



The Cell Painting Assay as a Screening Tool for the Discovery of Bioactivities in New Chemical Matter

Axel Pahl and Sonja Sievers

Abstract

Multiparametric phenotypic screening based on cellular morphology interrogates many biological pathways simultaneously and is therefore a valuable screening tool for the discovery of new biological activities. The cell painting assay stains various cellular features using six different dyes in one well. By automated image analysis, hundreds of parameters are calculated from the images which deliver a phenotypic profile of the cell. It has been shown that compounds with similar modes of action deliver similar phenotypic profiles. Using a reference set of compounds with known modes of action, it is possible to assign probable modes of action to new compounds and to discover compounds with potentially new modes of action.

Here we describe the cell painting assay as a screening tool using a hit identification workflow which has been implemented using open-source software.

Key words Cell painting assay, Phenotypic profile, Screening, Hit identification, Morphological profiling

1 Introduction

Traditionally, high-throughput screening (HTS) has been divided into target-based and phenotypic approaches. Target-based screens are often carried out in in vitro set ups and therefore suffer from a lack of physiological relevance, while offering excellent opportunities for hit optimization and hit-to-lead development especially when structural information on the target is available. Phenotypic screens are in most cases carried out in a target-agnostic manner and offer a higher degree of physiological relevance, especially when pluripotent stem cell models are used [1]. On the flipside, hit optimization may be difficult and target identification via for example proteomic methods is a huge effort requiring many resources [2]. While phenotypic screens offer the chance to analyze the cell as a whole, investigators still often choose to limit the readout to a single or only a few parameters, often picked because of special significance for a pathway of interest [3]. As a consequence, most of

the information which lies in the phenotype is disregarded in conventional phenotypic screening.

Multiparameter phenotypic screens based on cellular morphology hold promise in enabling unbiased screening for the discovery of (new) bioactivities and have the potential to replace a whole fleet of single parameter assays. Thus multiparametric phenotypic screening might be more efficient than conventional screening [4].

The comprehensive description of hundreds of cellular features generates a phenotypic profile of a chemical compound. By comparing phenotypic profiles between agents with related modes of action, it could be shown that similar perturbations lead to similar changes in cellular morphology. For example, structurally unrelated tubulin inhibitors or HDAC inhibitors, respectively, were found to induce particular phenotypes and cluster together after hierarchical clustering analysis [5, 6]. Multiparametric phenotypic screening also has the potential to distinguish between new and already known bioactivities when a set of annotated reference compounds is screened alongside the new compound collection. For screening compounds with a phenotypic profile similar to a reference profile, multiparametric screening can deliver information about possible targets or mechanisms of action. Any compound with a phenotypic profile that significantly differs from the DMSO control but shows little or no similarity to the reference profiles of known compounds potentially modulates a new kind of bioactivity. It must be noted that in every case the screening results depend heavily on the reference compound set used. Focused libraries with drugs or drug candidates like the LOPAC library, the Prestwick Chemical library or the US-Drug collection are mostly used for this purpose [7]. However, it should be pointed out that drugs often do not only have one cellular target but that therapeutic efficacy may be linked to the modulation of several targets [8, 9]. Compound annotations seldom contain this polypharmacology information. Thus, it may be very well the case that an observed phenotype might be derived from the modulation of a target other than the annotated one [10].

The cell painting assay (CPA) has been described as a morphological profiling tool for the generation of multiparametric profiles of cells [5]. While not being the first example of a morphological multiparametric profiling assay [11], CPA has the potential to be applied to large compound sets [12] because of its relative ease of use and fairly low costs associated with the used dyes as opposed to antibody-based staining in other approaches [6]. It is also highly versatile in the adaptation to different cell lines. CPA is highly multiplexed, it highlights various cellular features by staining eight cellular compartments (nucleus, nucleoli, mitochondria, ER, Golgi, plasma membrane, actin cytoskeleton, cytosolic RNA) in five different fluorescent channels in one well [13]. By subsequent image analysis hundreds of parameters are deduced

from every channel which are combined into the morphological profile of the cell.

We combined the cell painting assay with a hit identification workflow to efficiently identify bioactive compounds in large compound sets.

2 Materials

2.1 Cell Staining

1. U2OS cells.
2. DMEM containing 10% fetal bovine serum, 1% L-glutamine, 1% sodium pyruvate, and 1% nonessential amino acids.
3. Black clear-bottomed 384 well plates.
4. Automated Dispenser: Multidrop Combi (Thermo).
5. Acoustic dispensing machine: Echo 520 dispenser (Labcyte Inc.).
6. Automated cell washer: Elx405 (Biotek).
7. Dulbecco's Phosphate-buffered saline (PBS): 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄*2 H₂O, 2 mM KH₂PO₄ in ultrapure water H₂O, adjust pH to 7.2 to 7.4 and autoclave.
8. MitoTracker staining solution: dilute MitoTracker stock solution (1 mM) to 100 nM with DMEM containing 10% fetal bovine serum freshly every screening day.
9. Fixing solution: dilute formaldehyde (37% stock solution) to 18.5% with PBS freshly every screening day.
10. Permeabilization solution: 0.1% (v/v) Triton X-100 in PBS.
11. Staining solution: 5 µg/ml Hoechst 33342, 1.5 µg/ml WGA, 25 µg/ml concanavalin A, 5 µl/ml phalloidin solution, 1.5 µM SYTO 14, and 1% BSA in PBS (*see* **Notes 1** and **2**). Prepare freshly every screening day.
12. Self-adhesive aluminum foil.
13. Reference set of compounds: e.g., LOPAC (Library of Pharmacologically Active Compounds) (Sigma), Prestwick Chemical Library (Prestwick Chemical), US-Drug collection (MicroSource Discovery Systems), Selleckchem kinase inhibitor library (Selleckchem), the published kinase inhibitor set (PKIS [14]).
14. Automated microscope: ImageXpress Micro XL (Molecular Devices) with filter sets for DAPI, FITC, SybrGold, TxRed, and Cy5 (*see* Table 1).
15. CellProfiler (<http://cellprofiler.org/> and <https://github.com/CellProfiler/CellProfiler> last accessed 18 July 2017).

Table 1
Overview about the different dyes and filter sets used for the cell painting assay

Dye	Cellular component(s)	Filter set	Ex/nm	Em/nm
Hoechst 33342	Nucleus	DAPI	350–400	410–480
Concanavalin A/Alexa Fluor 488 conjugate	Endoplasmic reticulum	FITC	470–500	510–540
SYTO 14 green fluorescent nucleic acid stain	Nucleoli, cytoplasmic RNA	Spectrum Gold	520–545	560–585
Phalloidin/Alexa Fluor 568	F-actin cytoskeleton	TxRed	535–585	600–650
Wheat-germ agglutinin/Alexa Fluor 555	Golgi, plasma membrane	TxRed	535–585	600–650
MitoTracker Deep Red	Mitochondria	Cy5	605–650	670–715

2.2 Software

1. Python 2.7 (<https://www.python.org/>; for CellProfiler on the Cluster in an Anaconda (<https://www.continuum.io/downloads>) environment),
2. Python 3.5 (for the processing workflow).
3. Nim (<https://nim-lang.org/>; for small scripts on the cluster).
4. Jupyter (<http://jupyter.org/>; as execution environment for the data processing workflow).
5. Pandas (<http://pandas.pydata.org/>; Python Data Analysis Library).
6. RDKit (<http://rdkit.org/>; Cheminformatics toolkit for working with chemical structures).

All last accessed 18 July 2017.

3 Methods

3.1 Cell Culture and Compound Treatment

1. Grow U2OS cells in supplemented DMEM. Split twice a week, using 600 cells per cm² for reseedling of the cells.
2. Seed U2OS cells at 1600 cells/ well in 25 µl medium in 384 well plates using a multichannel pipette or automated dispensing and incubate for 4 h at 37 °C and 5% CO₂ (*see* **Notes 3–5**).
3. Add compounds to a final screening concentration of 10 µM, for example using acoustic dispensing (*see* **Note 6**).
4. Incubate cells with compounds for approx. 20 h at 37 °C with 5% CO₂.

3.2 Live Cell Staining

Mitochondrial staining has to be performed in live cells. MitoTracker dyes diffuse passively across membranes and accumulate in active mitochondria. Most of the MitoTracker dyes are also well retained in the mitochondria after fixation.

1. For the live cell staining, aspirate medium from the plates leaving ca. 10 μl of residual volume using an automated cell washer.
2. Add 25 μl of MitoTracker staining solution using a multichannel pipette or automated dispensing.
3. Incubate plates for 30 min in the dark at 37 °C.

3.3 Fixation and Permeabilization

1. Fix the cells by addition of 7 μl of fixing solution to give a final concentration of 3.1% (v/v) formaldehyde using a multichannel pipette or automated dispensing and incubate at room temperature for 20 min (*see Note 7*).
2. Wash the plates three times with 70 μl PBS with final aspiration using an automated washer.
3. Permeabilize the cells by addition of 25 μl of permeabilization solution using a multichannel pipette or automated dispensing and incubate in the dark at room temperature for 10–20 min.
4. Wash the plates three times with 70 μl PBS with final aspiration using an automated washer.

3.4 Staining

1. Stain the cells by addition of 25 μl of staining solution to each well using a multichannel pipette or automated dispensing and incubate in the dark at room temperature for 30 min.
2. Wash the plates three times with 70 μl PBS with final aspiration using an automated washer.
3. Seal the plate with self-adhesive aluminum foil.

3.5 Imaging

1. Image the plate at 20 \times magnification using an automated microscope with filter sets for the 5 wavelengths as shown in Table 1. Take nine pictures per well to cover a large number of cells (>1000 cells in total) (*see Notes 8 and 9*).

3.6 Image Analysis

CellProfiler is an open source image analysis software which is fully compatible with high-throughput screening [15]. CellProfiler offers a modular approach to single cell image analysis even to non-image analysis experts. Different modules for image processing, object detection and measurements are combined into a pipeline. Measured parameters include size, shape, intensity, and texture features of the defined objects. For a small number of plates CellProfiler can easily be run on a desktop PC. However, for HTS mode CellProfiler needs to be installed and run on a cluster which might require the help of IT experts.

1. Analyze the images with CellProfiler software using the analysis pipeline in <https://github.com/mpimp-comas/cellpainting>. This pipeline was based on Bray et al. [13] with slight modifications. First, images are loaded into the pipeline and an illumination correction is performed. Then, object detection is carried out: primary objects (nuclei) are detected in the Hoechst channel, secondary objects (cytoplasm) are identified from the TxRed channel using a watershed algorithm and tertiary objects (cells) are defined as the sum of primary and secondary objects. CellProfiler then measures more than 2300 parameters for each individual cell. During Cell Profiler analysis, parameters are aggregated by the median values of the microscope sites, resulting in 9 values per microtiter well for every parameter.
2. For high-throughput analysis, install CellProfiler on a cluster. In our case, the processing of each image set from one plate is distributed over 96 parallel jobs. The processing of one plate takes approx. 5 h. At the end of a distributed processing job on the cluster, the result files from the individual processing jobs are concatenated to one result file per plate of 3456 lines (384 wells \times 9 sites) and 1937 columns (~90 MB). The CellProfiler pipeline is exported and reused in every run on the cluster (*see* **Notes 10** and **11**).

3.7 Data Mining

To further develop CPA into a screening assay, we established a data mining procedure to transform phenotypic profiles into a direct measure of compound activity (“CPA activity score”) without losing similarity information gained by reference compounds (Pahl et al., manuscript in preparation). For the CPA activity score, first a phenotypic profile is established using median per well values of the 307 parameters that were found to be nonredundant (for selection of parameters *see* Subheading 2). For profile comparison, we assign equal weight to all parameters and therefore, the absolute change in the parameters is not taken into account. Rather, for every parameter it is checked if it is identical to the DMSO control (activity code = 1) or if it has a significantly ($>7.5 \times$ Median Absolute Deviation, MAD) lower or higher numeric value than the DMSO control (activity code = 0 or activity code = 2, respectively). All activity codes are concatenated resulting in a 307 digit-activity string. The CPA activity score is then calculated from the sum of parameters with an activity code different from 1 divided by the total number of parameters and is expressed in percent. Thus, the CPA activity score indicates in how many parameters the phenotypic profile of a compound differs from the DMSO profile.

The data mining functionalities (Parts 1–4) are implemented as a Python module and the workflows are run in Jupyter Notebooks. All scripts have been uploaded to GitHub (*see* **Note 10**).

1. Data preparation

First, the data is prepared for further analysis (*see* **Note 12**). In this script, the values for every parameter of each well are grouped on the median, skipped wells from the Echo compound transfer are removed, the well type (Control/Compound) is assigned by their position on the plate (DMSO controls are in columns 11 and 12) and results for toxic compounds are flagged (a compound is considered toxic when it has less than 50% of the median cell count of the controls) (*see* **Note 13**).

Finally, data from all plates of a screening batch is concatenated into one result.

2. Finding relevant parameters

In this step, the relevant, uncorrelated parameters that are used for calculation of the activity score are determined from the compound set, from which toxic compounds have been removed (*see* **Note 14**).

The measured parameters are first filtered for their relevance, that is, parameters are removed that among the controls have a standard deviation of less than 0.05% or more than 2.5%, effectively removing parameters with constant values or very small variability as well as parameters with high variance already in the controls which would make them unsuitable for characterizing test compounds.

For our data set this results in 773 parameters, which are then submitted to a correlation filter which filters out parameters that are highly correlated to each other. A correlation matrix is generated; the parameters which are correlated to the highest number of other parameters are analyzed and only the parameter with the lowest standard deviation is kept and the other parameters are removed. From the remaining parameters, a new correlation matrix is generated and the process is repeated until no correlated parameters remain in the data set, leading to a set of 307 uncorrelated parameters at a cut-off of 0.9 for the correlation factor.

3. Calculation of CPA activity score for the reference set

In this part the phenotypic profiles are calculated for the reference compounds from the set of relevant parameters. The phenotypic profile is a fingerprint-like property that is calculated from each of the relevant parameters as follows: if the value of a compound for a specific parameter is higher (or lower) than the median of the controls for that parameter ± 7.5 times the MAD, then the fingerprint at that position

gets assigned a 2 (or 0). Otherwise a 1 is assigned for inactivity. The resulting fingerprint is a string with the length of the number of relevant parameters that at each position has either a 0 (lower than control), a 1 (inactive) or a 2 (higher than control). Then, the activity score is calculated which consists of the absolute number of active parameters in the activity profile divided by the total number of parameters, and is expressed in percent.

This categorization of parameters makes it very easy to compare compounds by their activity profiles (*see Note 15*). The profile similarity score is calculated by the Tanimoto method [16] as follows: two activity profiles are compared position by position. For all significant parameters (with value 0 or 2) that have the same value at the same position, a counter is incremented. The similarity is then the final value of the counter divided by the total number of significant positions from both profiles, resulting in values from 0 (different in every position) to 1 (completely identical).

The script for the calculation of the CPA activity score of the reference compounds first processes the raw data as described in **step 1**. In addition, the plate layout is joined (linking plate position and Compound_ID), the structures are joined as Smiles, low-purity or toxic reference compounds are removed, activity profiles and activity scores are calculated for the determined relevant parameter set, annotations (trivial names, known activities) are joined to the reference compound set, the data set is saved to a TSV file (tab-separated).

4. Comparison of the screening compounds to the reference compounds

A screening plate containing test compounds is processed as in Subheading 3 whereby removal of impure and toxic compounds is omitted. For active compounds (minimum activity score of 5%), profile similarity scores are calculated.

Finally, a report is generated for each test compound. If a test compound shows similarity to a reference compound (minimum similarity score of 25%), a list of reference compounds and their annotations is also reported, otherwise it is noted that no similar references could be found.

4 Notes

1. As a cheaper alternative, Concanavalin A FITC conjugate can also be used. We used it at the same concentration as the Alexa 488 conjugate and achieved similar staining results.
2. As a cheaper alternative, MFP-DY-594-Phalloidin conjugate (Dyomics) can also be used. We used it at the same

concentration as the Alexa 594 conjugate and achieved similar staining results.

3. For most gentle cell seeding, standard tube cassettes should be used with dispensing speed set to “medium.”
4. Even cell seeding can be achieved by first dispensing 5 μ l medium only, followed by a second dispensing step of 20 μ l cell suspension. After that, the plate should be left at room temperature for 10–20 min before putting it into the incubator [17].
5. CellCarrier Ultra plates outperformed all other plates tested because their bottom is highly even and no drift in focusing was observed across the images.
6. When screening focused libraries of for example kinase inhibitors or certain classes of natural compounds, which tend to be toxic at high concentrations, the screening concentration should be reduced to 1–2 μ M. At a screening concentration of 10 μ M for known drugs and 2 μ M for kinase inhibitors, we found 39% of our reference compound set to be active (activity score >10%).
7. Formaldehyde should be freshly diluted to 18.5% from the 37% stock every screening day.
8. Exposure times may vary for the different dyes. Make sure to not saturate your images. We usually choose the brightest objects to have an intensity of 10,000 of the 64,000 grey shades.
9. If possible, use camera binning set to 2. This will quarter image size and reduce storage space as well as data transfer and processing times.
10. The setup of CellProfiler on a cluster may need the help of IT experts. It is recommended to install the CellProfiler Python program in an Anaconda virtual environment directly from the CellProfiler GitHub sources. Scripts and helper tools to run parallel jobs for the Sun Grid Engine (SGE) can be found on the COMAS GitHub page (<https://github.com/mpimp-comas/cellpainting>) and need to be adapted to the individual setup.

After the parallel processing of the images by CellProfiler and the concatenation of the results on the cluster, the result data is transferred to a normal desktop computer on which all downstream processing and reporting is performed.

11. CellProfiler offers multiple ways to handle configurations and to output results. For parallel processing on a cluster, we found the best solution to be working with *.cppipe configuration files and to generate the result data in plain CSV (or TSV) files. The CellProfiler pipeline (*.cppipe) is generated by

configuring and exporting a workflow in the graphical interface of CellProfiler which is accessed on the cluster via VNC client.

12. To construct a phenotypic profile for a compound from the single cell data of the primary image analysis, the data needs to be normalized (to account for plate and batch differences) and aggregated: different statistical methods can be used to combine data on the image level, the well level (several sites are imaged per well) and the compound level (if replicates are measured). Comparison of different statistical methods showed that using a simple mean or median value per well already delivers robust results, even for heterogeneous phenotypes [18, 19]. However, improvements in the accuracy of classification of a set of test compounds could be achieved using the median combined with median absolute deviation (MAD) or percentile values [18, 19].
13. We chose the toxicity cutoff at <50% of the median cell count of the DMSO control. Admittedly, a potent antimetabolic compound could also lead to such a reduction in cell count over 20 h compound incubation time. However, using a screening test set of 1400 proprietary compounds only 4% of compounds were flagged to be toxic and in our reference set of 4220 compounds only 3.3% of compounds were flagged to be toxic.
14. During image analysis, several hundred parameters are calculated for every cell delivering the basis for a phenotypic profile. Many approaches for the generation of phenotypic profiles (also known as fingerprints) have been described. Thereby, some approaches work with a high number of individual parameters; while others use dimension reduction or feature selection methods to come up with a screening result that can be biologically interpreted and is more accessible to the human mind [20, 21]. Reducing the number of highly correlated parameters can be beneficial in subsequent analysis, for example, in reducing processing times or because of higher importance of the chosen parameters which in turn can result in a more stable phenotypic classification [19]. However, dimension reduction or feature selection also bears the risk of losing information which could be valuable for phenotype discrimination. This is especially true for unbiased screening with an undefined phenotype; thus it is expected that for such approaches a high number of parameters should be included into the phenotypic profile to cover a broad range of phenotypes that might be induced by the screening compounds. In every case caution should be applied to preserve as much of the variance as possible.
15. Phenotypic profiles of the test compounds need to be compared to those of DMSO controls and reference compounds.

Because of the multiparametric nature of the phenotypic profiles, multivariate statistical measures are often applied to detect relevant differences and similarities between phenotypic profiles. For small molecule screening purposes, the Mahalanobis distance has often been used [22]. A study comparing different distance- and correlation-based methods found that correlation-based methods work best for different purposes while distance-based methods are stronger in detecting all phenotypes different from the neutral controls [23]. Multivariate similarity measures are also a prerequisite for any kind of clustering approaches which group similar phenotypic profiles together. However, clustering workflows usually do not implement an activity threshold which is needed for screening purposes to differentiate active from inactive compounds.

To overcome this problem, an arbitrary cutoff has been applied to the Mahalanobis distance to define a hit [22]. In another effort to create an activity score for phenotypic profiles, a “multidimensional perturbation value” (mp-value) has been designed to define a hit [24, 25]. Mp-values are significance measurements that are calculated from Mahalanobis distances by permutation analysis. Like for nominal significance tests, a cutoff of 0.05 can be applied as a cutoff for active compounds.

References

1. Vincent F, Loria P, Pregel M, Stanton R, Kitching L, Nocka K, Doyonnas R, Steppan C, Gilbert A, Schroeter T, Peakman MC (2015) Developing predictive assays: the phenotypic screening "rule of 3". *Sci Transl Med* 7(293):293ps215. <https://doi.org/10.1126/scitranslmed.aab1201>
2. Vendrell-Navarro G, Brockmeyer A, Waldmann H, Janning P, Ziegler S (2015) Identification of the targets of biologically active small molecules using quantitative proteomics. *Methods Mol Biol* 1263:263–286. https://doi.org/10.1007/978-1-4939-2269-7_21
3. Singh S, Carpenter AE, Genovesio A (2014) Increasing the content of high-content screening: an overview. *J Biomol Screen* 19(5):640–650. <https://doi.org/10.1177/1087057114528537>
4. Pennisi E (2016) IMAGING. 'Cell painting' highlights responses to drugs and toxins. *Science* 352(6288):877–878. <https://doi.org/10.1126/science.352.6288.877>
5. Gustafsdottir SM, Ljosa V, Sokolnicki KL, Anthony Wilson J, Walpita D, Kemp MM, Petri Seiler K, Carrel HA, Golub TR, Schreiber SL, Clemons PA, Carpenter AE, Shamji AF (2013) Multiplex cytological profiling assay to measure diverse cellular states. *PLoS One* 8(12):e80999. <https://doi.org/10.1371/journal.pone.0080999>
6. Reisen F, Sauty de Chalon A, Pfeifer M, Zhang X, Gabriel D, Selzer P (2015) Linking phenotypes and modes of action through high-content screen fingerprints. *Assay Drug Dev Technol* 13(7):415–427. <https://doi.org/10.1089/adt.2015.656>
7. Jones LH, Bunnage ME (2017) Applications of chemogenomic library screening in drug discovery. *Nat Rev Drug Discov* 16(4):285–296. <https://doi.org/10.1038/nrd.2016.244>
8. Boran AD, Iyengar R (2010) Systems approaches to polypharmacology and drug discovery. *Curr Opin Drug Discov Devel* 13(3):297–309
9. Hu Y, Bajorath J (2013) High-resolution view of compound promiscuity. *Fl1000 Res* 2:144. <https://doi.org/10.12688/fl1000research.2-144.v2>
10. Ursu A, Illich DJ, Takemoto Y, Porfetye AT, Zhang M, Brockmeyer A, Janning P, Watanabe N, Osada H, Vetter IR, Ziegler S, Scholer HR, Waldmann H (2016) Epiplastin A

- induces reprogramming of epiblast stem cells into embryonic stem cells by inhibition of casein kinase 1. *Cell Chem Biol* 23 (4):494–507. <https://doi.org/10.1016/j.chembiol.2016.02.015>
11. Perlman ZE, Slack MD, Feng Y, Mitchison TJ, Wu LF, Altschuler SJ (2004) Multidimensional drug profiling by automated microscopy. *Science* 306(5699):1194–1198. <https://doi.org/10.1126/science.1100709>
 12. Bray MA, Gustafsdottir SM, Ljosa V, Singh S, Sokolnicki KL, Bittker JA, Bodycombe NE, Dancik V, Hasaka TP, Hon CS, Kemp MM, Li K, Walpita D, Wawer MJ, Golub TR, Schreiber SL, Clemons PA, Shamji AF, Carpenter AE (2017) A dataset of images and morphological profiles of 30,000 small-molecule treatments using the Cell Painting assay. *Gigascience* 6 (12):1–5. <https://doi.org/10.1093/gigascience/giw014>
 13. Bray MA, Singh S, Han H, Davis CT, Borgeson B, Hartland C, Kost-Alimova M, Gustafsdottir SM, Gibson CC, Carpenter AE (2016) Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat Protoc* 11 (9):1757–1774. <https://doi.org/10.1038/nprot.2016.105>
 14. Knapp S, Arruda P, Blagg J, Burley S, Drewry DH, Edwards A, Fabbro D, Gillespie P, Gray NS, Kuster B, Lackey KE, Mazzafera P, Tomkinson NC, Willson TM, Workman P, Zuercher WJ (2013) A public-private partnership to unlock the untargeted kinome. *Nat Chem Biol* 9(1):3–6. <https://doi.org/10.1038/nchembio.1113>
 15. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin DA, Chang JH, Lindquist RA, Moffat J, Golland P, Sabatini DM (2006) CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* 7(10):R100. <https://doi.org/10.1186/gb-2006-7-10-r100>
 16. Rogers DJ, Tanimoto TT (1960) A Computer program for classifying plants. *Science* 132 (3434):1115–1118. <https://doi.org/10.1126/science.132.3434.1115>
 17. Lundholt BK, Scudder KM, Pagliaro L (2003) A simple technique for reducing edge effect in cell-based assays. *J Biomol Screen* 8 (5):566–570. <https://doi.org/10.1177/1087057103256465>
 18. Ljosa V, Caie PD, Ter Horst R, Sokolnicki KL, Jenkins EL, Daya S, Roberts ME, Jones TR, Singh S, Genovesio A, Clemons PA, Carragher NO, Carpenter AE (2013) Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment. *J Biomol Screen* 18 (10):1321–1329. <https://doi.org/10.1177/1087057113503553>
 19. Kummel A, Selzer P, Beibel M, Gubler H, Parker CN, Gabriel D (2011) Comparison of multivariate data analysis strategies for high-content screening. *J Biomol Screen* 16 (3):338–347. <https://doi.org/10.1177/1087057110395390>
 20. Twarog NR, Low JA, Currier DG, Miller G, Chen T, Shelat AA (2016) Robust classification of small-molecule mechanism of action using a minimalist high-content microscopy screen and multidimensional phenotypic trajectory analysis. *PLoS One* 11(2):e0149439. <https://doi.org/10.1371/journal.pone.0149439>
 21. Young DW, Bender A, Hoyt J, McWhinnie E, Chirn GW, Tao CY, Tallarico JA, Labow M, Jenkins JL, Mitchison TJ, Feng Y (2008) Integrating high-content screening and ligand-target prediction to identify mechanism of action. *Nat Chem Biol* 4(1):59–68. <https://doi.org/10.1038/nchembio.2007.53>
 22. Gerry CJ, Hua BK, Wawer MJ, Knowles JP, Nelson SD Jr, Verho O, Dandapani S, Wagner BK, Clemons PA, Booker-Milburn KI, Boskovic ZV, Schreiber SL (2016) Real-time biological annotation of synthetic compounds. *J Am Chem Soc* 138(28):8920–8927. <https://doi.org/10.1021/jacs.6b04614>
 23. Reisen F, Zhang X, Gabriel D, Selzer P (2013) Benchmarking of multivariate similarity measures for high-content screening fingerprints in phenotypic drug discovery. *J Biomol Screen* 18(10):1284–1297. <https://doi.org/10.1177/1087057113501390>
 24. Hutz JE, Nelson T, Wu H, McAllister G, Moutsatsos I, Jaeger SA, Bandyopadhyay S, Nigsch F, Cornett B, Jenkins JL, Selinger DW (2013) The multidimensional perturbation value: a single metric to measure similarity and activity of treatments in high-throughput multidimensional screens. *J Biomol Screen* 18 (4):367–377. <https://doi.org/10.1177/1087057112469257>
 25. Wawer MJ, Li K, Gustafsdottir SM, Ljosa V, Bodycombe NE, Marton MA, Sokolnicki KL, Bray MA, Kemp MM, Winchester E, Taylor B, Grant GB, Hon CS, Duvall JR, Wilson JA, Bittker JA, Dancik V, Narayan R, Subramanian A, Winckler W, Golub TR, Carpenter AE, Shamji AF, Schreiber SL, Clemons PA (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 (30):10911–10916. <https://doi.org/10.1073/pnas.1410933111>