis to leverage a robust phenotypic profiling approach to vascular modulation. *J. Biomol. Screen* 18(10), 1246–1259 (2013).

56 Vincent F, Loria P, Pregel M *et al.* Developing predictive assays: the phenotypic screening "rule of 3". *Sci. Transl. Med.* 7(293), 293ps215 (2015).

57 Zheng W, Thorne N, Mckew JC. Phenotypic screens as a renewed approach for drug discovery. *Drug Discov. Today* 18(21–22), 1067–1073 (2013).

58 Pollard SM, Yoshikawa K, Clarke ID *et al.* Glioma stem cell lines expanded in adherent culture have tumor-specific phenotypes and are suitable for chemical and genetic screens. *Cell Stem Cell* 4(6), 568–580 (2009).

59 Nobis M, Mcghee EJ, Morton JP *et al.* Intravital FLIM-FRET imaging reveals dasatinib-induced spatial control of src in pancreatic cancer. *Cancer Res.* 73(15), 4674–4686 (2013).

60 Hirschhaeuser F, Menne H, Dittfeld C, West J, Mueller-Klieser W, Kunz-Schughart LA. Multicellular tumor spheroids: an underestimated tool is catching up again. *J. Biotechnol.* 148(1), 3–15 (2010).

61 Tang H, Duggan S, Richardson PL, Marin V, Warder SE, Mcloughlin SM. Target identification of compounds from a cell viability phenotypic screen using a bead/lysate-based affinity capture platform. *J. Biomol. Screen.* 21(2), 201–211 (2016).

62 Mellacheruvu D, Wright Z, Couzens AL *et al.* The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat. Methods* 10(8), 730–736 (2013).

63 Trinkle-Mulcahy L, Boulon S, Lam YW *et al.* Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *J. Cell Biol.* 183(2), 223–239 (2008).

64 Lee J, Bogyo M. Target deconvolution techniques in modern phenotypic profiling. *Curr. Opin. Chem. Biol.* 17(1), 118–126 (2013).

65 Cong F, Cheung AK, Huang SM. Chemical genetics-based target identification in drug discovery. *Annu. Rev. Pharmacol. Toxicol.* 52 57–78 (2012).

66 Kunkel EJ, Dea M, Ebens A *et al.* An integrative biology approach for analysis of drug action in models of human vascular inflammation. *FASEB J.* 18(11), 1279–1281 (2004).

67 Broad Institute. Cancer Program Resource Gateway. Connectivity Map. www.broadinstitute.org/software/cprg/?q=node/12

68 Lee KH, Lo HL, Tang WC, Hsiao HH, Yang PM. A gene expression signature-based approach reveals the mechanisms of action of the Chinese herbal medicine berberine. *Sci. Rep.* 4, 6394 (2014).

69 Rees MG, Seashore-Ludlow B, Cheah JH *et al.* Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat. Chem. Biol.* 12(2), 109–116 (2016).

70 NIH LINCS Program. www.lincsproject.org/

71 Kunkel EJ, Plavec I, Nguyen D *et al.* Rapid structure-activity and selectivity analysis of kinase inhibitors by BioMAP analysis in complex human primary cell-based models. *Assay. Drug Dev. Technol.* 2(4), 431–441 (2004).

72 Berg EL, Yang J, Polokoff MA. Building predictive models for mechanism-of-action classification from phenotypic assay data sets. *J. Biomol. Screen.* 18(10), 1260–1269 (2013).

73 Charboneau L, Tory H, Chen T *et al.* Utility of reverse phase protein arrays: applications to signalling pathways and human body arrays. *Brief Funct. Genomic Proteomic* 1(3), 305–315 (2002).

74 Lee MJ, Ye AS, Gardino AK *et al.* Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell* 149(4), 780–794 (2012).

75 Van Oostrum J, Calonder C, Rechsteiner D *et al.* Tracing pathway activities with kinase inhibitors and reverse phase protein arrays. *Proteomics Clin. Appl.* 3(4), 412–422 (2009).

76 Akbani R, Becker KF, Carragher N *et al.* Realizing the promise of reverse phase protein arrays for clinical, translational, and basic research: a workshop report: the RPPA (Reverse Phase Protein Array) society. *Mol. Cell. Proteomics* 13(7), 1625–1643 (2014).

77 Skinner HD, Mccurdy MR, Echeverria AE *et al.* Metformin use and improved response to therapy in esophageal adenocarcinoma. *Acta Oncol.* 52(5), 1002–1009 (2013).

78 Honjo S, Ajani JA, Scott AW *et al.* Metformin sensitizes chemotherapy by targeting cancer stem cells and the mTOR pathway in esophageal cancer. *Int. J. Oncol.* 45(2), 567–574 (2014).

79 Ahern TP, Pedersen L, Tarp M *et al.* Statin prescriptions and breast cancer recurrence risk: a Danish nationwide prospective cohort study. *J. Natl Cancer Inst.* 103(19), 1461–1468 (2011).

80 Wolfe AR, Debeb BG, Lacerda L *et al.* Simvastatin prevents triple-negative breast cancer metastasis in pre-clinical models through regulation of FOXO3a. *Breast Cancer Res. Treat.* 154(3), 495–508 (2015).

81 Jia L, Jin H, Zhou J *et al.* A potential anti-tumor herbal medicine, Corilagin, inhibits ovarian cancer cell growth through blocking the TGF-beta signaling pathways. *BMC Complement Altern. Med.* 13, 33 (2013).

82 Park ES, Rabinovsky R, Carey M *et al.* Integrative analysis of proteomic signatures, mutations, and drug responsiveness in the NCI 60 cancer cell line set. *Mol. Cancer Ther.* 9(2), 257–267 (2010).

83 Cardnell RJ, Feng Y, Diao L *et al.* Proteomic markers of DNA repair and PI3K pathway activation predict response to the PARP inhibitor BMN 673 in small cell lung cancer. *Clin. Cancer Res.* 19(22), 6322–6328 (2013).

84 Taylor DL, Woo ES, Giuliano KA. Real-time molecular and cellular analysis: the new frontier of drug discovery. *Curr. Opin. Biotechnol.* 12(1), 75–81 (2001).

85 Gough AH, Chen N, Shun TY *et al.* Identifying and quantifying heterogeneity in high content analysis: application of heterogeneity indices to drug discovery. *PLoS ONE* 9(7), e102678 (2014).

• **The development of novel analysis methods for quantifying hetrogeneous responses across cell subpopulations in high-content phenotypic assays.**

86 Graf T, Stadtfeld M. Heterogeneity of embryonic and adult stem cells. *Cell Stem Cell* 3(5), 480–483 (2008).

87 Young DW, Bender A, Hoyt J *et al.* Integrating high-content screening and ligand-target prediction to identify mechanism of action. *Nat. Chem. Biol.* 4(1), 59–68 (2008).

•• **The intergation of chemical similarity with high-content phenotypic similarity methods to predict target mechanism.**

88 Tanaka M, Bateman R, Rauh D *et al.* An unbiased cell morphology-based screen for new, biologically active small molecules. *PLoS Biol.* 3(5), e128 (2005).

89 Caie PD, Walls RE, Ingleston-Orme A *et al.* High-content phenotypic profiling of drug response signatures across distinct cancer cells. *Mol. Cancer Ther.* 9(6), 1913–1926 (2010).

90 Ljosa V, Caie PD, Ter Horst R *et al.* Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment. *J. Biomol. Screen.* 18(10), 1321–1329 (2013).

91 Kandaswamy C, Silva LM, Alexandre LA, Santos JM. High-content analysis of breast cancer using single-cell deep transfer learning. *J. Biomol. Screen.* 21(3), 290–297 (2016).

92 Smith K, Horvath P. Active learning strategies for phenotypic profiling of high-content screens. *J. Biomol. Screen.* 19(5), 685–695 (2014).

93 Zhang JH, Chung TD, Oldenburg KR. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screen.* 4(2), 67–73 (1999).

94 Kozak K, Csucs G. Kernelized Z′ factor in multiparametric screening technology. *RNA Biol.* 7(5), 615–620 (2010).

95 Kummel A, Gubler H, Gehin P, Beibel M, Gabriel D, Parker CN. Integration of multiple readouts into the Z′ factor for assay quality assessment. *J. Biomol. Screen.* 15(1), 95–101 (2010).

96 Broad Institute Benchmark Collection. www.broadinstitute.org/bbbc/

97 Ljosa V, Sokolnicki KL, Carpenter AE. Annotated high-throughput microscopy image sets for validation. *Nat. Methods* 9(7), 637 (2012).

98 Pedemonte N, Tomati V, Sondo E, Galietta LJ. Influence of cell background on pharmacological rescue of mutant CFTR. *Am. J. Physiol. Cell Physiol.* 298(4), C866–C874 (2010).

99 Breinig M, Klein FA, Huber W, Boutros M. A chemical-genetic interaction map of small molecules using high-throughput imaging in cancer cells. *Mol. Syst. Biol.* 11(12), 846 (2015).