Article:
Pahl, A., Schölermann, B., Lampe, P. et al. (2023) Morphological subprofile analysis for bioactivity annotation of small molecules. Cell Chemical Biology, 30 (7). P839-P853.E7. ISSN: 2451-9456

https://doi.org/10.1016/j.chembiol.2023.06.003

CellPress

Cell Chemical Biology

Review

Morphological profiling of small molecules

Slava Ziegler,[1,*] Sonja Sievers,[1] and Herbert Waldmann[1,2,*]
[1]Max-Planck Institute of Molecular Physiology, Department of Chemical Biology, Otto-Hahn-Strasse 11, 44227 Dortmund, Germany
[2]Technical University Dortmund, Faculty of Chemistry and Chemical Biology, Otto-Hahn-Strasse 6, 44227 Dortmund, Germany
*Correspondence: slava.ziegler@mpi-dortmund.mpg.de (S.Z.), herbert.waldmann@mpi-dortmund.mpg.de (H.W.)
https://doi.org/10.1016/j.chembiol.2021.02.012

SUMMARY

Profiling approaches such as gene expression or proteome profiling generate small-molecule bioactivity profiles that describe a perturbed cellular state in a rather unbiased manner and have become indispensable tools in the evaluation of bioactive small molecules. Automated imaging and image analysis can record morphological alterations that are induced by small molecules through the detection of hundreds of morphological features in high-throughput experiments. Thus, morphological profiling has gained growing attention in academia and the pharmaceutical industry as it enables detection of bioactivity in compound collections in a broader biological context in the early stages of compound development and the drug-discovery process. Profiling may be used successfully to predict mode of action or targets of unexplored compounds and to uncover unanticipated activity for already characterized small molecules. Here, we review the reported approaches to morphological profiling and the kind of bioactivity that can be detected so far and, thus, predicted.

INTRODUCTION

A phenotype unites the observable characteristics of an organism or a cell, such as gene and protein expression, morphology, and biochemical properties, and results from the interaction of genotype and environment (Nussinov et al., 2019). Cell morphology has been linked to specific cellular states or cellular processes and, thus, has predictive value in the analysis of genetic, chemical, or disease-related perturbations. However, morphological alterations often are not obvious to the human eye, which may not be able to discern subtle changes independent of visualization tools, thus calling for the development of reliable and unbiased analysis methods.

There has been a high demand for detailed mapping of the bioactivity space, i.e., targets, off-targets, and mode of action (MoA) of small molecules in general and, more importantly, drug candidates in particular. Whereas available approaches to detect bioactivity mostly address the already known drug-target space, e.g., G-protein-coupled receptors (GPCRs), kinases, and enzymes, in general, ''omics'' approaches such as transcriptomics, proteomics, epigenomics, and metabolomics enable profiling by collecting all measurable parameters to obtain a holistic view of a given cellular state. Although omics studies rarely provide direct proof for small-molecule targets, the inherently rich data they deliver may inform about numerous altered cellular traits between two states, in particular when different omics approaches are combined. Moreover, to date, generic methods for target identification of small molecules are not available, and target identification often is labor- and time-intensive (Saxena, 2016; Wilkinson et al., 2020; Ziegler et al., 2013). Complementary strategies employ comparison of structural similarity (Awale and Reymond, 2019; Byrne and Schneider, 2019), similarity in gene expression profiles (Lamb et al., 2006; Subramanian et al., 2017), or selective cell toxicity profiles to deduce target hypotheses (Rees et al., 2016; Seashore-Ludlow et al., 2015). Similar compounds have been linked to the modulation of similar targets (Keiser et al., 2007, 2009) and similar gene expression patterns are provoked by compounds with common targets or targeted pathways (Lamb et al., 2006; Subramanian et al., 2017). By analogy, selective growth inhibition in a broad cell-line panel has been successfully employed for target/MoA studies (Rees et al., 2016; Seashore-Ludlow et al., 2015).

A general challenge at the heart of chemical biology is how to detect bioactivity in (synthesized) compound collections as early as possible and how to address various biological processes simultaneously that may inform compound optimization. *Omics* approaches per se are well suited for this purpose; however, they are not amenable to high-throughput screening.

The advent of automated high-content imaging and the development of algorithms for pattern recognition, feature extraction, and, thus, image data analysis paved the way for an additional pillar in the suite of profiling approaches, i.e., morphological profiling, for chemical biology research and drug discovery. High-content assays usually are designed to monitor a given biological process and detect a limited number of parameters. However, the extraction of hundreds of parameters independent of a particular process enables the exploration of small-molecule-related bioactivity in a more unbiased manner. For morphological profiling (Figure 1), cellular components are stained using fluorescent dyes or antibodies to detect pattern changes upon perturbation by small molecules. Alternatively, fluorescent tags can be genetically introduced into cells to directly monitor morphology markers. High-content imaging is employed to

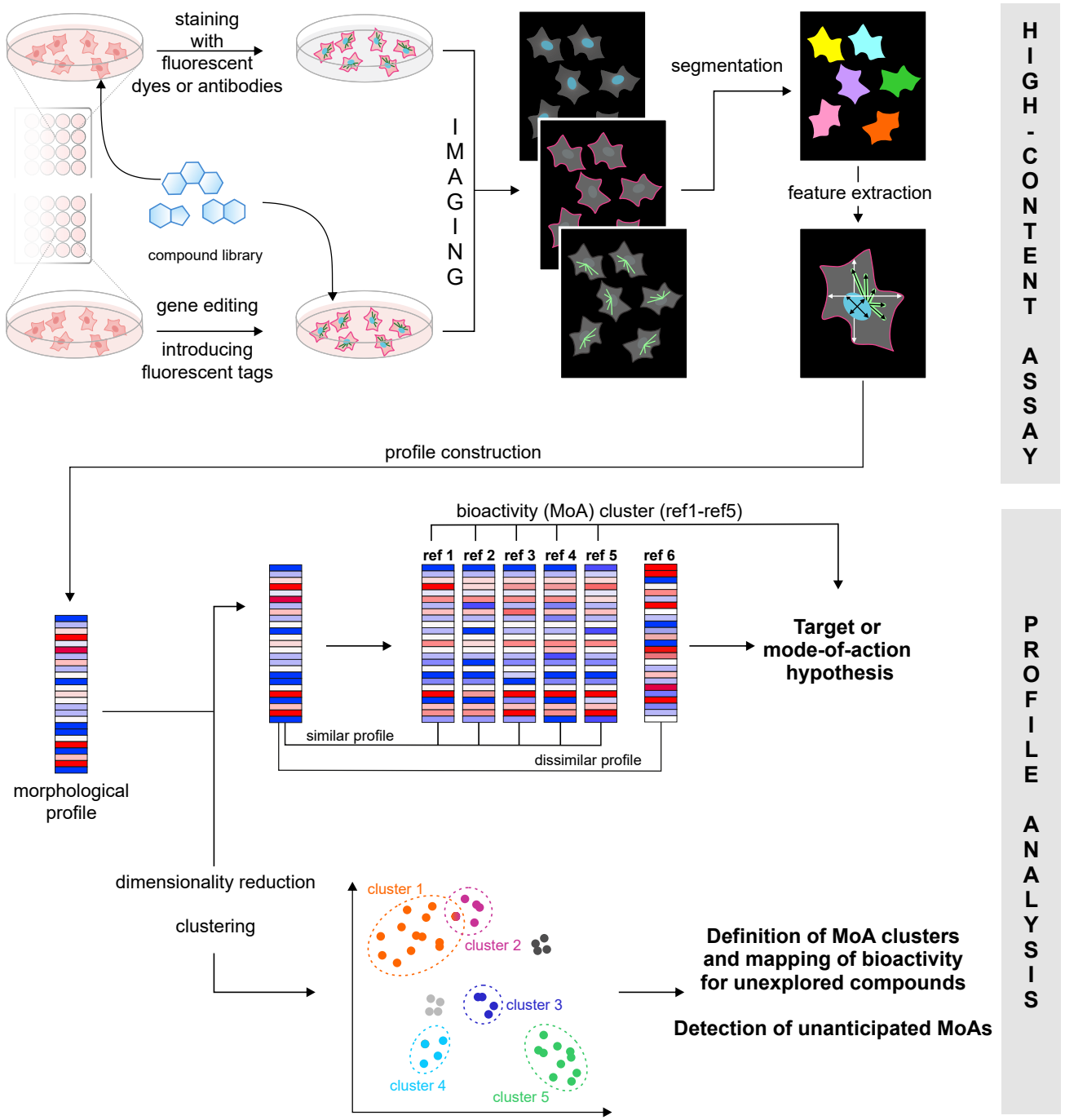


Figure 1. Workflow for morphological profiling

Upon treatment with compounds, cells are stained for cellular compartments or markers using fluorescent dyes or antibodies. Alternatively, cells expressing fluorescently tagged proteins for detection of cellular components can be used. High-content imaging and analysis are employed to extract multiple features and generate morphological profiles that can be compared with profiles of reference compounds. Profile similarity suggests a similar mode of action or even target. Dimensionality reduction assists clustering of similar profiles and assignment of bioactivity to unexplored compounds. Ref denotes reference compound, i.e., compound with known target or mode of action (MoA). See also Table S1.

capture images for each cellular component, and cell segmentation identifies cellular and subcellular regions (Bougen-Zhukov et al., 2017). Hundreds of numerical phenotypic descriptors are then extracted to generate morphological profiles that describe phenotypes (as compared with the non-perturbed state) (Bougen-Zhukov et al., 2017; Boutros et al., 2015; Caicedo et al., 2017; Grys et al., 2017). These features are related to size and shape of cells and organelles, intensity, and texture, among others. Profiles of unexplored compounds can then be compared with the profiles of agents with known targets and

mode of action, where profile similarity implies common MoA or target. Dimensionality reduction is usually performed to convert high-dimensional into low-dimensional profiles by removing redundant or irrelevant features while retaining the variance of the dataset (Bougen-Zhukov et al., 2017; Grys et al., 2017). Clusters of compounds with similar profiles and, thus, MoA, can be mapped using unsupervised machine learning, as the expected phenotypic output is unknown (Grys et al., 2017). Phenotypic categories that can be defined prior to the analysis can be used in supervised approaches to train models that will then assign new profiles to existing categories (Bougen-Zhukov et al., 2017; Grys et al., 2017). Thus, morphological profiling has the power to assign already known and unanticipated MoAs to annotated compounds, to detect bioactivity, and to predict MoAs for biologically uncharacterized small molecules.

This review aims to give an overview of the currently reported approaches to morphological profiling of small molecules according to the employed biological system and strategies to detect cellular components, the kind of bioactivity that can be assessed by these methods (see Table S1), and the lessons learned by their application. It does not focus on the computational approaches for extraction and analysis of morphological features (Bougen-Zhukov et al., 2017; Grys et al., 2017; Scheeder et al., 2018). We lay emphasis on the Cell Painting Assay (CPA) (Bray et al., 2016; Gustafsdottir et al., 2013), as it does not require genetic manipulation, can be used with a variety of cell lines, and has found broad acceptance and application in the chemical biology community and increasingly in the pharmaceutical industry.

MORPHOLOGICAL PROFILING IN MAMMALIAN CELLS

Automated detection of the major cellular compartments (endoplasmic reticulum [ER], Golgi, mitochondria, lysosomes, endosomes, the actin and tubulin cytoskeleton, nucleoli, and nucleus) was first reported by Boland and Murphy (2001) in HeLa cells. Upon staining of cells for one of the compartments and for DNA, 84 features per image were extracted which, for example, monitored image texture, patterns, object distance measures with respect to the cell center, object size, stain overlap with the nucleus, and others. The DNA pattern was employed as a standardization between cells as it is consistent among cells, and protein localization patterns were referred to the DNA stain as a common landmark. This pioneering work paved the way for the high-content analysis and image-based profiling techniques for small molecules that differ in the number and type of detected cellular components and the type of visualization, which is addressed in the following sections.

Staining of cellular components using dyes and antibodies

Staining of the DNA and DNA-related components

The Morphobase, a cell morphology database that employs DNA staining and bright-field images of *src*^{ts}-NRK and HeLa cells upon compound treatment (Futamura et al., 2012), detected clusters for several target classes such as tubulin, actin, DNA synthesis, histone deacetylase (HDAC), and heat-shock proteins (HSPs) (Table S1). Importantly, the clusters of HDAC and proteasome inhibitors could not be distinguished in *src*^{ts}-NRK, whereas RNA and protein synthesis inhibitors could not be separated in HeLa cells. However, when the perturbation profiles in both cell lines were considered, the activities could be clearly differentiated. The Morphobase linked the mitochondrial complex I inhibitor rotenone to tubulin, the cyclin-dependent kinase (CDK) inhibitor 3-ATA, and the polypharmacological compound resveratrol to DNA synthesis. As the observed activities had been previously associated with these compounds, these findings emphasize the need of complete annotation for reference compounds as mostly only the nominal target, i.e., the target most commonly associated with the compound (Moret et al., 2019), is annotated in public databases or vendors' websites. Further studies using the Morphobase identified inhibitors of tubulin and the proteasome also in combination with proteome profiling using the ChemProteobase (Futamura et al., 2012, 2013; Minegishi et al., 2015).

Markers for DNA synthesis (5-ethynyl-3-deoxyuridine [EdU]) and mitosis (phospho-histone H3) along with a DNA stain were employed to profile compounds in HeLa cells (Young et al., 2008) (Table S1). The most information-rich characteristics stemmed from the nuclear stain, i.e., size of the nucleus and DNA quantity. Profile analysis revealed bioactivity clusters around the vacuolar ATPase, of antimitotics, corticosteroids, and progesterone-related compounds. Interestingly, a cell-cycle-arrest-related cluster featured compounds with different MoAs such as cardiac glycosides, which target Na^+/K^+-ATPases, the protein translation inhibitors emetine and cycloheximide, and steroid hormones such as progesterone and danatrol. The authors compared the predictive nature of the phenotypic profiling with target prediction based on structural similarity. Overall, phenotypes were better correlated with the known or predicted targets than with compound structures.

Staining of the DNA and the cytoskeleton (with or without Golgi stain)

In addition to visualizing DNA, microtubules and Golgi stainings were employed for compound profiling in five different cell lines (SKOV3, A549, SF268, DU145, and HUVEC [human umbilical vein endothelial cell], see Table S1) (Adams et al., 2006; Tanaka et al., 2005). Subpopulation analysis was performed based on the nuclear stain to group cells in populations depending on the cell-cycle phase by employing the information on DNA content, morphology, and condensation. Different profiles were observed for two structurally related compounds, the Src inhibitor PP2 and hydroxy-PP, in A549 cells and HUVECs, whereas the effects of both compounds on DU145, SF268, and SKOV3 cells were similar. Like PP2, hydroxy-PP inhibited the kinases Fyn, p38-α, and protein kinases A and B, which may explain the profile similarity in DU145, SF268, and SKOV3 cells. The lack of similarity in A549 cells and HUVECs may stem from different expression levels of targeted proteins. Thus, a cell line set that represents diverse tissues and genetic backgrounds may be beneficial for the detection of subtle morphological changes.

Carragher's group employed DNA-, actin-, and tubulin-staining in four cell lines (Ovcar3, MiaPaCa2, MCF7, and MCF7 with truncated, dominant negative TP53 mutant) (Caie et al., 2010). They even defined subpopulations based on cell-shape descriptors and distinguished an epithelial from a mesenchymal phenotype in MCF7 cells. The assay differentiated clusters of inhibitors

of protein synthesis, proteases, DNA synthesis, and tubulin or actin modulators, as well as Aurora kinase and Eg5 inhibitors (Table S1) (Caie et al., 2010; Ljosa et al., 2013). Of note, tubulin-targeting compounds were separated in stabilizing (Taxol, epothilone B) and destabilizing (colchicine, nocodazole) agents. Further analysis revealed resistance and selectivity across the cell-line panel for some compounds, while other compounds induced similar phenotypes in all four cell lines. Overall, Ovcar3 cells were more resistant to several compound classes, whereas MiaPaCa2 cells were often most sensitive to compound perturbations (Caie et al., 2010).

DNA, tubulin, and phospho-histone H2A.X were used to classify compound activity in HeLa cells (Twarog et al., 2016). This approach identified clusters related to Aurora kinase, tubulin, proteasome, topoisomerases, HDAC, antimetabolites, and HSP90. However, mammalian target of rapamycin (mTOR) and phosphatidylinositol 3-kinase (PI3K) inhibitors were not classified due to low activity in this minimalist assay, in contrast to approaches that employ additional staining (Reisen et al., 2015).

Genotype-dependent profiling was performed upon DNA and actin staining in 12 isogenic cell lines based on the parental cell line HCT116 (Breinig et al., 2015). The cell lines harbored deletions in oncogenic mutations or knockouts (see Table S1). Both genotype-dependent and genotype-independent phenotypes were observed in compound-treated cells, and the number of drug-gene interactions varied across the cell-line panel. Drug-gene-phenotype interactions were associated with 15% of the compounds and provided information on crosstalk of signaling pathways and potential drug synergy. The analysis clustered compounds that target tubulin, MEK, p38, glucocorticoid receptor, DNA alkylation, or mitochondrial proton gradient. Clusters of connected biological processes were observed for an iron chelator, antifolates, and DNA methyltransferase inhibition. Uncoupling of the mitochondrial proton gradient was observed as unanticipated bioactivity for the protein kinase Cδ (PKCδ) inhibitor rottlerin (Figure 2A), which already had been reported in 2001 (Soltoff, 2001). In addition, proteasome inhibition was predicted by the chemical-genetic matrix for the aldehyde dehydrogenase (ALDH) inhibitor disulfiram, endothelial growth factor receptor (EGFR) inhibitor tyrphostin AG555, and the nuclear factor κB (NF-κB) inhibitor caffeic acid phenyl ester (CAPE), which was experimentally validated (see Figure 2B).

A more physiological setup for morphological profiling employed staining of DNA, actin, and membrane integrity to profile compounds in patient-derived organoids (PDOs) from 19 patients with colorectal cancer at different clinical stages (Betge et al., 2019). Analysis of compound-induced phenotypes revealed clusters of MEK, PI3K/AKT/mTOR, glycogen synthase kinase 3 (GSK3), EGFR, and CDK inhibitors (Table S1). Whereas phenotypic activity of MEK and CDK inhibitors was observed in all PDO lines, AKT and GSK3 activity clusters were detected only in a subset of PDOs. In addition, morphological responses provided mechanistic insights for MEK and GSK3 inhibitor classes and linked phenotypes to cancer mutations.

Staining using multiple staining sets

The use of multiple staining sets enables detection of various cellular components in parallel processed samples and is thus not limited by the number of spectrally separable dyes (Table S1).

A seminal work on morphological profiling of small molecules by the Altschuler and Wu groups explored 100 drugs mostly with known MoA in HeLa cells using a DNA stain and five different staining sets: (1) splicing factor SC35 and the cytokinesis protein anillin; (2) β-tubulin and actin; (3) phospho-p38 and phospho-extracellular signal-regulated kinase (ERK); (4) p53, c-Fos; and (5) phosphoadenosine 3,5-monophosphate response element-binding protein (CREB) and calmodulin (Perlman et al., 2004). For compounds that provoked strong responses, some descriptors changed differently at the different concentrations, implying differences of high and low concentrations in target engagement or interaction with several targets with different affinities. For example, camptothecin inhibits at low concentration topoisomerase I and causes an S-phase arrest, whereas at high concentration it inhibits transcription along with further cellular processes. The authors predicted the MoA of the poorly characterized compound austocystin that clustered with transcription and translation inhibitors. Already this first report made some essential observations: (1) compounds with common targets displayed similar profiles irrespective of the chemical structure; (2) compounds with common MoA may display very different profiles, and their clustering may be poor due to modulation of further targets; (3) clustering based on morphological profiles reflects similar MoA rather than chemical similarity.

Woehrmann et al. (2013) used two sets of stains for profiling in HeLa cells: (1) DNA, phospho-histone H3, and EdU; and (2) DNA, actin, and tubulin (Table S1). Filtering was applied to exclude weakly active (fatty acids, prostaglandins, and steroids) or toxic compounds that made up 54% of the initially profiled compounds. Several bioactivity clusters were defined based on profile similarity: G_1-S arrest supercluster (which could be further differentiated in subclusters around DNA antimetabolites, topoisomerases, transcription and translation, and proteasome inhibitors), actin, tubulin, phosphodiesterases, mitochondrial uncouplers, HSP90, kinases, calcium channels, calmodulin, and PAF (platelet-activating factor) ligand. For many compounds, profiles at different concentrations clustered together. Gene set enrichment analysis was performed using features instead of genes, feature categories instead of gene sets, annotated MoA classes instead of phenotypes, and compound-dose instances instead of samples. This enabled discrimination between DNA synthesis and protease inhibitors (belonging to the supercluster) as well as HSP and microtubule inhibitors. Importantly, the authors addressed unanticipated activity of some compounds: For example, SU-4312 (vascular endothelial growth factor receptor [VEGFR] inhibitor), SKF-96365 (calcium channel blocker), and GW-9662 (peroxisome proliferator-activated receptor [PPAR] ligand) clustered with microtubule poisons (Figure 2C). For the first three compounds, microtubule-targeting activity was either already reported or demonstrated by the authors. Tyrphostins such as tyrphostin 8 (AG10) and AG879 (Figure 2A), which inhibit multiple kinases, were assigned to the cluster of mitochondrial uncouplers. Several tyrphostin derivatives are indeed known to uncouple the mitochondrial proton gradient (Soltoff, 2004). To explore morphological clusters based on kinase inhibitors, the authors combined the cytological profiles of 24 kinase inhibitors with their inhibition profiles for 24 selected kinases. Kinase inhibitors often target

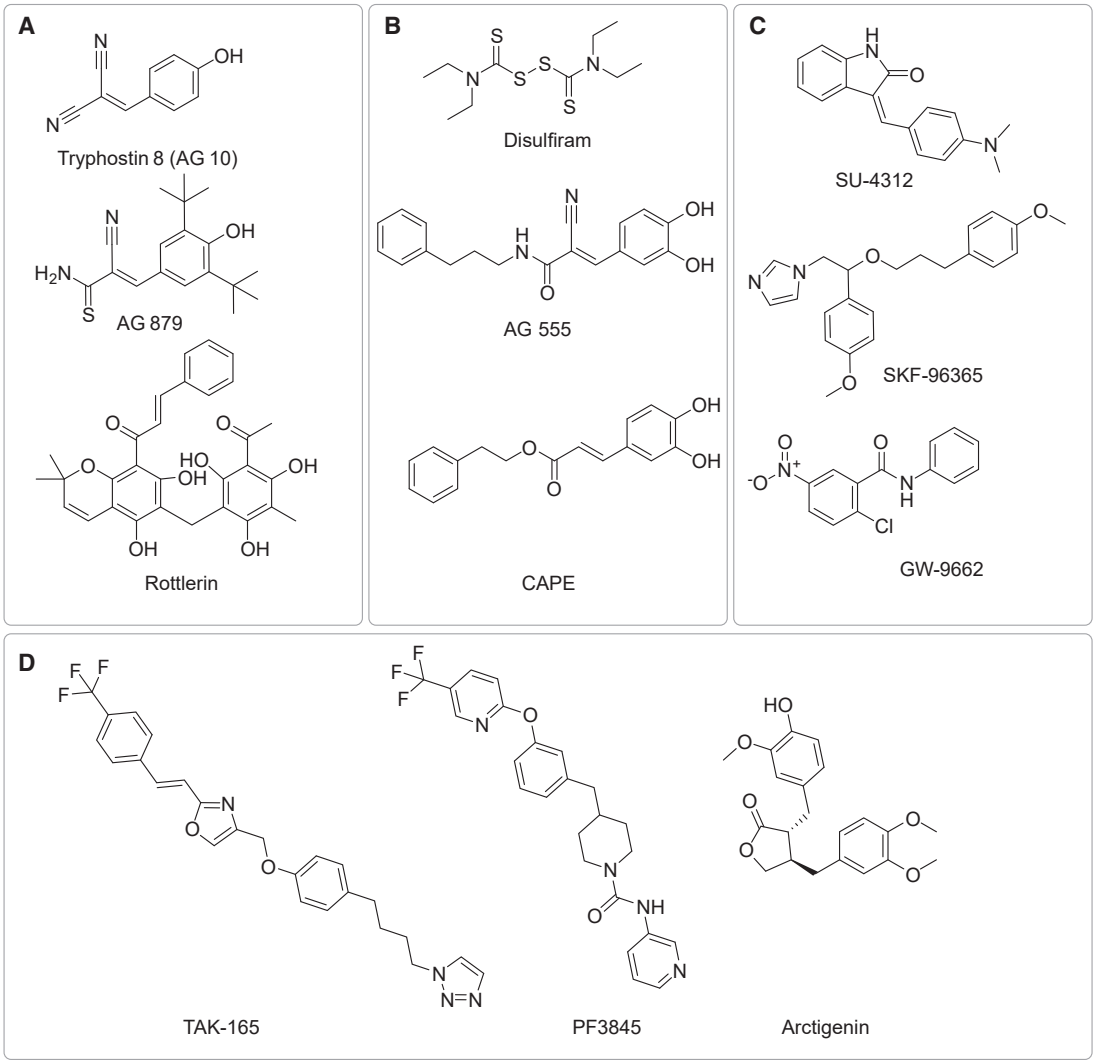


Figure 2. Small molecules explored using morphological profiling with unanticipated activity

(A) Structures of tyrosine kinase inhibitors tyrphostin 8 and AG 879 and the PKCδ inhibitor rottlerin that displayed unanticipated proton gradient uncoupling activity (Breinig et al., 2015; Woehrmann et al., 2013).

(B) Structures of disulfiram (ALDH inhibitor), AG 555 (EGFR inhibitor), and CAPE (NF-κB inhibitor) that displayed unanticipated proteasome targeting activity (Breinig et al., 2015).

(C) Structures of SU-4312 (VEGFR inhibitor), SKF-95365 (Ca^{2+} channel blocker), and GW-9662 (PPAR ligand) that displayed unanticipated tubulin-targeting activity (Woehrmann et al., 2013).

(D) Structures of TAK-165 (Erb inhibitor), PF3845 (FAAH inhibitor), and actigenin (MEK and IκB inhibitor) with unanticipated activity as mitochondrial toxins (Cox et al., 2020).

See also Table S2.

several kinases and their, in part broad, inhibitory profiles make it challenging to assign a profile or cluster to the inhibition of a given kinase. Merging the morphological and kinase inhibition profiles revealed clusters around AMPK (AMP-activated protein kinase), PKCα, Rho-kinase (ROCK), and phosphorylase kinase. A very similar investigation demonstrated that the morphological profiles of iron siderophores such as deferoxamine cluster with the profiles of DNA-interacting agents, and similarity to these profiles enabled identification of microferrioxamines as iron chelators (Schulze et al., 2013). Moreover, antimitotic activity and calcium channel modulation was suggested for two newly isolated natural products, respectively (Ochoa et al., 2015).

A Novartis team also selected sets of different stains for profiling in U2OS cells: (1) DNA, cytoplasm, tubulin; (2) DNA, cytoplasm, mitochondria; (3) DNA, cytoplasm, ER; and (4) DNA, cytoplasm, Golgi (Reisen et al., 2015). Compounds were selected to comprise different targets and processes with relevance to drug discovery, and 52% of the tested compounds were active. Compounds lacking profile similarity to other compounds were removed prior to cluster analysis. Clusters were analyzed for enrichment of targets or gene sets using public and commercial databases and could be linked to the following targets or processes: HDACs, tubulin, PI3K-AKT-mTOR, HMG-CoA reductase (HMGCR), platelet-derived growth factor

receptor B, CaM kinase, HSP90, TOP1, RAR, CaSR, HB-EGF-PKCβ-cyclin D1, sphingosine kinase 1, muscle organ development, and kinases.

A different study combined four staining sets—(1) actin, tubulin, DNA; (2) ER, lysosomes, membranes; (3) mitochondria and NF-κB; (4) p53 and caspase 9—for profiling in HeLa cells (Kremb and Voolstra, 2017). Bioactivity groups were detected for DNA-targeting agents, topoisomerase and tubulin as well for neurotransmitter-related compounds that accumulate in lysosomes and impair their function (Table S1). The authors explored manually the alterations of the 20 core features for single compounds or group of similar compounds to detect the changed parameters and, thus, map the modulated cellular components and biological processes (and relate alterations to structural modification within one compound class). Inhibition of protease activity was predicted for a poorly explored natural product, which was experimentally confirmed. This approach linked the activity of Red Sea algae fractions to modulation of CDKs or tubulin polymerization. Furthermore, some Red Sea algae fractions showed similar profiles to inhibitors of HIV-1 reverse transcriptase (RT), and HIV-1 RT inhibition was subsequently confirmed (Kremb et al., 2017).

Simultaneous staining of multiple components

The availability of fluorescent markers for different cellular components that can be spectrally separated enables simultaneous detection of various cellular components. To the best of our knowledge, the CPA, which was developed at the Broad Institute, is the only example thus far. It employs six different stains (DNA, mitochondria, Golgi, ER, actin, plasma membrane, and RNA) that are simultaneously detected in five different channels (Gustafsdottir et al., 2013). This approach has found numerous applications and thus will be discussed separately in detail (see below).

Morphological profiling using genetically encoded markers

The use of genetically encoded markers for visualization of cellular components enables imaging of live cells and omits the need for a staining procedure. The Altschuler and Wu groups generated a set of 93 triply labeled cell lines (called optimal reporter cell lines for annotating compound libraries, or ORACLs) (Kang et al., 2016). The cell lines were derived from A549 cells and genetically modified to carry the mCherry fluorescent protein (RFP) for image segmentation and CFP (cyan fluorescent protein)-histone 2B for nucleus detection. Each cell line additionally carried a different, randomly yellow fluorescent protein (YFP)-tagged protein. Live-cell phenotypic profiles were generated and compound signatures were further extended by concatenating profiles from different concentrations, time points, and ORACLs. The cell-line panel was treated with 30 compounds that target six mechanistic classes to assess the performance of the reporter cell lines. The X-ray repair and cross-complementing protein 5 (XRCC5)-tagged cell line demonstrated high accuracy (94%) for MoA prediction, whereas the reporter cell line carrying only RFP showed the lowest accuracy (ca. 60%). Compounds were profiled at one or three concentrations in the YFP-XRCC5 ORACL at 24 and 48 h, and data for both time points were merged in the final profiles. Clusters for compounds were detected that target DNA, microtubules, mTOR, proteasome, HDAC, HSP90, Aurora kinase, and ER. A literature search revealed clusters for glucocorticoids, Na^+/K^+-ATPase, and dihydrofolate reductase. Furthermore, two new proteasome inhibitors and one new mTOR inhibitor were identified using the ORACL approach, and these predictions were experimentally validated.

A more targeted approach to genetically label selected cellular components was employed by a team at Janssen Pharmaceutica (Table S1). The cell lines A549, HepG2, and WPMY1 were designed to carry TagBFP and nuclear-targeted TagBFP to distinguish between nucleus and cytosol, and GFP- and RFP-tagged organelles or proteins. Five variants of each cell line were generated to detect: (1) actin and early endosome; (2) ER and mitochondria; (3) Golgi and autophagosome; (4) tubulin and NF-κB signaling; and (5) non-homologous end joining-mediated repair of DNA double-strand breaks and clathrin-mediated endocytosis (Cox et al., 2020). Each profiled compound was manually annotated with one or more MoA. In addition, gene names of the respective targets were extracted from the ChEMBL and IUPHAR database. Fifty-three percent of the compounds showed activity in at least one cell line and 92% of the active compounds were detectable with only three reporter cell lines. Compounds whose targets were not expressed were active at higher concentrations, which is most likely attributed to off-target activity. No cell line from the panel outperformed all others. The number of distinct MoA increased with the number of screened cell lines, the number of cell types, and marker pairs. MoAs could be better differentiated when multiple concentrations were tested. Clusters were detected for the following targets or processes: PI3K/mTOR, Aurora kinases, bromodomains, glucocorticoid receptor, HDAC, HSP, HMGCR, kinesins, tubulin, ROCK, DNA synthesis, and MEK. Poor clustering was observed for ABL1 and VEGFR inhibitors. Interestingly, glucocorticoids displayed phenotypic activity only in A549-derived cells, even though the glucocorticoid receptor gene NR3C1 was expressed across the cell-line panel. Similar observations were made for compounds that induce mitochondrial toxicity, such as uncouplers and respiratory chain inhibitors. Thus, profiling compounds in different cell lines may be essential to map a certain MoA. The study classified several agents as mitochondrial toxins that had not been annotated for this type of activity: Erb inhibitor TAK-165, the fatty acid amide hydrolase (FAAH) inhibitor PF3845, and arctigenin, a MEK and IκB inhibitor (Figure 2D), for which the activity was then experimentally proved. Of note, PF3845 and arctigenin had been previously linked to impairment of mitochondria (Brecht et al., 2017; Romero et al., 2018).

MORPHOLOGICAL PROFILING IN BACTERIA AND YEAST

Small molecules address targets in different organisms and, thus, their bioactivity is not restricted to mammalian cells. Almost 10 years after the seminal work by Perlman et al., a bacterial cytological profiling (BCP) for MoA deconvolution of antibacterial agents was reported (Nonejuie et al., 2013). Bacteria are much smaller than eukaryotic cells and lack organelles, thus restricting the number of general cell morphology markers that can be detected. In the initially reported BCP, *Escherichia coli lptD4213* cells were treated for 2 h with 41 antibiotics, which are used in the clinic and

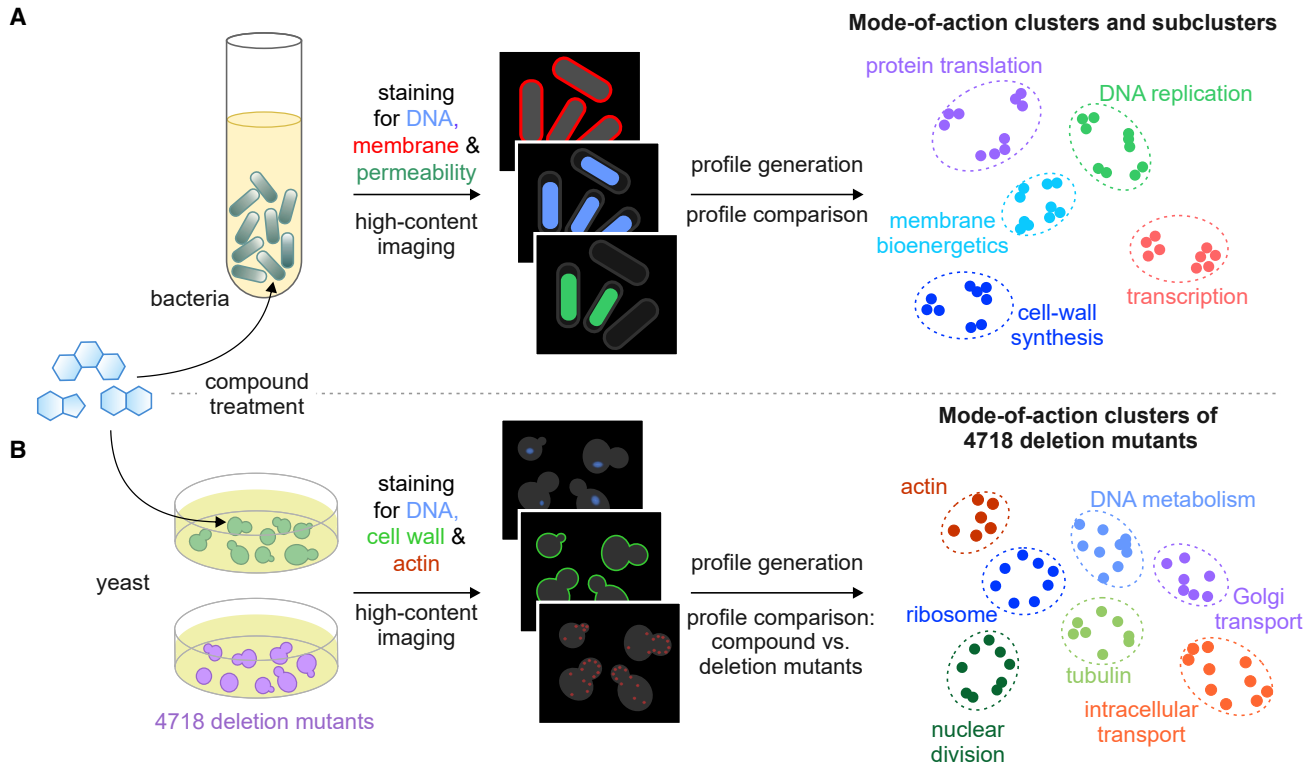
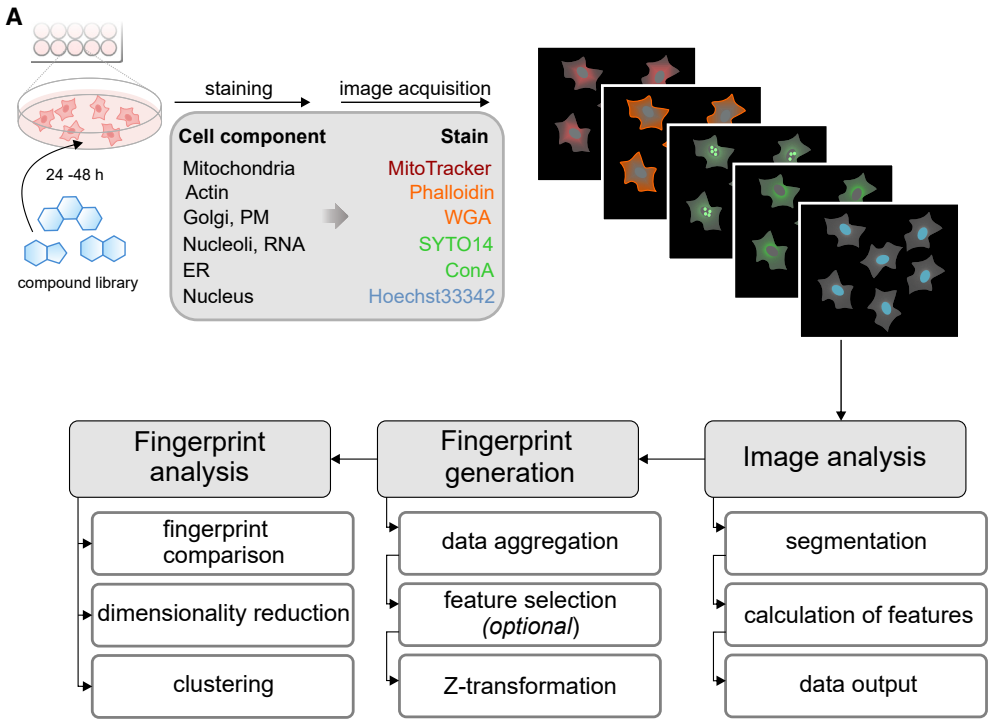
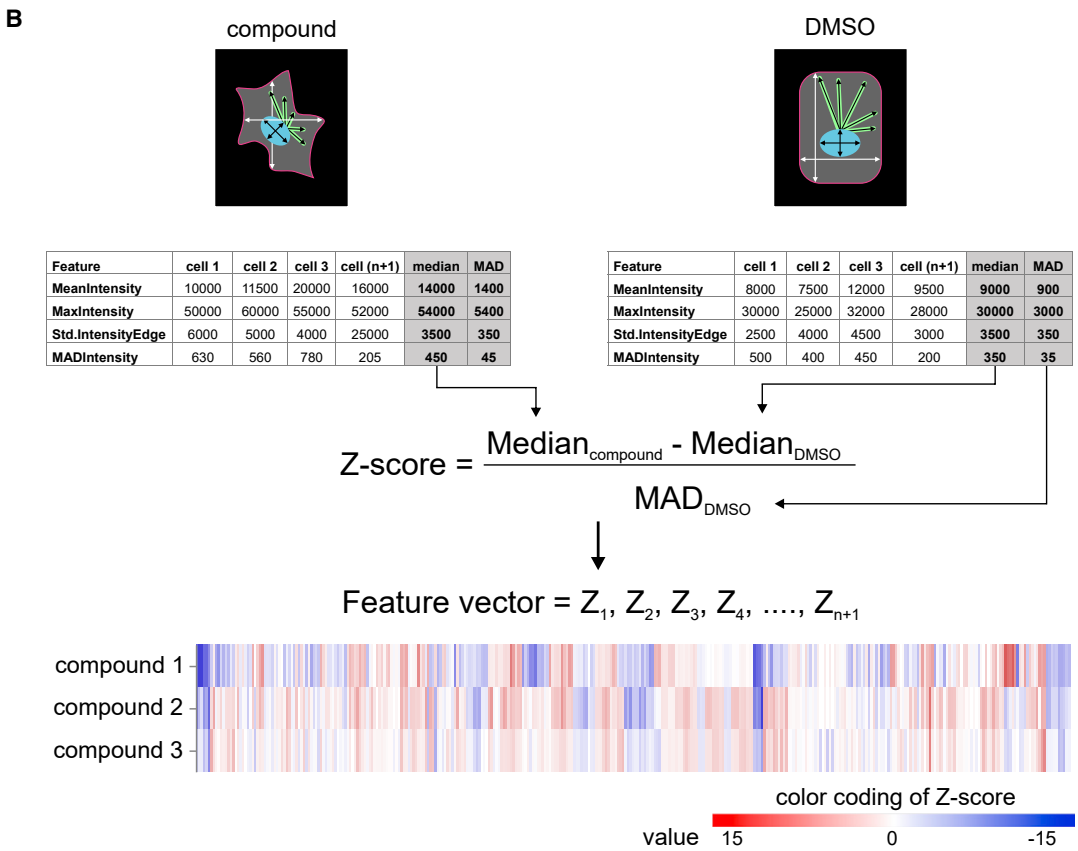


Figure 3. Morphological profiling of small molecules in bacteria or yeast (Nonejuie et al., 2013; Ohya et al., 2005)
Bacterial cells (A) or yeast (B) are exposed to small molecules prior to staining of cellular components, high-content imaging, and analysis to generate morphological profiles. Mode-of-action clusters are identified based on profile similarity. For yeast (B), morphological profiles are generated for 4718 deletion mutants, with which the profiles of small molecules are compared to deduce mode of action.

target translation, transcription, DNA replication, lipid synthesis, or peptidoglycans. Cells were stained for DNA, bacterial membrane, and membrane permeability (Figure 3A). Fourteen features were employed for the generation of cytological profiles. Dimensionality reduction assisted activity mapping and successfully separated the antibiotics into different bioactivity categories. BCP differentiated between distinct mechanisms of action for agents with common MoA. Three different subclasses of protein translation inhibitors were detectable: inhibitors of peptide elongation, compounds promoting mistranslation, or premature chain termination. BCP differentiated DNA-replication inhibitors into subclasses of intercalators, DNA-crosslinkers, and gyrase A or gyrase B inhibitors. BCP detected three subgroups of cell-wall synthesis inhibitors, four subgroups of membrane bioenergetics-interfering compounds, and two subclasses of RNA transcription inhibitors. The initial BCP has been further developed to enrich profiles with more features (Htoo et al., 2019; Lamsa et al., 2016), and was applied for different bacterial pathogens (Htoo et al., 2019; Lamsa et al., 2016). Moreover, a group at Roche adapted the assay for high-throughput use to identify small molecules that change bacterial phenotype as a novel strategy for the development of antibiotics (Zoffmann et al., 2019). The BCP target/MoA space can be further extended by the use of GFP reporter strains, as few reference antibiotics failed to alter the phenotype but changed gene expression. BCP was also employed for MoA deconvolution of less or unexplored compounds and to uncover novel MoAs for already known antibiotics (Htoo et al., 2019; Nonejuie et al., 2013; Zoffmann et al., 2019). In addition, BCP was coupled to rapid inhibition profiling (RIP) to generate cytological profiles upon degradation of essential bacterial proteins (Lamsa et al., 2016). This combined approach can detect similarity between profiles of antibiotics and profiles induced by protein degradation and may be especially suited to targets for which antibiotics do not exist. Coupling RIP to BCP demonstrated a complex profile upon degradation of cytidylate kinase (CMK), which is involved in DNA replication and wall teichoic acid biosynthesis. The cytological profile of CMK degradation matched the profile of simultaneous inhibition of both processes with ciprofloxacin and tunicamycin, and suggests complex profiles for putative CMK inhibitors due to targeting multiple biosynthetic pathways (Peters et al., 2018).

Morphological profiling can be also performed in yeast cells using staining of DNA, actin, and the cell surface manno-protein (Figure 3B) (Ohya et al., 2005). Thereby, morphological changes induced by 4,718 haploid mutants of non-essential genes were explored. Deletion of functionally related genes induced similar phenotypic signatures. In a follow-up study, hydroxyurea, concanamycin A, lovastatin, and echinocandin B were profiled in yeast (Ohnuki et al., 2010). These compounds were chosen to affect DNA metabolism, vacuolar ATPases, lipid metabolism, and cell wall, respectively. The obtained profiles were compared with the phenotypes of the deletion mutants using principal component analysis (PCA). Gene ontology (GO) term enrichment was performed for the subset of deletion





(legend on next page)

mutants that displayed similarity to a given compound. Thereby, hydroxyurea was linked to the GO term "DNA replication," and the highest ranked mutant *rnr4* codes for the target of this compound, which is ribonucleotide reductase. Concanamycin A was correlated with the activity of a mutant subgroup with the enriched GO term "vacuolar acidification," which was among the top five GO terms. In contrast, the top ten predicted MoAs of lovastatin and echinocandin B were not related to the corresponding target genes (*hmg1* and *fks1*, respectively), which may be attributed to the weak phenotypes induced by the respective mutant target genes. Although this approach was limited to the use of only four compounds and focused on non-essential genes only, it demonstrates the feasibility of morphological profiling in yeast and should be suitable for profiling further compounds (Gebre et al., 2015). A reference compound set would complement the genetic profile, as small molecules may modulate only one of many functions of a target protein and induced phenotypes may vary significantly from simply removing the protein by gene deletion.

CELL PAINTING

In this section we focus on the CPA, including all approaches that broadly follow the canonical protocol of Bray et al. (2016). In the CPA, subcellular compartments are stained with six different dyes in one well (Gustafsdottir et al., 2013): Hoechst 33342 (nuclei), concanavalin A (ER), SYTO 14 (nucleoli), phalloidin (actin), wheat germ agglutinin (Golgi), and MitoTracker Deep Red (mitochondria) (Figure 4A). By subsequent image analysis, hundreds of features are deduced from every channel, thereby generating numeric data from the images, and combined into a morphological profile. Often, a data pre-processing and feature selection step is included followed by a final data-mining step whereby fingerprints are further analyzed. The assay proceeds at reasonable medium throughput (several tens of thousands), employs affordable dyes, and has rapidly been developed into a widely applied morphological profiling method (Boyd et al., 2020; Hughes et al., 2020; Nyffeler et al., 2020; Rohban et al., 2017).

Image acquisition

CPA was originally established in the U2OS cell line; however, other cell lines are also suitable, for example, MCF7, HepG2, A549, HTB-9, ARPE-19, 3T3, HeLa, SH-SY5Y, HUVEC, HMVEC (human microvascular endothelial cell), primary human fibroblasts, and primary human hepatocyte/3T3-J2 fibroblast cocultures (Bray et al., 2016; Gerlach et al., 2019; Hippman et al., 2020; Willis et al., 2020). For adaptation to esophageal adenocarcinoma and different breast cancer cell lines, a live-cell MitoTracker staining had to be replaced by a fixed-cell protocol to avoid cell damage during the live-cell staining (Hughes et al., 2020; Warchal et al., 2020). Reduction of phalloidin and/or concanavalin A dye concentration with no loss in signal is possible (Christoforow et al., 2019; Svenningsen and Poulsen, 2019). Most studies incubated cells with the compounds for 48 h, although some groups chose 20–24 h to cover the more direct effects of compound treatment (Christoforow et al., 2019; Hippman et al., 2020). For image acquisition, both wide-field (ImageXpress XLS, Zeiss Celldiscoverer 7) and confocal systems (Opera Phenix High Content Screening System) were used. In the original protocol, five images were taken in five channels. However, as the laser-based confocal systems offer only four excitation wavelengths, the RNA and ER channels were either imaged separately at the same wavelength using different offsets or only four images were taken, again with multiplexing of the concanavalin and SYTO 14 dyes because the signals are spatially well separated (Melillo et al., 2018; Nyffeler et al., 2020).

Fingerprint generation

To arrive at numerical data, cells are analyzed with regard to their morphological features including size, shape, texture, and intensity. Therefore, mostly the CellProfiler software is used (Carpenter et al., 2006); however, further software tools are freely available as well (Bougen-Zhukov et al., 2017; Boutros et al., 2015). In addition, the Harmony software of the Opera Phenix HCS system was employed to deduce 1,300 cellular features (Nyffeler et al., 2020; Willis et al., 2020). In the CellProfiler analysis pipeline, image segmentation based on the Hoechst channel is carried out as a first step, followed by definition of the cytoplasm based on the SYTO14 channel and subsequent calculation of morphological features in the different channels.

Data aggregation on the median and normalization by *Z* transformation is suggested in the Bray protocol and is adopted by most authors. The fingerprint thus consists of an n-dimensional vector of the *Z* scores of all (n) features (Figure 4B). Fingerprints can be visualized as heatmaps or line plots, thereby allowing a direct visual comparison of resulting phenotypic changes upon treatment.

For fingerprint generation, all features can be used (Gustafsdottir et al., 2013) or an optional feature selection step may be included in the analysis pipeline to exclude features that carry no information (median absolute deviation [MAD] close to 0) or are highly redundant (Pearson correlation >0.9 or >0.95) (Hughes et al., 2020; Warchal et al., 2020). In addition to these technical ways of feature selection, methods to reduce data dimensionality such as PCA or factor analysis help to save computing time while preserving the variance in datasets.

To increase the robustness of the assay, a reproducibility measure as feature selection criterion may be applied

Figure 4. Morphological profiling using the Cell Painting Assay

(A) Cell Painting Assay (CPA) workflow. Compound-treated cells are stained for mitochondria, actin, Golgi, plasma membrane (PM), nucleoli/RNA, ER, and nucleus prior to automated image acquisition and analysis. Hundreds of numerical features are extracted to generate morphological profiles. Profiles are directly compared or dimensionality reduction is performed to assist further analysis.

(B) Fingerprint generation. Cellular features are quantified on a cell-by-cell basis and are aggregated on image and well level by calculation of the median (random numbers are shown). Data are then normalized by *Z* transformation. The *Z* score thereby represents a measure to what extent the median of compound-treated cells deviates from the median of DMSO-treated cells (x-fold the median absolute deviation [MAD] of DMSO-treated cells). Fingerprints may contain all features calculated during image analysis, or the feature set might has been reduced to exclude features that carry no information, are highly redundant, or highly variable. The fingerprint is represented by a feature vector of all features' *Z* scores and can be visualized as a color-coded heatmap. WGA, wheat germ agglutinin; ConA, concanavalin A.

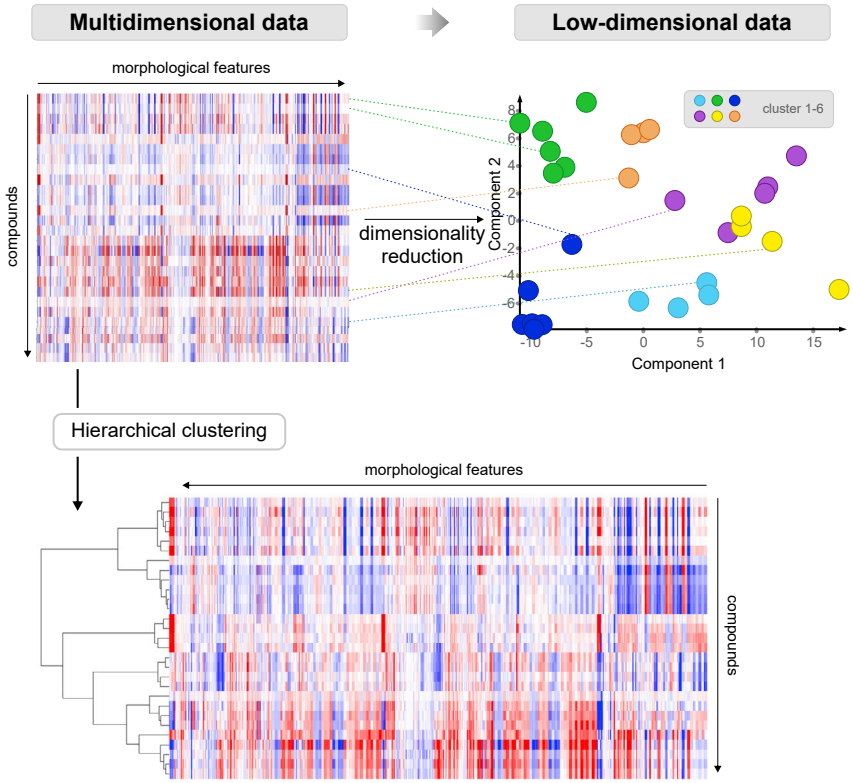


Figure 5. Approaches for bioactivity cluster analysis

Multivariate morphological profiles are transformed into low-dimensional profiles using dimensionality reduction, thus removing irrelevant or redundant features but retaining the variance of the datasets. Multivariate profiles can be directly subjected to hierarchical clustering, or clusters can be detected upon dimensionality reduction.

(Woehrmann et al., 2013). To this end, each feature's *Z* score is compared across a set of compounds—for example, a set of landmark compounds could be used to include multiple MoAs and thus cover a broad range of phenotypic changes. If the *Z* score shows a Pearson correlation $\geq$0.8 between repeats, the respective feature is considered robust. Thereby, for instance, a set of 579 robust features was defined (Christoforow et al., 2019; Foley et al., 2020; Laraia et al., 2020; Schneidewind et al., 2020; Zimmermann et al., 2019).

Fingerprint analysis

To generate knowledge from the morphological fingerprints, different data-mining approaches have been applied. It is of high interest to reveal whether a compound shows some bioactivity or not, particularly in the analysis of newly synthesized compound collections. Thus, a measure to decide the significance of a morphological change upon treatment is required. In addition, methods to compare fingerprints are needed to, for example, group treatments by class.

Compounds initially were defined as active based on the fingerprints' Euclidean distance to the median control fingerprint (Gustafsdottir et al., 2013). Subsequently, the multidimensional perturbation value was developed as a single metric to measure activity and similarity of treatments in high-throughput multidimensional screens (Gerlach et al., 2019; Gerry et al., 2016; Hippman et al., 2020; Hutz et al., 2013; Melillo et al., 2018; Nelson et al., 2016; Wawer et al., 2014). In this method, two treatment-specific subsets are compared (e.g., compound versus dimethyl sulfoxide [DMSO] or different compounds). To this end, several replicates are combined in a matrix, PCA is carried out, and components are scaled by the proportion of the total variance that they explain. In this adjusted PCA space, the Mahalanobis distance is calculated between the two treatment groups. The mp value is then calculated by permutating treatment labels 1,000 times (disallowing the original label configuration) and calculating Mahalanobis distances each time. The mp value is equivalent to the percentage of permuted Mahalanobis distances that are greater than the original. Usually, an mp value of less than 0.05 is deemed to be significant for a treatment to be different from another or for a compound to be active.

Hughes et al. (2020) used the Mahalanobis distance between compound-treated cells versus DMSO controls for primary hit detection in their different esophageal adenocarcinoma (EAC) lines. To this end, they used the first 15 principal components, which explain approximately 90% of the variation in the data across each cell line. Hits are defined somewhat arbitrarily as having a Mahalanobis distance of greater than 1,500 from the DMSO controls.

In an alternative approach to describe compound activity in the CPA, an induction value was introduced. Induction is calculated directly from the fingerprints by a feature-by-feature comparison of *Z* scores (Christoforow et al., 2019; Foley et al., 2020; Laraia et al., 2020; Schneidewind et al., 2020; Zimmermann et al., 2019). A significant change was defined as a deviation of more than three times the MAD from the median of the DMSO controls. The induction value is then determined for every compound as the fraction of significantly changed features (as a percentage). An induction of 5% or higher was considered a valid indication that the morphological change produced by the compound is meaningful.

Hierarchical clustering approaches are widely used to categorize perturbations according to their MoA and are especially powerful in providing an easy visualization of results as clustering trees, thereby grouping the most similar phenotypes together (Figure 5). Either the clustering is performed directly on the morphological fingerprints using the cosine distance and single linkage (Gustafsdottir et al., 2013), or Pearson or sometimes Spearman correlations of Mahalonobis distances between treatment pairs are used to generate the clusters (Gerlach et al., 2019; Hughes et al., 2020). Other visualization methods include correlation plots of the first two or three principal components after PCA (Figure 5) or t-distributed stochastic neighbor embedding analysis. Similar phenotypes are then located more closely

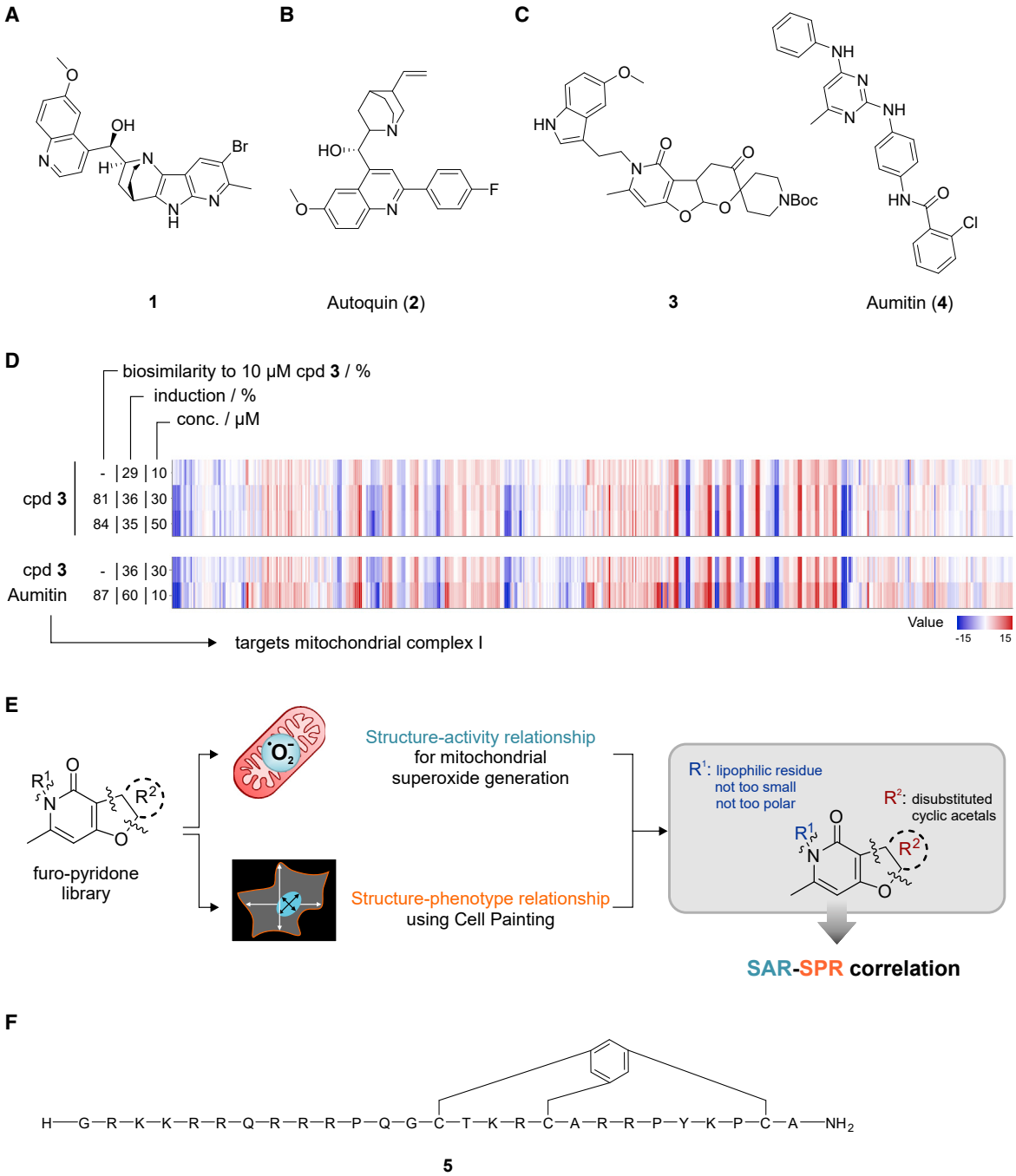


Figure 6. Examples of compounds profiled using CPA to map their MoA

(A and B) Chemical structures of cinchona derivative **1** (A) and autoquin (**2**) (B).

(C–E) CPA reveals mitochondria-targeting activity for furo-pyridones. (C) Chemical structures of furo-pyridone derivative **3** and aumitin (**4**). (D) Furo-pyridone derivative **3** displayed dose-dependent activity in CPA and similarity to aumitin (**4**), which targets mitochondrial complex I. Inhibition of complex I was also detected for **3**. (E) Increase in mitochondrial superoxide by **3** was employed to study the structure-activity relationship (SAR). A structure-phenotype relationship (SPR) was obtained for the same compound collection, which paralleled the SAR (Christoforow et al., 2019).

(F) Chemical structure of the bicyclic peptide **5** that targets RbAp48.

together than more dissimilar phenotypes in the two- or three-dimensional space.

Currently there is no standardized way of analyzing Cell Painting data. The data-mining step is an especially active area of research. It is to be expected that no "gold standard" will emerge but that the analysis workflows will also in the future be dependent on the specific application of the method.

Identification and characterization of small-molecule bioactivity by the CPA

The CPA has successfully been employed to identify MoAs or even molecular targets of small molecules. Foley et al. (2020) identified submicromolar pseudo-natural-product autophagy inhibitors by merging the cinchona alkaloid scaffold with the indole ring system to generate indocinchona alkaloids (see **1**, Figure 6A). Investigation of these compounds in the CPA indicated high fingerprint similarity with two common reference compounds. Fingerprint similarity was found with SAR405, an inhibitor of the PI3K VPS34 that is linked to autophagy. Inhibition of VPS34 was validated by a direct kinase assay ($IC_{50} \sim 350$ nM). Target engagement was proved using a cellular thermal shift assay, which showed a compound-induced thermal stabilization of VPS34. This study underlines the potential of CPA for target identification of small molecules. Similarly, the target hypothesis for the bacterial toxin protein phenomycin could be confirmed using a combination of CPA and direct measurement of cellular protein synthesis. The morphological profile of phenomycin was highly similar to those of inhibitors of the eukaryotic ribosome such as cycloheximide, emetine, and homoharringtonine. Inhibition of protein synthesis was confirmed using a puromycin-based assay that quantifies newly synthesized proteins (Hansen et al., 2020).

Laraia et al. (2020) discovered the cinchona alkaloid-derived autophagy inhibitor autoquin (Figure 6B) . All attempts to find the target via proteomic approaches failed. CPA revealed high similarity between the morphological fingerprints of autoquin and perphenazine (a non-selective GPCR ligand), loperamide (an opioid receptor agonist), and toremifene (an estrogen receptor ligand), which are known lysosomotropic compounds. The lysosomotropic action of autoquin was confirmed as well as the functional inhibition of sphingomyelinase and acid ceramidase. These findings highlight the suitability of the CPA for the identification of non-protein targets of small molecules, for which proteomic approaches are ineffective.

Likewise, the work of Christoforow et al. (2019) on pseudo-natural products showed that CPA can reveal an underlying mechanism as a common MoA. A collection of pyrano-furo-pyridones was analyzed by CPA. Fingerprint similarities to a plethora of reference compounds were identified to be involved in signaling pathways, such as Wnt or Hedgehog, or metabolic regulation, such as glucose uptake, autophagy, or mitochondrial respiration. As reactive oxygen species (ROS) might be involved in all of the pathways, the pyrano-furo-pyridones were analyzed concerning modulation of ROS formation. The pyrano-furo-pyridone **3** (Figure 6C) indeed induced mitochondrial superoxide formation and is a structurally novel inhibitor of mitochondrial complex I (Figures 6D and 6E). Thus, compounds with different annotated cellular targets may cause shared physiological responses if the cellular targets are involved in a common MoA.

CPA was used to identify compounds that showed similar morphological fingerprints compared with the iron chelator deferoxamine (Schneidewind et al., 2020). Interestingly, besides other iron-chelating compounds, inhibitors of topoisomerase, poly(ADP-ribose)-polymerase, lysine-specific histone demethylase 1, or folic acid analogs and DNA-intercalating agents displayed related morphological profiles. Closer analysis revealed modulation of the cell cycle as a unifying mechanism either via inhibition of cell-cycle regulators, the iron dependence of cell-cycle regulatory enzymes, or interference with DNA synthesis. Hierarchical clustering of morphological fingerprints nevertheless revealed subclusters that matched with the individual mechanism of action. In addition, CPA successfully identified novel cell-cycle modulators by screening an in-house library and searching for fingerprints with similarity to the fingerprint of deferoxamine. Likewise, screening of an in-house library revealed a new class of tubulin-targeting cyclic sulfonamides as novel mitotic inhibitors upon searching with the profiles of known tubulin binders (Zimmermann et al., 2019). The application of CPA has recently extended to investigation of new drug modalities (Valeur et al., 2017): CPA demonstrated bioactivity for a cell-permeable bicyclic peptide **5** (Figure 6F) that binds to retinoblastoma-binding protein RbAp48 and inhibits its interaction with the scaffold protein MTA1. Profile comparison revealed similarity to reference compounds that increase the levels of p53, which was also confirmed for the RbAp48-targeting peptide and is in line with epigenetic regulation ('t Hart et al., 2020).

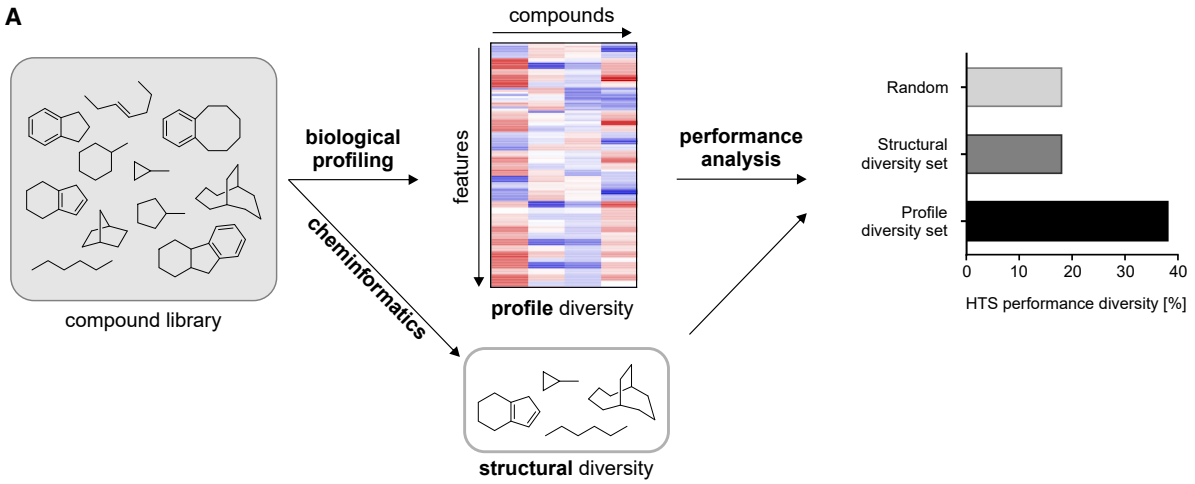
These examples demonstrate that it is possible to employ CPA as a screening tool to search for different chemotypes with desired MoA through comparing profiles with the data recorded for reference compounds. Following the same rationale, compounds could also be excluded from hit sets if they show similarity to profiles of unwanted mechanisms of action. For the inference of targets or mode of actions by CPA, the comparison with well-annotated reference libraries is of utmost importance. Commonly used reference sets include known drug collections such as the LOPAC (Library of Pharmacologically Active Compounds) or Prestwick Chemical library. However, as many known drugs display polypharmacology, which is often not reflected in the annotation of the compound, this deduction step may be complicated. Inclusion of target-specific chemical probes in profiling campaigns might thus offer additional benefit for CPA analysis. Likewise, integrating activity annotation, for example from the ChEMBL database in CPA analysis, could be a helpful step toward an efficient MoA disclosure.

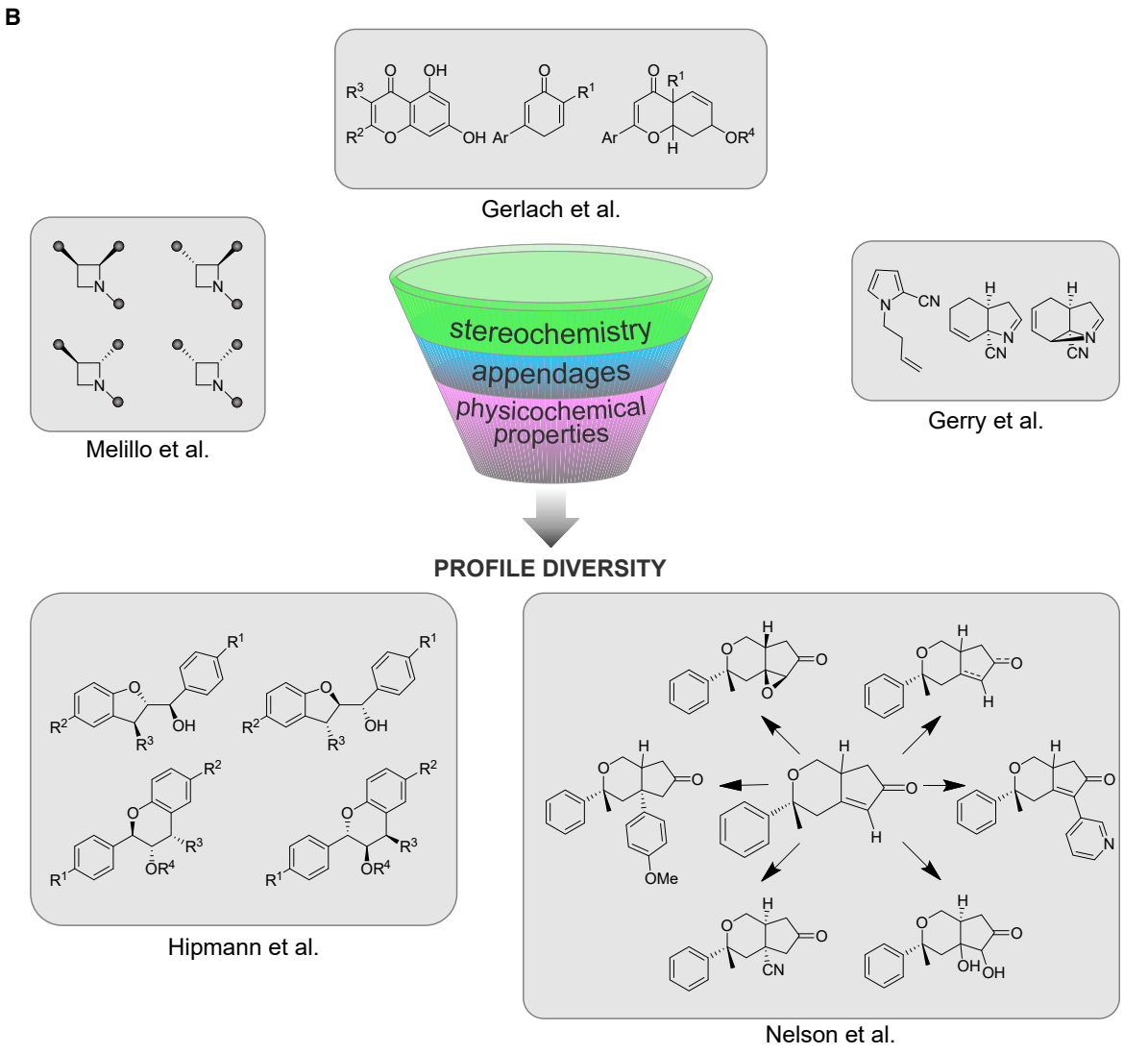
Relating phenotype to structure

Morphological fingerprints obtained via CPA can reveal structure-phenotype relationships (SPR) and even qualitative structure-activity relationships (SAR). Analogous to SAR, SPR describes differences in phenotypes brought about by structural features of compounds. Using dose-dependent measurements, Christoforow et al. (2019) found that most compounds showed an increase in the number of changed cellular features with increasing concentration. Potency trends revealed by SPR correlated with trends in SAR obtained in a functional assay (Figure 6E). Thus, through identification of SPRs, CPA can guide synthesis decisions to inform hit expansion and optimization.

Toward performance-diverse libraries

Efforts to maximize the efficiency of screening collections for cell-based and biochemical screening were guided by CPA and defined early applications of this assay for analysis of small-molecule bioactivity. Contrary to the assumption that chemical structure diversity automatically leads to a diverse biological performance of a library, CPA results suggested the use of morphological profile diversity to assemble libraries with





(legend on next page)

diverse assay-performance patterns (Figure 7A) (Wawer et al., 2014). Selection of compounds with diverse CPA profiles led to higher screening performance diversity than random selection or selection of diverse chemical structures. Similarly, the extent of morphological changes can be correlated to structural features of chemical compounds. Thus, triads of constitutional isomers or functionalized stereoisomeric azetidines or of reduced flavones were analyzed by CPA (Figure 7B) (Gerlach et al., 2019; Gerry et al., 2016; Melillo et al., 2018). Differential activity within the compound series could be detected with higher concentrations of compounds being associated with larger activity scores; for example, within the collection of isomeric triads, aziridines and amines were enriched in the active compound set as compared with pyrroles (Gerry et al., 2016).

By calculating Pearson correlation coefficients, differences in the morphological profiles of active compounds could be detected implying differences in MoA. Also, differences in morphological profiles arising from stereoisomerism were revealed—for example, profiles of *cis*-isomer azetidines were more highly correlated with each other than with *trans*-isomer profiles and vice versa (Melillo et al., 2018). PCA showed that appendage and stereochemical diversity synergize to increase performance diversity (Gerlach et al., 2019). In a recent follow-up study, the correlation of biological activity with chemical properties of the compounds was analyzed more extensively (Hippman et al., 2020). To this end, 82 novel nitrogen-containing benzopyran and benzofuran flavonoids, which displayed considerable diversity of physicochemical properties, were profiled. A correlation of activity values with clogP and tPSA was found for some series only. Upon pairwise comparison of profiles of 26 scaffold pairs and 31 stereoisomeric pairs, only 12 out of 31 stereoisomeric pairs were significantly dissimilar. Thus, variation in the scaffold seems more important for biological performance diversity than stereochemical diversity, and including molecules with diverse physicochemical properties in a library will likely maximize its performance diversity. Together, these results highlight that CPA has great potential to guide the design and characterization of the performance diversity of compound sets. Performance-diverse libraries potentially yield improved hit rates compared with larger compound libraries, and thus have the promise to greatly advance probe discovery efforts. Future trends will likely include algorithms to predict whether a compound is supposed to be active or not solely based on the chemical features of a compound via comparison with the chemical features of a profiled reference set. Published CPA image sets have already been used for this purpose (Bray et al., 2017). However, as experimental proof of the predicted compound activity is lacking, it is currently not possible to evaluate this approach further (Hofmarcher et al., 2019).

Personalized medicine

CPA has been employed to profile clinically relevant cancer cell lines with the goal of advancing drug repurposing and pharmacogenomic studies. To identify novel compounds with selective activity against EAC, the phenotypic response of six EAC and two tissue-matched control cell lines was profiled against nearly 20,000 compounds using CPA combined with machine learning (Hughes et al., 2020). From the annotated compound collection, HDAC inhibitors, microtubule disruptors, and antimetabolites were identified as hits in all of the EAC cell lines. Selective activity of antimetabolites against EAC cell lines relative to tissue-matched controls was confirmed using dose-response measurements in a nuclei count assay. From the diverse proprietary screening set, a number of hits could be retrieved that showed selective activity on a fraction of the EAC cell lines. Interestingly, these compounds did not cluster with the reference library and may thus be targeting novel esophageal cancer biology.

Warchal et al. (2020) profiled a set of eight genetically distinct human breast cancer cell lines, which were classified into three clinical subtypes, by CPA. The TCCS (Theta Comparative Cell Scoring) method was used to identify differential phenotypic responses between pairs of cell lines upon treatment. Among the top scoring compounds, a series of four known serotonin receptor modulators was revealed. Further analysis by expression profiling and pathway network analysis revealed regulators of the cell cycle as downregulated genes and an upregulation of the tumor necrosis factor receptor 1 (TNFR1) signaling pathway, suggesting an induction of apoptosis via TNF signaling. Thus, a potential drug-repurposing opportunity for serotonin modulators in breast cancers has been revealed by CPA.

The two studies described above used established cancer cell lines in CPA analysis. Ultimately, to reach toward a truly personalized approach with direct clinical relevance, it will be necessary to profile primary patient-derived cells. The existing results and developed methodologies, however, make it very probable that this step will be taken soon.

Environmental toxicology

CPA has been evaluated for the characterization of biological activity and potency of environmental chemicals for potential use in *in vitro* safety assessments.

A set of 462 chemicals from the ToxCast library were screened in U2OS cells in concentration-response mode. Nyffeler et al. (2020) then used concentration-response modeling to determine *in vitro* thresholds for chemical bioactivity. Therefore, 1,300 features were grouped by channel, module, and compartment into 49 categories. If 30% of features within one category were responsive to treatment, that category was defined as affected. Median benchmark concentrations of all affected categories were calculated, and the most sensitive category was defined as the point of departure. Point-of-departure values were then converted to administered equivalent doses (AEDs) using reverse dosimetry for the comparison with standard toxicology measurements. In many instances (68%), the CP-derived AEDs were either more conservative than or comparable with

Figure 7. Performance-diverse libraries

(A) Biological profiling can assist the generation of performance-diverse screening libraries. Wawer et al. (2014) compared compound subsets with maximal structural diversity and maximal biological profile diversity regarding their performance in various HTS campaigns. While the structural diversity subset did not perform better than a random subset, the profile diversity subset showed increased hit rates in a broad range of assays.

(B) Different libraries with diverse chemical properties were analyzed using CPA. The study of systematic variations of chemical features on morphological profile diversity may inform synthesis decisions in the future.

the *in vivo* effect values. Later, the method was transferred to a set of biologically distinct cancer cell lines, which turned out to show quite similar results when treated with 14 reference chemicals (Willis et al., 2020). CPA was found to be an efficient, cost-effective, and reproducible method for characterizing the biological activity and potency of environmental chemicals.

DISCUSSION AND OUTLOOK

Cellular profiling approaches have become indispensable sources of information for a holistic view on a given cellular state and have been extended in recent years to phenomics approaches, i.e., morphological profiling. Whereas analysis and interpretation of gene or protein expression profiling data in a biological context are well established, comprehension of morphological profiles is only emerging. First lessons concerning morphological profiling of small molecules have been learned, and the stage is set for extension to new horizons.

Morphological profiling is feasible for and accessible to numerous laboratories, and this applies to both the wet-lab procedure and computational analysis. The use of fluorescent dyes to stain cellular components instead of gene-encoded fluorescent tags simplifies the experimental protocol and the prework, although it is limited by the number of available dyes and the number of channels for their recording. Several algorithms have been developed for image-based feature extraction and have been made available to the public. Computational tools, such as machine learning and statistics methodologies, are well established and have been implemented in the data analysis workflow to map profile similarities and bioactivity clusters, and to predict MoA. In principle, profiling of large numbers of small molecules, assessment of their bioactivity, assignment to a given bioactivity cluster, and, thus, prediction of their targets and MoA are possible. Ideally, these predictions would subsequently be experimentally validated. However, in reality, unbiased mapping of bioactivity for each bioactive compound will remain challenging and limited in the near future. This is partly due to the limited target space (Finan et al., 2017) that is occupied by small molecules with known MoA. In addition, interpretation of morphological profiles and biological similarities even for annotated, well-explored reference compounds often is not conclusive. In fact, there is a crucial difference between morphological profiling and omics approaches: whereas the number of protein-coding genes and proteins that can be expressed has been determined for a variety of cases, the total number of relevant morphology-related parameters is not known. In addition, knowledge of regulation of gene and protein expression is available and thus can link altered gene or protein expression to upstream regulators. Unfortunately, a similar bottom-up deconvolution is much more difficult for connecting altered morphological parameters (as equivalent to altered gene or protein abundance) to upstream regulation. Nonetheless, informative and valid conclusions can be deduced from the reported morphological profiling approaches and are invaluable for further explorative work.

Detected cellular components

In morphological profiling, usually only few selected cellular components or targets are analyzed and their number is limited due to spectral overlap of dyes if concurrent visualization is used. The CPA is one of the most efficient and well-explored methods to obtain maximal information on the cellular state (Rose et al., 2018). Several approaches go beyond the restriction to available fluorescent probes and channels and use genetic methods to label and visualize endogenous proteins. However, this is not applicable to all cell lines and often requires extensive prework to generate the cell lines and the parallel use of several cell lines (Cox et al., 2020; Kang et al., 2016). Of note, some of these cell lines are commercially available (Cox et al., 2020).

The MoA prediction accuracy depends on the type and the number of employed markers (Kang et al., 2016), although the perturbing activity of small molecules can be detected even when the respective target is not visualized. For example, including a tubulin stain assists the detection of tubulin-targeting agents. However, this is feasible also in approaches that lack tubulin visualization such as using stains for DNA only (Futamura et al., 2012) or in combination with actin (Breinig et al., 2015), phospho-histone H3, and EdU (Young et al., 2008), among others (Gustafsdottir et al., 2013; Perlman et al., 2004). The use of a tubulin stain, however, apparently allows differentiation between microtubule stabilizers and destabilizers (Adams et al., 2006; Caie et al., 2010). A DNA stain alone enables classification of various reference compounds according to their MoA, and additional stains increase the accuracy of clustering (Reisen et al., 2015; Rose et al., 2018). Whereas one or few stains usually group compounds in broad clusters, detection of more cellular components may enable differentiation between distinct MoAs (Schneidewind et al., 2020).

Compound concentrations

The MoA of compounds depends on the employed concentration, the incubation time, and the genotype of the cell line (Adams et al., 2006). Compounds often display similar profiles at different concentrations. However, small molecules can provoke distinct morphological changes at different concentrations. Microtubule stabilizers could be distinguished from destabilizers only at high concentrations (Adams et al., 2006). Dissimilar profiles at different concentrations usually result from polypharmacology (Gerry et al., 2016). Cytochalasin A clustered at low concentrations with actin-targeting agents, while its profile at high concentration was similar to tubulin-interfering compounds (Adams et al., 2006). Rotenone clusters at low concentration with mitochondrial toxins, whereas it is part of the tubulin cluster at high concentrations (Cox et al., 2020). Thus, analyzing the profiles of the separate concentrations may provide additional bioactivity information (Gerry et al., 2016). Compared with unexplored small molecules, reference compounds often display morphological activity at lower concentration as they may be optimized regarding potency. If the respective target is absent in the used cell line, compounds are active at higher concentration due to off-target activity (Cox et al., 2020).

Cell lines

Compounds may lead to different morphological perturbations and profiles in cell lines of different origin and with different genotype. These differences may range from inactivity in certain cell lines (Cox et al., 2020) to completely dissimilar profiles (Tanaka et al., 2005), and are most likely attributable to the lack of the

engaged target, different target abundance, post-translational modifications, and pathway redundancy. Thus, compounds may have unique signatures or combinations of altered features across multiple cell lines (Adams et al., 2006). MoA prediction accuracy is cell-line dependent and increases when multiple cell lines are used (Futamura et al., 2012; Rose et al., 2018). Profiles may compile all changes induced in different cell lines to obtain comprehensive profiles; however, additional cell lines should be included if they expand the kind of detected MoA—for example, glucocorticoids and mitochondrial toxins cause more predictive phenotypes in A549 cells than in HepG2 or WPMY1 cells (Cox et al., 2020).

Morphological profiling can be applied not only to human cell lines and has been adapted to different bacterial strains (Htoo et al., 2019; Lamsa et al., 2016; Nonejuie et al., 2013) and yeast (Ohnuki et al., 2010; Ohya et al., 2005). These approaches will extend the targeted bioactivity space of small molecules to non-mammalian targets and may provide valuable starting points in the quest for antibiotics (Zoffmann et al., 2019). Moreover, the next step of morphological profiling in physiologically more relevant systems such as organoids has already been taken (Betge et al., 2019), and platforms have been developed to track changes over time (Cox et al., 2020; Kang et al., 2016).

Clustering and identified bioactivity clusters

Morphological profiles can be analyzed for similar patterns to define clusters of bioactivity. An unsupervised machine-learning approach is employed when the expected output is not known (Grys et al., 2017) and has been used in various studies (Gustafsdottir et al., 2013; Young et al., 2008). However, for correct MoA assignment to detected profile groups, clusters have to be evaluated and validated (Grys et al., 2017). To avoid misinterpretation, the bioactivity of each cluster is ideally defined by a rich set of reference compounds, which are structurally different but have common annotated bioactivity. It is expected that small molecules with similar activity should cluster together. One reference compound, even if profiled at different concentrations, is not sufficient to reliably draw biological conclusions without prior profiling knowledge. Structurally different landmark compounds for the same target should be enriched in the respective cluster to claim bioactivity. Usually, not all reference compounds with common target are grouped into the same cluster. Polypharmacological compounds will be assigned to the phenomically most dominant activity only or may not be linked to any activity cluster if reference compounds with similar profiles were not assayed (Cox et al., 2020). Morphological profiles of such compounds are complex, reflect modulation of multiple targets, and require deconvolution of their influence on specific targets (Cox et al., 2020). Apparently "misclassified" compounds most likely have different targets, and the assignment to a bioactivity cluster is due to activity different from the nominal target. In fact, the occurrence of such outliers within clusters is very informative and directly points toward an additional MoA of reference compounds. There is a high demand to annotate available reference compounds in as much detail as possible, e.g., in a collective community effort. Several studies have already implemented structure-based target prediction (Young et al., 2008) or ChEMBL bioactivity data (Cox et al., 2020; Hofmarcher et al., 2019) to better correlate the observed phenotypes and targets. However, only rarely is complete bioactivity information available in one database, and non-protein targets are often disregarded in structure-based target prediction approaches, e.g., iron chelators (Schneidewind et al., 2020).

Detected bioactivity strongly depends on the screened compounds. This is particularly important when comparing the bioactivity clusters that were identified by means of reference compounds (Table S3). The ability of the described approaches for morphological profiling to detect given bioactivity can only be assessed by careful inspection and comparison of the screened references between the different methods. Of note, not all reference compounds induce morphological changes and, thus, may be inactive in morphological profiling. Lack of activity may be attributed to the employed biological system (target not expressed or target present only in different species), lack of stimulus (that activate the modulated process), treatment time, concentration, or lack of tractability of phenotypes with the visualized cellular components.

Handling multidimensional data often requires dimensionality reduction to enable data reduction by retaining most of the information. This approach is employed for clustering of compounds with similar profiles. However, clusters defined by this method do not always reveal the underlying MoA. Whereas it is straightforward to assign the MoA to a cluster of compounds with common annotated target or MoA, it may be particularly challenging to define the bioactivity of a cluster that is composed of small molecules with no obvious common target or activity. The lower-dimensional feature space may be difficult to interpret and thus may not be suitable to foster understanding or interpretation of the detected cluster (Bougen-Zhukov et al., 2017). It is essential to define MoA clusters based on as many as possible compounds with common MoA. However, annotated compounds are limited in number, and often only few reference compounds for a given target or MoA are employed. Caution should be taken when assigning an MoA to a cluster based on one or few landmark compounds, and one should carefully consider already published MoA clusters and compounds associated therewith. Single outliers, however, should not question the MoA of a cluster if the cluster is composed of a sufficient number of structurally diverse reference compounds with common annotated targets or MoA (Cox et al., 2020).

Bioactivity clusters may not be based on single targets but rather on similar MoA (DNA synthesis inhibition) or pathway modulation (PI3K-mTOR). Several clusters were identified in most studies and are considered as "low-hanging fruits," e.g., tubulin, HDAC, mTOR, HSP90, proteasome, and DNA synthesis (see Table S3) (Kang et al., 2016; Moffat et al., 2014). Additional reported clusters contain compounds targeting actin, protein synthesis, kinases (MEK, p38, Aurora kinase), HMGCR, Na^+/K^+-ATPases, and the mitochondrial proton gradient (Table S1).

Profiling approaches are promising strategies for deorphanization of dark chemical matter (DCM), i.e., small molecules that show inactivity in more than 100 target- and cell-based assays (Wassermann et al., 2015, 2017). DCM inactivity may stem from the bias in employed screened targets and processes (Wassermann et al., 2015). DCM may be highly specific and non-promiscuous compounds with unprecedented biological profiles (Wassermann et al., 2015, 2017). Thus, morphological profiling in mammalian cells but also in bacteria or yeast is

suitable for the discovery of unique biological profiles using DCM across different species and for covering very broad target space that is not restricted to protein targeting.

Data integration

Morphological profiling can be regarded as complementary to omics technologies and adds bioactivity information to the activity catalog of small molecules that may encompass transcriptomics, proteomics, metabolomics, and low-content phenotypic or target-centered data. Integration of these data-rich approaches (Wawer et al., 2014) with target-based data, e.g., inhibition data for kinase inhibitors (Woehrmann et al., 2013) and target prediction based on chemical similarity (Young et al., 2008), will further inform the bioactivity-to-cluster assignment.

Structure-activity/phenotype correlation

Phenotypic profiles correlate more significantly with the predicted activity and target than morphological profiles and chemical similarity (Young et al., 2008). Thus, a structurally diverse compound collection will not necessarily lead to diversity in bioactivity (Kremb and Voolstra, 2017). The SPR can be deduced from morphological profiling (Christoforow et al., 2019) and may be physiologically more relevant, as small structural changes can influence various cellular processes (Gerry et al., 2016; Kremb and Voolstra, 2017; Melillo et al., 2018). It is essential to have quantitative measures for bioactivity and profile similarity, e.g., the proportion of altered parameters (Christoforow et al., 2019; Foley et al., 2020; Laraia et al., 2020; Schneidewind et al., 2020; Zimmermann et al., 2019). However, SPR data may be difficult to interpret, as time and dose may influence individual parameters differently; thus, extracting and analyzing specific features instead of whole profiles may be advantageous (Boyd et al., 2020).

Resource sharing

The profiling community would benefit from sharing data, methods, and software code (Caicedo et al., 2017), and initiatives for data sharing for reuse and cross-analysis (Chessel and Salas, 2019) (Bray et al., 2017) and studies derived thereof (Hofmarcher et al., 2019) have been already launched. Of note, the Broad Institute disclosed a Cell Painting image dataset for 30,616 compounds (Bray et al., 2017) and launched the academic-industry cell imaging consortium JUMP-CP (Joint Undertaking in Morphological Profiling with Cell Painting) to accelerate drug discovery by creating reference datasets for genetic and small-molecule perturbations using Cell Painting.

In conclusion, morphological profiling has gained a firm place in the analysis of small molecules for bioactivity and can deliver essential information in the evaluation of unexplored compounds by demonstrating morphological activity, predicting the MoA or even target. Moreover, it provides valuable evidence for polypharmacology and may link landmark compounds to a yet unanticipated bioactivity. The current challenge is to comprehensively annotate reference and other bioactive compounds, extend the number of predictable modes of action and targets, and explore outliers and compounds with unique profiles. In addition, it would be very interesting to evaluate the bioactivity of DCM using morphological profiling, which might yield unique profiles and, thus, lead to unprecedented targets and MoAs. Harvesting these "high-hanging" fruits promises to provide mechanistic understanding underlying morphological profiles.

SIGNIFICANCE

Bioactive small molecules are indispensable for medicinal and chemical biology research and are usually identified in biased, i.e., biological context-dependent, assays. Compounds are often characterized in a small number of target- and/or cell-based assays to confirm their activity and address off-target effects. Profiling approaches can detect hundreds of altered parameters that describe a perturbed cellular state and can assess activity of small molecules on different targets, pathways, or biological processes simultaneously. Similar profiles have been linked to similar targets or modes of action. Morphological profiling is an accessible and high-throughput technique to assess morphological changes and enables detection of bioactivity and prediction of target or mode of action early on in the development of compound collections. Application of this approach may further guide the design and synthesis of new bioactive compounds by exploring the structure-phenotype relationship. Various mode-of-action clusters have been mapped and employed to assign bioactivity to unexplored small molecules along with detection of unanticipated activity for well-characterized agents. Thus, morphological profiling holds promise to provide a more holistic view on the bioactivity of thousands of compounds and to prioritize compounds that are more selective with regard to the targeted bioactivity space.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.chembiol.2021.02.012.

ACKNOWLEDGMENTS

Research at the Max Planck Institute of Molecular Physiology was supported by the Max Planck Society (Germany) and was co-funded by the European Union (Drug Discovery Hub Dortmund, EFRE-0200481).

AUTHOR CONTRIBUTIONS

S.Z., S.S., and H.W. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Adams, C.L., Kutsyy, V., Coleman, D.A., Cong, G., Crompton, A.M., Elias, K.A., Oestreicher, D.R., Trautman, J.K., and Vaisberg, E. (2006). Compound classification using image-based cellular phenotypes. Methods Enzymol. *14*, 440–468.

Awale, M., and Reymond, J.L. (2019). Web-based tools for polypharmacology prediction. Methods Mol. Biol. *1888*, 255–272.

Betge, J., Rindtorff, N., Sauer, J., Rauscher, B., Dingert, C., Gaitantzi, H., Herweck, F., Miersch, T., Valentini, E., Hauber, V., et al. (2019). Multiparametric phenotyping of compound effects on patient derived organoids. bioRxiv, 660993, https://doi.org/10.1101/660993.

Boland, M.V., and Murphy, R.F. (2001). A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells. Bioinformatics *17*, 1213–1223.

Bougen-Zhukov, N., Loh, S.Y., Lee, H.K., and Loo, L.H. (2017). Large-scale image-based screening and profiling of cellular phenotypes. Cytometry A *91*, 115–125.

Boutros, M., Heigwer, F., and Laufer, C. (2015). Microscopy-based high-content screening. Cell *163*, 1314–1325.

Boyd, J., Fennell, M., and Carpenter, A. (2020). Harnessing the power of microscopy images to accelerate drug discovery: what are the possibilities? Expert Opin. Drug Discov. *15*, 639–642.

Bray, M.A., Gustafsdottir, S.M., Rohban, M.H., Singh, S., Ljosa, V., Sokolnicki, K.L., Bittker, J.A., Bodycombe, N.E., Dancik, V., Hasaka, T.P., et al. (2017). A dataset of images and morphological profiles of 30 000 small-molecule treatments using the Cell Painting assay. Gigascience *6*, 1–5.

Bray, M.A., Singh, S., Han, H., Davis, C.T., Borgeson, B., Hartland, C., Kost-Alimova, M., Gustafsdottir, S.M., Gibson, C.C., and Carpenter, A.E. (2016). Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat. Protoc. *11*, 1757–1774.

Brecht, K., Riebel, V., Couttet, P., Paech, F., Wolf, A., Chibout, S.D., Pognan, F., Krahenbuhl, S., and Uteng, M. (2017). Mechanistic insights into selective killing of OXPHOS-dependent cancer cells by arctigenin. Toxicol. Vitro *40*, 55–65.

Breinig, M., Klein, F.A., Huber, W., and Boutros, M. (2015). A chemical-genetic interaction map of small molecules using high-throughput imaging in cancer cells. Mol. Syst. Biol. *11*, 846, https://doi.org/10.15252/msb.20156400.

Byrne, R., and Schneider, G. (2019). In silico target prediction for small molecules. Methods Mol. Biol. *1888*, 273–309.

Caicedo, J.C., Cooper, S., Heigwer, F., Warchal, S., Qiu, P., Molnar, C., Vasilevich, A.S., Barry, J.D., Bansal, H.S., Kraus, O., et al. (2017). Data-analysis strategies for image-based cell profiling. Nat. Methods *14*, 849–863.

Caie, P.D., Walls, R.E., Ingleston-Orme, A., Daya, S., Houslay, T., Eagle, R., Roberts, M.E., and Carragher, N.O. (2010). High-content phenotypic profiling of drug response signatures across distinct cancer cells. Mol. Cancer Ther. *9*, 1913–1926.

Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. *7*, R100, https://doi.org/10.1186/gb-2006-7-10-r100.

Chessel, A., and Salas, R.E.C. (2019). From observing to predicting single-cell structure and function with high-throughput/high-content microscopy. Essays Biochem. *63*, 197–208.

Christoforow, A., Wilke, J., Binici, A., Pahl, A., Ostermann, C., Sievers, S., and Waldmann, H. (2019). Design, synthesis, and phenotypic profiling of pyranofuro-pyridone pseudo natural products. Angew. Chem. Int.Ed. Engl. *58*, 14715–14723.

Cox, M.J., Jaensch, S., Van de Waeter, J., Cougnaud, L., Seynaeve, D., Benalla, S., Koo, S.J., Van Den Wyngaert, I., Neefs, J.-M., Malkov, D., et al. (2020). Tales of 1,008 small molecules: phenomic profiling through live-cell imaging in a panel of reporter cell lines. Sci. Rep. *10*, https://doi.org/10.1038/s41598-020-69354-8.

Finan, C., Gaulton, A., Kruger, F.A., Lumbers, R.T., Shah, T., Engmann, J., Galver, L., Kelley, R., Karlsson, A., Santos, R., et al. (2017). The druggable genome and support for target identification and validation in drug development. Sci. Transl. Med. *9*, https://doi.org/10.1126/scitranslmed.aag1166.

Foley, D.J., Zinken, S., Corkery, D., Laraia, L., Pahl, A., Wu, Y.-W., and Waldmann, H. (2020). Phenotyping reveals targets of a pseudo-natural-product autophagy inhibitor. Angew. Chem. Int. Ed. Engl. 12570–12576.

Futamura, Y., Kawatani, M., Kazami, S., Tanaka, K., Muroi, M., Shimizu, T., Tomita, K., Watanabe, N., and Osada, H. (2012). Morphobase, an encyclopedic cell morphology database, and its use for drug target identification. Chem. Biol. *19*, 1620–1630.

Futamura, Y., Kawatani, M., Muroi, M., Aono, H., Nogawa, T., and Osada, H. (2013). Identification of a molecular target of a novel fungal metabolite, pyrrolizilactone, by phenotypic profiling systems. Chembiochem *14*, 2456–2463.

Gebre, A.A., Okada, H., Kim, C., Kubo, K., Ohnuki, S., and Ohya, Y. (2015). Profiling of the effects of antifungal agents on yeast cells based on morphometric analysis. FEMS Yeast Res. *15*, fov040.

Gerlach, E.M., Korkmaz, M.A., Pavlinov, I., Gao, Q., and Aldrich, L.N. (2019). Systematic diversity-oriented synthesis of reduced flavones from gamma-pyrones to probe biological performance diversity. ACS Chem. Biol. *14*, 1536–1545.

Gerry, C.J., Hua, B.K., Wawer, M.J., Knowles, J.P., Nelson, S.D., Jr., Verho, O., Dandapani, S., Wagner, B.K., Clemons, P.A., Booker-Milburn, K.I., et al. (2016). Real-time biological annotation of synthetic compounds. J. Am. Chem. Soc. *138*, 8920–8927.

Grys, B.T., Lo, D.S., Sahin, N., Kraus, O.Z., Morris, Q., Boone, C., and Andrews, B.J. (2017). Machine learning and computer vision approaches for phenotypic profiling. J. Cell Biol. *216*, 65–71.

Gustafsdottir, S.M., Ljosa, V., Sokolnicki, K.L., Anthony Wilson, J., Walpita, D., Kemp, M.M., Petri Seiler, K., Carrel, H.A., Golub, T.R., Schreiber, S.L., et al. (2013). Multiplex cytological profiling assay to measure diverse cellular states. PLoS One *8*, e80999.

Hansen, B.K., Larsen, C.K., Nielsen, J.T., Svenningsen, E.B., Van, L.B., Jacobsen, K.M., Bjerring, M., Flygaard, R.K., Jenner, L.B., Nejsum, L.N., et al. (2020). Structure and function of the bacterial protein toxin phenomycin. Structure *28*, 528–539 e529.

Hippman, R.S., Pavlinov, I., Gao, Q., Mavlyanova, M.K., Gerlach, E.M., and Aldrich, L.N. (2020). Multiple chemical features impact biological performance diversity of a highly active natural product-inspired library. Chembiochem. https://doi.org/10.1002/cbic.202000356.

Hofmarcher, M., Rumetshofer, E., Clevert, D.-A., Hochreiter, S., and Klambauer, G. (2019). Accurate prediction of biological assays with high-throughput microscopy images and convolutional networks. J. Chem. Inf. Model. *59*, 1163–1171.

Htoo, H.H., Brumage, L., Chaikeeratisak, V., Tsunemoto, H., Sugie, J., Tribuddharat, C., Pogliano, J., and Nonejuie, P. (2019). Bacterial cytological profiling as a tool to study mechanisms of action of antibiotics that are active against acinetobacter baumannii. Antimicrob. Agents Chemother. *63*, https://doi.org/10.1128/AAC.02310-18.

Hughes, R.E., Elliott, R.J.R., Munro, A.F., Makda, A., O'Neill, J.R., Hupp, T., and Carragher, N.O. (2020). High-content phenotypic profiling in esophageal adenocarcinoma identifies selectively active pharmacological classes of drugs for repurposing and chemical starting points for novel drug discovery. SLAS Discov. *25*, 770–782.

Hutz, J.E., Nelson, T., Wu, H., McAllister, G., Moutsatsos, I., Jaeger, S.A., Bandyopadhyay, S., Nigsch, F., Cornett, B., Jenkins, J.L., et al. (2013). The multidimensional perturbation value: a single metric to measure similarity and activity of treatments in high-throughput multidimensional screens. J. Biomol. Screen. *18*, 367–377.

Kang, J., Hsu, C.H., Wu, Q., Liu, S., Coster, A.D., Posner, B.A., Altschuler, S.J., and Wu, L.F. (2016). Improving drug discovery with high-content phenotypic screens by systematic selection of reporter cell lines. Nat. Biotechnol. *34*, 70–77.

Keiser, M.J., Roth, B.L., Armbruster, B.N., Ernsberger, P., Irwin, J.J., and Shoichet, B.K. (2007). Relating protein pharmacology by ligand chemistry. Nat. Biotechnol. *25*, 197–206.

Keiser, M.J., Setola, V., Irwin, J.J., Laggner, C., Abbas, A.I., Hufeisen, S.J., Jensen, N.H., Kuijer, M.B., Matos, R.C., Tran, T.B., et al. (2009). Predicting new molecular targets for known drugs. Nature *462*, 175–U148.

Kremb, S., Muller, C., Schmitt-Kopplin, P., and Voolstra, C.R. (2017). Bioactive potential of marine macroalgae from the central Red Sea (Saudi arabia) assessed by high-throughput imaging-based phenotypic profiling. Mar. Drugs *15*, https://doi.org/10.3390/md15030080.

Kremb, S., and Voolstra, C.R. (2017). High-resolution phenotypic profiling of natural products-induced effects on the single-cell level. Sci. Rep. *7*, 44472, https://doi.org/10.1038/srep44472.

Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N., et al. (2006). The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. Science *313*, 1929–1935.

Lamsa, A., Lopez-Garrido, J., Quach, D., Riley, E.P., Pogliano, J., and Pogliano, K. (2016). Rapid inhibition profiling in Bacillus subtilis to identify the mechanism of action of new antimicrobials. ACS Chem. Biol. *11*, 2222–2231.

Laraia, L., Garivet, G., Foley, D.J., Kaiser, N., Muller, S., Zinken, S., Pinkert, T., Wilke, J., Corkery, D., Pahl, A., et al. (2020). Image-based morphological profiling identifies a lysosomotropic, iron-sequestering autophagy inhibitor. Angew. Chem. Int. Ed. Engl. *59*, 5721–5729.

Ljosa, V., Caie, P.D., Ter Horst, R., Sokolnicki, K.L., Jenkins, E.L., Daya, S., Roberts, M.E., Jones, T.R., Singh, S., Genovesio, A., et al. (2013). Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment. J. Biomol. Screen. *18*, 1321–1329.

Melillo, B., Zoller, J., Hua, B.K., Verho, O., Borghs, J.C., Nelson, S.D., Jr., Maetani, M., Wawer, M.J., Clemons, P.A., and Schreiber, S.L. (2018). Synergistic effects of stereochemistry and appendages on the performance diversity of a collection of synthetic compounds. J. Am. Chem. Soc. *140*, 11784–11790.

Minegishi, H., Futamura, Y., Fukashiro, S., Muroi, M., Kawatani, M., Osada, H., and Nakamura, H. (2015). Methyl 3-((6-methoxy-1,4-dihydroindeno[1,2-c]pyrazol-3-yl)amino)benzoate (GN39482) as a tubulin polymerization inhibitor identified by MorphoBase and ChemProteoBase profiling methods. J. Med. Chem. *58*, 4230–4241.

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

Moret, N., Clark, N.A., Hafner, M., Wang, Y., Lounkine, E., Medvedovic, M., Wang, J., Gray, N., Jenkins, J., and Sorger, P.K. (2019). Cheminformatics tools for analyzing and designing optimized small-molecule collections and libraries. Cell Chem. Biol. *26*, 765–777.

Nelson, S.D., Jr., Wawer, M.J., and Schreiber, S.L. (2016). Divergent synthesis and real-time biological annotation of optically active tetrahydrocyclopenta[c]pyranone derivatives. Org. Lett. *18*, 6280–6283.

Nonejuie, P., Burkart, M., Pogliano, K., and Pogliano, J. (2013). Bacterial cytological profiling rapidly identifies the cellular pathways targeted by antibacterial molecules. Proc. Natl. Acad. Sci. U S A *110*, 16169–16174.

Nussinov, R., Tsai, C.J., and Jang, H. (2019). Protein ensembles link genotype to phenotype. PLoS Comput. Biol. *15*, https://doi.org/10.1371/journal.pcbi.1006648.

Nyffeler, J., Willis, C., Lougee, R., Richard, A., Paul-Friedman, K., and Harrill, J.A. (2020). Bioactivity screening of environmental chemicals using imaging-based high-throughput phenotypic profiling. Toxicol. Appl. Pharmacol. *389*, 114876, https://doi.org/10.1016/j.taap.2019.114876.

Ochoa, J.L., Bray, W.M., Lokey, R.S., and Linington, R.G. (2015). Phenotype-guided natural products discovery using cytological profiling. J. Nat. Prod. *78*, 2242–2248.

Ohnuki, S., Oka, S., Nogami, S., and Ohya, Y. (2010). High-content, image-based screening for drug targets in yeast. PLoS One *5*, e10177.

Ohya, Y., Sese, J., Yukawa, M., Sano, F., Nakatani, Y., Saito, T.L., Saka, A., Fukuda, T., Ishihara, S., Oka, S., et al. (2005). High-dimensional and large-scale phenotyping of yeast mutants. Proc. Natl. Acad. Sci. U S A *102*, 19015–19020.

Perlman, Z.E., Slack, M.D., Feng, Y., Mitchison, T.J., Wu, L.F., and Altschuler, S.J. (2004). Multidimensional drug profiling by automated microscopy. Science *306*, 1194–1198.

Peters, C.E., Lamsa, A., Liu, R.B., Quach, D., Sugie, J., Brumage, L., Pogliano, J., Lopez-Garrido, J., and Pogliano, K. (2018). Rapid inhibition profiling identifies a keystone target in the nucleotide biosynthesis pathway. ACS Chem. Biol. *13*, 3251–3258.

Rees, M.G., Seashore-Ludlow, B., Cheah, J.H., Adams, D.J., Price, E.V., Gill, S., Javaid, S., Coletti, M.E., Jones, V.L., Bodycombe, N.E., et al. (2016). Correlating chemical sensitivity and basal gene expression reveals mechanism of action. Nat. Chem. Biol. *12*, 109–116.

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

Rohban, M.H., Singh, S., Wu, X., Berthet, J.B., Bray, M.A., Shrestha, Y., Varelas, X., Boehm, J.S., and Carpenter, A.E. (2017). Systematic morphological profiling of human gene and allele function via Cell Painting. eLife *6*, https://doi.org/10.7554/eLife.24060.

Romero, N., Swain, P., Rogers, G.W., and Dranka, B.P. (2018). Determining Mechanisms of Mitochondrial Toxicity Using Agilent Seahorse XF Technology (EUROTOX).

Rose, F., Basu, S., Rexhepaj, E., Chauchereau, A., Del Nery, E., and Genovesio, A. (2018). Compound functional prediction using multiple unrelated morphological profiling assays. SLAS Technol. *23*, 243–251.

Saxena, C. (2016). Identification of protein binding partners of small molecules using label-free methods. Expert Opin. Drug Dis. *11*, 1017–1025.

Scheeder, C., Heigwer, F., and Boutros, M. (2018). Machine learning and image-based profiling in drug discovery. Curr. Opin. Syst. Biol. *10*, 43–52.

Schneidewind, T., Brause, A., Pahl, A., Burhop, A., Mejuch, T., Sievers, S., Waldmann, H., and Ziegler, S. (2020). Morphological profiling identifies a common mode of action for small molecules with different targets. Chembiochem *21*, 3197–3207.

Schulze, C.J., Bray, W.M., Woerhmann, M.H., Stuart, J., Lokey, R.S., and Linington, R.G. (2013). Function-first" lead discovery: mode of action profiling of natural product libraries using image-based screening. Chem. Biol. *20*, 285–295.

Seashore-Ludlow, B., Rees, M.G., Cheah, J.H., Cokol, M., Price, E.V., Coletti, M.E., Jones, V., Bodycombe, N.E., Soule, C.K., Gould, J., et al. (2015). Harnessing connectivity in a large-scale small-molecule sensitivity dataset. Cancer Discov. *5*, 1210–1223.

Soltoff, S.P. (2001). Rottlerin is a mitochondrial uncoupler that decreases cellular ATP levels and indirectly blocks protein kinase C delta tyrosine phosphorylation. J. Biol. Chem. *276*, 37986–37992.

Soltoff, S.P. (2004). Evidence that tyrphostins AG10 and AG18 are mitochondrial uncouplers that alter phosphorylation-dependent cell signaling. J. Biol. Chem. *279*, 10910–10918.

Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., et al. (2017). A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell *171*, 1437–1452 e1417.

Svenningsen, E.B., and Poulsen, T.B. (2019). Establishing cell painting in a smaller chemical biology lab—a report from the frontier. Bioorg. Med. Chem. *27*, 2609–2615.

't Hart, P., Hommen, P., Noisier, A., Krzyzanowski, A., Schüler, D., Porfetye, A.T., Akbarzadeh, M., Vetter, I.R., Adihou, H., and Waldmann, W. (2020). Structure based design of bicyclic peptide inhibitors of RbAp48. Angew. Chem. Int. Ed. Engl. https://doi.org/10.1002/anie.202009749.

Tanaka, M., Bateman, R., Rauh, D., Vaisberg, E., Ramachandani, S., Zhang, C., Hansen, K.C., Burlingame, A.L., Trautman, J.K., Shokat, K.M., et al. (2005). An unbiased cell morphology-based screen for new, biologically active small molecules. PLoS Biol. *3*, e128, https://doi.org/10.1371/journal.pbio.0030128.

Twarog, N.R., Low, J.A., Currier, D.G., Miller, G., Chen, T., and Shelat, A.A. (2016). Robust classification of small-molecule mechanism of action using a minimalist high-content microscopy screen and multidimensional phenotypic trajectory analysis. PLoS One *11*, e0149439, https://doi.org/10.1371/journal.pone.0149439.

Valeur, E., Gueret, S.M., Adihou, H., Gopalakrishnan, R., Lemurell, M., Waldmann, H., Grossmann, T.N., and Plowright, A.T. (2017). New modalities for challenging targets in drug discovery. Angew. Chem. Int. Ed. Engl. *56*, 10294–10323.

Warchal, S.J., Dawson, J.C., Shepherd, E., Munro, A.F., Hughes, R.E., Makda, A., and Carragher, N.O. (2020). High content phenotypic screening identifies serotonin receptor modulators with selective activity upon breast cancer cell cycle and cytokine signaling pathways. Bioorg. Med. Chem. *28*, 115209, https://doi.org/10.1016/j.bmc.2019.115209.

Wassermann, A.M., Lounkine, E., Hoepfner, D., Le Goff, G., King, F.J., Studer, C., Peltier, J.M., Grippo, M.L., Prindle, V., Tao, J.S., et al. (2015). Dark chemical matter as a promising starting point for drug lead discovery. Nat. Chem. Biol. *11*, 958–966.

Wassermann, A.M., Tudor, M., and Glick, M. (2017). Deorphanization strategies for dark chemical matter. Drug Discov. Today Technol. *23*, 69–74.

Wawer, M.J., Li, K., Gustafsdottir, S.M., Ljosa, V., Bodycombe, N.E., Marton, M.A., Sokolnicki, K.L., Bray, M.A., Kemp, M.M., Winchester, E., et al. (2014). Toward performance-diverse small-molecule libraries for cell-based phenotypic screening using multiplexed high-dimensional profiling. Proc. Natl. Acad. Sci. U S A *111*, 10911–10916.

Wilkinson, I.V.L., Terstappen, G.C., and Russell, A.J. (2020). Combining experimental strategies for successful target deconvolution. Drug Discov. Today *25*, 1998–2005.

Willis, C., Nyffeler, J., and Harrill, J. (2020). Phenotypic profiling of reference chemicals across biologically diverse cell types using the cell painting assay. SLAS Discov. *25*, 755–769.

Woehrmann, M.H., Bray, W.M., Durbin, J.K., Nisam, S.C., Michael, A.K., Glassey, E., Stuart, J.M., and Lokey, R.S. (2013). Large-scale cytological profiling for functional analysis of bioactive compounds. Mol. Biosyst. *9*, 2604–2617.

Young, D.W., Bender, A., Hoyt, J., McWhinnie, E., Chirn, G.W., Tao, C.Y., Tallarico, J.A., Labow, M., Jenkins, J.L., Mitchison, T.J., et al. (2008). Integrating high-content screening and ligand-target prediction to identify mechanism of action. Nat. Chem. Biol. *4*, 59–68.

Ziegler, S., Pries, V., Hedberg, C., and Waldmann, H. (2013). Target identification for small bioactive molecules: finding the needle in the haystack. Angew. Chem. Int. Ed. Engl. *52*, 2744–2792.

Zimmermann, S., Akbarzadeh, M., Otte, F., Strohmann, C., Sankar, M.G., Ziegler, S., Pahl, A., Sievers, S., and Kumar, K. (2019). A scaffold-diversity synthesis of biologically intriguing cyclic sulfonamides. Chemistry *25*, 15498–15503.

Zoffmann, S., Vercruysse, M., Benmansour, F., Maunz, A., Wolf, L., Blum Marti, R., Heckel, T., Ding, H., Truong, H.H., Prummer, M., et al. (2019). Machine learning-powered antibiotics phenotypic drug discovery. Sci. Rep. *9*, 5013, https://doi.org/10.1038/s41598-019-39387-9.