



Published in final edited form as:

Nat Protoc. 2016 September ; 11(9): 1757–1774. doi:10.1038/nprot.2016.105.

Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes

Mark-Anthony Bray¹, Shantanu Singh¹, Han Han², Chadwick T. Davis², Blake Borgeson², Cathy Hartland³, Maria Kost-Alimova³, Sigrun M. Gustafsdottir³, Christopher C. Gibson^{2,*}, and Anne E. Carpenter^{1,*}

¹Imaging Platform, Broad Institute of Harvard and MIT, Cambridge USA

²Recursion Pharmaceuticals, Salt Lake City USA

³Center for the Science of Therapeutics, Broad Institute of Harvard and MIT, Cambridge USA

Abstract

In morphological profiling, quantitative data are extracted from microscopy images of cells to identify biologically relevant similarities and differences among samples based on these profiles. This protocol describes the design and execution of experiments using Cell Painting, a morphological profiling assay multiplexing six fluorescent dyes imaged in five channels, to reveal eight broadly relevant cellular components or organelles. Cells are plated in multi-well plates, perturbed with the treatments to be tested, stained, fixed, and imaged on a high-throughput microscope. Then, automated image analysis software identifies individual cells and measures ~1,500 morphological features (various measures of size, shape, texture, intensity, etc.) to produce a rich profile suitable for detecting subtle phenotypes. Profiles of cell populations treated with different experimental perturbations can be compared to suit many goals, such as identifying the phenotypic impact of chemical or genetic perturbations, grouping compounds and/or genes into functional pathways, and identifying signatures of disease. Cell culture and image acquisition takes two weeks; feature extraction and data analysis take an additional 1-2 weeks.

INTRODUCTION

Phenotypic screening has been tremendously powerful for identifying novel small molecules as probes and potential therapeutics and for identifying genetic regulators of many biological processes¹⁻⁴. High-throughput microscopy has been a particularly fruitful type of phenotypic screening; it is often called high-content analysis because of the high information content that can be observed in images⁵. However, most large-scale imaging

*To whom correspondence should be addressed AEC: (617) 714-7750, CCG: (801) 587-1629.

AUTHOR CONTRIBUTION STATEMENTS

All authors contributed to writing this protocol. CTD, HH, and SMG contributed to describing the benchwork aspects of the protocol. CH, MKA, MAB and CTD contributed to updates to the experimental design. MAB and SS contributed most heavily to describing the computational aspects of the protocol. AEC, CCG, BB, and SS contributed most heavily to describing the rationale and experimental design.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests. Recursion Pharmaceuticals is a biotechnology company in which CCG, BB, CTD, HH, and AEC have real or optional ownership interest (see the HTML version of this article for details).

experiments extract only one or two features of cells⁶ and/or aim to identify just a few “hits” in a screen, meaning that vast quantities of quantitative data about cellular state remain unharnessed.

In this article, we detail a protocol for the Cell Painting assay, a generalizable and broadly-applicable method for accessing the valuable biological information about cellular state that is contained in morphology. Cellular morphology is a potentially rich data source for interrogating biological perturbations, especially in large scale^{5,7–10}. The techniques and technology necessary to generate these data have advanced rapidly, and are now becoming accessible to non-specialized laboratories¹¹. In this protocol, we discuss morphological profiling (also known as image-based profiling), contrast it with conventional image-based screening, illustrate applications of morphological profiling, and provide guidance, tips, and tricks related to the successful execution of one particular morphological profiling assay, the Cell Painting assay.

Broadly speaking, the term *profiling* describes the process of quantifying a very large set of features, typically hundreds to thousands, from each experimental sample in a relatively unbiased way. Significant changes in a subset of profiled features can thus serve as a “fingerprint” characterizing the sample condition. Some of the earliest instances of profiling involved the NCI-60 tumor cell line panel, where patterns of anticancer drug sensitivity were discovered to reflect mechanisms of action¹², and gene expression, in which signatures related to small molecules, genes, and diseases were identified¹³.

It is important to note that profiling differs from conventional screening assays in that the latter are focused on quantifying a relatively small number of features selected specifically because of a known association with the biology of interest. Profiling, on the other hand, casts a much wider net and avoids the intensive customization usually necessary for problem-specific assay development in favor of a more generalizable method. Therefore, taking an unbiased approach via morphological profiling offers the opportunity for discovery unconstrained by what we know (or think we know). It also holds the potential to be more efficient, as a single experiment can be mined for many different biological processes or diseases of interest.

In morphological profiling, measured features include staining intensities, textural patterns, size, and shape of the labeled cellular structures, as well as correlations between stains across channels, and adjacency relationships between cells and among intracellular structures. The technique enables single-cell resolution, enabling detection of perturbations even in subsets of cells. Morphological profiling has successfully been used to characterize genes and compounds in a number of studies. For instance, morphological profiling of chemical compounds has been used to determine their mechanism of action^{7,14–18}, identify their targets^{19,20}, discover relationships with genes^{20,21}, and characterize cellular heterogeneity²². Genes have been analyzed by creating profiles of cell populations where the gene is perturbed by RNA interference (RNAi), which in turn have been used to cluster genes^{23,24}, identify genetic interactions^{25–27}, or characterize cellular heterogeneity²⁸.

Development of the protocol

Until recently, most published profiling methods (such as those cited above) were performed using assays involving only three dyes. We sought to devise a single assay illuminating as many biologically relevant morphological features as possible, while still maintaining compatibility with standard high-throughput microscopes. We also wanted the assay to be feasible for large-scale experiments in terms of cost and complexity, so we chose dyes rather than antibodies. After considerable assay development, we selected six fluorescent stains imaged in five channels, revealing eight cellular components or compartments in a single microscopy-based assay²⁹ (Figure 1). We later dubbed the assay “Cell Painting”, given our aim to paint the cell as richly as possible with dyes. Automated image analysis pipelines extract ~1,500 morphological features from each stained and imaged cell to produce profiles (Figure 2). Profiles are then compared against each other and mined to address the biological question at hand. The Cell Painting assay described in this protocol has been successfully employed by multiple researchers. It was developed at the Broad Institute, where it was carried out in multiple laboratories, and later independently adopted at Recursion Pharmaceuticals; this protocol thus summarizes the implementation of the protocol at two independent sites and by more than ten different researchers. We refer interested readers to Gustafsdottir et al²⁹, Ljosa et al¹⁷, and Wawer et al³⁰ for details on assay development, relevant computational approaches to profiling, and application to compound library enrichment, respectively.

Applications of the method

Morphological profiling using the Cell Painting assay may be tremendously powerful for achieving a number of biological goals, only some of which have been demonstrated so far.

First, clustering small molecules by phenotypic similarity using the Cell Painting assay is effective. The first paper to use the protocol was a proof-of-principle study wherein cells were treated with various small molecules, stained and imaged using the Cell Painting assay, and the resulting profiles were clustered to identify which small molecules yielded similar phenotypic effects²⁹. Thus, the assay could be used to identify the mechanism of action or target of an unannotated compound (based on similarity to well-annotated compounds) or to “lead hop” to find additional small molecules with the same phenotypic effects but different structures (based on phenotypic similarity to compounds in a library with more favorable structural properties). As well, small-molecule hits from a screen could be clustered based on morphological profiles in order to reveal potential differences among hit classes in terms of mechanism as well as polypharmacology (e.g., off-target effects).

Likewise, by matching unannotated genes to known genes based on similar phenotypic profiles derived from the Cell Painting assay, similarities among genetic perturbations can reveal their biological functions. Due to caveats concerning RNAi off-target effects³¹ (see “Limitations”), in our current work, we instead overexpress genes and mine for similarities in the induced phenotypic profiles. In addition to mapping unannotated genes to known pathways based on profile similarity, overexpressing variant alleles is likely to enable discovery of the functional impact of a genetic variant by comparing the profiles induced by wild-type and variant versions of the same gene.

Cell Painting can also be used first to identify a phenotypic signature associated with disease, and then as a screen to revert that signature back to “wild-type”. The co-authors from Recursion Pharmaceuticals have implemented this approach in systematic fashion by simultaneously modeling hundreds of rare, monogenic loss-of-function diseases in human cells. The subset of disease models for which a strong disease-specific phenotype is uncovered in the Cell Painting assay are then systematically screened against a drug-repurposing library to identify drugs that can reduce the strength of the disease phenotype and thus rescue the putative disease-specific features of the profile. Ultimately, the goal is to find new indications for existing drugs; this general approach (using an assay of three stains rather than the Cell Painting assay) has already been used successfully to identify potential new uses of known drugs for the treatment of cerebral cavernous malformation, a hereditary stroke syndrome^{32,33}.

Lastly, both academic and pharmaceutical screening groups have an increasing desire to improve the efficiency of screening, particularly when assays are complex and expensive³⁴. Profiles generated by Cell Painting applied to a large set of small molecules can be used to identify a more efficient, enriched screening set that minimizes phenotypic redundancy. The benefit is to maximize profile diversity (and thus likelihood of diverse phenotypic effects) while simultaneously eliminating compounds that do not produce any measurable effects on the cell type of interest. In a recent study, morphological profiling by Cell Painting was more powerful for this purpose than choosing a screening set based on structural diversity or diversity in high-throughput gene expression profiles³⁴.

Comparison with other methods

Though diverse methods for generating rich (100+ measurement) profiles of biological samples exist, such as metabolomic or proteomic profiling, to our knowledge gene expression profiling by L1000 (<http://www.lincscloud.org/l1000/>,³⁵) is currently the only practical alternative to image-based morphological profiling in terms of throughput and efficiency⁴. Both profiling approaches (morphological profiling by Cell Painting and gene expression profiling by L1000) yield ~1000 raw features for each sample, and both are likely to capture a broad range of cellular states following perturbation. However, gene-expression profiling in high-throughput can be currently performed only by aggregating cell populations and not at the single cell level, whereas morphological profiles are obtained at the level of individual cells, potentially improving the ability to resolve changes in subpopulations of cells. Cell Painting is also currently substantially less costly per sample.

Many potentially interesting additional points of comparison have yet to be rigorously tested. Thus far, no quantitative comparison of the two profiling methods' reproducibility has been published. The two methods likely have quantitatively and qualitatively different information content; only one direct comparison has been published so far, the above-mentioned study indicating better predictive power for Cell Painting versus L1000 gene expression profiling, for the purposes of library enrichment³⁰. The study also indicated that orthogonal profiling approaches are capable of capturing a wider range of biological performance diversity than either technique alone. The fact that the two profiling methods yielded only partially overlapping library selections indicates the two modalities capture

distinct information about cell state and thus are likely quite complementary. No studies have yet been published combining the two orthogonal modalities into a single profile; we believe this approach would be extremely powerful.

An alternative, image-based approach to the Cell Painting assay described here is to use a different selection of stains. The choice of stains for the Cell Painting protocol was based on the desire to detect a broad range of phenotypic effects upon compound treatment or genetic perturbation, while keeping the assay inexpensive and straightforward to implement using conventional sample preparation and imaging equipment. Morphological profiling has proven to be powerful using a relatively diverse and unbiased set of stains, i.e., stains not selected to target particular pathways. Differences in phenotypic signature can be detectable even though the Cell Painting stains do not, a priori, seem likely to label a pathway targeted by a particular perturbation^{14,29}. Still, if an experiment's questions are targeted to a particular biological area of interest, it is worth considering replacing one or more of the original stains for a more targeted one; the same basic principles of morphological profiling will still apply. However, exchanging stains may entail significant optimization effort; for example, replacing the Concanavalin A/Alexa Fluor 488 label requires a fluorophore bright enough to mask the fluorescence of SYTO 14 in the 472/30 nm channel (see Table 1, note 1). The Cell Painting assay could also be adapted for living cells, imaged over time, by using an alternative set of live-cell compatible stains or expressing fluorescently tagged proteins.

Experimental Design

Cell type—We and our collaborators have successfully applied the Cell Painting assay to 13 different cell cultures, including cell lines, primary cells, and co-culture systems. Although we have used U2OS and A549 cells most commonly, the staining protocol has worked well in our hands for MCF-7, 3T3, HTB-9, HeLa, HepG2, HEKTE, SH-SY5Y, HUVEC, HMVEC, primary human fibroblasts, and primary human hepatocyte/3T3-J2 fibroblast co-cultures.

In many applications, the assay is intended to be unbiased and not targeted to a particular biological area of interest; in such cases, using one of these already-tested cell types is sensible. In profiling, it is anticipated that many very specific biological effects and pathways can be interrogated even in a relatively “generic” cell type^{13,29}. However, it is important to note that profiles produced from various cell types are likely to differ from each other. This variation may be mitigated to some degree by developing methods of normalization across different cell types, but substantive differences among the cell types (e.g., signaling pathways, gene expression levels, baseline morphology, etc.) make it unlikely that these dissimilarities can be resolved completely.

In other cases, it is worth considering selection of a cell line that is physiologically relevant and well-characterized in the biological area of study; for example, avoiding immortal cell lines with significant genetic alterations might be essential for certain studies. If data integration is a goal, it may also be valuable to choose a cell line for which there are additional sources of complementary data available. Generally, no adjustment of the staining protocol has been needed from cell type to cell type; the only change in protocol is to optimize the seeding density to adjust the confluency of the cells in accordance with the

perturbations applied and biology to be examined (see “Perturbations and/or timepoints“ below).

Because we do not consider it risky to apply the assay to new cell types, selecting a new cell type might be warranted, subject to the following criteria. A major criterion in choosing a cell line is that a seeding density can be identified at which individual cells do not substantially or frequently overlap each other in the final images, i.e., the cells form a monolayer. This will allow accurate measurements to be obtained from single cells upon image analysis. It should be noted, however, that it may still be possible to obtain rich and useful data *without* the accurate segmentation of individual cells. A segmentation-free approach has been demonstrated to work for some image classification applications^{36–38}, but whether this suffices for intensive morphological profiling applications remains untested.

A second major criterion is that the cell type should grow in a manner conducive to fluorescent imaging and analysis. Specifically, the cells should typically be adherent and grow reasonably flat (i.e., non-spheroid), without significant clumping under the culture conditions used. Cell types we have tested that fail to meet this criterion are SW480 and DLD-1; presumably, non-adherent cells that are grown in suspension would also be less than ideal. The more rounded a cell type is, or the more cells grow on top of each other, the less internal structure is clearly visible by microscopy. In such cases, the staining protocol itself will label the appropriate components and images can be produced and processed, but the information content is likely to be lower for cell types with a rounded morphology as compared to a more flattened one.

Plate layout and selection of replicates and controls—When selecting the plate layout (i.e., the pattern of treatments and controls across each multi-well plate), and the number of controls and replicates, the predominant concern is that phenotypic effects may be subtle and the assay is sensitive. Therefore, the experiment requires careful design to avoid the impact of systematic errors^{39,40}.

In order to test and compensate for systematic effects related to well position, biological replicates should not be present in the same well position on every plate (e.g., compound X should not always be present in well position G07, controls should not always be present on the top row, etc). This can be accomplished by spatially offsetting sample replicates (especially controls) with respect to each other, either by having two (or more) replicates on the same plate but spaced out as much as possible, or by having two (or more) plate layouts where the replicates are present in different well positions. Having all the replicates of a perturbation (or a control) lying on the edge of the plate should be especially avoided.

At least four biological replicates are recommended as we have observed a significant loss in data quality with fewer replicates. In practice, we have used five or more replicates to buffer against accidental sample loss. For larger-scale experiments (more than 1000 perturbations), we have used four replicates for cost reasons.

For morphological profiling, negative controls are used to normalize the image features, but positive controls can be included if they can be reliably defined for the experiment. For

compound library screening, typically the negative controls are ‘vehicle’-only conditions (e.g., DMSO). We have found that ~30 wells/plate designated for negative controls works well for a 384-well format assay. In cases where there is no obvious negative control, the untreated wells (i.e., wells containing cells but not subjected to drug vehicles nor gene perturbant delivery reagents) may serve as a substitute for this purpose. If more than one type of control is used, the positions should be interleaved or randomized. For gene profiling, multiple negative controls may be considered such as empty vectors, or control treatments towards genes that are irrelevant or non-native to the cell type. However, it should be kept in mind that in the case of gene knockdown using RNAi, even control hairpins targeting the same gene but containing different seed sequences can induce different morphological profiles; it is also the case that negative controls targeting gene sequences not present in a cell line yield different profiles³¹.

A perennial concern with assay development is that any technical sources of variation can impact all the wells and/or plates such that any biological signal are overwhelmed by systematic noise introduced by sample preparation. We advise mitigating this issue by spreading the position of replicate samples across the experimental axes with the most variability, i.e., the processing order of batches and/or plates, the spatial positioning of samples on a plate, etc. For example, as mentioned above, we recommend scattering the control and biological replicate wells across the plate rather than placing them adjacent to each other, e.g., clustered in neighboring wells or placed in a single row or column. The standard pattern that we use for compound screening is to place the control wells in a chevron pattern across the plate; other patterning variations are feasible. Ideally, some control/replicate wells should be placed in a pattern that can be distinguished if the plate is turned 180° inadvertently, e.g., placing a particular cytotoxic treatment in the top left well but not in the bottom right well, or omitting cells in a non-symmetric well location.

With automated plate handlers, groups of plates are typically washed and stained as a single batch. We recommend spreading the biological replicate plates across batches, rather than processing them within a single batch, as well as staggering the processing order of the replicate plates in each batch. As an example, consider an experiment set to be carried out in four-plate batches, consisting of five different assay plates (labeled A-E), with four biological replicates each (labeled 1-4), so that A₁– A₄ comprise all replicates of plate A, B₁– B₄ are the replicates of plate B, and so on. Ideally, only one particular assay plate should be processed within each batch, and a single batch should not process all four replicates together. A possible batch processing order could be [A₁,B₁,C₁,D₁], [E₁,A₂,B₂,C₂], [D₂,E₂,A₃,B₃], [C₃,D₃,E₃,A₄], [B₄,C₄,D₄,E₄], [A₅,B₅,C₅,D₅].

Perturbations and/or timepoints—Prior to carrying out a large-scale experiment, consider initially performing a smaller pilot with a small number of perturbations and/or timepoints. During assay development, we recommend culturing the cells until the desired timepoint, and then examining them for the degree of confluency. This assessment may be performed by fluorescent staining (i.e., following some or all of the regular protocol) or even by eye under brightfield; the latter is less effort and is feasible once a researcher has some experience with the assay. Of note, the pilot assay should replicate the conditions under which the cells will be analyzed in the full experiment as closely as possible; at a minimum,

this should include vehicle controls for small-molecule treatment and transfection or infection with control sequences for genetic screens, as such perturbations can have a significant effect on cellular growth and confluence. We have used both 24- and 48-hour^{29,30} small-molecule exposures prior to fixation and imaging; in our limited testing, the latter yields a higher proportion of small molecules yielding morphological profiles distinguishable from negative controls. For infection/transfection of RNAi and over-expression plasmids, we use 96-³¹ and 72-hour exposures, respectively. It may well be that a shorter or longer exposure time is optimal, particularly for certain types of biological processes of interest and certain perturbation types, but we have not extensively explored this area.

As noted above, some adjustment of the cell seeding density may be needed depending on the cell line used and the biological processes under examination. Because cell-cell junction interactions play a significant physiological role in endothelial and epithelial cell types, we recommend growing these cultures as confluent or near-confluent monolayers. If such biological processes are not of particular concern, we recommend optimizing the cell density while striking a balance between two considerations. Because the expected phenotypes are subtle, a low cell count will lead to a small sample size that is not truly representative of the phenotype. In cases where a fair number of perturbations may be cytotoxic, increased seeding density may mitigate the smaller numbers of surviving cells comprising the morphological sample. On the other hand, if the cell number is too high (or the cells form a confluent monolayer in the extreme case), the cells are too crowded for representative measurements of many phenotypes, particularly for image features derived from cell shape. Therefore, we recommend aiming for a seeding density that provides the cells enough space to exhibit their full-fledged morphological phenotypes while maintaining a high sample size for each phenotype expressed. Generally, we have found that ~80% confluency at the time of fixation provides a good balance.

The assay is theoretically amenable to evaluating any biological perturbation type. We have performed the Cell Painting assay with small-molecule treatments^{29,30}, viral infection³¹, transient transfection, and using selectable markers resulting in all surviving cells receiving the treatment.

Image analysis workflow for morphological feature extraction—An automated image analysis workflow is required for the image feature extraction portion of the Cell Painting assay, regardless of the experimental scale. While a number of bioimaging software packages (free and commercial) exist for morphological feature extraction¹¹, we have chosen CellProfiler for its broad range of applicable cell types, rich suite of morphological features, and optimization for analysis at large-scale and high-throughput. The image feature extraction workflow for Cell Painting is divided into three tasks, each of which is performed by a CellProfiler pipeline: (a) illumination correction, (b) quality control, and (c) morphological feature extraction.

Illumination correction serves to correct each image for spatial illumination heterogeneities introduced by the microscope optics, which can bias intensity-based measurements and impair cellular feature identification. The illumination correction pipeline aggregates the

fluorescent images on a per-plate basis to produce a post-hoc estimate of the 2-D illumination distribution, one for each channel, per plate. We have found that this corrective step improves the ability to detect subtle phenotypic differences in profiling applications⁴¹.

Quality control aims to identify and remove any aberrant images that might impair profiling quality. Because fluorescence artifacts can assume a wide variety of “phenotypes” of their own, we have chosen a supervised machine-learning approach to identify such images, using the CellProfiler Analyst software package. Machine-learning algorithms operate best when provided with a broad set of quantitative features to assign classes; we use a CellProfiler pipeline to measure a large suite of whole-image features previously validated for quality control⁴².

Finally, morphological image feature extraction provides the raw quantitative material for profiling. The third CellProfiler pipeline corrects each image with the previously-calculated illumination estimate, labels images which fail the previously-determined quality control criteria, and for each cell identifies the nucleus, cell body and cytoplasm, and makes measurements of morphology, intensity, texture, and adjacency for each cellular sub-compartment. The results are then exported for downstream analysis. When optimizing the image analysis workflow for a given cell line, attention should be given to ensuring that the cellular sub-compartments are identified robustly; beyond this, the selected morphological measurements need not be changed.

Level of expertise needed to implement the protocol

Experience with high-throughput automated equipment is required to carry out the full sample preparation portion of the protocol, although in certain cases the assay might be carried out at a smaller scale manually and using a non-automated microscope. High-content screening facilities will generally be well-equipped to aid laboratories in conducting the Cell Painting assay at larger scale. Prior image analysis experience is helpful but not necessary to carry out the image analysis procedure on a desktop computer. However, this solution is not suitable for large-scale Cell Painting assays with greater than ~1000 images. In such cases, we recommend using a computing cluster, which will likely require an information technology (IT) expert’s assistance. An active moderated forum (<http://forum.cellprofiler.org/>) exists for answering questions and troubleshooting issues that may arise using CellProfiler. Thus, despite some technical challenges, this protocol is accessible to nearly any laboratory which, at the least, has access to collaborators with some experience with high-throughput automated screening, some advanced computational skills, and possesses a willingness to learn.

Limitations

Although the Cell Painting assay is intended to be unbiased with regard to the cell type chosen, certain biological processes may simply not yield any relevant discernible morphological phenotypes, given the experimental conditions used (stains, cell type, time point, etc). In this case, augmenting the image-based profiles with additional or orthogonal assays may reveal additional biological effects that would be otherwise missed. In addition, alternate stains may be chosen to highlight the relevant cellular sub-compartments while

maintaining broad coverage of other organelles (see “Comparison with other methods” above for more details).

For RNAi experiments, the magnitude and prevalence of off-target effects in mammalian cells via the RNAi seed-based mechanism make the morphological profiles of RNAi reagents targeting the same gene rarely look any more similar than those targeting different genes³¹. This phenomenon has been observed in other multiparametric assays and is not specific to morphological profiling. Unfortunately, this effect impedes large-scale experiments using short RNAi reagents where the experimental design requires widespread comparisons across all samples. However, we note that it does not preclude experiments where the goal is to identify particular genes for which multiple RNAi reagents do yield a consistent profile, as is the case for our work identifying disease-associated phenotypes at Recursion Pharmaceuticals. An alternative gene suppression technique, CRISPR/Cas9, has not yet been extensively evaluated in conjunction with morphological profiling, but is likely to be effective.

Finally, there are several computational challenges associated with this assay. First, there are statistical challenges associated with the analyzing the high-dimensional feature space that results from Cell Painting. Similar to the case with gene expression data⁴³, issues such as the “curse of dimensionality”, model overfitting, spurious correlations, and multiple testing complicate the data analysis of this assay; these types of challenges are widely recognized in systems biology⁴⁴. Second, while in principle single-cell data is preferable over aggregated data, the former requires substantially more computational storage and processing resources; thus far, no routine analytical protocol has been established for this issue. Lastly, data analysis across separately-performed experiments is likely to be complicated, requiring proper control over the potentially substantial effects of differences in cell seeding, growth, and other batch-related or other systematic artifacts. Protocols for such cases have not yet been developed.

MATERIALS

REAGENTS

Critical: We have performed the Cell Painting assay using these specific catalog numbers. If planning on changing to a different catalog number or vendor for a given reagent, re-optimization of that reagent for the protocol may be necessary.

- Cell line of interest, e.g. we have previously used U2OS cells (ATCC, cat. no. HTB-96) or A549 cells (ATCC, cat. no. CCL-185)

Caution: Cell lines should be regularly checked to ensure that they are authentic and not infected with *Mycoplasma*.

- DMEM (Fisher Scientific, cat. no. MT-10-017-CV)
- Fetal bovine serum (FBS) (Life Technologies, cat. no. 10437-028)
- Penicillin-streptomycin (Fisher Scientific, cat. no. MT-30-002-CI)

- Trypsin, TrypLE™ Express Enzyme (Life Technologies, cat. no. 12605-036)
- Phosphate Buffered Saline pH 7.4 (PBS) (Life Technologies, cat. no. 10010-023)
- Lipofectamine RNAiMax (Life Technologies cat. no. 13778030)
- Silencer Select Pre-designed and custom siRNAs (Ambion)

Critical: Lengthy optimization is often required in new cell types for which siRNA transfection is performed. Specific conditions of transfection should be evaluated in pilot assays to confirm suitability.
- Optimem (Life Technologies, cat. no. 31985-070)
- Small-molecule libraries, typically 10 mM stock in DMSO (e.g., Chembridge Library or Maybridge Library)

Caution: Some small-molecule libraries contain toxic compounds; suitable precautions should be taken.
- MitoTracker Deep Red (Invitrogen, cat. no. M22426)

Caution: The MitoTracker stock solution is in DMSO. DMSO is a toxic chemical and easily penetrates the skin. One must avoid ingestion, inhalation and direct contact with skin and eyes. Use proper gloves to handle DMSO. Follow your institutional guidelines for using and discarding waste chemicals.
- Wheat germ agglutinin/Alexa Fluor 555 conjugate (Invitrogen, cat. no. W32464)
- Paraformaldehyde 16% (wt/vol), methanol free (Electron Microscopy Sciences, cat. no. 15710-S)

Caution: PFA is a very toxic chemical and one must avoid inhalation and/or direct contact with skin and eyes. Use proper gloves and a mask to handle PFA. Follow your institutional guidelines for using and discarding waste chemicals.
- Hank's Balanced Salt Solution (10x), HBSS (Invitrogen, cat. no. 14065-056)
- Triton X-100 (Sigma, cat. no. T8787)

Caution: Triton X-100 is a toxic chemical and one must avoid inhalation and/or direct contact with skin and eyes. Use proper gloves to handle Triton X-100. Follow your institutional guidelines for using and discarding waste chemicals.
- Phalloidin/Alexa Fluor 568 conjugate (Invitrogen, cat. no. A12380)

Caution: Phalloidin is a toxic chemical. One must avoid ingestion, inhalation and direct contact with skin and eyes. Follow your institutional guidelines for using and discarding waste chemicals.

- Concavalin A/Alexa Fluor 488 conjugate (Invitrogen, cat. no. C11252)
- Hoechst 33342 (Invitrogen, cat. no. H3570)
- SYTO 14 green fluorescent nucleic acid stain (Invitrogen, cat.no. S7576)
- Sodium bicarbonate (HyClone, cat. no. SH30033.01)
- Methanol (VMR, cat. no. BDH1135)

Caution: Methanol is a very toxic chemical and one must avoid ingestion, inhalation and/or direct contact with skin and eyes. Use proper gloves and a mask to handle methanol. Follow your institutional guidelines for using and discarding waste chemicals.

- BSA (Equitech-Bio, cat. no. BAH66)
- DMSO (Fisher Chemical, cat. no. D128-500)

Caution: DMSO is a toxic and flammable chemical and one must avoid ingestion, inhalation and/or direct contact with skin and eyes. Use proper gloves and a mask to handle DMSO. Follow your institutional guidelines for using and discarding waste chemicals.

EQUIPMENT

- Microplates: Corning 384-well black/clear flat bottom, fibronectin-coated (Corning, cat. no. 4585) or Corning 384-well black/clear flat bottom, TC-treated, bar-coded (Corning, cat. no. 3712BC). Other microplates that are compatible with the microscope will suffice, as long as they are validated for use in high-content imaging.
- Deep 384-well plates (USA Scientific, cat. no. 1884-2410)
- T-150 culture vessel (Corning, cat. no. 430825)
- Aluminum single tab foil, standard size (USA Scientific, cat. no. 2938-4100)
- Tissue culture incubator at 37 °C, 5% CO₂
- CyBi-Well 96/384-channel simultaneous pipettor (CyBio, cat. no. 3391 3 4112)
- Automated liquid handler: Multidrop Combi reagent dispenser (Thermo Scientific, cat. no. 5840300) or Freedom EVO with 384-channel arm (Tecan, cat no. MCA384)
- Plate washer: Biotek ELx405 HT

- Centrifuge: Allegra 6 (Beckman Coulter, cat. no. 366802) or PlateFuge (Benchmark Scientific, cat. no. C2000)
- ImageXpress Micro XLS epifluorescent microscope (Molecular Devices)
- CRS CatalySt Express robot microplate handler system (Thermo Scientific)
- Microscope light source: LED light engine (Lumencor)
- Access to a high-performance computer or a remote-host computing cluster (optional; recommended if planning to acquire >1,000 fields of view)
- CellProfiler and CellProfiler Analyst biological image analysis software. Available at <http://www.cellprofiler.org>.
- CellProfiler pipelines: We describe three pipelines in this protocol for Illumination correction, quality control, and feature extraction. The pipelines are available at https://github.com/carpenterlab/2016_bray_natprot/raw/master/supplementary_files/cell_painting_pipelines.zip and were created using CellProfiler 2.1.1. Please see the module notes for Cell Painting-specific documentation. Our Cell Painting wiki (https://github.com/carpenterlab/2016_bray_natprot) contains a static copy of all files used in the protocol, as well as updates to these files (e.g., to accommodate updated software versions or updated versions of the protocol).
- Raw image data from a RNAi Cell Painting knockdown experiment applied to U2OS cells³¹. Available at <https://www.broadinstitute.org/bbbc/BBBC025/>.
- ImageXpress microscope plate acquisition settings file. An example is available at https://github.com/carpenterlab/2016_bray_natprot/raw/master/supplementary_files/ImageXpress_CellPainting_plate_acquisition_settings.zip.
- Illumination correction images produced by an illumination correction pipeline applied to the U2OS image data, available at https://github.com/carpenterlab/2016_bray_natprot/raw/master/supplementary_files/illumination_correction_images.zip.
- A listing of per-cell image features generated by CellProfiler using the analysis pipeline, available at https://github.com/carpenterlab/2016_bray_natprot/raw/master/supplementary_files/cellprofiler_feature_listing.pdf.
- Files containing an example morphological profiling dataset, available at https://github.com/carpenterlab/2016_bray_natprot.
- Steps to run a Python script to produce per-well morphological profiles from single cell measurements, available as a Supplementary Method.

- A CSV of per-well profiles generated by running the profiling script, available at http://pubs.broadinstitute.org/bray_natprot_2016/suppl/online/profiles.zip.

REAGENT SETUP

MitoTracker Deep Red stock

- The product from Invitrogen (cat. no. M22426) contains 50 μg in each vial. Add 91 μL DMSO to one vial to make 1 mM solution. Store the solution at -20°C , protected from light, and use it within one month.

Wheat Germ Agglutinin (WGA)/Alexa Fluor 555 conjugate stock

- The product from Invitrogen (cat. no. W32464) contains 5 mg in each vial. Add 5 ml dH_2O to each to make 1 mg/ml solution. Store the solution at -20°C , protected from light, and use it within one month. We recommend centrifuging the WGA conjugate solution briefly before use, in order to remove any protein aggregates in solution which would contribute to nonspecific background staining.

Concanavalin A/Alexa Fluor 488 conjugate stock

- The product from Invitrogen (cat. no. C11252) contains 5 mg in each vial. Add 1 ml 0.1 M sodium bicarbonate to each vial to make 5 mg/ml solution. Store the solution at -20°C , protected from light, and use it within one month.

Phalloidin/Alexa Fluor 568 conjugate stock

- The product from Invitrogen (cat. no. A12380) contains 300 units in each vial. Add 1.5 ml methanol to each vial. Store the solution at -20°C , protected from light, and use it within one year.

SYTO 14 green fluorescent nucleic acid stain stock

- The product from Invitrogen (cat.no. S7576) is 5 mM solution in DMSO. Store the solution at -20°C , protected from light, and use it within one year.

Hoechst 33342 stock

- The product from Invitrogen (cat. no. H3570) is 10 mg/ml solution in water. Store the solution at 4°C , protected from light, and use within six months.

HBSS (1X)

- The product from Invitrogen (cat. no. 14065-056) is 10X. Add 100 ml HBSS (1X) to 900 ml water to make HBSS (1X). Filter the HBSS (1X) with 0.22 μm filter. The 1X solution should preferably be made fresh from the 10x stock solution, but can also be stored at 4°C .

BSA solution in HBSS

- Weigh 1g BSA and dissolve in 100 ml HBSS to make 1% (wt/vol) BSA solution. Filter the 1% (wt/vol) BSA solution with a 0.22 μm filter. Make fresh solution for each experiment.

Triton X-100 solution in HBSS

- Add 100 μL Triton X-100 into 100 ml HBSS to make 0.1% (vol/vol) Triton X-100 solution. Make fresh solution for each experiment.

Live cell MitoTracker staining solution

- Prepare MitoTracker staining solution by adding 25 μL MitoTracker Deep Red stock solution to 50 ml prewarmed media for a final concentration of 500 nM. Make fresh solution for each staining session.

Phalloidin, Concanavalin A, Hoechst, WGA, and SYTO 14 staining solution

- Prepare a 5 $\mu\text{L}/\text{ml}$ phalloidin solution, 100 $\mu\text{g}/\text{ml}$ Concanavalin A, 5 $\mu\text{g}/\text{ml}$ Hoechst 33342, 1.5 $\mu\text{g}/\text{ml}$ WGA, and 3 μM SYTO 14 green fluorescent nucleic acid stain solution in 1x HBSS, 1% (wt/vol) BSA. To make 50 ml stain solution, add 250 μL phalloidin stock solution, 1 ml Concanavalin A stock solution, 25 μL Hoechst stock solution, 75 μL WGA stock solution, and 30 μL SYTO 14 green fluorescent nucleic acid stain stock solution to 48.7 ml 1% (wt/vol) BSA solution in HBSS.

Compound library

- Dissolve the compounds in DMSO to yield the desired molarity; the final concentration should be such that the density is equivalent to the cell culture media. Seal and store at $-20\text{ }^{\circ}\text{C}$ for long-term storage or at RT for up to 6 months; other common compound management solutions may also be used⁴⁵.

EQUIPMENT SETUP**Microscope selection**

- The Cell Painting assay has been applied using both wide-field and confocal microscopy. Confocal microscopes are typically faster instruments than their wide-field counterparts and are able to achieve higher image contrast (and hence increased cellular feature definition and improved object segmentation) by rejecting light originating from out-of-focus planes of field. However, as compared to wide-field, confocal microscopes possess a limited number of excitation wavelengths available for use, typically higher purchase prices which may be prohibitive for smaller research groups and are traditionally lower throughput.
- We have used an ImageXpress Micro XLS epifluorescent wide-field microscope (Molecular Devices) for most of our Cell Painting assays. The

images are captured in five fluorescent channels given in Table 1. See the EQUIPMENT section for a link to a sample ImageXpress plate acquisition settings file that can be used in this protocol.

- We have also used an Opera Phenix high content screening microscope (Perkin Elmer) for Cell Painting, which is capable of imaging in both wide-field and confocal modes. However, since the Phenix uses four excitation lasers in confocal mode and the Cell Painting assay requires five channels to capture all stains, the microscope must be used wide-field mode in order to use the same filters as the ImageXpress XLS. A comparison between the excitation and emission wavelengths used for Cell Painting between the ImageXpress XLS and Phenix microscope is provided in the Supplementary Note.

Critical: The same microscope should be used for imaging all microtiter plates during an experiment. We do not recommend switching microscopes mid-stream because lamp intensities, filter patterns and other subtleties can be quite different even between supposedly identical microscope setups.

- If multiple microscopes must be used, we recommend imaging one full replicate all on one microscope, as opposed to arbitrarily assigning plates to different instruments as the experiment proceeds. The rationale is to avoid imager-induced batch effects. If the differences between perturbations are dramatic, then post-acquisition normalization will probably be effective (see “Normalize morphological features across plates” in the PROCEDURE for more details). However, if the morphological effects to be measured are subtle, normalization may not be sufficient, and the similarities in the collected image features will more likely reflect the different image acquisition than the underlying biological perturbations.

Automated image acquisition settings

- The images should be acquired with the maximum bit-depth possible, in a “lossless” image format so as to preserve all the information captured by the light sensor. Each channel should be captured as an individual grayscale image. For the ImageXpress XLS system, a 16-bit grayscale TIF is sufficient. No further pre-processing should be performed on the images prior to analysis.
- The choice of objective magnification is important as there is a trade-off between increased image feature resolution at higher magnifications (therefore enabling more specific quantification of certain organelles) versus a smaller field of view and hence less cells imaged (therefore decreasing throughput and statistical power for profile generation). Acquiring more fields of view can mitigate the latter consideration, but at the cost of a substantial increase in image acquisition and computational

processing time, especially for those who do not have access to computing cluster resources. We have found that using a 20× water immersion objective sufficiently balances all competing issues.

- Typically, nine sites are collected per well in a 3 × 3 site layout, at 20× magnification and 2× binning. Time permitting, more sites can be imaged in order to increase well coverage and improve sample statistics; it is best to capture as many cells as possible.

Critical: Avoid capturing the edges of the well in the images, particularly if a large number of sites/well are imaged. While it is feasible to remove the well edges from the images post-acquisition using image processing approaches, such methods are challenging and best avoided. One helpful approach is to reduce the field of view size in order to avoid the well edges; this setting (expressed as a percentage) is accessible through the “Sites to Visit” tab in the MetaXpress software.

- The order that the channels are imaged may have an impact on the likelihood of photobleaching during the experiment; photobleaching manifests as a decay in the fluorescence signal intensity over time with repeated illumination. Since the emission wavelengths for the chosen fluorophores are broad and in close proximity to each other, photobleaching may occur for the low-intensity dyes as they are irradiated by the lower-wavelength light. To mitigate this effect, we recommend imaging the five channels in order of decreasing excitation wavelength; more details on this point for both confocal and wide-field microscopes are shown in the Supplementary Note.
- If using an ImageXpress Micro XLS microscope, use laser-based focusing with image recovery for autofocus. The autofocus can be applied for only the first site of each well, or alternately for all sites; the latter is a minimal increase in time and is recommended for those using glass-bottom plates to decrease focussing problems.
- Use the Hoechst channel for the image recovery, with the focus binning set to 3 and a Z-offset for the other channels. The choice of optimal Z-offset will depend on the cell line and should be set by looking for the optimal focus by eye.
- Exposure times for each channel should be optimized for each experiment. Higher exposure times will yield a larger dynamic range but will increase automated image acquisition time.

Critical: Be sure that the images are not saturated. Generally, set exposure times such that a typical image uses roughly 50% of the dynamic range. For example, because the pixel intensities will range from 0 to 65,535 for a 16-bit image, a rule-of-thumb is for the typical sample to yield a maximum intensity of ~32,000. This guideline will prevent saturation (i.e.,

reaching the value 65,535) from samples that are brighter due to a perturbation.

- Do not use shading correction, as the background illumination heterogeneities will be corrected post-acquisition using the CellProfiler software.
- Prior to beginning the complete imaging run, it is useful to capture images from 3 to 5 wells at a few different locations across the plate, in order to confirm that the microscope is operating as expected and the acquisition settings are optimal for the experiment and cell line at hand.
- We recommend exporting the microscope configuration for future use once the optimal settings have been determined. See the EQUIPMENT section for a link to a sample ImageXpress plate acquisition settings file that can be used in this protocol.
- See the EQUIPMENT section for a link to an example image data set for an RNAi Cell Painting study of U2OS cells that can be used in this protocol.

Image processing software

- CellProfiler biological image analysis software is used to extract per-cell morphology feature data from the Cell Painting images, as well as per-image quality control metrics. The software and associated pipelines are designed to handle both low- and high-throughput analysis but we routinely run this software as part of this protocol on thousands, even millions, of imaged fields of view.
- To download and install the open-source CellProfiler software, go to <http://www.cellprofiler.org>, follow the download links for CellProfiler and follow the installation instructions. The current version at the time of writing is 2.1.1.
- This protocol assumes basic knowledge of the CellProfiler image analysis software package. Extensive online documentation and tutorials can be found at <http://www.cellprofiler.org/>. Also, the '?' buttons within CellProfiler's interface provide detailed help. The pipelines used here are compatible with CellProfiler version 2.1.1 and above.
- This protocol uses three CellProfiler pipelines to perform the following tasks: illumination correction, quality control; and morphological feature extraction. See the EQUIPMENT section for a link to the CellProfiler pipelines used in this protocol.
- Each module of the pipelines is annotated with details on the purpose of the module and considerations in making adjustments to the settings. The annotations may be found at the top of the settings, in the panel labeled "Module notes".

- The pipelines are configured assuming that the image files follow the nomenclature of the ImageXpress microscope system, in which the plate/well/site metadata is encoded as part of the filename. The plate and well metadata in particular are essential because CellProfiler uses the plate metadata in order to process the images on a per-plate basis, and the plate and well metadata are needed for linking the plate layout information with the images for the downstream profiling analysis. Therefore, images coming from a different acquisition system may require adjustments to the Metadata module to capture this information; please refer to the help for this module for more details.
- The quality control and morphological feature extraction pipelines are set to write out cellular features to a MySQL database which is recommended for analyses involving >1,000 images; see “Computing system” for details. If using a smaller number of images, the pipelines can be adjusted to output the measured features to a comma-delimited file (CSV) using the *ExportToSpreadsheet* module. Third-party data analysis tools may be more amenable to importing data from a CSV-formatted file than from a MySQL database. However, the scripts provided to generate per-well profiles from the extracted features are MySQL-only; see the EQUIPMENT section for a link to the Python scripts used in this protocol.

Computing system

- If the number of images to analyze is sufficiently large that a single computer would take too long to process them (e.g., more than ~1,000), we recommend using a computing cluster if available, such as a high-performance server farm or a cloud-computing platform such as Amazon AWS. Carrying out this step requires significant setup effort and will probably require enlisting the help of your IT department. Please refer to our GitHub webpage <https://github.com/CellProfiler/CellProfiler/wiki/Adapting-CellProfiler-to-a-LIMS-environment> for more details.
- We recommend setting up a MySQL database to allow multiple CellProfiler processes to write out cellular features in parallel; doing so will probably require enlisting the help of your IT department. This database will be used by the CellProfiler module *ExportToDatabase* to create data tables as described under PROCEDURE, as well as by the scripts listed under EQUIPMENT to generate per-well profiles from the extracted features.

Image data exploration software

- The CellProfiler Analyst data exploration software may be used to explore the data or for quality control⁴².
- To download and install the CellProfiler Analyst software, go to <http://cellprofiler.org>, follow the download link for CellProfiler Analyst and

follow the installation instructions. The current version of CellProfiler Analyst at the time of writing is 2.2.0.

PROCEDURE

Cell Culture

Timing: variable; 2 - 3 d—CRITICAL The following cell-plating procedure is validated for many cell types; each step may need adjustment depending on local conditions or alternate cell types. We have included recommended optional steps for experiments involving small molecule library treatment and siRNA transfection.

CRITICAL Check the wiki at GitHub for any updates to the Cell Painting protocol: https://github.com/carpenterlab/2016_bray_natprot.

- 1** Prepare cells for seeding according to known best practices for the cell type of choice. For most high-content applications, a black plate with a clear, flat bottom for cell culture is appropriate. The following protocol is validated for use on A549 cells in Corning 384-well 200 nm thick glass-bottomed plates.
- 2** Grow cells to near confluence (~80%) in a T-150 culture vessel.
- 3** Optional: If performing experiments involving addition of compounds (step 8A), prepare the compound library according to the instructions in REAGENT SETUP. If performing experiments involving siRNA transfection (step 8B), prepare the siRNA transfection reagent mixture according to Box 1.
- 4** Rinse cells with PBS without Ca^{2+} or Mg^{2+} .
- 5** Add 6 ml TrypLE Express and incubate at room temperature (21°C) for 30 seconds. Remove the TrypLE Express and add 1 ml fresh TrypLE Express. Incubate at 37°C until the cells have detached. This should occur within 3 - 5 minutes.
- 6** Add 10 ml growth medium to deactivate trypsin, and determine the live cell concentration using standard methods (hemocytometer or cell counter).
- 7** Dilute A549 cells to 50,000 live cells/ml in media, and dispense 40 μl (2,000 live A549 cells) to each well of the 384-well plates. For large-scale Cell Painting assays, we recommend the use of an automated liquid handling system. Different cell types and growth conditions will require variations in seeding density; typical ranges will vary from 1,500 to 3,000 cells/well.

Critical step: Adequately resuspend the cell mixture to ensure a homogeneous cell suspension prior to each dispense. It is not uncommon for cells to rapidly settle in their reservoir resulting in plate-to-plate variation in cell numbers. If utilizing a liquid handler with a multi-dispense function, be sure to adequately prime the dispensing cassette and/or dispense at least 10 μL of cell suspension back into the reservoir prior to dispensing the cells into culture plates; the latter is helpful if cells or reagent are sticking to the tubing.

Critical step: When handling liquid for many plates with one set of tips, confirm that no residual bubbles within the tips touch the head of the liquid handler during aspiration in order to ensure accurate liquid dispensation.

Box 1**Prepare siRNA transfection reaction mixture**

The following protocol is written for a 10 nM final siRNA concentration in 0.1% Lipofectamine RNAiMAX which achieves >70% knockdown for diverse targets in A549, U2OS, and HUVEC cells.

1. Thaw and dilute siRNAs to a concentration of 2 μ M in sterile molecular-grade water.
2. Dilute Lipofectamine RNAiMAX with FBS-free Opti-MEM medium (1:100) in appropriate RNase-free deep 384-well plates.
3. Mix 0.2 μ L diluted siRNA with 10 μ L diluted Lipofectamine RNAiMAX.
4. Ensure that siRNA-Lipofectamine mix has incubated at room temperature for at least 10 min before proceeding to step 8 of the main procedure.

Treatment with small molecule library or siRNA transfection

- 8 If performing treatments with a small molecule library, please follow option A. If performing siRNA transfection, follow option B.

A. Small molecule library

Timing: variable; approximately 2 - 3 d for one batch experiment of 384-well plates

- i. Allow plates to set on a flat, level surface at room temperature for 1 – 2 hrs to reduce plate edge effects⁴⁶.
- ii. Put the plates into the incubator (37°C, 5% CO₂, 90–95% humidity). To reduce plate edge effects produced by incubator temperature variations and media evaporation, we recommend either spacing out the plates in the incubator or using racks with “dummy” plates filled with liquid placed on the top and bottom. We also recommend rotating the plates/stacks within the incubator to avoid positional effects.
- iii. Replace the culture medium with 50 μ L of 2% (vol/vol) FBS in DMEM 24 hours after seeding. Perform the aspiration steps using a plate washer such as the BioTek ELx405 microplate washer or equivalent. Reducing FBS concentration minimizes the risk of overgrowth at the time of fixation. Optimal FBS concentrations may vary depending on the cell type and transfection reagent selected.

- iv. Add compounds to cells using a pin tool or liquid handler. We have added compounds either 24 or 48 hours prior to staining and fixation, but the timing should be adjusted depending on the growth rate of each cell type and the biological processes under consideration. Recursion Pharmaceuticals typically adds compounds to cells in an environment that is antibiotic-free (to avoid perturbations arising from complex antibiotic-drug interactions) and low-serum (to synchronize cell state). To ensure adequate mixture of compounds in solution, we recommend that compounds are mixed well in the culture medium before adding to the cells.

B. siRNA transfection

Timing: approximately 5 d for one batch experiment of 384-well plates

- i. Add the siRNA-transfection reagent mixture to the 384-well plate using the CyBi-Well simultaneous pipettor, at 10.2 μ L /well.
- ii. Allow plates to set on a flat, level surface at room temperature for 1 – 2 hrs to reduce plate edge effects⁴⁶.
- iii. Put the plates into the incubator (37°C, 5% CO₂, 90–95% humidity). To reduce plate edge effects produced by incubator temperature variations and media evaporation, we recommend either spacing out the plates in the incubator or using racks with “dummy” plates filled with liquid placed on the top and bottom. We also recommend rotating the plates/stacks within the incubator to avoid positional effects.
- iv. Replace the culture medium with 50 μ l of 2% (vol/vol) FBS in DMEM 24 hours after seeding. Perform the aspiration steps using a plate washer such as the BioTek ELx405 microplate washer or equivalent. Reducing FBS concentration minimizes the risk of overgrowth at the time of fixation. Optimal FBS concentrations may vary depend on the cell type and transfection reagent selected.
- v. Incubate the cells for 3 more days prior to fixation and staining.
- vi. Optional: Add starvation medium (0.1% (vol/vol) FBS in DMEM) to the cells approximately 24 hours prior to staining and fixation in order to synchronize cell growth rate. Cell cycle synchronization may improve profile quality because it reduces variability in whole-culture cell cycle stage; however, profiling of asynchronous cell populations may enable capture of phenotypes affecting all stages of the cell cycle.

Staining and Fixation

Timing: variable; 2.5–3 h for one batch experiment of 384-well plates—Critical:

If using a cell line for the first time, we suggest testing the staining protocol on a pilot plate in order to manually confirm visibility of the cellular features (see Figures 1 and 3 of this protocol, and Figure S1 of Gustafsdottir et al²⁹).

- 9 Prepare the live cell MitoTracker staining solution for all plates.
- 10 Remove media from plates, set the aspiration height in the plate washer to leave 10 μ L of residual volume to minimize the disturbance to the live cells from the pins and media turbulence.
- 11 Add 30 μ L of MitoTracker staining solution.
- 12 Centrifuge the plate (500 g at room temperature for 1 min) after adding stain solutions and ensure there are no bubbles in the bottom of the wells.
- 13 Incubate the plates for 30 min in the dark at 37 °C.
- 14 Prepare the phalloidin, Concanavalin A, Hoechst, WGA, and SYTO 14 staining solution for all plates.
- Critical Step:** Prepare the working stain solution before use. Do not store the working stain solution exposed to light or over long periods, to maintain fluorescence.
- 15 To fix the cells, add 10 μ L of 16% (wt/vol) methanol-free paraformaldehyde for a final concentration of 3.2% (vol/vol).
- Critical Step:** We recommend performing the fixation and subsequent permeabilization and staining steps with no pauses. In our hands, halting between steps, e.g., between the fixing/permeabilizing steps and the staining step, results in degradation of the SYTO 14 staining quality.
- Critical Step:** Having FBS or BSA present during fixation may help prevent cellular retraction.
- 16 Centrifuge the plate (500 g at room temperature for 1 min) after adding stain solutions and ensure there are no bubbles in the bottom of the wells.
- 17 Incubate the plates in the dark at room temperature for 20 min.
- 18 Wash the plates once with 70 μ L 1X HBSS.
- 19 To permeabilize the cells, remove HBSS and add 30 μ L of 0.1% (vol/vol) Triton X-100 solution to the wells.
- 20 Centrifuge the plate (500 g at room temperature for 1 min) after adding stain solutions and ensure there are no bubbles in the bottom of the wells.
- 21 Incubate the plates in the dark at room temperature for 10–20 min.
- Critical Step:** Once the MitoTracker solution is added, take special to keep the cells dark for the rest of the experiment.
- 22 Wash the wells twice with 70 μ L 1x HBSS.
- 23 Remove HBSS and add 30 μ L of the phalloidin, Concanavalin A, Hoechst, WGA, and SYTO 14 staining solution to each well.
- 24 Centrifuge the plate (500 g at room temperature for 1 min) after adding stain solutions and ensure there are no bubbles in the bottom of the wells.

- 25 Incubate the plates in the dark at room temperature for 30 min.
- 26 Wash cells three times with 70 μ L 1x HBSS, with no final aspiration.
- 27 Seal plates with adhesive foil and store at 4°C in the dark.

Automated image acquisition

Timing: variable; 1–3.5 hours per 384-well plate

- 28 Mount the microtiter plates into the automated microscopy system for imaging. For large-scale Cell Painting assays, we recommend the use of an automated microplate handling system.
- 29 Set up the microscope acquisition settings as described in EQUIPMENT SETUP.
- 30 Start the automated imaging sequence according to the microscope manufacturer's instructions.

(TROUBLESHOOTING)

Morphological image feature extraction from microscopy data

Timing: variable; 20 h per batch of 384-well plates

- 31 Illumination correction to improve fluorescence intensity measurements. Start CellProfiler.
- 32 Load the illumination correction pipeline into CellProfiler by selecting *File > Import > Pipeline from File* from the CellProfiler main menu and selecting *illumination.cppipe*.

Critical step: Non-homogeneous illumination introduced by microscopy optics can result in errors in cellular feature identification and can degrade the accuracy of intensity-based measurements. This is an especially important problem in light of the subtle phenotypic signatures that morphological profiling aims to capture. Non-homogeneous illumination can occur even when fiber-optic light sources are used and even if the automated microscope is set up to perform illumination correction. The use of a uniformly fluorescent reference image (“white-referencing”), while common, is not suitable to high-throughput screening. A retrospective method to correct all acquired images on a per-channel, per-plate basis is therefore recommended⁴¹; the illumination pipeline takes this approach.

- 33 Select the *Images* input module in the 'Input modules' panel to the top-left of the interface. From your file browser, drag and drop the folder(s) containing your raw images into the 'File list' panel. See the EQUIPMENT section for a link to raw image files that can be used as an example in this protocol.
- 34 Click the 'View output settings' button at the bottom-left of the interface. In the settings panels, select an appropriate 'Default Output Folder' where the illumination correction images will be saved.

- 35 Save the current settings to a project (.cpproj) file containing the pipeline, the list of input images and the output location by selecting *File > Save Project*. Enter the desired project filename in the dialog box that appears.
- 36 Press the 'Analyze Images' button at the bottom-left of the interface. A progress bar in the bottom-right will indicate the estimated time of completion. The end result of this step will be a collection of illumination correction images in the Default Output Folder, one for each plate and channel. We have provided an example set of images for comparison on our Cell Painting wiki (see EQUIPMENT for details).

Critical step: This step assumes that you will be running the illumination correction pipeline locally on your computer. If your institution has a shared high-performance computing cluster, we recommend executing the pipeline on the cluster as a batch process, i.e., a series of smaller processes entered at the command line; this will result in much more efficient processing. Enlist the help of your institution's IT department to find out whether this is an option and what resources are available. If so, carry out instructions in Box 2, describing modifications to the pipeline to run it as a batch process.

- 37 **Quality control to identify and exclude aberrant images.** Start CellProfiler, if not already running.
- 38 Load the quality control pipeline into CellProfiler by selecting *File > Import > Pipeline from File* from the CellProfiler main menu and selecting *qc.cppipe*.

Critical step: As mentioned above, high-quality images are essential for robust downstream analysis of Cell Painting data. Therefore, we recommend implementing quality control (QC) measures. The approach detailed here uses CellProfiler to analyze the data using quality control metrics that do not require cell identification⁴². However, the same goal can be met with other analytical approaches after cell identification and measurement.

- 39 Select the *Images* input module in the 'Input modules' panel to the top-left of the interface. From your file browser, drag and drop the folder(s) containing your raw images into the 'File list' panel (a shortcut for this step is to simply use the same project file as the illumination step above and load the QC pipeline to replace the illumination pipeline, while retaining the same image list).
- 40 Select the *ExportToDatabase* module. The setting for "Database name" is highlighted in red because it is waiting for a proper value to be provided. Change the fields for "Database name", "Database host", "Username" and "Password" to their respective values appropriate for your MySQL database server; the red text will disappear at that point. Once done, you can press the "Test connection" button to confirm that the settings are correct. The setting "Table Prefix" should be changed to a different value such that a new table is created and used for this experiment.

Critical step: For smaller number of images (e.g., less than 1,000) or instances when a MySQL database is not available, the *ExportToSpreadsheet* module may instead be used to export the collected measurements to a comma-delimited file (CSV) format. Add this module by selecting *Edit > Add Module > File Processing > ExportToSpreadsheet* from the CellProfiler main menu, and position it as the last module in the pipeline by using the “^” or “v” buttons beneath the pipeline. Disable the *ExportToDatabase* module by clicking the green checkmark next to the module name; the green checkmark will then be grayed out to indicate its status. Select the *ExportToSpreadsheet* module, select “No” for “Export all Measurement types?”, and then select “Image” from the “Data to export” drop-down drop that appears. This will place a CSV file containing the per-image data in the ‘Default Output Folder’.

- 41 Click the 'View output settings' button at the top-left of the interface. In the settings panels, select an appropriate 'Default Output Folder' where the QC data will be saved.
- 42 Save the current settings to a project (.cproj) file containing the pipeline, the list of input images and the output location by selecting *File > Save Project*. Enter the desired project filename in the dialog box that appears.
- 43 Press the 'Analyze Images' button at the bottom-left of the interface. A progress bar in the bottom-left will indicate the estimated time of completion.

Critical step: This step assumes that you will be running the quality control pipeline locally on your computer. If your institution has a shared high-performance computing cluster, we recommend executing the pipeline on the cluster as a batch process, i.e., a series of smaller processes entered at the command line; this will result in much more efficient processing. Enlist the help of your institution's IT department to find out whether this is an option and what resources are available. If so, carry out instructions in Box 2, describing modifications to the pipeline to run it as a batch process.

- 44 When the QC processing run is completed, apply the workflow described in Bray et al⁴² to use CellProfiler Analyst to explore the data and select QC image features and thresholds in order to exclude out-of-focus and saturated images from further analysis.
- 45 **Image analysis to extract morphological features.** Start CellProfiler, if not already running.
- 46 Load the analysis pipeline into CellProfiler by selecting *File > Import > Pipeline from File* from the CellProfiler main menu and selecting *analysis.cppipe*.
- 47 Select the *Images* input module in the 'Input modules' panel to the top-left of the interface. From your file browser, drag and drop the folder(s) containing your raw images into the 'File list' panel (A shortcut for this step is to simply use the same project file as the illumination step above and load the analysis

pipeline to replace the illumination pipeline while retaining the same image list). For this step, you should also drag and drop the folder containing your illumination correction images into the 'File list' panel.

- 48** Select the *FlagImages* module, which is used to label images with a metadata tag of 0 or 1, depending on whether features from particular image channels pass or fail chosen QC criteria, respectively. Two sample measurements and thresholds are provided in the pipeline; the choice of QC image feature(s) and threshold(s) should be adjusted to reflect your results from step 44. To add more QC image features to an existing flag, press the "Add another measurement" button, and select "ImageQuality" as the category, the desired measurement and image from the respective drop-box boxes, whether the image is flagged based on a high or low threshold values, and the actual threshold value for the measurement. You can also add more flags by pressing the "Add another flag" button, give the flag a name and specify whether the image needs to pass any or all of the criteria to be flagged; you may add as many flags and/or features to a flag as needed. If you do not wish to use this module for QC, you can disable the module by clicking the green checkmark to the left of the module name; the checkmark is grayed out when the module is disabled.
- 49** Select the *ExportToDatabase* module, which is used to write image-based feature measurements to a MySQL database. The setting for "Database name" is highlighted in red because it is waiting for a proper value to be provided. Change the fields for "Database name", "Database host", "Username" and "Password" to their respective values appropriate for your MySQL database server; the red text will disappear at that point. For provenance purposes, we recommend that the "Database name" field should be the same as that used for the quality control step above. Once done, you can press the "Test connection" button to confirm that the settings are correct. The setting "Table Prefix" should be changed to a different value such that a new table is created and used for this step.

Critical step: For smaller number of images (e.g., less than 1,000) or instances when a MySQL database is not available, the *ExportToSpreadsheet* module may be used to export the collected measurements to a comma-delimited file (CSV) format. Add this module by selecting *Edit > Add Module > File Processing > ExportToSpreadsheet* from the CellProfiler main menu, and position it as the last module in the pipeline by using the "∧" or "∨" buttons beneath the pipeline. Disable the *ExportToDatabase* module by clicking the green checkmark next to the module name; the green checkmark will then be grayed out to indicate its status. Select the *ExportToSpreadsheet* module, select "No" for "Export all Measurement types?", and then select "Image" for the "Data to export" drop-down drop that appears. Click the "Add another data set" button and select "Nuclei" for the "Data to export" drop-down drop that appears. Click the "Add another data set" button again and select "Cells" for the "Data to export" drop-down drop that appears, and select "Yes" for the

“Combine these object measurements with those of the previous object?” setting. Click the “Add another data set” button again and select “Cytoplasm” for the “Data to export” drop-down drop that appears, and select “Yes” for the “Combine these object measurements with those of the previous object?” setting. This will place two CSV files, containing the per-image and per-cell data, in the Default Output Folder.

- 50 Click the 'View output settings' button at the top-left of the interface. In the settings panels, select an appropriate 'Default Output Folder' where the analysis data will be saved.
- 51 Save the current settings to a project (.cproj) file containing the pipeline, the list of input images and the output location by selecting *File > Save Project*. Enter the desired project filename in the dialog box that appears.
- 52 Use CellProfiler’s Test mode functionality (accessible from the “Test” menu item) to carry out analysis and visually inspect results from a small sample of images from across the experiment for accuracy of nuclei and cell body identification. Adjust image analysis pipeline parameters within CellProfiler as needed. The CellProfiler website contains resources and tutorials on how to optimize an image analysis pipeline. The Anticipated Results section outlines the expected nuclei and cell identification quality.

Critical step: Because capturing subtle phenotypes is important for profiling, accurate nuclei and cell body identification is essential for success. Examine the outputs of *IdentifyPrimaryObjects* and *IdentifySecondaryObjects* for a few images to make sure the boundaries generally match expectations. Under the “Test” menu item, there are options for selecting sites for examination. We recommend either randomly sampling images for inspection (via “Random Image set”) and/or selecting specific sites (via “Choose Image Set”) from negative control wells or specific treatment locations from the plates. The rationale is to check a wide variety of treatment-induced phenotypes to ensure that the pipeline will generate accurate results.

(TROUBLESHOOTING)

- 53 Press the 'Analyze Images' button at the bottom-left of the interface. A progress bar in the bottom-left will indicate the estimated time of completion. The pipeline will identify the nuclei from the Hoechst-stained image (referred to as “DNA” in CellProfiler), use the nuclei to guide identification of the cell boundaries using the SYTO 14-stained image (“RNA” in CellProfiler), and then use both of these features to identify the cytoplasm. The pipeline then measures the morphology, intensity, texture, and adjacency statistics of the nuclei, cell body and cytoplasm, and outputs the results to a MySQL database. See the EQUIPMENT section for a link to a listing of the image features measured for each cell.

Critical step: This step assumes that you will be running the image analysis pipeline locally on your computer, which generally is only recommended for

experiments with less than 1,000 fields of view. If your institution has a shared high-performance computing cluster, we recommend executing the pipeline on the cluster as a batch process, i.e., a series of smaller processes entered at the command line; this will result in much more efficient processing. Enlist the help of your institution's IT department to find out whether this is an option and what resources are available. If so, carry out instructions in Box 2, describing modifications to the pipeline to run it as a batch process.

Box 2

Configuring pipelines for batch processing on a computer cluster

We recommend using a computing cluster for analyzing Cell Painting experiments to speed processing, especially for experiments with greater than 1,000 fields of view. The typical batch processing workflow is to distribute smaller subsets of the acquired images to run on individual computing nodes. Each subset is run using CellProfiler in “headless” mode, i.e., from the command line without the user interface. The headless runs are executed in parallel, with a concomitant decrease in overall processing time.

Carrying out this step requires significant setup effort and will probably require enlisting the help of your IT department. Please refer to our GitHub webpage <https://github.com/CellProfiler/CellProfiler/wiki/Adapting-CellProfiler-to-a-LIMS-environment> for more details.

1. Insert the CreateBatchFiles module into the pipeline by pressing the ‘+’ button, and selecting the module from the “File Processing” category. Move this module to the end of the pipeline by selecting with your mouse and using the ‘^’ or ‘v’ buttons at the bottom-left of the interface.
2. Configure the CreateBatchFiles module by setting the ‘Local root path’ and ‘Cluster root path’ settings. If your computer mounts the file system differently than the cluster computers, CreateBatchFiles can replace the necessary parts of the paths to the image and output files. For instance, a Windows machine might access files images by mounting the file system using a drive letter, e.g., C:\your_data\images and the cluster computers access the same file system using /server_name/your_name/your_data/images. In this case, the local root path is C:\ and the cluster root path is /server_name/your_name. You can press the ‘Check paths’ button to confirm that the path mapping is correct.
3. Press the ‘Analyze Images’ button at the bottom-left of the interface.
4. The end result of this step will be a ‘Batch_data.h5’ (HDF5 format) file. This file contains the pipeline plus all information needed to run on the cluster.

5. This file will be used as input to CellProfiler on the command line, in order for CellProfiler to run in “headless” mode on the cluster. There are a number of command line arguments to CellProfiler that allow customization of the input and output folder locations, as well as which images are to be processed on a given computing node. Enlist an IT specialist to specify the mechanism for sending out the individual CellProfiler processes to the computing cluster nodes. Please refer to our GitHub webpage <https://github.com/CellProfiler/CellProfiler/wiki/Adapting-CellProfiler-to-a-LIMS-environment> for more details.

Normalize morphological features across plates

Timing: < 5 min per 384-well plate

- 54 The extracted features need to be normalized to compensate for variations across plates. For each feature, compute the median and median absolute deviation for all reference cells within a plate. The reference cells need not be a perfect negative control but instead simply provide a baseline from which other treatments can be measured. In RNAi and overexpression experiments, we have found untreated cells to be an effective baseline for normalization. For chemical experiments, we have found DMSO-treated cells to be an effective baseline.
- 55 Normalize the feature values for all the cells (both treated and untreated) in the plate by subtracting the median and dividing by the median absolute deviation (MAD) 1.4826 (multiplying the MAD by this factor provides a good estimate of the standard deviation for normal distributions).
- 56 Exclude features having MAD = 0 in any plate, because when this is the case, all samples have the exact same value for that feature and thus the feature does not carry any sample-specific information. See the EQUIPMENT section for sample profiling scripts to perform the normalization in this protocol.

Create per-well profiles

Timing: 1 h per 384-well plate

- 57 There are a number of possible approaches to creating per-well profiles from the individual cell measurements from each image/site within the well. We have published a comparison of several such methods for creating morphological profiles⁴⁷. Here, we describe the approach of population-averaging profiles, which has been shown to be effective. For each well, compute the median for each of the n features across all the cells in the well. This produces an n -dimensional data vector per well.
- 58 (Optional) Use principal components analysis (PCA) to reduce the dimensionality of the data. To do this, collect all the n -dimensional data vectors corresponding to all the k wells in the experiment, and produce an $n \times k$ dimensional data matrix. Perform PCA on this data matrix to obtain a lower

dimensional representation of the per-well profiles. For instance, in one of our recent papers²⁹, the dimensionality of the data vectors was reduced from 1301 to 205, while preserving 99% of the variance in the data. Other methods to reduce dimensionality or select a subset of features in morphological profiling data include: factor analysis^{16,17}, stepwise feature selection to remove linear dependencies^{25,27}, and SVM recursive feature elimination¹⁵. We refer the reader to a review on feature selection methods⁴⁸ to evaluate the advantages and disadvantages of these approaches.

- 59 See the Supplementary Methods for sample profiling scripts to create the per-well profiles in this protocol.

Data analysis

Timing: variable

- 60 Use the per-well profiles to analyze patterns in the data. How to do so is an area of active research and is customized to the biological question at hand. For example, morphological profiles were used to discover compounds that induce similar phenotypes using clustering²⁹, to identify compound sets with high rates of activity and diverse biological performance in combination with high-throughput gene-expression profiles³⁰, and to determine the dominance of seed-sequence driven off-target effects in RNAi-induced gene knockdown studies³¹. See the cited publications for example analyses and code; our own laboratory is developing an R package for this purpose at <https://github.com/CellProfiler/cytominr>. A typical profiling data analysis workflow begins with the per-well profiles; for most applications a key step is measuring the similarity (or, equivalently, distance) between each sample's profile and all other profiles in the experiment. Methods often used for measuring similarity or distance are Pearson correlation, Spearman correlation, Euclidean distance, and cosine distance. For quality control purposes, it is customary to check that replicates of the same sample yield small distances. If positive controls are available (that is, samples that are known to yield similar phenotypes), their replicates can also be checked for producing small distances relative to random pairs of samples. Samples are often clustered using hierarchical clustering, although other clustering methods may also be used.

TIMING

Cell culture

It typically takes about 2–3 days for the cells to reach appropriate confluency depending on cell type and growth conditions. Harvesting the cells (steps 4–6) takes 30 minutes and seeding the cells (step 7) takes 20 minutes. Optional transfection of siRNA takes 3 hours for one batch experiment of 384-well plates, including reagent preparation and media change (steps 3 and 8), and 2 days for siRNAs to achieve appropriate knockdown. Optional addition of a compound library takes about 3 hours for one batch experiment of 384-well plates,

including reagent preparation and media change (steps 3 and 8), and 1 to 2 days for compound incubation. After seeding, the cells are cultured for 2–5 days before staining.

Staining and fixation

Approximately 2.5 – 3 hours including reagent preparation. The total timing will vary depending the number of plates in the experiment and the automation available. We have found that up to 12 plates can be simultaneously fixed and stained as one batch in this span of time, although we recommend no more than 4 – 5 plates per batch due to the increased likelihood of sample preparation error by the researcher.

Automated image acquisition

About 3.5 hours per 384-well plate, for 9 fields-of-view per well and typical exposure times (and as little as 1 hour per plate for smaller numbers of fields-of-view). The total time may vary depending on the number of sites imaged per plate and exposure time for each channel.

Morphological image feature extraction from microscopy data

Approximately 10 minutes per plate for CellProfiler to scan the inputs folder(s) after manually drag/dropping the needed images into the CellProfiler interface. The pipeline execution time will depend on the computing setup; run times on a single compute node of 20 sec (illumination correction), 30 sec (quality control), and 10 min (analysis) per field of view are typical. A substantial time-savings can be achieved if running the feature extraction and quality control pipelines on a distributed computing cluster, which massively parallelizes the processing as compared to running on a single local computer. Performing the quality control workflow using CellProfiler Analyst takes approximately 4 hrs of hands-on time, though this time can be significantly shortened if cutoffs are re-used from experiment to experiment.

Normalize morphological features across plates

Less than five minutes of processing time per 384-well plate. Additional time is required to set up data access infrastructure prior to this step. Create per-well profiles: Up to an hour of processing time per 384-well plate, depending on data access methods.

Data analysis

Approximately one hour for basic analysis of replicate quality and signature strength. Time for additional analysis varies significantly depending on the problem at hand.

TROUBLESHOOTING

For troubleshooting advice, see Table 2.

ANTICIPATED RESULTS

The automated imaging protocol will produce a large number of acquired images in 16-bit TIF format; each resultant image will be $1,080 \times 1,080$ pixels ($0.656 \mu\text{m}/\text{pixel}$) and ~ 2.3 MB in size. The total number of images generated equals (number of samples tested) \times (number

of sites imaged per well) \times (5 channels imaged). In terms of data storage, a single 384-well microplate will produce 3,456 fields of view, or 17,280 images total across all channels, for a total of ~40 GB/microplate. Results from a typical Cell Painting experiment (an siRNA study using HUVEC cells) are shown for an untreated negative control well and a treated well (Figure 3).

In addition, the illumination correction pipeline will yield five illumination correction images per plate, one for each channel. One microplate's worth of illumination correction images will occupy ~23 MB of storage space. The quality control pipeline will produce a set of numerical measurements extracted at the image level, and export them to a MySQL database. These measurements can optionally be used for removing images that are unacceptable for further processing due to focal blur or saturation artifacts.

The image analysis pipeline will produce several outputs. Generally, the pipeline is not configured to save any processed images (to conserve data storage space) but the *SaveImages* module can be used for this purpose if desired, e.g., for saving outlines such as those in the last column of Figure 3. The pipeline also produces the raw numerical image features extracted from the cell images, which are deposited to a MySQL database. The database contains one row for each cell in each image, and ~1,500 columns containing the values for the different morphological features that have been measured for that cell. The combination of data tables for the quality control and feature extraction pipelines typically total ~6.4 GB/microplate.

The quality of the extracted image features and downstream profiling will depend on accurate nuclei and cell body segmentation. The last column of Figure 3 contains overlays of the nuclei and cell body identification (i.e., segmentation) to highlight the differences in cellular morphology between the two treatments. First, the nuclei are identified from the Hoechst image because it is a high-contrast stain for a well-separated organelle; subsequently, the nucleus along with an appropriate channel is used to delineate the cell body⁴⁹. We have found the SYTO 14 image is the most amenable for finding cell edges, as it has fairly distinct boundaries between touching cells. We have found that very little adjustment of the analysis pipeline is needed to achieve good quality data, even across multiple cell lines. Even so, it is important to understand the key segmentation parameters of the pipeline in order to optimize the output if needed (see Troubleshooting note for step 60, "Automated image acquisition").

After running the profiling scripts to normalize the image features across plates and create the per-well profiles, the output will be a morphological profile file in comma-delimited format (CSV). Each row of this file represents data vector for an individual plate and well, with each column containing the median for each of the ~1,500 image features across all the cells in that well.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Research reported in this publication was supported in part by NIH R44 TR001197 (CCG), NSF RIG DBI 1119830 (MAB), NIH R01 GM089652 (AEC) and NSF CAREER DBI 1148823 (AEC). The RNAi Cell Painting knockdown experiment used in this publication was previously published³¹ and was supported in part by the Slim Initiative for Genomic Medicine, a project funded by the Carlos Slim Foundation in Mexico. The authors thank the original developers of earlier versions of the protocol, who are the authors of the original paper describing the assay²⁹; these authors are: Sigrun M. Gustafsdottir, Vebjorn Ljosa, Katherine L. Sokolnicki, J. Anthony Wilson, Deepika Walpita, Melissa M. Kemp, Kathleen Petri Seiler, Hyman A. Carrel, Todd R. Golub, Stuart L. Schreiber, Paul A. Clemons, Anne E. Carpenter, Alykhan F. Shamji. For enabling this work, Recursion Pharmaceuticals thanks the University of Utah Core Facilities (John Phillips), specifically the Drug Discovery Core (Bai Luo) and the Fluorescent Imaging Core (Chris Rodesch). The Broad thanks members of the Broad Institute's Center for the Development of Therapeutics, especially Thomas P. Hasaka for technical assistance. We also thank Alison Kozol for proofreading the equipment and reagents and for testing the image analysis workflow, and Alice Berger and Xiaoyun Wu for offering helpful comments and suggestions during manuscript preparation.

References

1. Swinney DC, Anthony J. How were new medicines discovered? *Nat Rev Drug Discov.* 2011; 10:507–519. [PubMed: 21701501]
2. Swinney DC. The Contribution of Mechanistic Understanding to Phenotypic Screening for First-in-Class Medicines. *J Biomol Screen.* 2013; 18:1186–1192. [PubMed: 23983234]
3. Moffat JG, Joachim R, David B. Phenotypic screening in cancer drug discovery — past, present and future. *Nat Rev Drug Discov.* 2014; 13:588–602. [PubMed: 25033736]
4. Johannessen CM, Clemons PA, Wagner BK. Integrating phenotypic small-molecule profiling and human genetics: the next phase in drug discovery. *Trends Genet.* 2015; 31:16–23. [PubMed: 25498789]
5. Bickle M. The beautiful cell: high-content screening in drug discovery. *Anal Bioanal Chem.* 2010; 398:219–226. [PubMed: 20577725]
6. Singh S, Carpenter AE, Genovesio A. Increasing the content of high-content screening: an overview. *J Biomol Screen.* 2014; 19:640–650. [PubMed: 24710339]
7. Perlman ZE, et al. Multidimensional drug profiling by automated microscopy. *Science.* 2004; 306:1194–1198. [PubMed: 15539606]
8. Danuser G. Computer Vision in Cell Biology. *Cell.* 2011; 147:973–978. [PubMed: 22118455]
9. Altschuler SJ, Wu LF. Cellular heterogeneity: do differences make a difference? *Cell.* 2010; 141:559–563. [PubMed: 20478246]
10. Snijder B, Pelkmans L. Origins of regulated cell-to-cell variability. *Nat Rev Mol Cell Biol.* 2011; 12:119–125. [PubMed: 21224886]
11. Eliceiri KW, et al. Biological imaging software tools. *Nat Methods.* 2012; 9:697–710. [PubMed: 22743775]
12. Paull KD, et al. Display and analysis of patterns of differential activity of drugs against human tumor cell lines: development of mean graph and COMPARE algorithm. *J Natl Cancer Inst.* 1989; 81:1088–1092. [PubMed: 2738938]
13. Lamb J, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science.* 2006; 313:1929–1935. [PubMed: 17008526]
14. Adams CL, et al. Compound classification using image-based cellular phenotypes. *Methods Enzymol.* 2006; 414:440–468. [PubMed: 17110206]
15. Loo LH, Wu LF, Altschuler SJ. Image-based multivariate profiling of drug responses from single cells. *Nat Methods.* 2007; 4:445–453. [PubMed: 17401369]
16. Young DW, et al. Integrating high-content screening and ligand-target prediction to identify mechanism of action. *Nat Chem Biol.* 2008; 4:59–68. [PubMed: 18066055]
17. Ljosa V, et al. Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment. *J Biomol Screen.* 2013; 18:1321–1329. [PubMed: 24045582]

18. Reisen F, et al. Linking Phenotypes and Modes of Action Through High-Content Screen Fingerprints. *Assay Drug Dev Technol.* 2015; 13:415–427. [PubMed: 26258308]
19. Futamura Y, et al. Morphobase, an encyclopedic cell morphology database, and its use for drug target identification. *Chem Biol.* 2012; 19:1620–1630. [PubMed: 23261605]
20. Sundaramurthy V, et al. Integration of chemical and RNAi multiparametric profiles identifies triggers of intracellular mycobacterial killing. *Cell Host Microbe.* 2013; 13:129–142. [PubMed: 23414754]
21. Castoreno AB, et al. Small molecules discovered in a pathway screen target the Rho pathway in cytokinesis. *Nat Chem Biol.* 2010; 6:457–463. [PubMed: 20436488]
22. Loo LH, et al. An approach for extensively profiling the molecular states of cellular subpopulations. *Nat Methods.* 2009; 6:759–765. [PubMed: 19767759]
23. Fuchs F, et al. Clustering phenotype populations by genome-wide RNAi and multiparametric imaging. *Mol Syst Biol.* 2010; 6:370. [PubMed: 20531400]
24. Collinet C, et al. Systems survey of endocytosis by multiparametric image analysis. *Nature.* 2010; 464:243–249. [PubMed: 20190736]
25. Laufer C, Fischer B, Billmann M, Huber W, Boutros M. Mapping genetic interactions in human cancer cells with RNAi and multiparametric phenotyping. *Nat Methods.* 2013; 10:427–431. [PubMed: 23563794]
26. Liberali P, Snijder B, Pelkmans L. A hierarchical map of regulatory genetic interactions in membrane trafficking. *Cell.* 2014; 157:1473–1487. [PubMed: 24906158]
27. Fischer B, et al. A map of directional genetic interactions in a metazoan cell. *Elife.* 2015; 4
28. Yin Z, et al. A screen for morphological complexity identifies regulators of switch-like transitions between discrete cell shapes. *Nat Cell Biol.* 2013; 15:860–871. [PubMed: 23748611]
29. Gustafsdottir SM, et al. Multiplex cytological profiling assay to measure diverse cellular states. *PLoS One.* 2013; 8:e80999. [PubMed: 24312513]
30. Wawer MJ, et al. Toward performance-diverse small-molecule libraries for cell-based phenotypic screening using multiplexed high-dimensional profiling. *Proc Natl Acad Sci U S A.* 2014; 111:10911–10916. [PubMed: 25024206]
31. Singh S, et al. Morphological profiles of RNAi-induced gene knockdown are highly reproducible but dominated by seed effects. *PLoS One.* 2015; 10:e0131370. [PubMed: 26197079]
32. Gibson CC, et al. Strategy for identifying repurposed drugs for the treatment of cerebral cavernous malformation. *Circulation.* 2015; 131:289–299. [PubMed: 25486933]
33. MacRae CA. A new phenotypic lexicon for accelerated translation: rise of the machines. *Circulation.* 2015; 131:234–236. [PubMed: 25601949]
34. Petrone PM, et al. Biodiversity of small molecules—a new perspective in screening set selection. *Drug Discov Today.* 2013; 18:674–680. [PubMed: 23454345]
35. Peck D, et al. A method for high-throughput gene expression signature analysis. *Genome Biol.* 2006; 7:R61. [PubMed: 16859521]
36. Rajaram S, Pavie B, Wu LF, Altschuler SJ. PhenoRipper: software for rapidly profiling microscopy images. *Nat Methods.* 2012; 9:635–637. [PubMed: 22743764]
37. Hartwell KA, et al. Niche-based screening identifies small-molecule inhibitors of leukemia stem cells. *Nat Chem Biol.* 2013; 9:840–848. [PubMed: 24161946]
38. Uhlmann V, Singh S, Carpenter AE. CP-CHARM: segmentation-free image classification made accessible. *BMC Bioinformatics.* 2016; 17:51. [PubMed: 26817459]
39. Bray, M-A.; Carpenter, A. *Assay Guidance Manual*. Sittampalam, GS., et al., editors. Eli Lilly & Company and the National Center for Advancing Translational Sciences; 2013.
40. Iversen, PW., et al. *Assay Guidance Manual*. Sittampalam, GS., et al., editors. Eli Lilly & Company and the National Center for Advancing Translational Sciences; 2012.
41. Singh S, Bray MA, Jones TR, Carpenter AE. Pipeline for illumination correction of images for high-throughput microscopy. *J Microsc.* 2014; 256:231–236. [PubMed: 25228240]
42. Bray MA, Fraser AN, Hasaka TP, Carpenter AE. Workflow and metrics for image quality control in large-scale high-content screens. *J Biomol Screen.* 2012; 17:266–274. [PubMed: 21956170]

43. Clarke R, et al. The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. *Nat Rev Cancer*. 2008; 8:37–49. [PubMed: 18097463]
44. Feng Y, Mitchison TJ, Bender A, Young DW, Tallarico JA. Multi-parameter phenotypic profiling: using cellular effects to characterize small-molecule compounds. *Nat Rev Drug Discov*. 2009; 8:567–578. [PubMed: 19568283]
45. Janzen WP, Popa-Burke IG. Advances in improving the quality and flexibility of compound management. *J Biomol Screen*. 2009; 14:444–451. [PubMed: 19483148]
46. Lundholt BK, Scudder KM, Pagliaro L. A simple technique for reducing edge effect in cell-based assays. *J Biomol Screen*. 2003; 8:566–570. [PubMed: 14567784]
47. Ljosa V, Sokolnicki KL, Carpenter AE. Annotated high-throughput microscopy image sets for validation. *Nat Methods*. 2012; 9:637. [PubMed: 22743765]
48. Guyon I, Elisseeff A. An Introduction to Variable and Feature Selection. *J Mach Learn Res*. 2003; 3:1157–1182.
49. Carpenter AE, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol*. 2006; 7:R100. [PubMed: 17076895]

Editor Summary

Cell Painting is a high-content screening assay that uses multiplexed fluorescent dyes for image-based profiling of ~1500 morphological features. Image analysis with CellProfiler automatically identifies and extracts data from individual cells.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

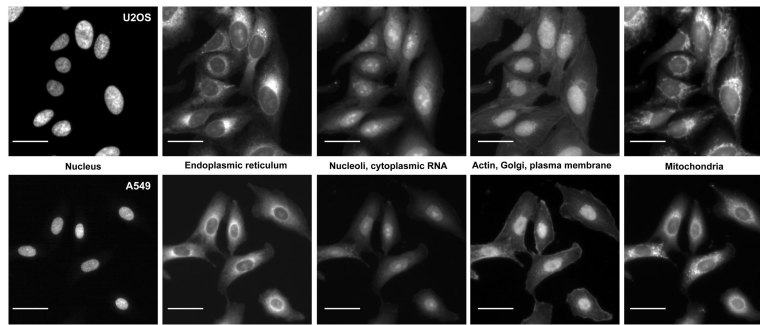


Figure 1. The Cell Painting assay, in A549 and U2OS cells

The columns display the five channels imaged in the Cell Painting assay protocol (left to right) as imaged using the ImageXpress XLS microscope: Hoechst 33342 (DNA), concanavalin A (ER), SYTO 14 (nucleoli, cytoplasmic RNA), phalloidin (actin) and WGA (Golgi, plasma membrane), and MitoTracker Deep Red (mitochondria). Scale bar is 20 μm . See Table 1 for additional details about the stains and channels imaged.

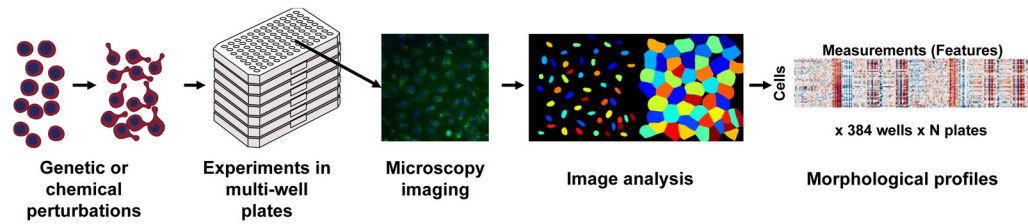


Figure 2. Overview of the strategy of morphological profiling using an image-based assay
After perturbing, staining, and imaging cells, the open-source software CellProfiler is used to extract 1,000+ morphological features of each cell. The collection of features is known as a *profile*: it reflects the phenotypic state of the cells in that sample, and can be compared to other profiles to make inferences.

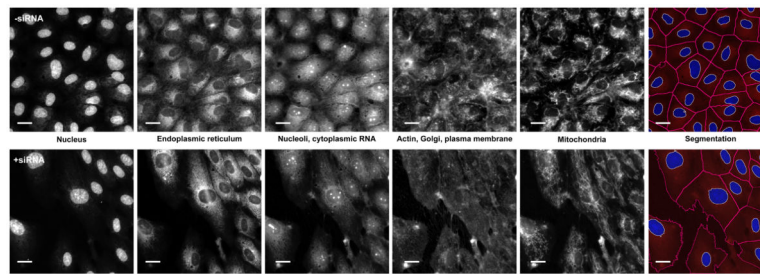


Figure 3. Sample images from a small-molecule Cell Painting experiment using HUVEC cells
 Images are shown from a well treated with transfection reagent alone (negative control, top row) and an siRNA-treated well (bottom row), as imaged on an ImageXpress XLS microscope. The first five columns display the five channels imaged in the Cell Painting assay protocol, while the last column illustrates a merged image of the DNA stain (blue) and the nucleolar/cytoplasmic RNA stain (red), with the identified (i.e., segmented) nuclei and cell body outlines resulting from image analysis overlaid in white and magenta, respectively. Scale bar is 20 μm . See Table 1 for details about the stains and channels imaged. Each panel is 6% of a full image; for each well, 9 images are acquired in a 3×3 grid, representing ~52% of the well area in total.

Table 1

Details of the ImageXpress Micro XLS channels and stains imaged in the Cell Painting assay.

Dye	Filter (Excitation)	Filter (Emission)	Organelle or cellular component	Channel name, in CellProfiler
Hoechst 33342	387/11 nm	417 – 477 nm	Nucleus	DNA
Concanavalin A/Alexa Fluor488 conjugate	472/30 nm ¹	503 – 538 nm ¹	Endoplasmic reticulum	ER
SYTO 14 green fluorescent nucleic acid stain	531/40 nm	573 – 613 nm	Nucleoli, cytoplasmic RNA ²	RNA
Phalloidin/Alexa Fluor 568 conjugate, wheat germ agglutinin (WGA)/Alexa Fluor 555 conjugate	562/40 nm	622 – 662 nm ³	F-actin cytoskeleton, Golgi, plasma membrane	AGP
MitoTracker Deep Red	628/40 nm	672 – 712 nm	Mitochondria	Mito

¹ Alternately, a FITC (482/536) filter may be used.

² The reagent SYTO 14 was selected after comparing a number of RNA-staining SYTO dyes because it appeared to have the highest nucleolar RNA affinity and the lowest cytoplasmic RNA affinity. Though unbound SYTO 14 fluoresces primarily in the green spectrum, the excitation/emission maxima for SYTO 14 bound to RNA is 521/547 nm, making the RNA channel more appropriate to use for nucleoli detection than the ER channel. Some cytoplasmic staining is still noticeable in the RNA channel and nucleolar staining is noticeable in the ER channel.

³ In Gustafsdottir et al²⁹, the TexasRed filter was incorrectly listed as excitation/emission (562/642 nm); it is actually (562/624 nm).

Table 2

Troubleshooting Table

Step	Problem	Possible reason	Solution
30	The images contain bright, slender or punctate artifacts that appear in multiple wells, across multiple channels. Too many of these artifacts can adversely affect nuclei and cell body identification and measurement.	The washing reagents are contaminated with fibers, e.g. from clothing or dust.	Filter the washing solutions and diluents before use. Prepare plates in a clean, dust-minimal environment.
52	The identified nuclei or cell bodies do not reflect the actual boundaries of the stained nuclei or cells in the image.	The settings in the <i>IdentifyPrimaryObjects</i> or <i>IdentifySecondaryObjects</i> modules (for nuclei and cell identification, respectively) were optimized for U2OS cells imaged on a particular microscope at a particular magnification, and may be inappropriate for different experimental conditions.	Cell lines with different morphological features may require additional optimization of the pipeline identification modules. After launching CellProfiler and loading the feature extraction pipeline, see the Module Notes in the main window of CellProfiler for more details on relevant settings for each module. Visual inspection is needed to confirm that the settings conform to expected results. If you encounter difficulties in adjusting the pipeline settings for this task, we recommend consulting the moderated forum at http://forum.cellprofiler.org/ for assistance.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript