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

Mark-Anthony Bray[1], Shantanu Singh[1], Han Han[2], Chadwick T Davis[2], Blake Borgeson[2], Cathy Hartland[3], Maria Kost-Alimova[3], Sigrun M Gustafsdottir[3], Christopher C Gibson[2] & Anne E Carpenter[1]

[1]Imaging Platform, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. [2]Recursion Pharmaceuticals, Salt Lake City, Utah, USA. [3]Center for the Science of Therapeutics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. Correspondence should be addressed to C.C.G. (chris.gibson@recursionpharma.com) or A.E.C. (anne@broadinstitute.org).

Published online 25 August 2016; doi:10.1038/nprot.2016.105

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, which is a morphological profiling assay that multiplexes six fluorescent dyes, imaged in five channels, to reveal eight broadly relevant cellular components or organelles. Cells are plated in multiwell plates, perturbed with the treatments to be tested, stained, fixed, and imaged on a high-throughput microscope. Next, an automated image analysis software identifies individual cells and measures ~1,500 morphological features (various measures of size, shape, texture, intensity, and so on) to produce a rich profile that is suitable for the detection of 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 2 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 processes[1–4]. 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 images[5]. However, most large-scale imaging experiments extract only one or two features of cells[6], and/or aim to identify just a few 'hits' in a screen, meaning that vast quantities of quantitative data about cellular state remain untapped.

In this article, we detail a protocol for the Cell Painting assay, which is 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 at a large scale[5,7–10]. The techniques and technology that are necessary to generate these data have advanced rapidly, and they are now becoming accessible to nonspecialized laboratories[11]. 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, tricks, and scripts (see **Supplementary Method**) 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. Substantial changes in a subset of profiled features can thus serve as a 'fingerprint' that characterizes the sample condition. Some of the earliest instances of profiling involved the NCI-60 tumor cell line panel—in which it was discovered that patterns of anticancer drug sensitivity reflect mechanisms of action[12]—and gene expression—in which signatures related to small molecules, genes, and diseases were identified[13].

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 that is 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, allowing detection of perturbations even in subsets of cells. Morphological profiling has been successfully 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 heterogeneity[22]. Genes have been analyzed by creating profiles of cell populations in which the gene is perturbed by RNA interference (RNAi); these profiles, in turn, have been used to cluster genes[23,24], identify genetic interactions[25–27], or characterize cellular heterogeneity[28].

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 assay[29] (**Fig. 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 (**Fig. 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 used 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 researchers. We refer interested readers to Gustafsdottir *et al.*[29], Ljosa *et al.*[17], and Wawer *et al.*[30] 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 effects[29]. 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). In addition, 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).

Similarly, 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 effects[31] (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 comparison of 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 coauthors from Recursion Pharmaceuticals have implemented this approach in a systematic manner 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)

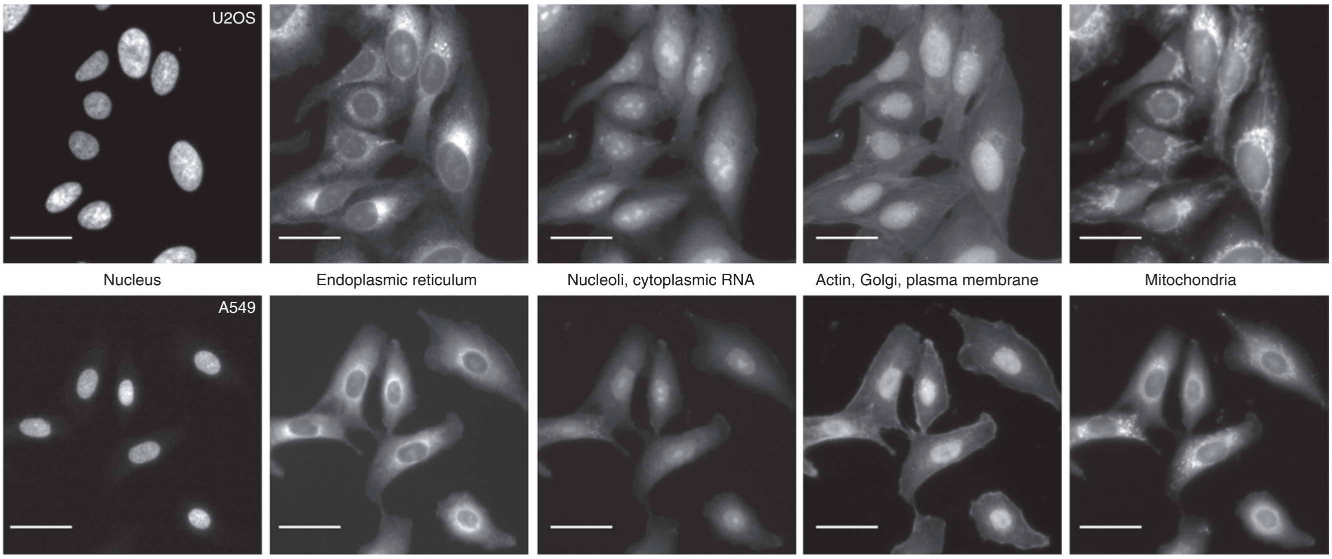

Figure 1 | The Cell Painting assay in U2OS and A549 cells. The columns display the five channels imaged in the Cell Painting assay protocol (left to right) as imaged using the ImageXpress XLS microscope: Hoechst 33342 (DNA), concanavalin A (endoplasmic reticulum), SYTO 14 (nucleoli and cytoplasmic RNA), phalloidin (actin) and WGA (Golgi and plasma membrane), and MitoTracker Deep Red (mitochondria). Scale bars, 20 μm. See **Table 1** for additional details about the stains and channels imaged.

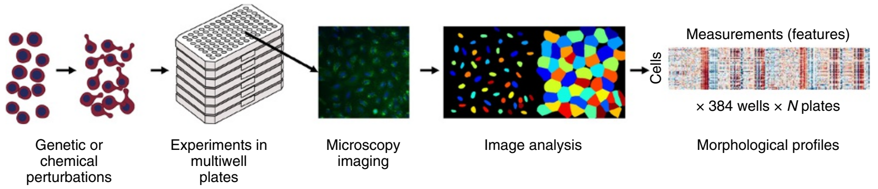

Figure 2 | Overview of the strategy of morphological profiling using an image-based assay. After perturbing, staining and imaging cells, the open-source software CellProfiler is used to extract ~1,500 morphological features of each cell. The collection of features is known as a *profile*: it reflects the phenotypic state of the cells in that sample, and it can be compared with other profiles to make inferences.

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].

Last, both academic and pharmaceutical screening groups have an increasing desire to improve the efficiency of screening, particularly when assays are complex and expensive[34]. 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 goal is to maximize profile diversity (and thus likelihood of diverse phenotypic effects) while simultaneously eliminating compounds that do not produce any measurable effect 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 profiles[34].

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/ (ref. 35)) is currently the only practical alternative to image-based morphological profiling in terms of throughput and efficiency[4]. Both profiling approaches (morphological profiling by Cell Painting and gene expression profiling by L1000) yield ~1,000 raw features for each sample, and both are likely to capture a broad range of cellular states after 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, which potentially improves the ability to resolve changes in subpopulations of cells. Cell Painting also costs substantially less 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 enrichment[30]. The study also indicated that these 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 that the two modalities capture distinct information about cell state, and thus they are probably quite complementary. No studies have yet been published combining the two orthogonal modalities into a single profile; we believe that 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 been 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]. Nevertheless, 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 with a more targeted one; the same basic principles of morphological profiling will still apply. However, exchanging stains may entail substantial optimization effort; for example, replacing the concanavalin A/Alexa Fluor 488 label requires a fluorophore that is bright enough to mask the fluorescence of SYTO 14 in the 472/30 nm channel (**Table 1**, footnote a). 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 cells 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 coculture systems. Although we have used U2OS and A549 cells most commonly, the staining protocol has worked well in our experience for MCF-7, 3T3, HTB-9, HeLa, HepG2, HEKTE, SH-SY5Y, HUVEC (**Fig. 3**), HMVEC, primary human fibroblasts, and primary human hepatocyte/3T3-J2 fibroblast cocultures.

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, and so on) 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 substantial 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 additional sources of complementary data are available.

TABLE 1 | Details of the ImageXpress Micro XLS channels and stains imaged in the Cell Painting assay.

Dye	Filter (excitation; nm)	Filter (emission; nm)	Organelle or cellular component	CellProfiler channel name
Hoechst 33342	387/11	417–477	Nucleus	DNA
Concanavalin A/Alexa Fluor 488 conjugate	472/30[a]	503–538[a]	Endoplasmic reticulum	ER
SYTO 14 green fluorescent nucleic acid stain	531/40	573–613	Nucleoli, cytoplasmic RNA[b]	RNA
Phalloidin/Alexa Fluor 568 conjugate, wheat-germ agglutinin/Alexa Fluor 555 conjugate	562/40	622–662[c]	F-actin cytoskeleton, Golgi, plasma membrane	AGP
MitoTracker Deep Red	628/40	672–712	Mitochondria	Mito

ER, endoplasmic reticulum.
[a]Alternatively, an FITC (482/536) filter may be used. [b]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 are 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.
[c]In Gustafsdottir *et al.*[29], the TexasRed filter was incorrectly listed as having excitation/emission at 562/642 nm; the actual numbers are 562/624 nm.

Generally, no adjustment of the staining protocol has been needed from cell type to cell type; the only change in the 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 time points' 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. However, it should be noted 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 fluorescence imaging and analysis. Specifically, the cells should typically be adherent and grow reasonably flat (i.e., nonspheroid), without substantial clumping, under the culture conditions used. The cell types we have tested that fail to meet this criterion are SW480 and DLD-1; presumably, nonadherent 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 reduced for cell types with a rounded morphology as compared with 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 multiwell plate), and the number of controls and replicates, the predominant concerns are 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].

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, and so on). 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 in which 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 substantial loss of data quality with fewer replicates. In practice, we have used five or more replicates to buffer against accidental sample loss. For larger-scale experiments (>1,000 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 per plate designated for negative controls works well for a 384-well format assay. In cases in which 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 toward 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 that are not present in a cell line yield different profiles[31].

A perennial concern with assay development is that any technical sources of variation can have an impact on all the wells and/or plates such that any biological signals are overwhelmed by systematic noise introduced by sample preparation. We advise mitigating this issue by spreading the positions 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, and so on. 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 nonsymmetric 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 A_1–A_4 comprise all replicates of plate A, B_1–B_4 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 [A_1,B_1,C_1,D_1], [E_1,A_2,B_2,C_2], [D_2,E_2,A_3,B_3], [C_3,D_3,E_3,A_4], [B_4,C_4,D_4,E_4], and [A_5,B_5,C_5,D_5].

Perturbations and/or time points. Before carrying out a large-scale experiment, consider initially performing a smaller pilot with a small number of perturbations and/or time points. During assay development, we recommend culturing the cells until the desired time point, and then examining them for the degree of confluency. This assessment may be performed by fluorescence staining (i.e., following some or all of the regular protocol) or even by eye under bright field; the latter requires 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 substantial effect on cellular growth and confluence. We have used both 24- and 48-h (refs. 29,30) small-molecule exposures before fixation and imaging; in our limited testing, the latter yields a higher proportion of small molecules with morphological profiles distinguishable from negative controls. For infection/transfection of RNAi and overexpression plasmids, we use 96- (ref. 31) and 72-h 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 have a substantial 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 phenotypic variations may be subtle, a low cell count will lead to a small sample size that is not truly representative of the phenotype. In cases in which 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 will be 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 infection[31], and transient transfection, as well as 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. Although a number of bioimaging software packages (free and commercial) exist for morphological feature extraction[11], 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: (i) illumination correction, (ii) quality control (QC), and (iii) 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 2D 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 applications[41].

QC 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 QC[42].

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 that fail the previously determined QC criteria, and, for each cell, identifies the nucleus, cell body, and cytoplasm, and makes measurements of morphology, intensity, texture, and adjacency for each cellular subcompartment. 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 subcompartments 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 nonautomated microscope. High-content screening facilities will generally be well equipped to aid laboratories in conducting the Cell Painting assay at a 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 >~1,000 images. In such cases, we recommend using a computing cluster or cloud computing resource, which will probably 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 that, at the least, has access to collaborators with some experience with high-throughput automated screening and some advanced computational skills, as well as 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, and so on). 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, alternative stains may be chosen to highlight the relevant cellular subcompartments while maintaining broad coverage of other organelles (see 'Comparison with other methods' above for more details).

For RNAi experiments in mammalian cells, the magnitude and prevalence of off-target effects (due largely to the RNAi seed-based mechanism) cause the morphological profiles of RNAi reagents targeting the same gene to look more dissimilar than those targeting different genes[31]. This phenomenon has been observed in other multiparametric assays, and it is not specific to morphological profiling. Unfortunately, this effect impedes large-scale experiments using short RNAi reagents in which the experimental design requires widespread comparisons across all samples. However, we note that it does not preclude experiments in which 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. Alternative gene suppression techniques, e.g., using CRISPR (clustered regularly interspaced short palindromic repeats) elements and Cas (CRISPR-associated) proteins, have not yet been extensively evaluated in conjunction with morphological profiling but are likely to be effective.

Finally, there are several computational challenges that are associated with this assay. First, there are statistical challenges associated with analyzing the high-dimensional feature space that results from Cell Painting. Similar to the case with gene expression data[43], 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 biology[44]. Second, although in principle single-cell data are preferred over aggregated data, the former require substantially more computational storage and processing resources; thus far, no routine analytical protocol has been established for this issue. Last, 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

- Cell line of interest—e.g., we have previously used U2OS cells (American Type Culture Collection (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 that they are not infected with *Mycoplasma*.
- DMEM (Fisher Scientific, cat. no. MT-10-017-CV)
- FBS (Life Technologies, cat. no. 10437-028)
- TrypLE Express Enzyme (Life Technologies, cat. no. 12605-036)
- PBS, pH 7.4 (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)
- 16% (wt/vol) Paraformaldehyde (PFA), 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.
- HBSS (10×; Invitrogen, cat. no. 14065-056)
- Triton X-100 (Sigma-Aldrich, 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.
- Concanavalin 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.

  ▲ CRITICAL We have performed the Cell Painting assay using the specific catalog numbers listed here. If you are planning on changing to a different product or vendor for a given reagent, reoptimization of that reagent for the protocol may be necessary.

EQUIPMENT

- Microplates: Corning 384-well black/clear flat bottom, fibronectin coated (Corning, cat. no. 4585) or Corning 384-well black/clear flat bottom, tissue culture (TC)-treated, barcoded (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
- 0.22-μm Filter
- 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)
- Cytomat 5C tissue culture incubator at 37 °C, 5% CO_2 (Thermo Fisher Scientific, cat. no. 50128822)
- CyBi-Well 96/384-channel simultaneous pipettor (CyBio, cat. no. 3391 3 4112)
- Automated liquid handler: Multidrop Combi Reagent Dispenser (Thermo Fisher 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 epifluorescence microscope (Molecular Devices)
- CRS CataLyst Express robot microplate handler system (Thermo Fisher 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, QC, 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 an RNAi Cell Painting knockdown experiment applied to U2OS cells[31]. These data are 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_acqusition_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 data set, 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 comma-delimited file (CSV) of per-well profiles generated by running the profiling script, available at https://raw.githubusercontent.com/carpenterlab/2016_bray_natprot/master/supplementary_files/profiles.csv.zip

REAGENT SETUP

MitoTracker Deep Red stock The product from Invitrogen (cat. no. M22426) contains 50 μg in each vial. Add 91 μl of DMSO to one vial to make a 1 mM solution. Store the solution at −20 °C, protected from light, and use it within 1 month.

Wheat-germ agglutinin/Alexa Fluor 555 conjugate stock The product from Invitrogen (cat. no. W32464) contains 5 mg in each vial. Add 5 ml of dH_2O to each to make a 1 mg/ml solution. Store the solution at −20 °C, protected from light, and use it within 1 month. We recommend centrifuging the wheat-germ agglutinin (WGA) conjugate solution at 10,000*g* for 30 s at RT before use, in order to remove any protein aggregates in the 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 of 0.1 M sodium bicarbonate to each vial to make a 5 mg/ml solution. Store the solution at −20 °C, protected from light, and use it within 1 month.

Phalloidin/Alexa Fluor 568 conjugate stock The product from Invitrogen (cat. no. A12380) contains 300 units in each vial. Add 1.5 ml of methanol to each vial. Store the solution at −20 °C, protected from light, and use it within 1 year.

SYTO 14 green fluorescent nucleic acid stain stock The product from Invitrogen (cat. no. S7576) is a 5 mM solution in DMSO. Store the solution at −20 °C, protected from light, and use it within 1 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 it within 6 months.

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

BSA solution in HBSS Weigh 1 g of BSA and dissolve it in 100 ml of HBSS to make a 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 of Triton X-100 to 100 ml of HBSS to make a 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 of MitoTracker Deep Red stock solution to 50 ml of prewarmed cell culture medium 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 μl/ml phalloidin, 100 μg/ml concanavalin A, 5 μg/ml Hoechst 33342, 1.5 μg/ml WGA, and 3 μM SYTO 14 green fluorescent nucleic acid stain solution in 1× HBSS and 1% (wt/vol) BSA. To make 50 ml of stain solution, add 250 μl of phalloidin stock solution, 1 ml of concanavalin A stock solution, 25 μl of Hoechst stock solution, 75 μl of WGA stock solution, and 30 μl of SYTO 14 green fluorescent nucleic acid stain stock solution to 48.7 ml of 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 that of the cell culture medium. Seal and store it at −20 °C for long-term storage or at room temperature (RT; 21 °C) for up to 6 months; other common compound management solutions may also be used[45].

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 they 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 with wide-field microscopes, confocal microscopes possess a limited number of excitation wavelengths available for use, have typically higher purchase prices, which may be prohibitive for smaller research groups, and are traditionally of lower throughput.

We have used an ImageXpress Micro XLS epifluorescence 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 (PerkinElmer)— which is capable of imaging in both wide-field and confocal modes—for Cell Painting. However, as 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 in the wide-field mode in order to use the same filters as the ImageXpress XLS. A comparison of the excitation and emission wavelengths used for Cell Painting with the ImageXpress XLS and Phenix microscopes is provided in the **Supplementary Note**.

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 marked, then postacquisition normalization will probably be effective (see 'Normalization of morphological features across plates' in the PROCEDURE section 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. ▲ CRITICAL The same microscope should be used for imaging all microtiter plates during an experiment. We do not recommend switching microscopes midstream because lamp intensities, filter patterns, and other subtleties can be quite different even between supposedly identical microscope setups.

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 before 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 fewer 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. If time permits, more sites can be imaged in order to increase well coverage and to improve sample statistics; it is best to capture as many cells as possible.

The order in which 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. As 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 you are using an ImageXpress Micro XLS microscope, use laser-based focusing with image recovery for autofocusing. The autofocusing can be applied for only the first site of each well, or alternatively for all sites; the latter is a minimal increase in time and is recommended for those using glass-bottom plates to decrease focusing 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 it 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.

Do not use shading correction, as the background illumination heterogeneities will be corrected post acquisition using the CellProfiler software.

Before beginning the complete imaging run, it is useful to capture images from three to five 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. ▲ CRITICAL Avoid capturing the edges of the well in the images, particularly if a large number of sites per well are imaged. Although 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. ▲ 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 because of a perturbation.

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 QC 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/. In addition, 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, QC, 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 about 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 on the assumption that the image files follow the nomenclature of the ImageXpress microscope system, in which the plate/well/site metadata are 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 function for this module for more details.

The QC 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 you are using a smaller number of images, the pipelines can be adjusted to output the measured features to a CSV file 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., >~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 substantial setup effort, and it 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 in the PROCEDURE, as well as by the scripts listed in the Equipment section 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 QC[42].

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 has been validated for many cell types; each step may need adjustment depending on local conditions or alternative 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 has been validated for use on A549 cells in Corning 384-well 200-nm-thick glass-bottom plates.

2| Grow cells to near confluence (~80%) in a T-150 culture vessel.

3| (Optional) If you are performing experiments that involve the addition of compounds (Step 8A), prepare the compound library according to the instructions in the Reagent Setup. If you are performing experiments that involve siRNA transfection (Step 8B), prepare the siRNA transfection reagent mixture according to **Box 1**.

4| Rinse the cells with PBS without Ca^{2+} or Mg^{2+}.

5| Add 6 ml of TrypLE Express and incubate the cells at RT for 30 s. Remove the TrypLE Express and add 1 ml of fresh TrypLE Express. Incubate the mixture at 37 °C until the cells have detached. This should occur within 3–5 min.

6| Add 10 ml of growth medium to deactivate the trypsin, and determine the live-cell concentration using standard methods (hemocytometer or cell counter).

7| Dilute A549 cells to 50,000 live cells per ml in media, and dispense 40 µl (2,000 live A549 cells) into 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 per well.

▲ CRITICAL STEP Adequately resuspend the cell mixture to ensure a homogeneous cell suspension before each dispensation. It is not uncommon for cells to rapidly settle in their reservoir, resulting in plate-to-plate variation in cell numbers. If you are using a liquid handler with a multidispense function, be sure to adequately prime the dispensing cassette and/or dispense at least 10 µl of cell suspension back into the reservoir before dispensing the cells into culture plates; the latter is helpful if cells or reagents are sticking to the tubing.

▲ CRITICAL STEP When you are 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.

Treatment with a small-molecule library or siRNA transfection

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

(A) Addition of a small-molecule library ● TIMING variable; ~2–3 d for one batch experiment of 384-well plates

(i) Allow plates to sit on a flat, level surface at RT for 1–2 h to reduce plate edge effects[46].

Box 1 | Preparation of the siRNA transfection reaction mixture

The following procedure 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 µM 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 µl of diluted siRNA with 10 µl of diluted Lipofectamine RNAiMAX.
4. Ensure that the siRNA-Lipofectamine mix has incubated at RT for at least 10 min before proceeding to Step 8 of the main procedure.

(ii) Put the plates into the incubator (37 °C, 5% CO_2, 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 μl of 2% (vol/vol) FBS in DMEM 24 h after seeding. Perform the aspiration steps using a plate washer such as the BioTek ELx405 microplate washer or equivalent. Reducing the 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 h before 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 be mixed well in the culture medium before adding them to the cells.

(B) siRNA transfection ● TIMING ~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 a volume of 10.2 μl per well.

(ii) Allow the plates to sit on a flat, level surface at RT for 1–2 h to reduce plate edge effects[46].

(iii) Put the plates into the incubator (37 °C, 5% CO_2, 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 μl of 2% (vol/vol) FBS in DMEM 24 h after seeding. Perform the aspiration steps using a plate washer such as the BioTek ELx405 microplate washer or equivalent. Reducing the 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.

(v) Incubate the cells for 2-3 more days (depending on the gene and the cell type) before fixation and staining.

(vi) (Optional) Add starvation medium (0.1% (vol/vol) FBS in DMEM) to the cells ~24 h before 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 the 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 you are 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 **Figs. 1** and **3** of this protocol, and Fig. S1 of Gustafsdottir *et al.*[29]).

9| Prepare the live-cell MitoTracker staining solution for all plates.

10| Remove media from plates, and set the aspiration height in the plate washer to leave 10 μl of residual volume to minimize the disturbance to the live cells from the pins and media turbulence.

11| Add 30 μl of MitoTracker staining solution.

12| Centrifuge the plate (500*g* at RT for 1 min) after adding stain solutions, and ensure that there are no bubbles in the bottoms 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 expose the working stain solution to light during storage or over long periods, in order to maintain fluorescence.

15| To fix the cells, add 10 μl of 16% (wt/vol) methanol-free PFA 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 experience, 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 RT for 1 min) after adding stain solutions, and ensure that there are no bubbles in the bottoms of the wells.

17| Incubate the plates in the dark at RT for 20 min.

18| Wash the plates once with 70 μl of 1× HBSS.

19| To permeabilize the cells, remove the HBSS and add 30 μl of 0.1% (vol/vol) Triton X-100 solution to the wells.

20| Centrifuge the plate (500*g* at RT for 1 min) after adding stain solutions, and ensure that there are no bubbles in the bottoms of the wells.

21| Incubate the plates in the dark at RT for 10–20 min.
▲ **CRITICAL STEP** Once the MitoTracker solution is added, take special care to keep the cells in the dark for the rest of the experiment.

22| Wash the wells twice with 70 μl of 1× HBSS.

23| Remove the HBSS and add 30 μl of the phalloidin, concanavalin A, Hoechst, WGA, and SYTO 14 staining solution to each well.

24| Centrifuge the plate (500*g* at RT for 1 min) after adding stain solution, and ensure that there are no bubbles in the bottoms of the wells.

25| Incubate the plates in the dark at RT for 30 min.

26| Wash the cells three times with 70 μl of 1× HBSS, with no final aspiration.

27| Seal the plates with adhesive foil and store them at 4 °C in the dark.

Automated image acquisition ● TIMING variable; 1–3.5 h per 384-well plate

28| Place the microtiter plates onto 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 the 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 then selecting *illumination.cppipe*.
▲ **CRITICAL STEP** Nonhomogeneous 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. Nonhomogeneous 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'), although common, is not suitable for high-throughput screening. A retrospective method to correct all acquired images on a per-channel, per-plate basis is therefore recommended[41]; 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' in which 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 choosing this option, carry out the instructions in **Box 2**, describing modifications to the pipeline to run it as a batch process.

37| QC to identify and exclude aberrant images: start CellProfiler, if you are not already running it.

38| Load the QC 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 QC measures. The approach detailed here uses CellProfiler to analyze the data using QC metrics that do not require cell identification[42]. 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 that for 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' should be 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

Box 2 | Configuration of the 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 >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 substantial 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” 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.

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 a smaller number of images (e.g., <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 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' 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' in which the QC data will be saved.

42| 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.

43| Press the 'Analyze Images' button at the bottom-left of the interface. A progress bar in the bottom-left of the interface will indicate the estimated time of completion.
▲ **CRITICAL STEP** This step assumes that you will be running the QC 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 choosing this option, carry out instructions in **Box 2**, which describes 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 al.*[42] 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 you are not already running it.

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 that for 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 the 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. Select the desired measurement for the Measurement dropdown box, and the desired image for the Image dropdown box. Select whether the image is flagged based on a high or low threshold value, and then select the actual threshold value for the measurement. You can also add more flags by pressing the 'Add another flag' button, giving the flag a name and specifying 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 be the same as that used for the QC 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 a smaller number of images (e.g., <1,000) or instances when a MySQL database is not available, the *ExportToSpreadsheet* module may be used to export the collected measurements to a 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; 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; 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' in which the analysis data will be saved.

51| 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.

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 the 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 that 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), then it will use the nuclei to guide identification of the cell boundaries using the SYTO 14–stained image ('RNA' in CellProfiler), and then it will 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 it 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 recommended only for experiments with <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 choosing this option, carry out the instructions in **Box 2**, which describes modifications to the pipeline to run it as a batch process.

Normalization of 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 (MAD) 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 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.

Creation of 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 per site within the well. We have published a comparison of several such methods for creating morphological profiles[47]. Here, we describe the population-averaging profile approach, 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 component 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 papers[29], the dimensionality of the data vectors was reduced from 1,301 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 (support vector machine) recursive feature elimination[15]. We refer the reader to a review of feature-selection methods[48] to evaluate the advantages and disadvantages of these approaches.

59| See the **Supplementary Method** 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 clustering[29], to identify compound sets with high rates of activity and diverse biological performance in combination with high-throughput gene-expression profiles[30], and to determine the dominance of seed sequence–driven off-target effects in RNAi-induced gene knockdown studies[31]. 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's correlation, Spearman's correlation, Euclidean distance, and cosine distance. For QC 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.

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 2**.

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 *IdentifyPrimaryObjects* or *IdentifySecondaryObjects* 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

● TIMING

Steps 1–7, cell culture: it typically takes ~2–3 d for the cells to reach appropriate confluency, depending on cell type and growth conditions. Harvesting the cells (Steps 4–6) takes 30 min, and seeding the cells (Step 7) takes 20 min. After seeding, the cells should be cultured for 2–5 d before staining.

Step 8(A), addition of a small-molecule library (optional): variable; approximately 2–3 d for one batch experiment of 384-well plates. Addition of a compound library takes ~3 h for one batch experiment of 384-well plates, including reagent preparation and media change, and 1–2 d for compound incubation.

Step 8(B), siRNA transfection (optional): ~5 d for one batch experiment of 384-well plates. Transfection of siRNA takes 3 h for one batch experiment of 384-well plates, including reagent preparation and media change, and 2 d for siRNAs to achieve appropriate knockdown.

Steps 9–27, staining and fixation: ~2.5–3 h including reagent preparation for one batch experiment of 384-well plates. The total timing will vary depending on 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 including no more than 4–5 plates per batch because of the increased likelihood of sample preparation error by the researcher.

Steps 28–30, automated image acquisition): variable; ~3.5 h per 384-well plate, for nine fields of view per well and typical exposure times (and as little as 1 h per plate for smaller numbers of fields of view). The total time may vary depending on the number of sites imaged per plate and the exposure time for each channel.

Steps 31–53, morphological image feature extraction from microscopy data): variable; 20 h per batch of 384-well plates. It takes ~10 min per plate for CellProfiler to scan the inputs folder(s) after manually dragging/dropping the needed images into the CellProfiler interface. The pipeline execution time will depend on the computing setup; run times on a single computing node of 20 s (illumination correction), 30 s (QC), and 10 min (analysis) per field of view are typical. A substantial time savings can be achieved if you run the feature extraction and QC pipelines on a distributed computing cluster, which massively parallelizes the processing as compared with running it on a single local computer. Performing the QC workflow using CellProfiler Analyst takes ~4 h of hands-on time, although this time can be substantially shortened if cutoffs are reused from experiment to experiment.

Steps 54–56, normalization of morphological features across plates: <5 min of processing time per 384-well plate.

Steps 57–59, creation of per-well profiles: up to an hour of processing time per 384-well plate, depending on data access methods.

Step 60, data analysis: variable; ~1 h for basic analysis of replicate quality and signature strength. Time for additional analysis varies substantially depending on the problem at hand.

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 × 1,080 pixels (0.656 μm/pixel) and ~2.3 MB in size. The total number of images generated equals (number of samples tested) × (number of sites imaged per well) × (five 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 per 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 (**Fig. 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 QC pipeline will produce a set of numerical measurements extracted at the image level, and will export them to a MySQL database. These measurements can optionally be used for removing images that are unacceptable for further processing because of 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 in 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 QC and feature extraction pipelines typically total ~6.4 GB per 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 body[49]. We have found that the SYTO 14 image is the most amenable for finding cell

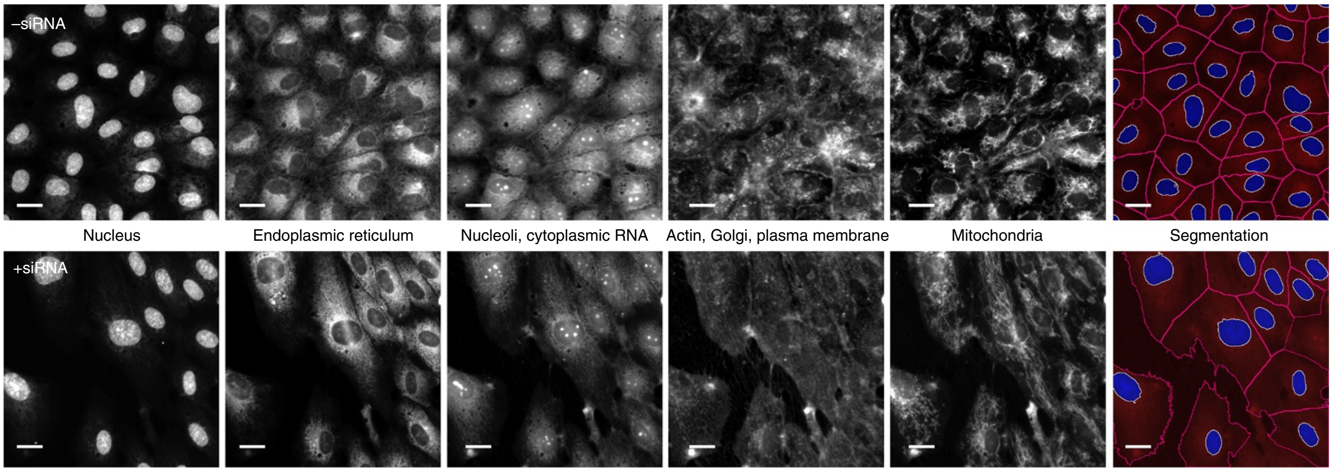

Figure 3 | Sample images from a small-molecule Cell Painting experiment using HUVEC cells. Images are shown from a well treated with transfection reagent alone (negative control, top row) and an siRNA-treated well (bottom row), as imaged on an ImageXpress XLS microscope. The first five columns display the five channels imaged in the Cell Painting assay protocol, whereas 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 bars, 20 μm. See **Table 1** for details about the stains and channels imaged. Each panel is 6% of a full image; for each well, nine images are acquired in a 3 × 3 grid, representing ~52% of the well area in total.

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 the TROUBLESHOOTING table (**Table 2**) for Step 30 of the 'Automated image acquisition' section).

After running the profiling scripts to normalize the image features across plates and to create the per-well profiles, the output will be a morphological profile file in CSV format. Each row of this file represents a 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.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS Research reported in this publication was supported in part by NIH R44 TR001197 (C.C.G.), NSF RIG DBI 1119830 (M.-A.B.), NIH R01 GM089652 (A.E.C.), and NSF CAREER DBI 1148823 (A.E.C.). The RNAi Cell Painting knockdown experiment used in this paper was previously published[31] 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 assay[29]; these authors are S.M. Gustafsdottir, V. Ljosa, K.L. Sokolnicki, J.A. Wilson, D. Walpita, M.M. Kemp, K. Petri Seiler, H.A. Carrel, T.R. Golub, S.L. Schreiber, P.A. Clemons, A.E. Carpenter, and A.F. Shamji. For enabling this work, Recursion Pharmaceuticals thanks the University of Utah Core Facilities (J. Phillips), specifically the Drug Discovery Core (B. Luo) and the Fluorescent Imaging Core (C. Rodesch). We thank members of the Broad Institute's Center for the Development of Therapeutics, especially T.P. Hasaka, for technical assistance. We also thank A. Kozol for proofreading the Equipment and Reagents sections and for testing the image analysis workflow, and A. Berger and X. Wu for offering helpful comments and suggestions during manuscript preparation.

AUTHOR CONTRIBUTIONS All authors contributed to writing this protocol. C.T.D., H.H., and S.M.G. contributed to describing the benchwork aspects of the protocol. C.H., M.K.-A., M.-A.B., and C.T.D. contributed to updates to the experimental design. M.-A.B. and S.S. contributed most heavily to describing the computational aspects of the protocol. A.E.C., C.C.G., B.B., and S.S. contributed most heavily to describing the rationale and the experimental design.

COMPETING FINANCIAL INTERESTS The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Swinney, D.C. & Anthony, J. How were new medicines discovered? *Nat. Rev. Drug Discov.* **10**, 507–519 (2011).
2. Swinney, D.C. The contribution of mechanistic understanding to phenotypic screening for first-in-class medicines. *J. Biomol. Screen.* **18**, 1186–1192 (2013).
3. Moffat, J.G., Joachim, R. & David, B. Phenotypic screening in cancer drug discovery — past, present and future. *Nat. Rev. Drug Discov.* **13**, 588–602 (2014).
4. Johannessen, C.M., Clemons, P.A. & Wagner, B.K. Integrating phenotypic small-molecule profiling and human genetics: the next phase in drug discovery. *Trends Genet.* **31**, 16–23 (2015).
5. Bickle, M. The beautiful cell: high-content screening in drug discovery. *Anal. Bioanal. Chem.* **398**, 219–226 (2010).
6. Singh, S., Carpenter, A.E. & Genovesio, A. Increasing the content of high-content screening: an overview. *J. Biomol. Screen.* **19**, 640–650 (2014).
7. Perlman, Z.E. *et al.* Multidimensional drug profiling by automated microscopy. *Science* **306**, 1194–1198 (2004).
8. Danuser, G. Computer vision in cell biology. *Cell* **147**, 973–978 (2011).
9. Altschuler, S.J. & Wu, L.F. Cellular heterogeneity: do differences make a difference? *Cell* **141**, 559–563 (2010).
10. Snijder, B. & Pelkmans, L. Origins of regulated cell-to-cell variability. *Nat. Rev. Mol. Cell Biol.* **12**, 119–125 (2011).
11. Eliceiri, K.W. *et al.* Biological imaging software tools. *Nat. Methods* **9**, 697–710 (2012).
12. Paull, K.D. *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.* **81**, 1088–1092 (1989).
13. Lamb, J. *et al.* The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).
14. Adams, C.L. *et al.* Compound classification using image-based cellular phenotypes. *Methods Enzymol.* **414**, 440–468 (2006).
15. Loo, L.-H., Wu, L.F. & Altschuler, S.J. Image-based multivariate profiling of drug responses from single cells. *Nat. Methods* **4**, 445–453 (2007).

16. Young, D.W. *et al.* Integrating high-content screening and ligand-target prediction to identify mechanism of action. *Nat. Chem. Biol.* **4**, 59–68 (2008).
17. Ljosa, V. *et al.* Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment. *J. Biomol. Screen.* **18**, 1321–1329 (2013).
18. Reisen, F. *et al.* Linking phenotypes and modes of action through high-content screen fingerprints. *Assay Drug Dev. Technol.* **13**, 415–427 (2015).
19. Futamura, Y. *et al.* Morphobase, an encyclopedic cell morphology database, and its use for drug target identification. *Chem. Biol.* **19**, 1620–1630 (2012).
20. Sundaramurthy, V. *et al.* Integration of chemical and RNAi multiparametric profiles identifies triggers of intracellular mycobacterial killing. *Cell Host Microbe* **13**, 129–142 (2013).
21. Castoreno, A.B. *et al.* Small molecules discovered in a pathway screen target the Rho pathway in cytokinesis. *Nat. Chem. Biol.* **6**, 457–463 (2010).
22. Loo, L.-H. *et al.* An approach for extensibly profiling the molecular states of cellular subpopulations. *Nat. Methods* **6**, 759–765 (2009).
23. Fuchs, F. *et al.* Clustering phenotype populations by genome-wide RNAi and multiparametric imaging. *Mol. Syst. Biol.* **6**, 370 (2010).
24. Collinet, C. *et al.* Systems survey of endocytosis by multiparametric image analysis. *Nature* **464**, 243–249 (2010).
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* **10**, 427–431 (2013).
26. Liberali, P., Snijder, B. & Pelkmans, L. A hierarchical map of regulatory genetic interactions in membrane trafficking. *Cell* **157**, 1473–1487 (2014).
27. Fischer, B. *et al.* A map of directional genetic interactions in a metazoan cell. *Elife* **4**, e05464 (2015).
28. Yin, Z. *et al.* A screen for morphological complexity identifies regulators of switch-like transitions between discrete cell shapes. *Nat. Cell Biol.* **15**, 860–871 (2013).
29. Gustafsdottir, S.M. *et al.* Multiplex cytological profiling assay to measure diverse cellular states. *PLoS One* **8**, e80999 (2013).
30. Wawer, M.J. *et al.* Toward performance-diverse small-molecule libraries for cell-based phenotypic screening using multiplexed high-dimensional profiling. *Proc. Natl. Acad. Sci. USA* **111**, 10911–10916 (2014).
31. Singh, S. *et al.* Morphological profiles of RNAi-induced gene knockdown are highly reproducible but dominated by seed effects. *PLoS One* **10**, e0131370 (2015).
32. Gibson, C.C. *et al.* Strategy for identifying repurposed drugs for the treatment of cerebral cavernous malformation. *Circulation* **131**, 289–299 (2015).
33. MacRae, C.A. A new phenotypic lexicon for accelerated translation: rise of the machines. *Circulation* **131**, 234–236 (2015).
34. Petrone, P.M. *et al.* Biodiversity of small molecules--a new perspective in screening set selection. *Drug Discov. Today* **18**, 674–680 (2013).
35. Peck, D. *et al.* A method for high-throughput gene expression signature analysis. *Genome Biol.* **7**, R61 (2006).
36. Rajaram, S., Pavie, B., Wu, L.F. & Altschuler, S.J. PhenoRipper: software for rapidly profiling microscopy images. *Nat. Methods* **9**, 635–637 (2012).
37. Hartwell, K.A. *et al.* Niche-based screening identifies small-molecule inhibitors of leukemia stem cells. *Nat. Chem. Biol.* **9**, 840–848 (2013).
38. Uhlmann, V., Singh, S. & Carpenter, A.E. CP-CHARM: segmentation-free image classification made accessible. *BMC Bioinformatics* **17**, 51 (2016).
39. Bray, M.-A. & Carpenter, A. in *Assay Guidance Manual* (eds. Sittampalam, G.S. *et al.*) (Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2013).
40. Iversen, P.W. *et al.* in *Assay Guidance Manual* (eds. Sittampalam, G.S. *et al.*) (Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2012).
41. Singh, S., Bray, M.-A., Jones, T.R. & Carpenter, A.E. Pipeline for illumination correction of images for high-throughput microscopy. *J. Microsc.* **256**, 231–236 (2014).
42. Bray, M.-A., Fraser, A.N., Hasaka, T.P. & Carpenter, A.E. Workflow and metrics for image quality control in large-scale high-content screens. *J. Biomol. Screen.* **17**, 266–274 (2012).
43. Clarke, R. *et al.* The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. *Nat. Rev. Cancer* **8**, 37–49 (2008).
44. Feng, Y., Mitchison, T.J., Bender, A., Young, D.W. & Tallarico, J.A. Multi-parameter phenotypic profiling: using cellular effects to characterize small-molecule compounds. *Nat. Rev. Drug Discov.* **8**, 567–578 (2009).
45. Janzen, W.P. & Popa-Burke, I.G. Advances in improving the quality and flexibility of compound management. *J. Biomol. Screen.* **14**, 444–451 (2009).
46. Lundholt, B.K., Scudder, K.M. & Pagliaro, L. A simple technique for reducing edge effect in cell-based assays. *J. Biomol. Screen.* **8**, 566–570 (2003).
47. Ljosa, V., Sokolnicki, K.L. & Carpenter, A.E. Annotated high-throughput microscopy image sets for validation. *Nat. Methods* **9**, 637 (2012).
48. Guyon, I. & Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **3**, 1157–1182 (2003).
49. Carpenter, A.E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).