



Optimizing the Cell Painting assay for image-based profiling

Beth A. Cimini¹, Srinivas Niranj Chandrasekaran¹, Maria Kost-Alimova², Lisa Miller², Amy Goodale³, Briana Fritchman³, Patrick Byrne², Sakshi Garg⁴, Nasim Jamali¹, David J. Logan⁵, John B. Concannon⁶, Charles-Hugues Lardeau⁷, Elizabeth Mouchet⁸, Shantanu Singh¹, Hamdah Shafqat Abbasi¹, Peter Aspesi Jr⁶, Justin D. Boyd⁵, Tamara Gilbert⁵, David Gnutt⁹, Santosh Hariharan¹⁰, Desiree Hernandez³, Gisela Hormel⁴, Karolina Juhani⁸, Michelle Melanson², Lewis H. Mervin⁸, Tiziana Monteverde⁸, James E. Pilling⁸, Adam Skepner¹⁰, Susanne E. Swalley¹¹, Anita Vrcic², Erin Weisbart¹, Guy Williams⁸, Shan Yu¹², Bolek Zapiec⁴ and Anne E. Carpenter¹✉

In image-based profiling, software extracts thousands of morphological features of cells from multi-channel fluorescence microscopy images, yielding single-cell profiles that can be used for basic research and drug discovery. Powerful applications have been proven, including clustering chemical and genetic perturbations on the basis of their similar morphological impact, identifying disease phenotypes by observing differences in profiles between healthy and diseased cells and predicting assay outcomes by using machine learning, among many others. Here, we provide an updated protocol for the most popular assay for image-based profiling, Cell Painting. Introduced in 2013, it uses six stains imaged in five channels and labels eight diverse components of the cell: DNA, cytoplasmic RNA, nucleoli, actin, Golgi apparatus, plasma membrane, endoplasmic reticulum and mitochondria. The original protocol was updated in 2016 on the basis of several years' experience running it at two sites, after optimizing it by visual stain quality. Here, we describe the work of the Joint Undertaking for Morphological Profiling Cell Painting Consortium, to improve upon the assay via quantitative optimization by measuring the assay's ability to detect morphological phenotypes and group similar perturbations together. The assay gives very robust outputs despite various changes to the protocol, and two vendors' dyes work equivalently well. We present Cell Painting version 3, in which some steps are simplified and several stain concentrations can be reduced, saving costs. Cell culture and image acquisition take 1–2 weeks for typically sized batches of ≤ 20 plates; feature extraction and data analysis take an additional 1–2 weeks.

This protocol is an update to *Nat. Protoc.* 11, 1757–1774 (2016): <https://doi.org/10.1038/nprot.2016.105>

Introduction

Since the advent of the digital camera, computationally minded biologists and biologically minded computationalists have realized that microscopy images represent an incredibly rich source of data; a camera that produces images that are 1,000 pixels on each side produces a million quantitative data points with every acquisition. The idea that scientific answers could be extracted from those pixels underlies the field of image-based profiling (also known as 'morphological profiling'): its premise is that this pixel information, the details of which are often invisible to the human eye, are sufficient to cluster samples to draw meaningful conclusions^{1,2}.

The Cell Painting assay (Fig. 1) is a major driver of this field's success—the creation of an assay that could be done simply and inexpensively on commonly available laboratory equipment provided researchers in both academia and industry the ability to try morphological profiling on their own data

¹Imaging Platform, Broad Institute of MIT and Harvard, Cambridge, MA, USA. ²Center for Development of Therapeutics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. ³Genetic Perturbation Platform, Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁴Merck Healthcare KGaA, Darmstadt, Germany. ⁵Internal Medicine Research Unit, Pfizer Worldwide Research, Development and Medical, Cambridge, MA, USA. ⁶Chemical Biology & Therapeutics Department, Novartis Institutes for Biomedical Research, Cambridge, MA, USA. ⁷Janssen Pharmaceutica N.V., Beerse, Belgium. ⁸AstraZeneca BioPharmaceuticals R&D, Cambridge, UK. ⁹Bayer AG, Research & Development, Pharmaceuticals, Wuppertal, Germany. ¹⁰Discovery Sciences, Pfizer, Groton, CT, USA. ¹¹Biogen, Inc., Cambridge, MA, USA. ¹²Takeda Development Center Americas, Inc., San Diego, CA, USA.

✉e-mail: anne@broadinstitute.org

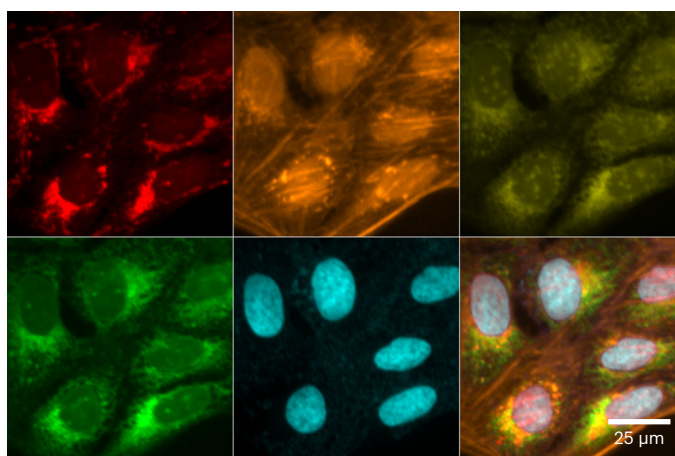


Fig. 1 | Visualization of cells in the Cell Painting assay. Dimethyl sulfoxide-treated U2OS cells treated with the Cell Painting assay; the six dyes in five channels stain eight cellular compartments. Top row (left to right): mitochondrial staining; actin, Golgi, and plasma membrane staining; and nucleolar and cytoplasmic RNA staining. Bottom row (left to right): endoplasmic reticulum staining, DNA staining, a montage of all five channels.

and an opportunity to standardize feature sets across laboratories. The two initial papers^{3,4} describing the assay have been cited more than 500 times combined. Briefly, Cell Painting involves staining the cells the scientist wishes to characterize with six small-molecule dyes that together mark eight organelles or cell compartments (DNA, cytoplasmic RNA, nucleoli, actin, Golgi apparatus, plasma membrane, endoplasmic reticulum (ER) and mitochondria). Once the stained cells are imaged, three major compartments (whole cells, nuclei and cytoplasm) are outlined or ‘segmented’ using image analysis software, and hundreds or thousands of image measurements (such as sizes, shapes, intensity distributions and textures) are calculated from each cell. These measurements can then be used in various ways, but one very common approach is to aggregate the single-cell data per well, perform normalization and feature reduction and then calculate the overall similarity in ‘feature space’ between pairs of wells to search for unexpected or interesting very-high- or very-low-similarity samples⁵. Cell Painting has been used to profile mutations in lung cancer⁶, optimize performance diversity of chemical libraries^{7,8}, find treatments for coronavirus disease 2019 (ref. ⁹) and help explore the toxicity of environmental chemicals¹⁰, among dozens of other applications¹. Scientists have also started creating their own variants, swapping in stains for lysosomes rather than mitochondria¹¹ or lipid droplets rather than ER¹². We therefore believe that new stain combinations can be powerful for particular biological areas of interest, but we also recognize the value of a standardized assay performed in many laboratories and on many biological questions across the community to share data.

We recently became interested in creating a very large public Cell Painting data set because of the success of multiple approaches in which image data can be used to predict not just particular phenotypes about the parts of the cell that were stained, but entirely orthogonal data such as gene expression data or the outcome of a biochemical assay^{13,14}. Such a database could also allow queries for compounds whose effects match those of genetic perturbations¹⁵ and vice versa. It stands to reason that a large, well-constructed set of image-based phenotypic screen data could not only inspire new computational tools to mine such data, but also serve as a resource for researchers to compare their own data against, accelerating discovery for thousands of scientists globally. The Joint Undertaking for Morphological Profiling (JUMP) Cell Painting Consortium was formed in 2019 to create such a dataset, with a goal to optimize and select a single set of staining conditions before creating more than 136,000 perturbation profiles across a dozen sites around the world¹⁶.

With these goals in mind, in this work we describe the updated procedure for these optimized staining conditions along with the experiments done over more than a year to select these optimal conditions for the JUMP Consortium’s production data set. We describe our optimization findings and decisions so that a researcher can consider these questions when setting up their own Cell Painting experiments. More than two dozen different parameters (such as cell type, dye concentration and microscope settings; see below for details) were optimized, often testing several options per parameter and in the context of multiple combinations, yielding 300 staining-plate-imaging runs or ‘logical plates’; given that many parameters are interconnected and that a full matrix of every parameter would be greater than 50 million logical plates, not all possible combinations are present.

A table describing all conditions tested is available as Supplementary Data 1; the barcodes of plates contributing to each figure are recorded into each figure's Source Data file as either column headings or as their own column.

Through these optimizations, we have found that Cell Painting is remarkably robust. Most of our attempted optimizations admittedly did not dramatically change any parameter, with the vast majority of conditions leading to consistently similar results. This robustness did not yield an exciting optimization process but speaks well of the stability of the assay and how likely it is to work across various laboratories worldwide, even if a given laboratory may need to adjust some parameters for their local conditions.

The major changes between our previous protocol and the current recommendations are as follows (presented in order of the protocol steps):

- 1 No media removal before the addition of MitoTracker, to simplify the protocol and minimize the loss of cells
- 2 Our recommendation for MitoTracker staining concentration remains at 500 nM, but previous versions of the protocol used instructions that unintentionally led to a lower final concentration (375 nM). The current protocol ensures a 500 nM concentration after dilution.
- 3 Combining permeabilization and staining steps to make the process more automation friendly
- 4 Reduction of phalloidin fourfold, from 5 µl/ml (33 nM) to 1.25 µl/ml (8.25 nM), to save reagent costs
- 5 Reduction of Hoechst fivefold, from 5 to 1 µg/ml, to save reagent costs
- 6 Increase of SYTO 14 twofold, from 3 to 6 µM, to improve its signal
- 7 Reduction of concanavalin A 20-fold, from 100 to 5 µg/mL, to save reagent costs
- 8 Overall reduction of post-fixation staining volumes from 30 to 20 µl/well, to save reagent costs

We additionally have updated the protocol to contain procedures for lentiviral transduction of open reading frame (ORF) overexpression constructs and CRISPR guide RNA constructs, as well as updated computational recommendations (including cloud computing options).

Optimization setup

A conventional assay is optimized by assessing each variant of a protocol for the optimal separation between a particular positive control and negative control. However, because the profiling assay aims to measure hundreds to thousands of readouts, this creates a challenge for optimization. In our prior efforts at optimization, we relied on assessing signal quality by eye^{3,4}. Here, we aimed to perform a comprehensive quantitative optimization of the Cell Painting assay. We evaluated each variant of the assay by using 90 compounds, selected as detailed later, to cover a broad spectrum of biological activities.

We selected optimal parameter settings based on two metrics calculated on image-based profiles derived from Cell Painting cells treated with those compounds: percent replicating and percent matching. Both describe how often a given pair of wells that should be similar actually are similar; specifically, how often is their pairwise correlation across features greater than the 95th percentile of a null distribution of the similarities of 10,000 pairs of random (non-matching) wells (for a detailed description, see ref. ¹⁷). In percent replicating, the two wells that are compared have undergone an identical treatment and should perfectly correlate, if not for technical variations. In percent matching, the two wells have undergone different treatments that are believed (because of outside knowledge, i.e., ground truth) to produce similar (or opposite) biological impact because they act on the same biological system. Percent replicating expects identical treatments to produce positive correlations, so it is scored in a one-tailed fashion (fraction of profile pairs above the 95th percentile of correlation values); by contrast, pairs of treatments might be expected to correlate or anti-correlate, so percent matching can be calculated as the fraction of treatments below the 5th percentile, the fraction above the 95th percentile, or both. These metrics have limitations, as elaborated previously¹⁸, but were sufficient at the time for the purposes here, namely selecting the optimal assay conditions given the presence of replicates in different well positions on our compound control plate, and with identical numbers of replicates of each condition. When comparing conditions in a given experiment, percent replicating has the advantage of being calculated for a higher number of conditions, because it is computed for each treatment individually. By contrast, percent matching has the advantage of better approximating an application of image-based profiling: matching samples that have undergone different yet biologically related treatments. However, because percent matching can be calculated only for pairs of samples, there are fewer classes and thus less statistical power. It is also a harder task

because most compounds annotated as having a target in common do not in fact produce similar morphological profiles, particularly due to polypharmacology¹⁷ (see Supplementary Methods) and because ground truth annotations are incomplete and imperfect. Nevertheless, because samples that are supposed to match each other are in different positions within our compound control plate layout, this metric is usually less influenced by plate position effects, such as edge effects, which can unfairly improve percent replicating when replicates are in identical well positions within the plate layout.

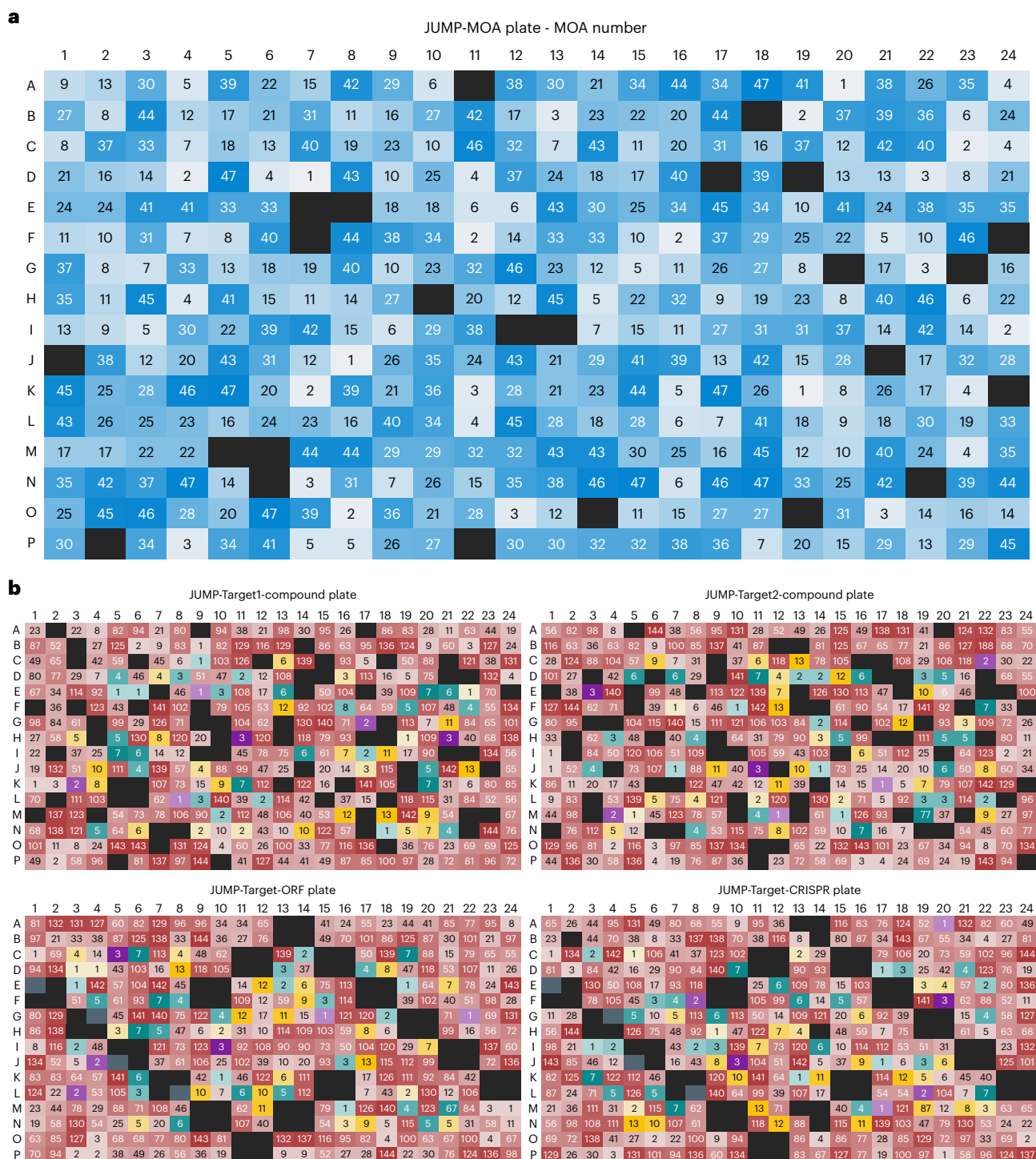
The percent replicating and percent matching of an entirely untreated or negative control plate would be expected to be 5% (or 10%, if both tails of the distribution are assessed). Our first step was to select a set of standardized controls to determine whether treated wells were matching more than would be expected by chance. For initial optimizations, a compound plate known as the JUMP-MOA (mechanism of action) plate was created (Fig. 2a and Extended Data Fig. 1); in 384 wells, it contains 24 dimethyl sulfoxide (DMSO) wells and 90 compounds with four replicate wells each. The 90 compounds are from 47 diverse MOA classes, with 43 classes having 2 compounds each and 4 classes having only a single compound. This allows for testing percent replicating for each of the 90 compounds and percent matching for each of the 43 multi-compound MOA classes within a single plate, allowing each plate to serve as its own ‘batch’; see Supplementary Methods for more information. This plate layout was used for most optimizations of staining reagents and conditions, imaging conditions and feature measurements; percent matching for these plates is calculated in a one-tailed fashion (>95th percentile).

To determine whether optimal conditions would differ across the different kinds of treatments planned for the JUMP consortium (compound treatments, ORF overexpression and CRISPR knockout), a second set of plates known as the JUMP-Target plates were produced (Fig. 2b). These are described elsewhere¹⁷ and online (<https://github.com/jump-cellpainting/JUMP-Target>); briefly, they contain either compounds, ORFs or CRISPR knockdowns related to a set of >175 genes thought to have strong and/or diverse phenotypic effects. Because this set of samples is so large, to reach the same four replicates per treatment found in JUMP-MOA, one must create four identical treated plates of the JUMP-Target-Compound source plate (there are two versions of this plate with the same set of 306 compounds but in different layouts: JUMP-Target-1-Compound and JUMP-Target-2-Compound; hereafter, JUMP-Target1 and JUMP-Target2). The JUMP-Target-CRISPR source plate, like the JUMP-Target-Compound plates requires four identical treated plates; it contains two guides for most genes, arrayed in separate wells on each plate. By contrast, the JUMP-Target-ORF plate has 130 ORFs duplicated on the plate (because typically only a single ORF is available per gene) and thus needs only two treatment plate copies of the source plate to get a sufficient number of wells with four replicates. Because JUMP-Target treatments may be expected to increase or decrease the function of a target gene, percent matching is calculated in a two-tailed fashion (<5th percentile or >95th percentile).

Optimization of cell line selection, treatment and culture conditions

The furthest-upstream task that the consortium needed to consider was which cell line to use in data production, given our desire to have all data in one cell line to maximize matching across treatments. In addition to typical imaging-assay concerns such as flatness, we wanted specifically to know for each candidate line: (1) how many diverse phenotypes could we detect by using the Cell Painting assay; (2) how well it overexpressed exogenously introduced genes, given our plans to analyze ORFs; and (3) how well particular Cas9-expressing clonal lines could knock down genes, given our plans to analyze CRISPR reagents. Especially when reproducibility across sites is critical, one additionally may wish to consider the availability of Cas9-expressing clones (some lines carry restrictive licenses, making them unavailable to the public).

Once a line was chosen, we also wanted to know the effects of several variables in the Cell Painting assay protocol: (4) how long a given treatment should be applied to the cells; (5) how sensitive the assay would be to changes in plating density; (6) how, if at all, using Cas9-expressing clonal lines for CRISPR experiments and parental lines for compounds would affect our ability to match treatments affecting the same mechanisms of action; (7) how, if at all, using drug selection in our CRISPR and ORF conditions would affect their phenotypes and therefore our ability to match these conditions to compound treatment plates; and (8) whether or not cell conditions that some treatments would be exposed to (polybrene for improving lentiviral introduction of ORF and CRISPR reagents) should be applied to even the cells that did not need to experience them (such as the compound-treatment plates).



We focused on assessing the relative performance of A549 versus U2OS, because both are lines in which large public Cell Painting data sets already exist¹⁹; dozens of other cell lines have performed well for Cell Painting experiments, and researchers should choose a line that demonstrates phenotypes in which they are interested. We tested parental populations and several Cas9-expressing clonal lines for each cell type (one polyclonal of each line, with one additional monoclonal line for U2OS and three monoclonal lines for A549) by using the JUMP-MOA plates to assess the ability to detect multiple phenotypes (Fig. 3a); a line in which fewer phenotypes were morphologically distinct would by definition have a lower percent replicating and percent matching because fewer treatments would fall outside the null distribution. The Cas9-editing efficiency of the Cas9-expressing lines was measured in parallel (Supplementary Table 1 and Supplementary Methods). In U2OS, the polyclonal line displayed much greater CRISPR efficiency as well as better percent matching. In A549, any, if not all, of the lines could have been suitable on the basis of efficiency and percent matching; we therefore chose on the basis of the line with the fewest restrictions on sharing between partners, which for the particular four clones in question was the polyclonal line. Additional experiments (quantified by cell count only) helped decide final cell densities, viral amounts and other conditions (Supplementary Methods).

We next carried the parental and polyclonal lines forward to a single large experiment that we called ‘CPJUMP1’¹⁷, by using the JUMP-Target plates to address questions 1–7 above; question 8 was assessed in a second, later experiment. Briefly, CPJUMP1’s main experiment involved treating both A549 and U2OS cells with the JUMP-Target-Compound or JUMP-Target-ORF plates and treating polyclonal-Cas9-expressing A549 and U2OS lines with the JUMP-Target-CRISPR plates; these were treated and stained in parallel at various treatment time points to test both percent replicating and the ability to cross-match wells from different plate types (i.e., whether overexpression of Gene A looks phenotypically opposite from CRISPR knockdown of Gene A, and whether ORFs, CRISPRs or both look phenotypically similar to treatment with compounds that have previously been shown to target Gene A). Additional A549 JUMP-Target-Compound plates were prepared to assess the effects of altering plating density by $\pm 20\%$ on ability to match, and additional plates of the polyclonal Cas9-expressing A549 line were prepared with JUMP-Target-Compound plates to see if screening the compounds in the Cas9-expressing lines would improve overall matching performance. Finally, an additional A549 plate treated with JUMP-Target-ORF and additional A549-polyclonal plates treated with JUMP-Target-CRISPR were created so that they could undergo drug selection to remove uninfected cells, to see if drug selection harmed matching, improved matching or had no visible effect.

Neither line performed poorly in any of our experiments, and on the whole, none of the cross-modality matching experiments (Fig. 3b, top row center and right panels, bottom row middle panel) performed much beyond what would be randomly expected (10%). U2OS had a higher percent replicating in two out of three modalities (Extended Data Fig. 2a), and on the whole, U2OS displayed a slightly higher cross-modality matching and higher genetic perturbation-across-time point matching (both of which would hopefully increase the ability to match to other datasets taken with different perturbation types and/or at variable time points) (Fig. 3b, bottom row left and right panels). Although U2OS performed worse at across-time point matching (Fig. 3b, top left panel), it performed better in our previous experiment at within-plate compound matching (Fig. 3a), making that comparison in our minds a wash. Finally, U2OS had a greater amount of publicly available data; thus, by a slim margin, U2OS was selected to be used for producing the Joint Undertaking for Morphological Profiling Cell Painting (JUMP-CP) Consortium’s full dataset (questions 1–3). On the basis of the highest percent replicating, we decided that compound treatments would be run at 48 h of treatment (Extended Data Fig. 2a); on the basis of comparing the changes in percent replicating (Extended Data Fig. 2a) and the ability to match to 48-h compound treatment (Fig. 3b), it was decided to run ORF overexpression experiments at 48 h of treatment and CRISPR knockdowns at 96 h of treatment (question 4).

We further analyzed the rest of CPJUMP1 to address questions 5 and 6; although these analyses were done in A549 rather than U2OS, we believe the results are likely to hold in both types because of their generally similar behavior. We saw little effect of profile sensitivity to plating density changes of $\pm 20\%$ (Extended Data Fig. 2b). We also saw little difference on modality matching when running compound experiments in the parental line versus the polyclonal Cas9 line (Extended Data Fig. 2c) and as such chose to have all partners running compound experiments screen their compounds in the commercially available parental U2OS lines.

To answer question 7, we needed to assess for our ORF and CRISPR conditions how selection of the lentiviral reagents by resistance markers in the viral backbones would affect our ability to match

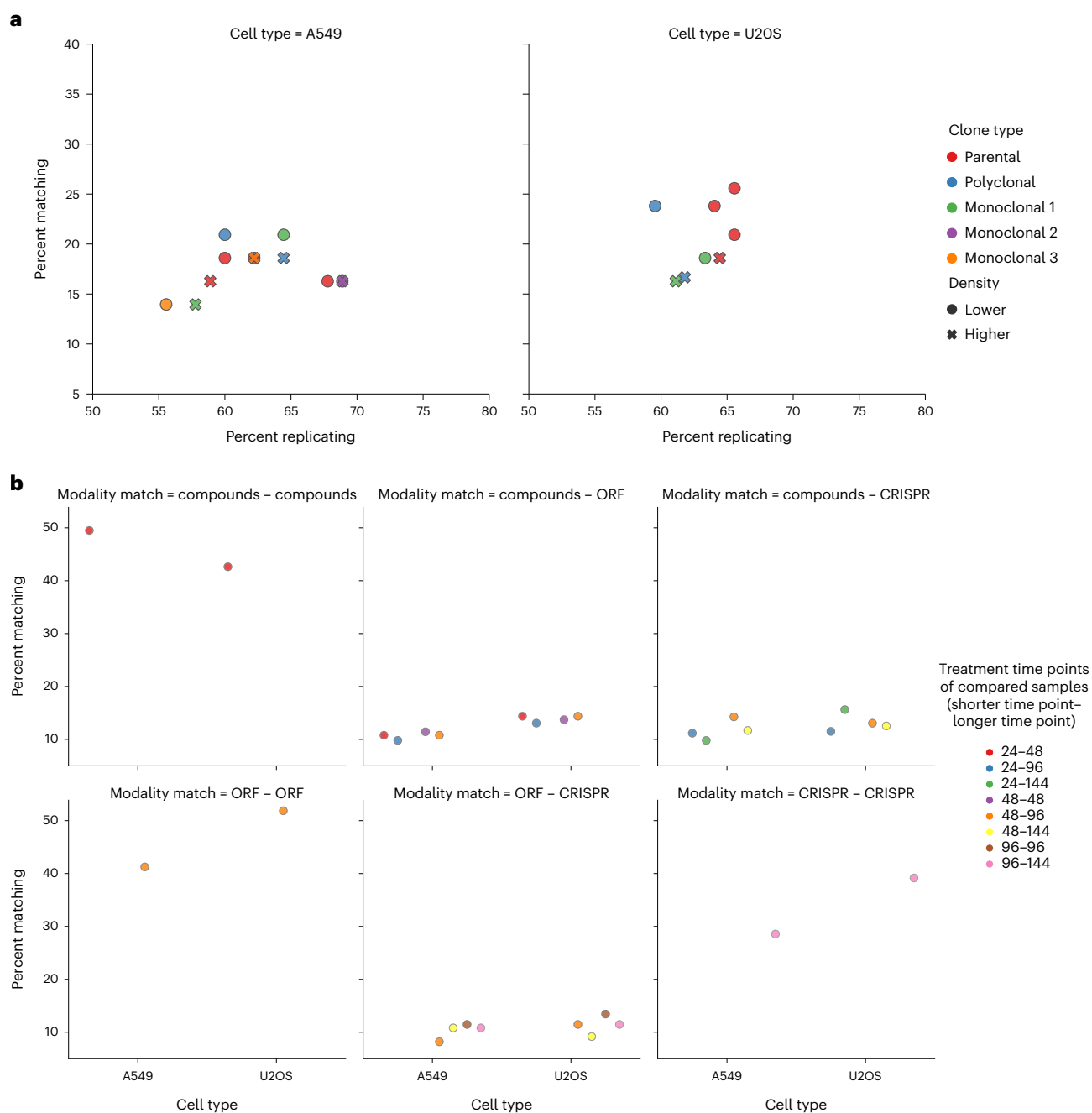


Fig. 3 | Assessment of cell type suitability for Cell Painting. **a**, Assessment of compound percent replicating and percent matching by using JUMP-MOA plates in A549 and U2OS parental, Cas9-polyclonal and Cas9-monoclonal lines. In U2OS, the polyclonal line outperforms the single monoclonal line tested; in A549, performance is more varied. **b**, Assessment of the ability to match compounds, overexpressed genes and CRISPR knockdown genes across treatment conditions in A549 and U2OS. The top-left, bottom-left and bottom-right panels reflect the ability to match across two time points of compound treatment, gene overexpression and gene knockdown, respectively; the remaining panels describe the ability to match to perturbations of a different modality but thought to be targeting the same gene. U2OS has slightly worse compound-to-compound-across-time-point percent matching than A549 but performs better across time points at matching overexpressions to overexpressions and knockdowns to knockdowns. Although no cell/time point cross-modality comparisons performed noticeably above random chance (10%), U2OS performed comparably or better than A549 at the final treatment time points chosen (purple dots in the top center panel, orange dots in the top-right and bottom-center panels). For more information including the plate(s) represented by each data point, see the Source Data file for this figure; expanded experimental details for each plate may be found in Supplementary Data File 1.

between those conditions and our larger compound panel. Selection could possibly alter the results in either direction, improving them by removing cells not expressing the treatment and/or harming them by introducing a second ‘drug selection’ signature that might perturb biological signals. Drug selection may have had a small deleterious effect on percent replicating (Extended Data Fig. 3b),

although we cannot rule out that this is due to fewer replicates for the selected plates than the unselected ones. In addition, our assessment of cross-modality percent matching suggests a potential small deleterious effect from drug selection, especially in ORFs (Extended Data Fig. 3a). This is understandable, because in our case, ORFs have a vector with a slower selectable marker (blasticidin) than CRISPR (puromycin) and are treated for a shorter period of time (96 versus 144 h). We therefore chose not to perform drug selection in our final protocol recommendations.

Finally, for question 8, we wanted to know whether we should include polybrene, which is used in aiding lentiviral transduction in our ORF and CRISPR plates, in our compound treatment production plates; as with the question of drug selection, one could a priori imagine it either helping or harming cross-modality matching. In this case, we saw a strong deleterious effect on profiling results from a 24-h treatment with 4 μ M polybrene, with cross-plate replication between polybrene-treated versus untreated plates dramatically lower than the intra-treatment cross-plate replication of either treated or untreated plates (Extended Data Fig. 3c). Polybrene addition also did not improve the ability to match Target2 plates to a previous batch of ORF overexpression plates (Extended Data Fig. 3d), causing us to recommend against including it.

Optimization of plates, staining reagents and conditions

Once a researcher has picked their cell line and treatment conditions, the next thing they must do is get their cells onto imaging plates and stain them. We saw no consistent difference in percent replicating or percent matching between plates from two manufacturers, one of which contained ‘barrier wells’, also known as reservoirs at the edges of the plates that hold liquid in them to try to create even humidification of the whole plate surface (Extended Data Fig. 4b). They may prove beneficial in other experimental contexts or research environments, especially for those new to running sensitive high-content assays like Cell Painting, but we decided against using them because of increased cost.

To try to minimize disruption to the cells, time and buffer spent washing plates, we have introduced two washing-related changes in this protocol versus our 2016 recommendations. First is to not remove any medium from the wells before the addition of MitoTracker; this creates higher costs for this particular reagent, because staining is done in a larger volume, but it omits a step (removing the medium) and decreases the likelihood of precious cells becoming detached before fixation. Second, the original protocol involved permeabilizing the cells, washing them twice and then adding the other (non-MitoTracker) dyes; now, we recommend performing the permeabilization and staining simultaneously. This leads to a shorter and easier protocol, with no apparent negative consequences to profile quality (Extended Data Fig. 4a). Although the total volume for the MitoTracker staining step has gone up from 30 to 60 μ l, all other dyes are now added in a smaller volume (20 rather than 30 μ l).

The next thing a researcher setting up a Cell Painting assay must decide is exactly which stains to use and at what concentrations. Although variations on Cell Painting can be powerful for investigating particular areas of biology^{11,12}, for optimization of the canonical assay for this public dataset, the consortium considered only the original six dyes used in earlier versions of the assay (with the exception of DAPI for Hoechst; see Extended Data Fig. 5b). We did briefly qualitatively assess if there were any benefits of moving wheat-germ agglutinin (WGA) to an ultraviolet dye or MitoTracker to an orange dye but quickly dismissed them: no immediate large improvement was shown (data not shown), and it would require reoptimization of the entire stain panel. A large number of stain concentration adjustments were tested; they are broken out comprehensively in Extended Data Fig. 5. We found that reducing Hoechst from its original 5 μ g/ml to 1 μ g/ml and diluting phalloidin from 33 to 8.25 nM had no ill effect and possibly even a positive one (Extended Data Fig. 5b), so we adopted these changes to reduce reagent costs and waste.

All the changes in the preceding paragraphs form what we call the Cell Painting version 2.5 (v2.5) protocol; this protocol was used for the CPJUMP1 experiment and represents an improvement in percent replicating and percent matching from the version 2 protocol from 2016 (ref. ³) (Fig. 4a). However, some consortium members reported higher than optimal bleed-through in version 2.5, specifically that on their microscopes, there was so much cross-talk between the ER (concanavalin A) and RNA (SYTO 14) stains that it affected detection of the RNA signal. We therefore did one additional round of optimization to yield the version 3 (v3) protocol, incorporating all the improvements from v2.5 plus reducing by 20 \times the amount of concanavalin A (the most expensive dye in the panel) and doubling the amount of SYTO 14 (Fig. 4a). These changes brought the two channels better into balance and became our final recommendation, v3. There are two major sources for the complete set of Cell Painting

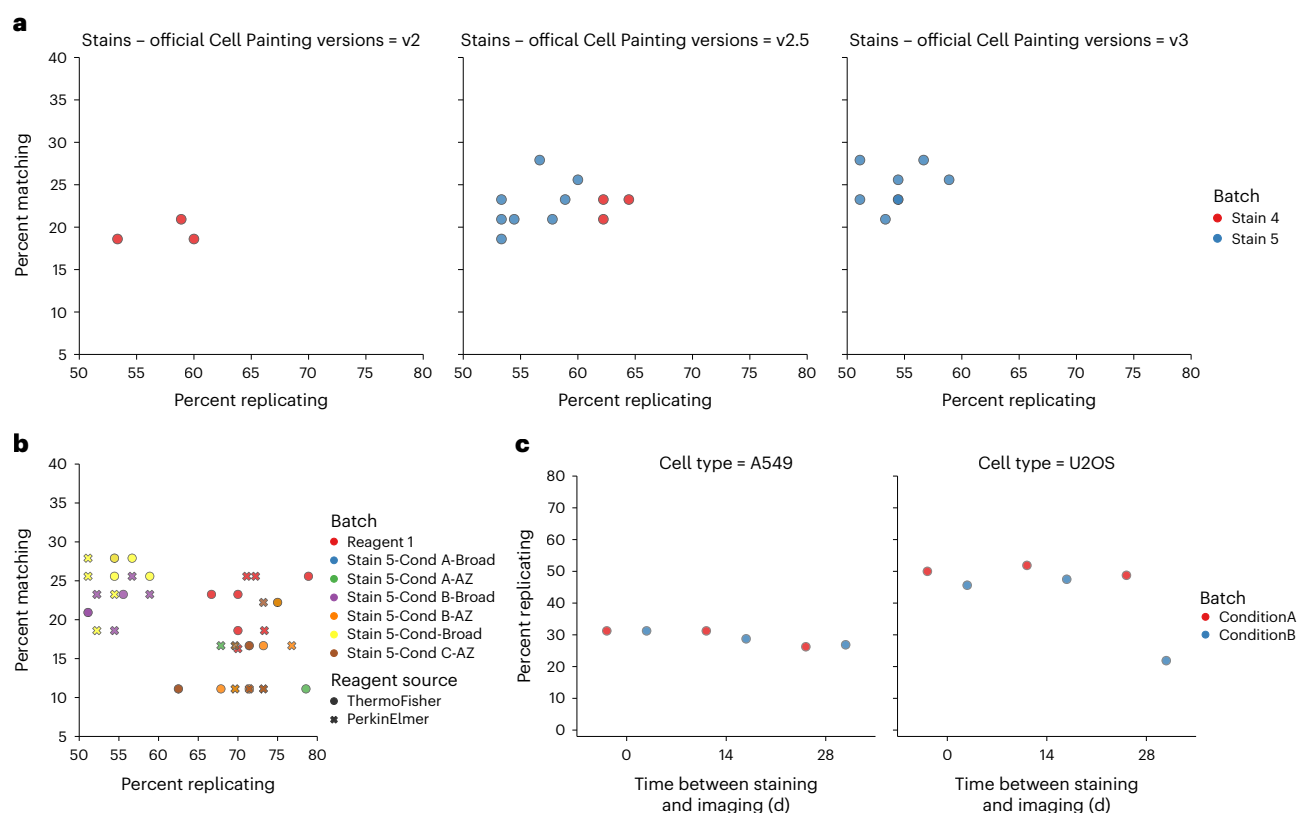


Fig. 4 | Assessment of staining conditions for use in the JUMP Consortium. **a**, Comparison of percent replicating and percent matching for three versions of Cell Painting: version 2 (v2; ref. ³ conditions) versus version 2.5 (v2.5; ref. ³ plus introduction changes 1–6) versus version 3 (v3; this paper's final recommendations). The move from v2 to v2.5 seemed to improve both percent replicating and percent matching in the Stain 4 pilot experiment (red dots); the move from v2.5 to v3 decreases reagent cost while maintaining comparable, if not slightly improved, percent matching in the Stain 5 pilot experiment (blue dots). Note that Stain 5 experiments (blue dots) were performed by using only half the compound dose used in Stain 4 experiments (red dots). **b**, Comparison of reagents from two different vendors across multiple stain conditions and microscopes. Performance is extremely similar between vendors in all conditions tested. **c**, Assessment of persistence of Cell Painting plate quality over storage time. Percent replicating seems to be decreasing by day 28 but is quite similar to initial values at day 14. For more information including the plate(s) represented by each data point, see the Source Data file for this figure; expanded experimental details for each plate may be found in Supplementary Data File 1. AZ, AstraZeneca; Cond, condition.

dyes: ThermoFisher and PerkinElmer. We tested both dye sets in a number of batches and across various conditions and found their performance to be equivalent (Fig. 4b).

The Cell Painting assay uses fixed cells, reducing the need for precise timing of the image-acquisition step. Because microscopes occasionally break or are booked, we finally wanted to test the timescale of deterioration for a fixed and stained Cell Painting plate, in terms of overall profile quality; we see no measurable degradation of profile percent replicating between plates imaged on the day they were stained and plates imaged 14 d later. There appears to be a consistent decline between weeks 2 and 4, however, so if possible we recommend that imaging be completed within 14 d after staining (Fig. 4c).

Optimization of imaging conditions

Once the staining conditions are finalized, the next step is to optimize the image-acquisition conditions. As in other areas of optimization, we had a number of questions. First, which kind of microscope should we use? Second, should we take one Z plane or several? Third, should the camera be set to 1 × 1 or 2 × 2 pixel binning? Fourth, how high should our exposures be? Fifth, how many fields of view do we need to take? Sixth, if a researcher needs to image a plate more than once (during optimization or because of a technical failure), how much deterioration of signal can they expect?

In our testing, the answer to most of these questions was simply 'probably anything will produce comparable results'. We saw no consistent difference between images taken in wide-field versus confocal microscopy (Fig. 5a), even when those modes were on two entirely different microscopes (Extended Data Fig. 6a). The only plate in which we saw decreased performance for confocal

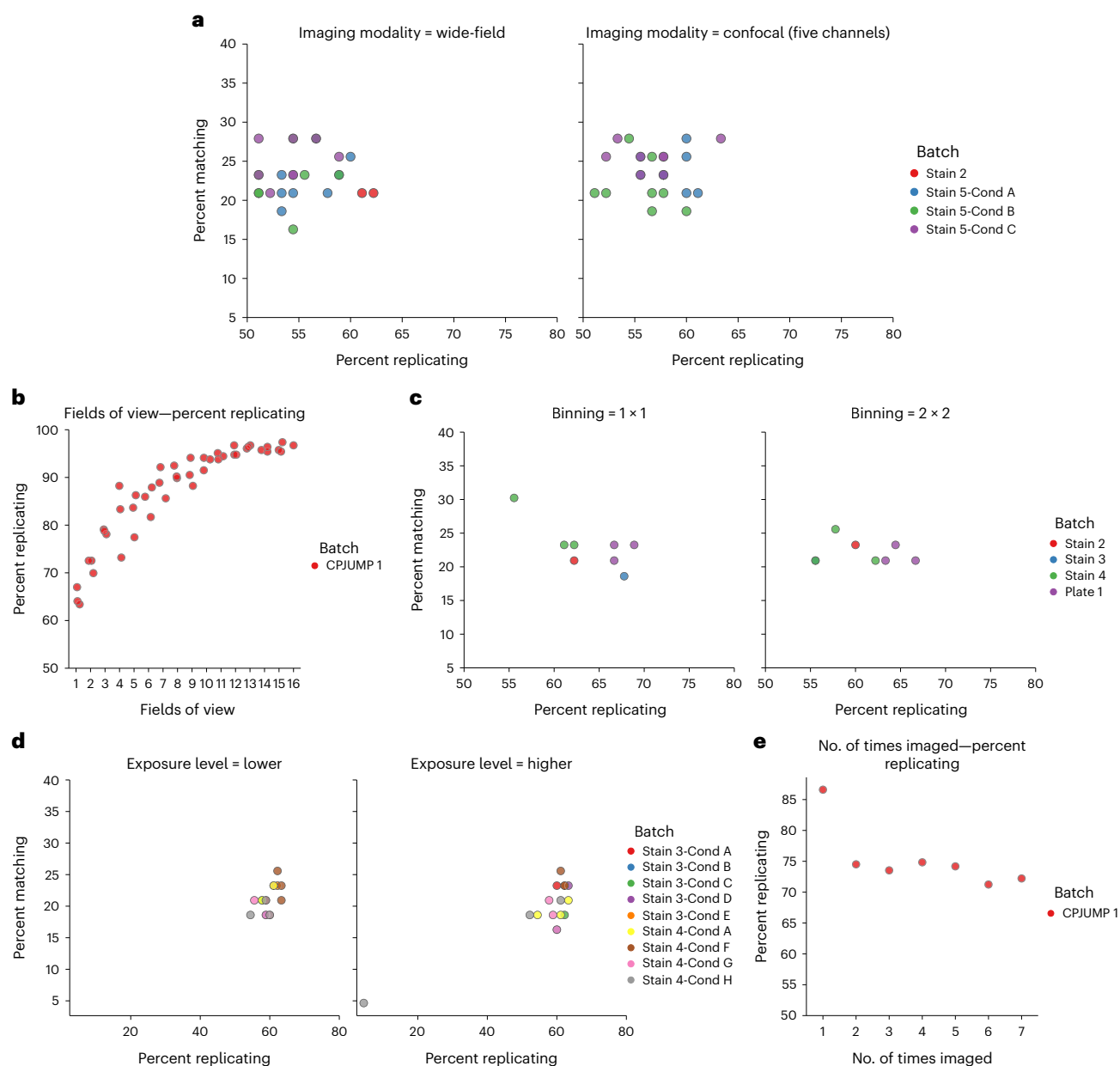


Fig. 5 | Assessment of imaging conditions for use in the JUMP Consortium. **a**, Cell Painting works similarly well with wide-field and confocal microscopy. **b**, Relationship between fields of view captured and percent replicating: increasing the number of fields of view increases percent replicating up until ~10 fields of view. At this magnification and plating density, each field of view contains ~145 cells. **c**, Cell Painting performs well when images are captured with either 1×1 or 2×2 binning. **d**, Cell Painting performs similarly when plates are imaged at a lower or two to four times higher (but still below saturation) laser power and/or exposure time. **e**, Assessment of the effect of re-imaging on percent replicating: a drop between first and second imaging is observed and a potential small continued decrease thereafter. For more information including the plate(s) represented by each data point, see the Source Data file for this figure; expanded experimental details for each plate may be found in Supplementary Data File 1.

microscopy was in a condition in which the microscope could not create a filter set match that would separate the ER and RNA data (Table 1) and thus captured them only as a single channel. To our surprise, there was essentially no profile quality loss in our hands by switching from capturing images with 1×1 binning to 2×2 binning (Fig. 5c), leading us to choose 2×2 binning, because it reduces the data storage price by four times and computational costs significantly. We also did not see any change between lower and higher exposures for each of several staining conditions (Fig. 5d), suggesting that as long as the exposure times are set reasonably enough to maximize dynamic range while minimizing saturation, the exact values are less important, and one can potentially save imaging time by using slightly shorter exposures.

Table 1 | Details of the ImageXpress Micro Confocal 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	377/54	447/60	Nucleus	DNA
Concanavalin A/Alexa Fluor 488 conjugate	475/34	536/40	Endoplasmic reticulum	ER
SYTO 14 nucleic acid stain	531/40	593/40	Nucleoli, cytoplasmic RNA	RNA
Phalloidin/Alexa Fluor 568 conjugate, WGA/Alexa Fluor 555 conjugate	560/32	624/40	F-actin cytoskeleton, Golgi, plasma membrane	AGP
MitoTracker Deep Red	631/28	692/40	Mitochondria	Mito

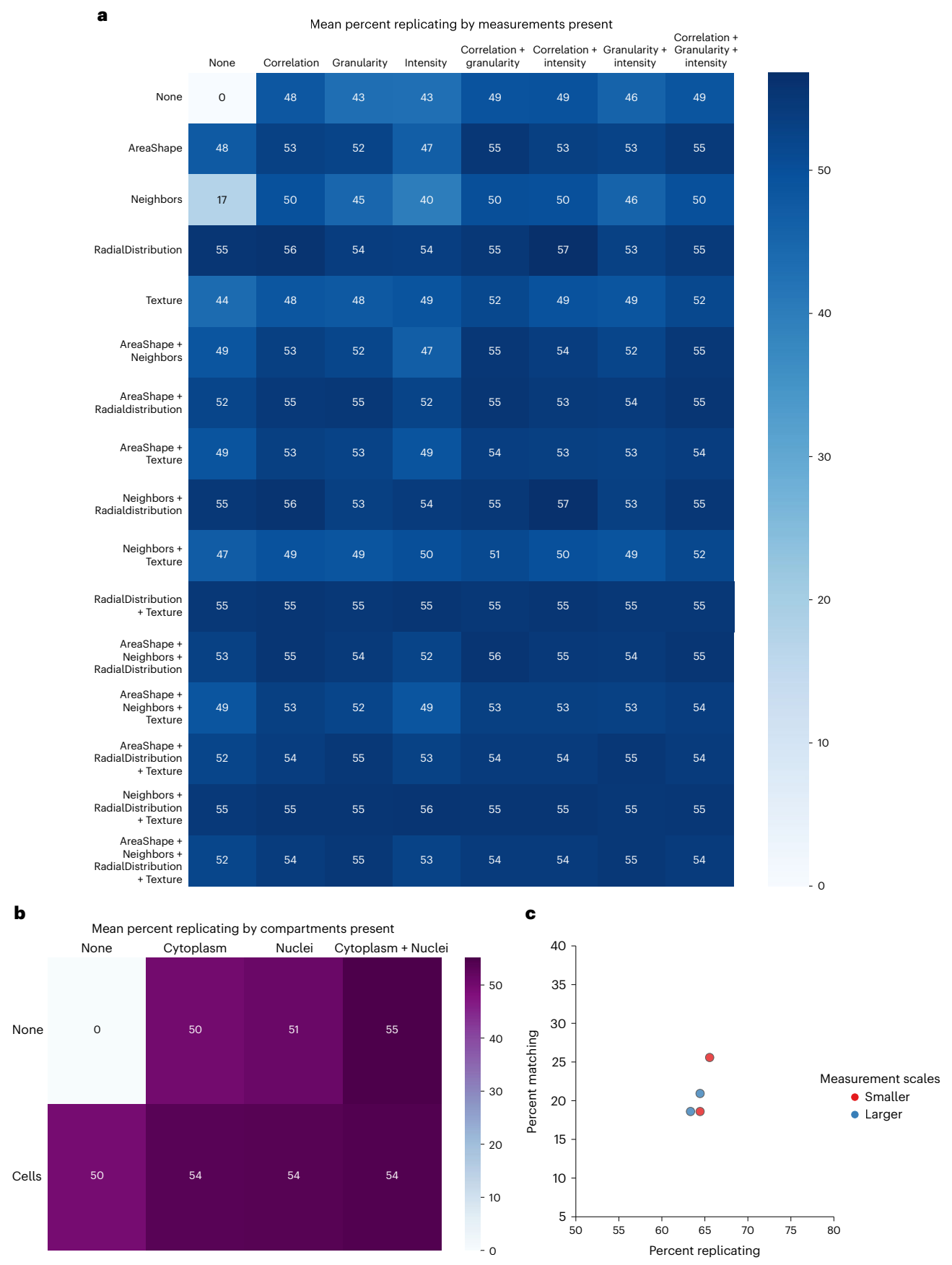
A few parameter changes do seem to produce measurable results; we saw small increases in percent replicating (but not percent matching) in two experiments in which we tested imaging the same plate in several fluorescent Z planes followed by maximum projection, versus a single plane. The increase was small enough (Extended Data Fig. 6b) that we chose not to complicate our workflow but may be worth considering in other circumstances, such as when using confocal imaging or when using cells that have varying Z heights, such as neurons. We also found that the number of fields of view acquired was important, with each additional field of view leading to an increase in percent replicating out to at least 10 fields of view (Fig. 5b) at our magnification and plating density (~145 cells per field of view). We suspect this has more to do with the total number of cells imaged than the number of fields of view per se. We saw a roughly 10% drop in percent replicating signal when plates had been imaged twice versus a single time and minor subsequent losses thereafter up to six (Fig. 5e). We therefore recommend that plates are imaged only once, although 10% loss may be acceptable in many contexts, relative to the cost of repeating sample preparation for a given plate or batch; because the initial percent replicating value was extremely high, a 10% loss is probably a worst-case scenario.

Optimization of image-analysis conditions

In our experience, creating accurate segmentations is crucial but difficult to optimize in a rule-driven way without substantial effort to create a ‘ground truth’ for evaluating changes in the pipeline’s parameters²⁰. We have described elsewhere²¹ a number of resources for learning to do so effectively, and the workflow described here includes steps for iterating on segmentation to ensure that it is optimal before feature extraction. As deep learning segmentation tools such as StarDist²² and Cellpose²³ become more popular, they may help solve many difficult segmentation issues. These tools may be used independently to create segmentations (with object label matrices subsequently brought into CellProfiler), used via their ImageJ implementations by using CellProfiler’s RunImageJMacro²⁴ or directly via their CellProfiler plugins (<https://github.com/CellProfiler/CellProfiler-plugins>) if CellProfiler is installed from the source.

Once objects are segmented, we generally recommend measuring as many image-based features as is practical in your image-analysis software, where ‘practical’ has a couple considerations, such as (i) the amount of time needed to generate these measurements and (ii) the limitation on your output size based on file constraints—unless manually compiled, SQLite allows only 2,000 columns in a given table, for example.

An open question about Cell Painting is how much each stain contributes to the information content of the assay. Likewise, the contributions of different measurement categories are unknown. Therefore we analyzed the relative contributions toward percent replicating of different measurement types (Fig. 6a), cell compartments (Fig. 6b) and channels (Extended Data Fig. 7) in the context of the 90 compounds present in the JUMP-MOA plate. As with the staining and imaging, for these phenotypes, the data we collect are extremely robust to changes in the measured features and channels—most subsamples of channels and/or features and/or compartments (Fig. 6a,b; Extended Data Fig. 8; and Extended Data Fig. 9) will still lead to a high-quality final analysis. We note, however, that these breakdowns describe only the phenotypes present in the JUMP-MOA plates; any specific phenotype(s) of interest may crucially depend on compartments, stains or features that are less critical for these 47 mechanisms of action, and as such, we always recommend capturing as many features as practicably possible. We saw only small differences when the scales of several CellProfiler features were adjusted, indicating that there is at least a reasonable range of tolerances for these parameters as well (Fig. 6c).



◀ **Fig. 6 | Assessment of image-analysis feature options for use in the JUMP Consortium.** **a**, Mean percent replicating of eight JUMP-MOA plates stained with the final staining conditions after dropping out all possible combinations of features from the seven major feature categories before performing feature selection and calculation of percent replicating. To create a sufficiently compact data representation, the seven categories present were split three onto the X (Correlation, Granularity and Intensity) and four onto the Y (AreaShape, Neighbors, RadialDistribution and Texture) axes; this allows visualization of the 127 possible unique combinations. A channel-by-channel breakdown of the importance of the feature categories is provided as Extended Data Fig. 9. **b**, A parallel analysis of the same experiment as in **a**, but with the three compartments present, two on the X axis (Cytoplasm and Nuclei) and Cells on the Y axis. **c**, Assessment of the effect of varying the measurement scales in CellProfiler (i.e., measuring Texture at 3- and 5-pixel spacings (Smaller) versus 5- and 10-pixel spacings (Larger)) on percent replicating for two plates of the CPJUMP1 experiment. Using larger measurement scales in the MeasureGranularity, MeasureTexture and MeasureObjectNeighbors modules seemed to produce a very small decrease in percent replicating but no major effect. For more information including the plate(s) represented by each data point, see the Source Data file for this figure; expanded experimental details for each plate may be found in Supplementary Data File 1.

Limitations

Like any assay, phenotypic or otherwise, Cell Painting will not be able to detect every phenotype—previous work found phenotypes for 50% of overexpressed genes²⁵. The percent replicating values found in this work suggest that, even in our handpicked treatments, we conclusively identify identically treated wells in different plate positions only ~50–70% of the time (see the *x* axes of non-CPJUMP1-related plates, such as in Figs. 3a, 4a,b; and 5a,c,d), although we hope that this will improve as better computational methods become available. We hope that this work will help guide researchers who want to adapt Cell Painting to a new system or create variants of Cell Painting that will definitely detect their phenotypes of interest to understand the necessary experiments needed to find the right conditions for their system. We also hope that the control plates designed for this work can make optimizations easier for future researchers.

In creating such a large public data set and making it easier to reproduce our exact technique, our goal is to help researchers better match the public set to their own samples of interest. However, major work still remains in being able to match samples across different experiments¹⁵; a major goal of the JUMP Consortium is to develop such methods. We hope that future technical advancements will make it possible to match samples across widely variant assays, but until then, matching our protocol to the degree possible represents the best chance of being able to match to the public set. Future technical advancements may make it possible to predict the results of this specific assay from brightfield images²⁶, but it is not yet clear how transferable models trained on data made at one location will be on data from another location.

Materials

Biological materials

• Cell line of interest—we used U2OS cells (American Type Culture Collection, cat. no. HTB-96; RRID: [CVCL_0042](#)) or A549 cells (American Type Culture Collection, cat. no. CCL-185; RRID: [CVCL_0023](#)) for most of our large screens. We are also aware of successful Cell Painting experiments using MCF-7, 3T3, ES2, HCC44, HTB-9, HeLa, HepG2, HEKTE, SH-SY5Y, HUVEC, HMVEC, Ocy454, primary human fibroblasts, primary human hepatocytes, primary human adipose-derived mesenchymal stem cells and primary human hepatocyte/3T3-J2 fibroblast cocultures. **! CAUTION** Cell lines should be regularly checked to ensure that they are authentic and that they are not infected with mycoplasma. **▲ CRITICAL** If using CRISPR reagents with the vector recommended below, Cas9 is not encoded in the vector, and you therefore must use a version of your preferred cell line that either permanently or temporarily expresses Cas9. Lentiviral vectors that co-express Cas9 may be used instead but may have lower titer²⁷.

Reagents

! CAUTION This protocol optionally uses replication-incompetent lentivirus, which should be handled carefully to avoid exposure. Contact your biosafety office about institutional guidelines and any required training for working with lentivirus. **▲ CRITICAL** We have performed the Cell Painting assay by 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.

- DMEM (Corning, cat. no. 10-013-CV)
- McCoy's 5A medium (Life Technologies, cat. no. 16600108)
- FBS (Sigma-Aldrich, cat. no. F2442-500ML)
- 0.25% (wt/vol) trypsin-EDTA (Corning, 25-053-CI)
- PBS (Sigma-Aldrich, cat. no. D8537-6X500ML)
- Pen-strep (Life Technologies, cat. no. 15140163)

- Ethanol (Decon Labs, cat. no. V1401)
- Polybrene (Sigma-Aldrich Inc., cat. no. 28728-55-4)
- (Optional) Blasticidin S HCl 10 mg/ml stock solution (Life Technologies, cat. no. A1113903)
! CAUTION Blasticidin is a toxic chemical and harmful if swallowed. Always wear gloves, a laboratory coat and safety glasses when handling blasticidin-containing solutions and wash your hands thoroughly after handling.
- (Optional) Puromycin 10 mg/ml stock solution (Sigma-Aldrich, cat. no. P9620-10ML)
- (Optional) Small-molecule libraries, typically 10 mM stock in DMSO (e.g., Chembridge library or Maybridge library; the compounds from the JUMP-Target-1 and JUMP-MOA plates are available as Pre-Plated Cell Painting libraries from Specs: <https://www.specs.net/index.php?page=2019041215290210#preplatedsets>)
! CAUTION Some small-molecule libraries contain toxic compounds; suitable precautions should be taken. 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.
- (Optional) 384 well plated ORF lentivirus (recommended vector: pLX_304 (Addgene, cat. no. 25890))
- (Optional) 384 well plated CRISPR lentivirus (recommended vector: pXPR_003 (Addgene, cat. no. 52963))
- (Optional) Positive control compounds for spike-in controls on viral plates. The following compounds were selected because they are known to introduce diverse, measurable phenotypes in the Cell Painting assay; adding a consistent set of diverse positive control compounds to each plate appears to improve the ability to align plates across experimental batches (Arevalo et al., manuscript in preparation): AMG-900, 10 mM in DMSO (Selleckchem, cat. no. S2719); LY2109761, 10 mM in DMSO (Selleckchem, cat. no. S2704); quinidine, 10 mM in DMSO (MedChemExpress, cat. no. HY-B1751); TC-S 7004 (Tocris, cat. no. 5088/10).
- (Optional) CellTiter-Glo (Promega, cat. no. G7573)
- 32% (wt/vol) paraformaldehyde (PFA), methanol free (Electron Microscopy Sciences, cat. no. 15740-S)
! CAUTION PFA is a very toxic chemical, and one must avoid inhalation and/or direct contact with skin and eyes. Handle in a fume hood and use proper gloves and a mask. Follow your institutional guidelines for using and discarding waste chemicals.
- Hank's balanced salt solution (HBSS; 10×; Invitrogen, cat. no. 14065-056)
- Triton X-100 (Sigma-Aldrich, cat. no. T9284)
- DMSO (Millipore Sigma, cat. no. D5879-100ML) **! CAUTION** 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.
- (Optional) Sodium azide (NaN_3 ; Sigma-Aldrich, cat. no. 106688) **! CAUTION** NaN_3 is a very toxic chemical, and one must avoid inhalation and/or direct contact with skin and eyes. Handle in a fume hood and use proper gloves and a mask. Follow your institutional guidelines for using and discarding waste chemicals.

Dyes

▲ CRITICAL Dyes can be obtained either as a kit or individually.

- PhenoVue Cell Painting kit (PerkinElmer, cat. no. PING22) **▲ CRITICAL** As in our prior version of the protocol³, ThermoFisher reagents performed comparably (Fig. 4b).
- Individual dyes: MitoTracker Deep Red (Invitrogen, cat. no. M22426) or PhenoVue 641 mitochondrial stain (PerkinElmer, cat. no. CP3D1), WGA/Alexa Fluor 555 conjugate (Invitrogen, cat. no. W32464) or PhenoVue Fluor 555 - WGA (PerkinElmer, cat. no. CP15551), phalloidin/Alexa Fluor 568 conjugate (Invitrogen, cat. no. A12380) or PhenoVue Fluor 568 - phalloidin (PerkinElmer, cat. no. CP25681), concanavalin A/Alexa Fluor 488 conjugate (Invitrogen, cat. no. C11252) or PhenoVue Fluor 488 concanavalin A (PerkinElmer, cat. no. CP94881), Hoechst 33342 (Invitrogen, cat. no. H3570) or PhenoVue Hoechst 33342 (PerkinElmer, cat. no. CP71), SYTO 14 green fluorescent nucleic acid stain (Invitrogen, cat. no. S7576) or PhenoVue 512 nucleic acid stain (PerkinElmer, cat. no. CP61), and PhenoVue Dye Diluent A 5×

Additional reagents when using ThermoFisher dyes

- Sodium bicarbonate (Sigma-Aldrich, cat. no. S571)
- Methanol (Honeywell, cat. no. 34966) **! 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 (Sigma-Aldrich, cat. no. 05470)

Equipment

- Microplates: PhenoPlate (previously Cell Carrier-384 Ultra) microplates, tissue culture treated, black, 384 wells with lid (PerkinElmer, cat. no. 6057300) **▲CRITICAL** Other microplates that are compatible with the microscope will suffice, as long as they are validated for use in high-content imaging (typically plates with minimal offset from the plate ‘skirt’ to the well bottom surface (e.g., ‘Low Base’ or ‘Ultra-Low Base’)). **▲CRITICAL** Order all 384-well plates from the same lot.
- (Optional) If performing CellTiter-Glo, 384-well, tissue culture-treated, white plate with clear, flat bottom, with lid, with barcode label (Corning, cat. no. 8793BC)
- T-175 culture vessel (Greiner Bio-One, cat. no. 661160)
- 50-ml Falcon centrifuge tubes (Corning, cat. no. 352070)
- 250-ml polypropylene centrifuge tubes with plug seal Cap (Corning, cat. no. 430776)
- Serological pipettes, various sizes (Greiner Bio-One)
- Aluminum single-tab foil, standard size (USA Scientific, cat. no. 2938-4100)
- Evaporation-resistant microplate lids or gas-permeable adhesive plate seals
- Tissue culture microscope (Olympus, model CKX41SF)
- Automated cell counter: Beckman Z1 particle counter (Beckman Coulter, Model Z1)
- Cytomat 5C tissue culture incubator at 37 °C, 5% CO₂ (Thermo Fisher Scientific, cat. no. 50128822) or Liconic Instruments Liconic incubator model STX-220 HR (with stacker and rotating plate carousel) or Thermo Heracell VIOS 160i CO₂ incubator at 37 °C, 5% CO₂ (Thermo Fisher Scientific, model Vios 160i)
- Automated liquid handler: Multidrop Combi reagent dispenser (Thermo Fisher Scientific, cat. no. 5840300), Freedom EVO with 384-channel arm (Tecan, cat. no. MCA384), Janus (PerkinElmer, model JANUSMPD) or BioTek EL406 washer dispenser (Agilent, cat. no. 406PSUB3SB). Pressurized valve dispensing systems (GNF Washer/Dispenser II (WDII) & OneTip Dispenser (OTD)) are more consistent than peristaltic systems (Combi/EL406)
- (Optional) Automated liquid handler metal tip cassette accessory (such as Agilent, cat. no. 7170014 for the EL406)
- 30- μ l filter tips, sterile for Janus (custom request from PerkinElmer)
- 235- μ l filter tips, sterile for Janus (PerkinElmer, cat. no. 6001289)
- Plate washer: Biotek ELx405 HT
- Centrifuge: Allegra 6 (Beckman Coulter, cat. no. 366802) or PlateFuge (Benchmark Scientific, cat. no. C2000)
- Centrifuge microplate carriers: uPlate Carrier for Rotor SX4750 (Beckman Coulter)
- High-content imager; see Tables 1–3 for microscopes used by the JUMP Consortium during these optimization experiments. Our partners also report that the Nikon Eclipse TI2, Yokogawa CV7000, Yokogawa CQ1 and PerkinElmer Operetta CLS microscopes have also been used successfully for imaging Cell Painting experiments. See ref. ²⁸ for more information.
- Plate shaker: Multi-purpose rotator (Barnstead Lab-Line, model 2314)
- (Optional) Plate reader: EnVision multilabel plate reader (PerkinElmer, model 21030010)

Computational equipment

- Standard desktop computer
- Access to a remote-host computing cluster or cloud-computing platform (optional; recommended if planning to acquire >1,000 fields of view; see Box 1 for details)
- CellProfiler biological image-analysis software²⁴ (free and open-source: <http://www.cellprofiler.org>)
- *CellProfiler pipelines*. We describe four pipelines in this protocol for illumination correction, segmentation, quality control (QC; optional) and feature extraction. The pipelines are available at https://github.com/broadinstitute/imaging-platform-pipelines/tree/master/JUMP_production and were created by using CellProfiler 4.1.3. Please see the module notes within the pipelines for Cell Painting-specific documentation. Our Cell Painting wiki (<https://broad.io/CellPaintingWiki>) 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 stored in the Cell Painting Gallery on the Registry of Open Data on Amazon Web Services (AWS) (<https://registry.opendata.aws/cellpainting-gallery/>) from chemical perturbations applied to U2OS cells. See download information at https://github.com/carpenter-singh-lab/2023_Cimini_NatureProtocols/blob/main/README.md. Note that this plate of images is much larger than we recommend running on a standard desktop computer. To use these data on a local computer, we recommend filtering down the images in CellProfiler to a single row, for example. This can be done by

Table 2 | Details of the PerkinElmer Opera Phenix 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	405	435–480	Nucleus	DNA
Concanavalin A/Alexa Fluor 488 conjugate	488	500–550	Endoplasmic reticulum	ER
SYTO 14 nucleic acid stain	488	570–630 (wide-field), 500–550 (confocal)	Nucleoli, cytoplasmic RNA	RNA
Phalloidin/Alexa Fluor 568 conjugate, WGA/Alexa Fluor 555 conjugate	561	570–630	F-actin cytoskeleton, Golgi, plasma membrane	AGP
MitoTracker Deep Red	640	650–760	Mitochondria	Mito

Table 3 | Details of the Yokogawa CV8000 channels and stains imaged in the Cell Painting assay

Dye	Filter (excitation; nm)	Filter (emission; nm)	Organelle or cellular component	Cellprofiler channel name	Recommended channel acquisition order
Hoechst 33342	405	445/45	Nucleus	DNA	5
Concanavalin A/Alexa Fluor 488 conjugate	488	525/50	Endoplasmic reticulum	ER	4
SYTO 14 nucleic acid stain	488	600/37	Nucleoli, cytoplasmic RNA	RNA	3
Phalloidin/Alexa Fluor 568 conjugate, WGA/Alexa Fluor 555 conjugate	561	600/37	F-actin cytoskeleton, Golgi, plasma membrane	AGP	2
MitoTracker Deep Red	640	676/29	Mitochondria	Mito	1

using the Images module. Replace the rule criterion ‘Extension Is the Extension of an Image File’ with ‘File Does Contain r01’, for example.

- Illumination correction images produced by an illumination-correction pipeline applied to the U2OS image data. See download information at https://github.com/carpenter-singh-lab/2023_Cimini_NatureProtocols/blob/main/README.md.
- A listing of per-cell image features generated by CellProfiler by using the analysis pipeline (https://github.com/carpenter-singh-lab/2023_Cimini_NatureProtocols/blob/main/CellProfiler_features.csv)
- A .csv of per-well profiles generated by running the profiling script, created from the provided sample images. See download information at https://github.com/carpenter-singh-lab/2023_Cimini_NatureProtocols/blob/main/README.md.
- A Python²⁹ or Conda installation with pycytominer³⁰ and cytominer-database³¹ installed per repository instructions
- (Optional) KNIME data analytics software³² with HCS Tools extension³³ and sample workflow (see Box 2 for details)
- (Optional) A graphical user interface (GUI)-based SQLite reader, such as DB Browser for SQLite (<https://sqlitebrowser.org/>)
- (Optional) Plate reader software, such as PerkinElmer EnVision Manager
- (Optional) Data analysis software, such as PRISM from GraphPad Software

Reagent setup

Reagent setup for using PhenoVue Cell Painting JUMP kit PING 22

PhenoVue 641 mitochondrial stain

The product from PerkinElmer (cat. no. PING22) contains 50 µg/vial. Add 90 µl of DMSO to one vial to make a 1 mM solution. Spin down at 3,200g for 10 s. Store the solution at –20 °C, protected from light, and use it over three freeze–thaw cycles. **!CAUTION** 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.

Box 1 | Configuration of the pipelines for batch processing on a computer cluster

We recommend using a computing cluster, such as a high-performance server farm, or a cloud-computing platform, such as AWS, 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 by 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.

Initial preparation to run on a cluster or cloud requires expert setup and will probably require enlisting the help of your IT department or local cloud-computing expert. Please refer to our Distributed CellProfiler GitHub repository documentation (<https://distributedscience.github.io/Distributed-CellProfiler/overview.html>) for more details on cloud computing using AWS and our command line startup guide <https://github.com/CellProfiler/CellProfiler/wiki/Getting-started-using-CellProfiler-from-the-command-line> for more details on cluster computing. Additional information on batch processing on a cluster or in the cloud is available on our video Headless CellProfiler/DistributedCellProfiler Tutorial (<https://youtu.be/LuJxIGGhRek>).

When running batch processes, we recommend exporting data to spreadsheets because it is easier to aggregate a large number of spreadsheets than databases. Uncheck or remove any ExportToDatabase modules in your pipelines and add instead ExportToSpreadsheet modules.

There are two ways to create batch information for CellProfiler: LoadData and CreateBatchFiles.

LoadData: LoadData is our preferred method, especially for use with Distributed CellProfiler and cloud computing. Detailed instructions can be found at <https://cytominer.github.io/profiling-handbook/>.

- 1 Insert the LoadData module into the pipeline by pressing the '+' button and selecting the module from the 'File Processing' category. Move this module to the beginning of the pipeline and configure it to find your file.
- 2 Create LoadData.csvs. If you are using a PerkinElmer microscope, you can use our pe2loaddata script, available at <https://github.com/broadinstitute/pe2loaddata>, to create the LoadData.csv file. Alternatively, if you have loaded your images into CellProfiler by using the Input modules, you can export a file list from CellProfiler by using *File* → *Export* → *Image Set Listing*.
- 3 The end result of this step will be a LoadData.csv file. This file, when used in conjunction with a CellProfiler pipeline (.cppipe), contains the information needed to run in 'headless' mode on the cluster or in the cloud.

CreateBatchFiles: If you have already created a CellProfiler project (.cpproj) with fully populated Input modules, you can create a batch file as follows:

- 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.
- 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 file images by mounting the file system by using a drive letter (e.g., C:\your_data\images), and the cluster computers access the same file system by 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 or in the cloud. 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.

PhenoVue Fluor 555 – WGA

The product from PerkinElmer (cat. no. PING22) contains 0.2 mg/vial. Add 1.3 ml of dH₂O to each to make a 0.15 mg/ml solution. Use a P1000 tip to break up any precipitates in the solution. Store the solution at –20 °C, protected from light, and use it over three freeze–thaw cycles.

PhenoVue Fluor 488 – concanavalin A

The product from PerkinElmer (cat. no. PING22) contains 1 mg/vial. Add 0.5 ml of dH₂O to each vial to make a 2 mg/ml solution. Use a P1000 tip to break up any precipitates in the solution. Store the solution at –20 °C, protected from light, and use it over three freeze–thaw cycles.

PhenoVue Fluor 568 – phalloidin

The product from PerkinElmer (cat. no. PING22) contains 1 nmol/vial. Add 150 µl of DMSO to each vial to make a 6.6 µM stock solution. Spin down at 3,200g for 10 s. Store the solution at –20 °C, protected from light, and use it over three freeze–thaw cycles.

PhenoVue 512 nucleic acid stain

The product from PerkinElmer (cat. no. PING22) contains 800 nmol/vial. It is a 5 mM solution in DMSO. Spin down at 3,200g for 10 s. Store the solution at –20 °C, protected from light, and use over three freeze–thaw cycles.

PhenoVue Hoechst 33342 nuclear stain

The product from PerkinElmer (cat. no. PING22) contains 140 µg/vial. It is a 1 mg/ml solution in H₂O. Spin down at 3,200g for 10 s. Store the solution at –20 °C, protected from light, and use over three freeze–thaw cycles.

Box 2 | Quality Control (QC) with CellProfiler

Although quantifying the quality of your images is not strictly necessary to obtain profiles, you may find it critical to ensuring that your data is of high enough staining and imaging quality to measure your phenotypes of interest; running it alongside each batch of data can help discover workflow issues that can be addressed before subsequent runs. You can integrate these quality measurements into your analysis pipeline, or you can run them as a separate pipeline before analysis, as we have in this protocol. We have found that the per-image cell count and the per-image median intensity are particularly well suited to use as QC metrics.

- 1 Start CellProfiler.
- 2 Load the QC pipeline into CellProfiler by selecting *File* → *Import* → *Pipeline from File* from the CellProfiler main menu and then selecting qc.cppipe or dragging and dropping the pipeline from your file browser into the pipeline panel on the left of the interface.
- 3 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 Computational equipment for a link to raw image files that can be used as an example in this protocol.
- 4 Click the 'Output settings' button at the bottom-left of the interface. Select an appropriate 'Default Output Folder' into which the QC measurements will be saved.
- 5 Save the current settings to a project (.cproj) file containing the pipeline, the list of input images and the output location by selecting *File* → *Save Project*. Enter the desired project file name in the dialog box that appears. This is not necessary for the running of this step; however, it functions as a snapshot of your complete setup, allowing you to directly replicate your work by loading this .cproj file instead of the .cppipe file with which you started (which contains the pipeline only).
- 6 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.
- 7 Visualize the output from the CellProfiler QC pipeline; our recommendation is to use KNIME, an open-source data analytics tool.
 - Download and install the current version of KNIME (<https://www.knime.com/downloads>).
 - Within the KNIME client, install the HCS Tools extension (<https://www.knime.com/community/hcs-tools>).
 - Download the KNIME workflow JUMP_QC_Plate-CV_v1.knwf from https://github.com/broadinstitute/imaging-platform-pipelines/tree/master/JUMP_production.
- 8 Run KNIME and load KNIME workflow:
 - Right-click the CSV Reader Node 'Top Line Per-Image', select 'Configure...' and set the File path to the TopLineImage.csv output from the CellProfiler workflow.
 - Right-click the CSV Reader Node 'Top Line Per-Object', select 'Configure...' and set the File path to the TopLineCells.csv output from the CellProfiler workflow.
 - Run the KNIME workflow (green, double-arrow button in the menu bar).
 - Inspect the plots and right-click the final nodes (right-click and select 'View: [Line Plot | Plate Heatmap Trellis]').
 - If the coefficient of variation (CV) of the cell count and/or the median intensity of any channels is >15% in ≥10% of the DMSO wells, nominate the plate for rejection.

PhenoVue Dye Diluent A (5×)

The product from PerkinElmer (cat no. PING22) contains 80 ml of a 5× stock solution. Dilute five fold in dH₂O to create a 1× HBSS solution with 1% (wt/vol) BSA. The 1× solution should preferably be made fresh but may be stored at −4 °C for ≤2 d.

Triton X-100 solution in HBSS

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

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×). The 1× solution should preferably be made freshly from the 10× stock solution, but it can also be stored at 4 °C for ≤1 week. If storing, filter the HBSS (1×) with a 0.22-µm filter.

Live-cell mitochondrial staining solution

Prepare mitochondrial staining solution by adding 150 µl of PhenoVue 641 mitochondrial 1 mM stock solution to 100 ml of prewarmed cell culture medium (preferred method) or HBSS for a working concentration of 1.5 µM to end with a final staining concentration of 500 nM. Make fresh solution for each staining session. We recommend not increasing the concentration of the staining solution any further, to avoid small changes in dispensing volume leading to high variability in final effective concentration. Keep the solution wrapped in foil and away from light.

Staining and permeabilization solution

To make 100 ml of stain solution, add 125 µl of 6.6 µM phalloidin stock solution, 250 µl of 2 mg/ml concanavalin A stock solution, 100 µl of 1 mg/ml Hoechst stock solution, 1 ml of 0.15 mg/ml WGA stock solution and 120 µl of 5 mM 512 nucleic acid stain stock solution in 1× HBSS (1% (wt/vol) BSA and 0.1% (vol/vol) Triton X-100). Keep the solution wrapped in foil and away from light. Make fresh solution for each staining session.

Fixation solution

To create a 16% (wt/vol) solution of PFA, dilute 32% (wt/vol) PFA 1:1 with distilled water. This solution is ideally made fresh but can be stored for up to a few weeks at room temperature (RT; 21 °C)

in a fume hood. **! CAUTION** PFA is a very toxic chemical, and one must avoid inhalation and/or direct contact with skin and eyes. Handle in a fume hood and use proper gloves and a mask. Follow your institutional guidelines for using and discarding waste chemicals.

NaN₃ solution

To create a 10% (wt/vol) NaN₃ stock solution, dissolve 5 g of NaN₃ in 50 ml of MilliQ water; this stock solution can be stored at RT for ≤1 month. To prepare final 0.05% (wt/vol) NaN₃ solution, dilute 10% (wt/vol) NaN₃ stock solution 1:200 in 1× HBSS; this solution should be made fresh.

! CAUTION NaN₃ is a very toxic chemical, and one must avoid inhalation and/or direct contact with skin and eyes. Handle in a fume hood and use proper gloves and a mask. Follow your institutional guidelines for using and discarding waste chemicals.

Reagent setup for perturbation treatment

Reagent setup for compound library

Dissolve the compounds in DMSO to yield the desired molarity. Do not exceed a final DMSO concentration of 0.5% (vol/vol) in the destination well. Seal and store it at −20 °C for long-term storage or at RT for ≤6 months in a dessicator; other common compound management solutions may also be used³⁴.

Reagent setup for lentiviral transduction

Growth medium

To 500 ml of DMEM or McCoy's base medium (consult the source of your cell line for the optimal medium) add 50 ml of FBS and 5 ml of pen-strep. Growth medium can be stored for 2–3 weeks at 4 °C.

Polybrene stock solution

Dissolve polybrene powder in sterilized, ultrapure water at a concentration of 8 mg/ml, pass through a 0.22-μm filter and then freeze in small quantities at −20 °C until needed. Avoid freeze–thaw cycles; rather, keep at 4 °C once thawed for ≤2 weeks.

Polybrene medium solution

Prepare fresh growth medium containing 16 μg/ml polybrene from a polybrene stock of 8 mg/ml on the basis of the total volume needed to seed 10 μl per 384 wells. Different cell types will require varying concentrations of polybrene for aid in lentiviral transduction, and varying concentrations should be assessed for cell toxicity; see Supplementary Methods.

(Optional) Blasticidin medium solution

This solution is for use with lentiviral expression vectors containing a blasticidin resistance cassette. Prepare fresh growth medium containing a final concentration of 16 μg/ml blasticidin from a blasticidin stock of 10 mg/ml on the basis of the total volume needed to seed 50 μl per 384 wells. Different cell types will require varying concentrations of blasticidin for complete antibiotic selection; see Supplementary Methods.

(Optional) Puromycin medium solution

This solution is for use with lentiviral expression vectors containing a puromycin resistance cassette. Prepare fresh growth medium containing a final concentration of 0.75 μg/ml puromycin from a puromycin stock of 10 mg/ml on the basis of the total volume needed to seed 50 μl per 384 wells. Different cell types will require varying concentrations of puromycin for complete antibiotic selection; see Supplementary Methods.

Control compound spike-ins

Initially prepare 10 mM TC-S 7004 stock in DMSO from powder. Distribute 5-μl aliquots of 10 mM stocks of compounds into PCR strip tubes for single use to avoid freeze–thaw cycles; then, store AMG-900, LY2109761 and TC-S 7004 at −20 °C and quinidine at −80 °C. For manual spike-in to virally treated 384-well plates, add freshly on each addition day 1 μl of 10 mM DMSO stock + 1 μl of 8 mg/ml polybrene to 498 μl of growth medium and mix; the final concentration is 20 μM. Then, add 10 μl of 20 μM compound stock to each 384-well plate seeded at 30 μl/well for a final compound concentration of 5 μM.

(Optional) CellTiter-Glo

Bring to RT the CellTiter-Glo reagent components, including the CellTiter-Glo substrate and CellTiter-Glo buffer. Once the buffer and substrate have come to RT, aspirate the total volume of buffer present in the tube (i.e., 10, 100 or 500 ml), depending on the size kit that was purchased. Dispense that volume into the container containing the powdered substrate. Mix by gently pipetting up and down until the powder has thoroughly dissolved. It is recommended to then cover the container or transfer to another container covered in foil, because the reagents are light sensitive. Mixed solution can be refrozen at −20 °C and thawed up to 5–10 times.

Equipment setup

Microscope selection

Cell Painting assay samples have been imaged by using both wide-field and confocal microscopy. Confocal microscopes 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, 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 three high-content imaging systems in the pilot experiments described here: an ImageXpress Micro XLS (Molecular Devices), an Opera Phenix (PerkinElmer) and Yokogawa CV8000. The images are captured in five fluorescent channels given in Tables 1–3. Note that although both the Opera Phenix and the ImageXpressMicro XLS are capable of imaging in both wide-field and confocal modes, while in confocal mode, the Opera Phenix outputs only four channel images rather than five because it cannot adjust its emission filter such that SYTO 14 can be captured separately from Alexa 488. 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 substantial, then post-acquisition normalization will probably be effective (see Creation, normalization and feature reduction of per-well profiles in the Procedure for more details). However, if the morphological effects to be measured are subtle, normalization may not be sufficient, and the similarities in the collected image features will more likely reflect the different image acquisition than the underlying biological perturbations, significantly increasing the technical difficulty of the analysis.

Automated image-acquisition settings

Each channel should be captured as an individual grayscale image. No further pre-processing should be performed on the images before analysis. The choice of objective magnification is important because there is a trade-off between increased image feature resolution at higher magnifications (therefore enabling more specific quantification of certain organelles) and 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 (numerical aperture: 1.0) sufficiently balances all competing issues. Typically, nine sites/fields of view are collected per well in a 3 × 3 site layout, at 20× magnification and 2× binning. If time permits, more sites can be imaged to increase well coverage and improve sample statistics; it is best to capture as many cells as possible.

If possible on your microscope, adjusting the relative acquisition heights of each fluorescent channel may lead to optimal capture of each channel's relevant cell structures. The final ORF overexpression data produced by the JUMP Consortium after the pilots detailed in this manuscript had 2 μm of total Z difference between channels captured at the lowest Z position and channels captured at the highest Z position.

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. Because the emission wavelengths for the chosen fluorophores are broad and in close proximity to each other, photobleaching may occur for the low-intensity dyes because they are irradiated by the lower-wavelength light. To mitigate this effect, we recommend imaging the five channels in order of decreasing excitation wavelength. For Opera Phenix instruments equipped with more than one camera, we recommend separating all channels.

Do not use shading correction if you are using the recommended CellProfiler workflow for image analysis, because the background illumination heterogeneities will be corrected after acquisition by 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, to confirm that the microscope is operating as expected and that the acquisition settings are optimal for the experiment and cell line at hand. See Computational equipment for a link to an example image data set. **▲ 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.

We recommend including the acquisition of at least one brightfield channel, which can be used in feature extraction but may also be useful for downstream applications such as fluorescence channel prediction. The JUMP ORF production captured three Z positions: one equal to the lowest fluorescence position, one 5 μm above that position and one 5 μm below that position. **▲ 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 after acquisition by using image-processing approaches; this adds a step and is best avoided.

Image processing software

CellProfiler biological image analysis software is used to extract per-cell morphology feature data from the Cell Painting images and can optionally be used to extract per-image QC metrics (see Box 2). 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 4.2.1.

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 4.1.3 and above.

This protocol uses three CellProfiler pipelines to perform the following tasks: illumination correction, segmentation and morphological feature extraction. In addition, you can use an optional QC pipeline to quantify image quality. See Computational equipment 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 PerkinElmer Opera Phenix system, in which the plate/well/site metadata are encoded as part of the file name. The plate and well metadata in particular are essential because CellProfiler uses the plate metadata 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 or our video tutorial walkthrough of the Input modules (https://youtu.be/Z_pUWuOV06Q).

The QC and morphological feature extraction pipelines are set to write out cellular features to a .csv for each well and site, respectively, by using the ExportToSpreadsheet module. We provide Python²⁹ scripts to generate per-well profiles from the extracted features.

Computing system

Small batches of images can be processed on most modern desktop computers. If the number of images to analyze is sufficiently large (e.g., more than ~1,000 images), processing time on a single computer becomes resource limiting. For large batches of images, we recommend using a computing cluster if available, such as a high-performance server farm or a cloud-computing platform such as AWS. Substantial setup effort is required for both cluster computing and cloud-distributed computing and will probably require enlisting the help of your information technology (IT) department. Please refer to our GitHub webpage (<https://github.com/CellProfiler/CellProfiler/wiki/Adapting-CellProfiler-to-a-LIMS-environment>) for more details on cluster computing and our Distributed CellProfiler GitHub repository documentation (<https://distributedscience.github.io/Distributed-CellProfiler/overview.html>) for more details on cloud computing using AWS.

Procedure

Cell culture ● Timing Variable; 2–7 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, ORF overexpression and CRISPR knockdown.

▲ **CRITICAL** Check the wiki at GitHub for any updates to the Cell Painting protocol: <https://broad.io/CellPaintingWiki>.

▲ **CRITICAL** 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 U2OS cells in Corning 384-well, 200-nm-thick glass-bottom plates. (Optional) White plates with a clear, flat bottom for cell culture are also appropriate for any cell viability assay using CellTiter-Glo.

- 1 Grow cells to near confluence (~80%) in a T-175 tissue culture flask.
- 2 (Optional) If you are performing experiments that involve the addition of compounds (Step 10A), prepare the compound library according to the instructions in Reagent setup for compound treatment. If you are performing experiments that involve the overexpression or CRISPR knockdown of genes via lentiviral transduction (Step 10, C and B, respectively), prepare these reagents per Reagent setup for lentiviral transduction.
- 3 Aspirate medium and rinse the cells with enough PBS without Ca^{2+} or Mg^{2+} to cover the cells (~10 ml).
- 4 Add 3 ml of trypsin-EDTA solution and incubate the cells at 37 °C until the cells have detached. This should occur within 3–5 min.
- 5 Add 4 ml of growth medium to deactivate the trypsin and collect the cell suspension into a conical tube.
- 6 Wash the tissue culture flask with an additional 5 ml of growth medium, add the wash to the same conical tube with the cell suspension and mix thoroughly but gently.
- 7 Determine the live-cell concentration by using standard methods (hemocytometer or cell counter).
- 8 Dilute U2OS cells to 50,000 live cells per ml in medium and dispense 30 μl (2,000 live U2OS 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. (Optional) Dispense cells into white plates if any CellTiter-Glo assays are to be performed.

▲ **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 multi-dispense function, be sure to adequately sterilize and prime the dispensing cassette and/or dispense $\geq 10 \mu\text{l}$ 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 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, to ensure accurate liquid dispensation.

? TROUBLESHOOTING

- 9 Keep cell plates at RT for ~1–2 h before proceeding with the small molecule, viral ORF overexpression or viral CRISPR knockdown, because this reduces plate edge effects³⁵. If not proceeding directly to compound addition or lentiviral transduction, after the initial RT period, transfer plates into the incubator (37 °C, 5% CO_2 , 90–95% humidity).

▲ **CRITICAL STEP** To reduce plate edge effects produced by incubator temperature variations and medium evaporation, during all 37 ° incubation steps, we recommend stacking plates no more than three plates high and 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 every 24 h to avoid positional effects.

Treatment with a small-molecule library, viral ORF overexpression library or viral CRISPR knockdown library

- 10 If you are performing treatments with a small-molecule library, please follow option A. If you are performing viral CRISPR knockdown, please follow option B. If you are performing viral ORF overexpression, please follow option C. For instructions on using siRNA transfection, please refer to the 2016 protocol. To facilitate alignment of data across batches, no matter which modality you use for your own data, we recommend in each batch of Cell Painting plates that you prepare both an additional negative control plate (using only DMSO as the treatment for cells) and an additional

positive control plate to be handled per option A (we recommend a control compound plate such as the JUMP-Target-1 or JUMP-Target-2 plates).

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

- (i) Add compounds to cells by using a pin tool or other liquid handler (e.g., the Agilent Bravo automated pipetting platform or the Beckman Coulter Echo acoustic liquid-handling system). Compounds may be added either 24 or 48 h before staining and fixation, but this timing should be adjusted depending upon the growth rate of the cell types being used and the biological processes under consideration, because the ability to ‘catch’ a particular desired phenotype at the right time is much more important than an exact match to our protocol; per Fig. 3b; Extended Data Fig. 2b,c; and Extended Data Fig. 3a, matching of 24-h compound treatment to 48-h compound treatment (top left sub-panel, red dot(s)) is typically quite good. 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).
- (ii) Place the cell plates in an incubator (37 °C, 5% CO₂).

▲ **CRITICAL STEP** Ensure that the liquid handler aspiration and dispensation speeds are set to the slowest possible setting, to avoid distributing the cells after the initial seeding.

(B) Addition of viral CRISPR knockdown constructs ● Timing Variable; ~4–6 d for one batch experiment of 384-well plates

- (i) CRISPR knockdown lentivirus should be pre-prepared in 384-well plates at –80 °C. Let plates thaw on wet ice and then quickly pulse-spin to prevent any virus from clinging to the plate seal. Place back on ice until needed.
- (ii) Using a liquid handler, add 10 µl of growth medium containing polybrene into each well of the 384-well plate that will receive a viral perturbation as described in Reagent setup.
▲ **CRITICAL STEP** Ensure that the liquid handler aspiration and dispensation speeds are set to the slowest possible setting, to avoid distributing the cells after the initial seeding.
- (iii) Immediately after polybrene addition, add the appropriate volume of CRISPR knockdown lentivirus per 384-well plate by using a liquid handler. Different cell types and lentiviral expression vectors will require variations in lentivirus volume; typical ranges will vary from 0.5 to 4 µl per well (see the Supplementary Methods CP-JUMPC section to show how to determine the correct amount for a given lentivirus).
- (iv) (Optional) Control compounds can be spiked into wells that do not contain CRISPR knockdown perturbations if desired. See Reagent setup.
- (v) Gently tap the plates to ensure an even distribution of cells across each well.
- (vi) After lentiviral transduction, centrifuge the 384-well plate(s) for 30 min at 1,178g and 37 °C.
- (vii) After centrifugation, place the cell plates in an incubator (37°C, 5% CO₂).
- (viii) 24 h after lentiviral transduction, remove 50 µl of medium and replace it with 50 µl of fresh growth medium and return to the incubator. (Optional) If assessing the efficiency of the lentiviral transduction as a QC measure, for two white cell plates, replace the growth medium in one with 50 µl of medium and the other with 50 µl of medium containing puromycin as described in Reagent setup.
- (ix) The timing after viral CRISPR knockdown transduction and before downstream assays may be either 96 or 144 h, but this timing should be adjusted depending upon the growth rate of the cell types being used and the biological processes under consideration. If not performing Cell Painting or optional CellTiter-Glo until 144 h after perturbation, at 96 h after initial viral CRISPR knockdown transduction, remove 50 µl of medium and replace it with 50 µl of fresh growth medium or (optional) 50 µl of medium containing puromycin in a white cell plate and return the plate to the incubator. After the predetermined number of hours, proceed with the appropriate subsequent steps for cell fixation, staining and imaging or (optional) determining the lentiviral transduction efficiency.
- (x) (Optional) The transduction efficiency is determined by adding 10 µl of RT CellTiter-Glo per 384-well plate to two white cell plates, one with puromycin treatment and one without puromycin treatment. Cover the plates with aluminum foil and put them on a shaker at low speed for 15 min. Then, image the plates by using a standard plate reader such as an Envision multilabel reader (PerkinElmer).

(C) **Addition of viral overexpression constructs** ● **Timing Variable; ~2–4 d for one batch experiment of 384-well plates**

- (i) ORF overexpression lentivirus should be pre-prepared in 384-well plates at -80°C . Let thaw on wet ice and then quickly pulse-spin the plate to prevent any virus from clinging to the plate seal. Place back on ice until needed.
- (ii) Using a liquid handler, add 10 μl of growth medium containing polybrene into each well of the 384-well plate that will receive a viral perturbation, as described in Reagent setup.
▲ **CRITICAL STEP** Ensure that the liquid handler aspiration and dispensation speeds are set to the slowest possible setting, to avoid distributing the cells after the initial seeding.
- (iii) Immediately after polybrene addition, add the appropriate volume of ORF overexpression lentivirus per 384-well plate by using a liquid handler. Different cell types and lentiviral expression vectors will require variations in lentivirus volume; typical ranges will vary from 0.5 to 4 μl per well.
- (iv) (Optional) Control compounds can be spiked into well(s) that do not contain ORF overexpression perturbations if desired. See Reagent setup.
- (v) Gently tap the plates to ensure an even distribution of cells across each well.
- (vi) After lentiviral transduction, centrifuge the 384-well plate(s) for 30 min at 1,178g and 37°C .
- (vii) After centrifugation, place the cell plates in an incubator (37°C , 5% CO_2).
- (viii) 24 h after lentiviral transduction, remove 40 μl of medium and replace it with 40 μl of fresh growth medium and return to the incubator. (Optional) If assessing the efficiency of the lentiviral transduction as a QC measure, for two white cell plates, replace the growth medium in one with 50 μl of medium and the other with 50 μl of medium containing blasticidin as described in Reagent setup.
- (ix) The timing after viral ORF overexpression transduction and before downstream assays may be either 48 or 96 h, but this timing should be adjusted depending upon the growth rate of the cell types being used and the biological processes under consideration. After the predetermined number of hours, proceed with the appropriate subsequent steps for cell fixation, staining and imaging or (optional) determining the lentiviral transduction efficiency.
- (x) (Optional) The transduction efficiency is determined by adding 10 μl of RT CellTiter-Glo per 384-well plate to two white cell plates, one with blasticidin treatment and one without blasticidin treatment. Cover the plates with aluminum foil and put them on a shaker at low speed for 15 min. Then, image the plates by using a standard plate reader such as an Envision multilabel reader (PerkinElmer).

Staining and fixation ● **Timing Variable; ~2.5–3 h for one batch experiment of 384-well plates**

- 11 Prepare the following for all plates in advance of initiating the staining process:
 - The live-cell PhenoVue 641 mitochondrial staining solution in cell medium or HBSS
 - The fixation solution containing 16% (wt/vol) PFA in distilled water
 - The staining and permeabilizing solution containing PerkinElmer PhenoVue dyes Hoechst 33342, Fluor 488 concanavalin A, 512 nucleic acid stain, Fluor 555 WGA and Fluor 568 phalloidin in 1× PhenoVue Dye Diluent A with 0.1% (vol/vol) Triton
- 12 Leaving growth medium in place to minimize disturbance to the live cells, add 20 μl of the mitochondrial staining solution over the top of each well to a final well volume of 60 μl .
 - Place a stir bar in the source bottle to prevent dye from settling and causing adverse plate patterns.
 - We recommend using a metal tip cassette to prevent bubbling and allow for a more even dispensing of dye.
 ▲ **CRITICAL STEP** If using a liquid handler such as a Combi multidrop, be aware that dye can settle in tubing if there is a significant lapse of time between dispensings (>2 min/plate). This may cause adverse plate patterns. We recommend introducing a repriming step to clear the dyes from the tube and introduce fresh dye from the source bottle. If plate patterns persist, the addition of an initial blank plate (no cells present) may alleviate the intensity of patterns.
- 13 If necessary, centrifuge the plate (500g at RT for 1 min) after adding stain solutions to ensure that there are no bubbles in the bottoms of the wells.
- 14 Incubate the plates for 30 min in the dark at 37°C .
▲ **CRITICAL STEP** Once mitochondrial staining solution is added, keep the plates in the dark for the rest of the experiment.

- 15 To fix the cells, add 20 μ l of 16% (wt/vol) methanol-free PFA on top of each well to bring each well to a final volume of 80 μ l with final concentration of 4% (wt/vol) PFA.
 - 16 Incubate the plates in the dark at RT for 20 min.
 - 17 Wash the plates four times with 80 μ l of 1 \times HBSS. Include a final aspiration step.
 - 18 To each well, add 20 μ l of the staining and permeabilizing solution (see Reagent setup).
 - Place a stir bar in the source bottle to prevent dye from settling and causing adverse plate patterns.
 - If experiencing bubbling, we recommend using a metal-tipped cassette.
 - 19 Incubate the plates in the dark at RT for 30 min.
 - 20 Wash the plates four times with 80 μ l of 1 \times HBSS. Leave a final volume of 80 μ l of 1 \times HBSS in each well.
- ? TROUBLESHOOTING**
- 21 Seal the plates with adhesive foil and store them at 4 °C in the dark.

! CAUTION NaN_3 may cause damage to organs through prolonged or repeated exposure and is fatal if swallowed, if it comes in contact with skin or if it is inhaled.

■ PAUSE POINT Plates can be stored for the long term, with 0.05% (wt/vol) NaN_3 added to mitigate contamination by replacing the final wash step above with 60 μ l of 0.05% (wt/vol) NaN_3 in HBSS, followed by a brief centrifugation before sealing as above.

Automated image acquisition ● Timing Variable; 1–3.5 h per 384-well plate

- 22 Acquire images from the microtiter plates by using the high-content imager. For large-scale Cell Painting assays, we recommend the use of an automated microplate handling system.
- 23 Set up the microscope acquisition settings as described in Equipment setup.
- 24 Start the automated imaging sequence according to the microscope manufacturer's instructions.

? TROUBLESHOOTING

Morphological image feature extraction from microscopy data ● Timing Variable; 24 h–1 week per batch of 384-well plates

Perform illumination correction to improve fluorescence intensity measurements

- 25 Start CellProfiler.
- 26 Load the illumination correction pipeline into CellProfiler by selecting *File* \rightarrow *Import* \rightarrow *Pipeline* from *File* from the CellProfiler main menu and then selecting illumination.cppipe or dragging and dropping the pipeline from your file browser into the pipeline panel on the left of the interface.

▲ 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 image-based 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 recommended³⁶; the illumination pipeline takes this approach.
- 27 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 Computational equipment for a link to raw image files that can be used as an example in this protocol.
- 28 Click the 'Output settings' button at the bottom-left of the interface. Select an appropriate 'Default Output Folder' into which the illumination correction images will be saved.
- 29 Save the current settings to a project (.cpproj) file containing the pipeline, the list of input images and the output location by selecting *File* \rightarrow *Save Project*. Enter the desired project file name in the dialog box that appears. This is not necessary for the running of this step; however, it functions as a snapshot of your complete setup, allowing you to directly replicate your work by loading this .cpproj instead of the .cppipe with which you started (which contains the pipeline only).
- 30 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, with one image created for each plate and channel. We have provided an example set of images for comparison on our Cell Painting wiki (see Computational 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, such as a high-performance server farm, or a cloud-computing platform, such as AWS, 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 1, describing modifications to the pipeline to run it as a batch process.

Configure the segmentation pipeline to optimize cell and nuclei segmentation

- 31 Start CellProfiler, if you are not already running it.
- 32 Load the Segmentation pipeline into CellProfiler by selecting *File* → *Import* → *Pipeline* from *File* from the CellProfiler main menu and selecting segmentation.cppipe or dragging and dropping the pipeline from your file browser into the pipeline panel on the left of the interface.
- 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 both your raw images and your illumination-correction images generated in Step 30 into the 'File list' panel.
- 34 Enter CellProfiler's Test mode by using the 'Start Test Mode' button. 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 both negative control wells and 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. The CellProfiler website contains resources and tutorials on how to optimize an image-analysis pipeline.

▲ CRITICAL STEP Because capturing subtle phenotypes is important for profiling, accurate nuclei and cell body identification is essential for success. This pipeline enables troubleshooting of nuclei and cell segmentation across an entire batch by outputting an image per well with nuclei and cell objects overlaid for visual inspection. The optimal parameters determined in this pipeline must be manually transferred to the analysis pipeline. Although effort has been made to make the provided pipelines as robust as possible, segmentation must always be assessed for each experiment and may need to be adjusted on a batch-to-batch basis.

? TROUBLESHOOTING

- 35 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 images with nuclei and cell segmentations overlaid on them in a 'Segmentations' folder in the Default Output Folder. We have provided an example set of images for comparison on our Cell Painting wiki (see Computational equipment for details).

▲ CRITICAL STEP This step assumes that you will be running the segmentation pipeline locally on your computer. If your institution has a shared high-performance computing cluster, such as a high-performance server farm, or a cloud-computing platform, such as AWS, 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 1, describing modifications to the pipeline to run it as a batch process.

- 36 Visually inspect your output segmentation images. Open the segmented images in your preferred image viewer (we recommend Fiji/ImageJ³⁷) and decide if the computationally determined Nuclei and Cell objects correspond to what you see with your eyes across your batch. Perfection is nigh impossible, but you should agree with the vast majority of segmentations. A rule of thumb is that you should see roughly the same number of objects over-segmented (one object labeled as two or more) as under-segmented (two or more objects labeled as one). If not satisfied with the segmentation, repeat the segmentation pipeline workflow with different segmentation parameters. To simplify image visualization, particularly in larger datasets, you may wish to create whole-plate montages of your images by using methods described by Dobson et al.²¹; images with annotations can be saved as SVG files via a Fiji plugin³⁸. The Anticipated results section outlines the expected nuclei and cell identification quality.

Run QC to extract image-quality measurements

- 37 (Optional) See Box 2 to run this step.

Run the final analysis pipeline to extract morphological features

- 38 Start CellProfiler, if you are not already running it.
- 39 Load the analysis pipeline into CellProfiler by selecting *File* → *Import* → *Pipeline from File* from the CellProfiler main menu and selecting analysis.cppipe or dragging and dropping the pipeline from your file browser into the pipeline panel on the left of the interface.
- 40 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. This may be slow for large numbers of images.
- 41 If you needed to change segmentation parameters in your Segmentation pipeline for accurate object identification, make those same changes to the IdentifyPrimaryObjects and IdentifySecondaryObjects modules for the identification of NucleiIncludingEdges and CellsIncludingEdges, respectively.
- 42 If you did not run a separate QC pipeline and would like to measure image-level quality metrics, check the two MeasureImageQuality modules at the beginning of the pipeline to include those measurements in the analysis pipeline.
- 43 Click the 'Output settings' button at the bottom-left of the interface. Select an appropriate 'Default Output Folder' into which the analysis data will be saved.
- 44 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 file name in the dialog box that appears. This is not necessary for the running of this step; however, it functions as a snapshot of your complete setup, allowing you to directly replicate your work by loading this .cpproj file instead of the .cppipe file you started with (which contains the pipeline only).
- 45 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 pipeline will identify the nuclei from the Hoechst-stained image (referred to as 'DNA' in CellProfiler), then use the nuclei to guide identification of the cell boundaries by using the SYTO 14-stained image ('RNA' in CellProfiler) and then use both of these features to identify the cytoplasm. The pipeline then measures the morphology, intensity, texture and adjacency statistics of the nuclei, cell body and cytoplasm, and it outputs the results to a series of .csv files. See Equipment 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, such as a high-performance server farm, or a cloud-computing platform, such as AWS, 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 1, which describes modifications to the pipeline to run it as a batch process.

Creation, normalization and feature reduction of per-well profiles ● Timing Variable; 8–16 h per 384-well plate

▲ **CRITICAL** All steps in this section are covered in more detail and are continuously updated in the profiling handbook³⁹.

- 46 Set up a terminal for bash scripting as well as a language in which you are comfortable writing scientific code; most of our tooling is in Python²⁹ because it has a large number of useful packages that you can use to write^{40–46} or execute⁴⁷ code per our templates, but it is not mandatory.
- 47 Create per-well mean aggregated profiles from the individual cell measurements from each site within the well by using the pycytominer³⁰ command 'collate.py'. If you have run your CellProfiler pipeline in a cluster-computing environment, you can also use this same script to aggregate multiple CSV files into a single SQLite database before aggregation by using the cytominer-database³¹ package. This step can take 8–16 h per plate to run; if running multiple plates, using a parallel⁴⁸ or similar tool is strongly recommended to parallelize profile creation.
- 48 Create a metadata file describing the treatments on each plate of your experiment. Instructions for creating this file are available as part of the profiling recipe⁴⁹ and profiling handbook.
- 49 Follow the instructions in the profiling template⁵⁰ and profiling handbook to create a new template-analysis repository and weld a copy of the profiling recipe into it.
- 50 Follow the instructions in the profiling recipe to set up your profiling configuration file for your current batch of data.
- 51 Execute the profiling recipe script to create annotated, normalized and/or feature-reduced profiles. This step will take a few minutes per plate.

Table 4 (continued)

Step	Problem	Possible reason	Solution
	Bad image quality for the whole plate or parts of the plate	<p>Plate manufacturing errors</p> <p>Imager astigmatism/optical alignment</p> <p>Debris on objectives/in light path</p>	<p>Ensure that autofocus is working well on the imagers to avoid artefacts due to drifts in plate bottom thickness. Use Z-projection to prevent well-to-well drifts. When possible, buy plates from a single lot to avoid batch-to-batch quality issues</p> <p>Regular profiling of the imaging system should be performed by using fluorescent beads; the system should be recalibrated as needed per the manufacturer's instructions. Special care should be taken on systems with multiple cameras. Acquiring multiple z-planes can help to counteract focal artefacts due to issues with optical alignment or sample tilt/flatness. Ensuring that fields of view are close to the center of each well in the microplate can also help to minimize variations in the bottom surface of the microplate well</p> <p>Perform preventative maintenance and conduct visual inspection of sample images before running larger-scale campaigns and cleaning according to the manufacturer's recommendations</p>
34	The identified nuclei or cell bodies do not reflect the actual boundaries of the stained nuclei or cells in the image	<p>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</p> <p>Cell confluency can have a profound impact on your ability to determine cell boundaries and thus segment cells. Segmentation algorithms can struggle if there is no background (i.e., blank space) in an image, and particularly confluent cells may grow on top of each other, making segmentation impossible</p>	<p>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.image.sc/ for assistance</p> <p>Experiment with plating your cells at different concentrations to find an appropriate balance between maximizing data and leaving some background in your images while avoiding cells growing on top of each other. Although ideally fixed before acquisition, if your images do have heavily confluent regions preventing accurate segmentation, you may wish to add into your pipelines an additional IdentifyPrimaryObjects module to detect the confluent region and then mask the confluent region from your images by using MaskImage modules before identifying your nuclei and cell objects</p>

Timing

Steps 1–9, cell culture: typically, ~2–3 d for the cells to reach appropriate confluency, depending on cell type and growth conditions. Harvesting the cells (Steps 3–7) takes 20 min, and seeding the cells (Step 8) takes 20 min. After seeding, the cells should be cultured for 2–5 d before staining.

Step 10A, addition of a small-molecule library (optional): variable; ~2–3 d for one batch experiment of 384-well plates. Addition of a compound library takes ~1 h for one batch experiment of 384-well plates, including reagent preparation and medium change, and 1–2 d for compound incubation.

Step 10B, CRISPR knockdown (optional): variable; ~4–6 d for one batch experiment of 384-well plates. Addition of polybrene and a lentiviral CRISPR library takes ~2 h for one batch experiment of 384-well plates, including centrifugation. Medium change(s) take ~1 h for one batch experiment of 384-well plates, and incubation takes an additional 3–5 d.

Step 10C, ORF overexpression (optional): variable; ~2–4 d for one batch experiment of 384-well plates. Addition of polybrene and a lentiviral ORF library takes ~2 h for one batch experiment of 384-well plates, including centrifugation. Medium change takes ~1 h for one batch experiment of 384-well plates, and incubation takes an additional 1–3 d.

Steps 11–21, 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. We recommend having all staining and fixation solutions prepared before beginning mitotracker staining to allow for sufficient time to prepare equipment during incubation steps. Steps 22–24, automated image acquisition: variable; 1–3.5 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 varies depending on the number of sites imaged per plate, the number of channels acquired per site, the number of Z planes in which each channel is imaged and the exposure time for each channel.

Steps 25–45, morphological image feature extraction from microscopy data: variable, very dependent on computing setup; 24 h–1 week per 384-well plate. It takes ~10 min per plate for CellProfiler to scan the images in the input folder(s) after manually dragging and dropping them into the CellProfiler interface. Optimizing the segmentation pipeline before running it can take from minutes to hours depending on the diversity of the input images and how different they are from the starting conditions. The pipeline execution time will depend on the computing setup; run times on a single computing node of 20 s (illumination correction), 20 s (segmentation), 30 s (QC) and 10 min (analysis) per field of view are typical; these equate to ~650 CPU hours of processing per plate, most of which can be fully parallelized. Handling of QC results, once a visualization tool like the open-source KNIME is set up, takes 20 s per plate (see Box 2). A substantial time saving can be achieved if you run the feature extraction, segmentation and QC pipelines on a distributed computing cluster, which massively parallelizes the processing compared with running it on a single local computer, as well as removing slowdowns in CellProfiler that can accumulate after several hundreds of images are processed in a single run.

Steps 46–51, creation, normalization and feature reduction of per-well profiles: variable; collation of CSVs (if CellProfiler is run in parallel with ExportToSpreadsheet rather than on a single machine with ExportToDatabase) and aggregation into per-well mean profiles can take 8–16 unattended hours per 384-well plate; plates may be run in parallel. Metadata CSV creation and repository setup will take ~30 min for an experienced user and may take 2–3 h for an inexperienced user. Execution of the profiling recipe steps takes <5 min of processing time per 384-well plate.

Step 52, 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 single-channel images in 16-bit TIF format; each resultant image will be ~1,000 × 1,000 pixels and ~2.5 MB in size. The total number of images generated equals (number of samples tested) × (number of sites imaged per well) × (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. Each additional brightfield channel acquired adds another 3,456 images to the total image count and thus an additional ~8 GB of storage.

The illumination-correction pipeline will yield an illumination-correction image per plate and channel. Each image is ~4.5 MB; thus, one microplate's worth of illumination-correction images with five fluorescent channels and one brightfield channel will occupy ~26.5 MB of storage space.

The optional QC pipeline will produce a set of numerical measurements extracted at the image level and will export them to a series of .csv files. These measurements can optionally be used for comparing quality between batches or for flagging images that you may want to remove from downstream processing because of focal blur or saturation artefacts.

The segmentation pipeline outputs a single image for each well with segmented nuclei and cell objects overlaid onto the images. The quality of the image features extracted by the analysis pipeline and downstream profiling will depend on accurate nuclei and cell body segmentation determined in the segmentation pipeline. 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. We have found that the SYTO 14 image is the most amenable for finding cell edges, because it has fairly distinct boundaries between touching cells. For help in optimizing the output, if needed, see Table 4.

The image-analysis pipeline produces the raw numerical features extracted from nuclei, cell and cytoplasm objects identified in the images. If run locally, these are saved as SQLite database files with one table per object. If run distributed, these are saved into per-compartment .csv files (i.e., Cells.csv, Cytoplasm.csv and Nuclei.csv) in a unique folder for each plate-well-site combination. The .csv files contain one row for each individual object in each image and ~2,000 columns containing the values for the different

morphological features that have been measured for that object. The pipeline also outputs Experiment.csv and Image.csv files containing information about the experiment and whole-image measurements, respectively. The pipeline also saves cell and nuclei object outlines, although these can optionally be toggled off. Generally, the pipeline is not configured to save any processed images (to conserve data storage space), but additional SaveImages modules can be used for this purpose if desired. The combination of data tables and object outlines from the feature-extraction pipeline is typically ~5 MB per site.

After running the profiling scripts to create the per-well profiles, the output will be a number of image-based profile files in CSV format. Each row of this file represents a data vector for an individual plate and well, with each column containing a per-well mean measurement of a given image feature. Initial profiles will contain 6,000 columns of raw features; normalized profile files will contain the same number of columns but will have had each column independently normalized to the median and median absolute deviation (MAD) of the data distribution.

Data availability

All images, single-cell profiles and processed profiles are available at the Cell Painting Gallery at <https://registry.opendata.aws/cellpainting-gallery/> under accession numbers cpg0000-jump-pilot and cpg0001-cellpainting-protocol. Processed profiles, metadata and a detailed description of all plates are available in the paper's GitHub repository (https://github.com/carpenterlab/2022_Cimini_NatureProtocols) and the data repositories found therein (<https://github.com/jump-cellpainting/pilot-cpjump1-data>, <https://github.com/jump-cellpainting/pilot-cpjump1-fov-data> and <https://github.com/jump-cellpainting/pilot-data-public>). Source data underlying all figures are provided with this paper.

Code availability

All code necessary to reproduce these analyses is available in the paper's GitHub repository (https://github.com/carpenter-singh-lab/2023_Cimini_NatureProtocols) and is archived to Zenodo (<https://doi.org/10.5281/zenodo.7267354>).

References

1. Chandrasekaran, S. N., Ceulemans, H., Boyd, J. D. & Carpenter, A. E. Image-based profiling for drug discovery: due for a machine-learning upgrade? *Nat. Rev. Drug Discov.* **20**, 145–159 (2021).
2. Pratapa, A., Doron, M. & Caicedo, J. C. Image-based cell phenotyping with deep learning. *Curr. Opin. Chem. Biol.* **65**, 9–17 (2021).
3. Bray, M.-A. et al. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.* **11**, 1757–1774 (2016).
4. Gustafsdottir, S. M. et al. Multiplex cytological profiling assay to measure diverse cellular states. *PLoS ONE* **8**, e80999 (2013).
5. Garcia-Fossa, F. et al. Interpreting image-based profiles using similarity clustering and single-cell visualization. *Curr. Protoc.* **3**, e713 (2023).
6. Caicedo, J. C. et al. Cell Painting predicts impact of lung cancer variants. *Mol. Biol. Cell* **33**, ar49 (2022).
7. Grigalunas, M. et al. Natural product fragment combination to performance-diverse pseudo-natural products. *Nat. Commun.* **12**, 1883 (2021).
8. 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).
9. Heiser, K. et al. Identification of potential treatments for COVID-19 through artificial intelligence-enabled phenomic analysis of human cells infected with SARS-CoV-2. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.04.21.054387> (2020).
10. Nyffeler, J. et al. Bioactivity screening of environmental chemicals using imaging-based high-throughput phenotypic profiling. *Toxicol. Appl. Pharmacol.* **389**, 114876 (2020).
11. Carey, K. L. et al. TFEB transcriptional responses reveal negative feedback by BHLHE40 and BHLHE41. *Cell Rep.* **33**, 108371 (2020).
12. Laber, S. et al. Discovering cellular programs of intrinsic and extrinsic drivers of metabolic traits using LipocyteProfiler. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.07.17.452050> (2021).
13. Simm, J. et al. Repurposing high-throughput image assays enables biological activity prediction for drug discovery. *Cell Chem. Biol.* **25**, 611–618.e3 (2018).
14. Moshkov, N. et al. Predicting compound activity from phenotypic profiles and chemical structures. *Nat. Commun.* **14**, 1967 (2023).
15. Rohban, M. H. et al. Virtual screening for small-molecule pathway regulators by image-profile matching. *Cell Syst.* **13**, 724–736.e9 (2022).
16. Chandrasekaran, S. N. et al. JUMP Cell Painting dataset: morphological impact of 136,000 chemical and genetic perturbations. Preprint at *bioRxiv* <https://doi.org/10.1101/2023.03.23.534023> (2023).

17. Chandrasekaran, S. N. et al. Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.01.05.475090> (2022).
18. Way, G. P. et al. Morphology and gene expression profiling provide complementary information for mapping cell state. *Cell Syst.* **13**, 911–923.e9 (2022).
19. Haghighi, M., Singh, S., Caicedo, J. & Carpenter, A. High-dimensional gene expression and morphology profiles of cells across 28,000 genetic and chemical perturbations. *Nat. Methods* **19**, 1550–1557 (2022).
20. Caicedo, J. C. et al. Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. *Nat. Methods* **16**, 1247–1253 (2019).
21. Dobson, E. T. A. et al. ImageJ and cellProfiler: complements in open-source bioimage analysis. *Curr. Protoc.* **1**, e89 (2021).
22. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell detection with star-convex polygons. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. 265–273 (Springer International, 2018).
23. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
24. Stirling, D. R. et al. CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinformatics* **22**, 1–11 (2021).
25. Rohban, M. H. et al. Systematic morphological profiling of human gene and allele function via Cell Painting. *Elife* **6**, e24060 (2017).
26. Cross-Zamirski, J. O. et al. Label-free prediction of cell painting from brightfield images. *Sci. Rep.* **12**, 10001 (2022).
27. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
28. Jamali, N. et al. Assessing the performance of the Cell Painting assay across different imaging systems. Preprint at *bioRxiv* <https://doi.org/10.1101/2023.02.15.528711> (2023).
29. Van Rossum, G. & Drake, F. L. *Python 3 Reference Manual: (Python Documentation Manual Part 2)*. (CreateSpace Independent Publishing Platform, 2009).
30. Way, G. et al. Pycytominer: Data processing functions for profiling perturbations. *GitHub* <https://github.com/cytomining/pycytominer> (2023).
31. Singh, S. et al. cytominer-database. *GitHub* <https://github.com/cytomining/cytominer-database> (2022).
32. Berthold, M. R. et al. KNIME: The Konstanz Information Miner. In *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)* (Springer, 2007).
33. Stöter, M. et al. CellProfiler and KNIME: open source tools for high content screening. In *Target Identification and Validation in Drug Discovery: Methods and Protocols* (eds Moll, J. & Colombo, R.) 105–122 (Humana Press, 2013).
34. Janzen, W. P. & Popa-Burke, I. G. Advances in improving the quality and flexibility of compound management. *J. Biomol. Screen* **14**, 444–451 (2009).
35. 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).
36. 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).
37. Schindelin, J., Rueden, C. T., Hiner, M. C. & Eliceiri, K. W. The ImageJ ecosystem: an open platform for biomedical image analysis. *Mol. Reprod. Dev.* **82**, 518–529 (2015).
38. Brocher, J. biovoxxel/BioVoxel-Figure-Tools: BioVoxel-Figure-Tools_1.2.1b. *Zenodo* <https://zenodo.org/record/7268128> (2022).
39. Cimini, B. A. et al. Broad Institute Imaging Platform Profiling Handbook. *GitHub* <https://github.com/cytomining/profiling-handbook> (2023).
40. Reback, J. et al. pandas-dev/pandas: Pandas 1.3.4. *Zenodo* <https://zenodo.org/record/5574486/export/hx#.ZFmvRezMIq0> (2021).
41. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
42. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
43. Waskom, M. seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).
44. van der Walt, S. et al. scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014).
45. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
46. Satopaa, V., Albrecht, J., Irwin, D. & Raghavan, B. Finding a “Kneedle” in a haystack: detecting knee points in system behavior. In 2011 31st International Conference on Distributed Computing Systems Workshops 166–171 (Institute of Electrical and Electronics Engineers, 2011).
47. Kluyver, T. et al. Jupyter Notebooks—a publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (eds Loizides, F. & Schmidt, B.) 87–90 (IOS Press, 2016).
48. Tange, O. GNU Parallel 2018 (Lulu.com, 2018).
49. Chandrasekaran, S. N., Weisbart, E., Way, G., Carpenter, A. & Singh, S. Broad Institute Imaging Platform Profiling Recipe. *GitHub* <https://github.com/cytomining/profiling-recipe> (2022).
50. Chandrasekaran, S. N., Way, G., Carpenter, A. & Singh, S. Broad Institute Imaging Platform Profiling Template. *GitHub* <https://github.com/cytomining/profiling-recipe> (2022).
51. Caicedo, J. C. et al. Data-analysis strategies for image-based cell profiling. *Nat. Methods* **14**, 849–863 (2017).
52. Assay Guidance Manual. (Eli Lilly and the National Center for Advancing Translational Sciences, 2012).

Acknowledgements

The authors gratefully acknowledge a grant from the Massachusetts Life Sciences Center Bits to Bytes Capital Call program for funding the data production. We appreciate funding to support data analysis and interpretation from members of the JUMP Cell Painting Consortium and from the National Institutes of Health (NIH MIRA R35 GM122547 to A.E.C.). We would like to acknowledge the Consortium's Supporting Partner PerkinElmer for providing an in-kind contribution of the PhenoVue Cell Painting JUMP kit. The authors also gratefully acknowledge the use of the PerkinElmer Opera Phenix high-content/high-throughput imaging system at the Broad Institute, funded by the S10 Grant NIH OD-026839-01. H.S.A. was supported by a postdoctoral scholarship from the Knut and Alice Wallenberg Foundation. This project has been made possible in part by grant number 2020-225720 to B.A.C. from the Chan Zuckerberg Initiative Donor-Advised Fund, an advised fund of the Silicon Valley Community Foundation. The authors thank B. Marion and other members of the automation team in the Broad Institute's Center for the Development of Therapeutics. The authors appreciate the more than 100 scientists who have contributed to the organization and scientific direction of the JUMP Cell Painting Consortium. The authors also thank the original developers of earlier versions of the protocol, who contributed to prior papers describing the assay^{3,4}; those who are not also authors on this paper are: M. A. Bray, H. Han, C. T. Davis, B. Borgeson, C. Hartland, S. M. Gustafsdottir, C. C. Gibson, 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 and A. F. Shamji.

Author contributions

B.A.C., S.N.C., M.K.-A., L.M., A.G., B.F., S.G., J.B.C., C.-H.L., E.M., S.S., J.D.B., T.G., D.G., T.M., J.E.P., A.S., S.E.S., A.V., G.W., S.Y., B.Z. and A.E.C. contributed to experimental design. M.K.-A., L.M., A.G., B.F., P.B., J.B.C., E.M., P.A.Jr, J.D.B., T.G., D.H., G.H., K.J., M.M., T.M., J.E.P., A.S., A.V., G.W. and S.Y. conducted laboratory experiments. B.A.C., S.N.C., D.J.L., N.J., H.S.A., S.H., L.H.M., G.W. and B.Z. performed image and/or data analysis. B.A.C., S.N.C., M.K.-A., L.M., P.B., S.G., J.B.C., C.-H.L., E.M., S.S., J.D.B., G.H., M.M., L.H.M., T.M., A.S., S.E.S., G.W., B.Z. and A.E.C. contributed to data interpretation. B.A.C., D.J.L., J.B.C., P.A.Jr, S.H., E.W. and A.E.C. contributed to protocol development. B.A.C., S.N.C., M.K.-A., L.M., A.G., B.F., S.G., J.B.C., S.S., P.A.Jr, D.G., E.W. and A.E.C. wrote the manuscript, with all authors contributing to editing and giving manuscript approval. Supervision and administration were carried out by B.A.C., S.S. and A.E.C.

Competing interests

S.S. and A.E.C. serve as scientific advisors for companies that use image-based profiling and Cell Painting (A.E.C.: Recursion; S.S.: Waypoint Bio and Dewpoint Therapeutics) and receive honoraria for occasional talks at pharmaceutical and biotechnology companies. D.G. is an employee of Bayer, AG, Pharmaceuticals. S.G., B.Z. and G.H. are employees of Merck Healthcare KGaA, Darmstadt, Germany. J.D.B. and T.G. were employed at Pfizer for the duration of this work. S.Y. was employed at Takeda for the duration of this work. S.E.S. was employed at Biogen for the duration of her contributions to this work. C.-H.L. was employed at Janssen Pharmaceutica at the time of writing. J.B.C. and P.A.Jr are employees of the Novartis Institutes for Biomedical Research, Cambridge, MA, USA and declare no competing interests. E.M., G.W., T.M., L.M. and J.E.P. are employees of AstraZeneca, Cambridge, UK. K.J. was employed at AstraZeneca for the duration of this work. D.J.L. and S.H. are employees of Pfizer, Inc.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41596-023-00840-9>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41596-023-00840-9>.

Correspondence and requests for materials should be addressed to Anne E. Carpenter.

Peer review information *Nature Protocols* thanks Marc Bickle, Joshua Harrill and Sonja Sievers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Received: 13 July 2022; Accepted: 28 March 2023;

Published online: 21 June 2023

Related links

Key references using this protocol

Nyffeler, J. et al. *Toxicol. Appl. Pharmacol.* **389**, 114876 (2020): <https://doi.org/10.1016/j.taap.2019.114876>

Rohban, M. H. et al. *Cell Syst.* **13**, 724–736.e9 (2022): <https://doi.org/10.1016/j.cels.2022.08.003>

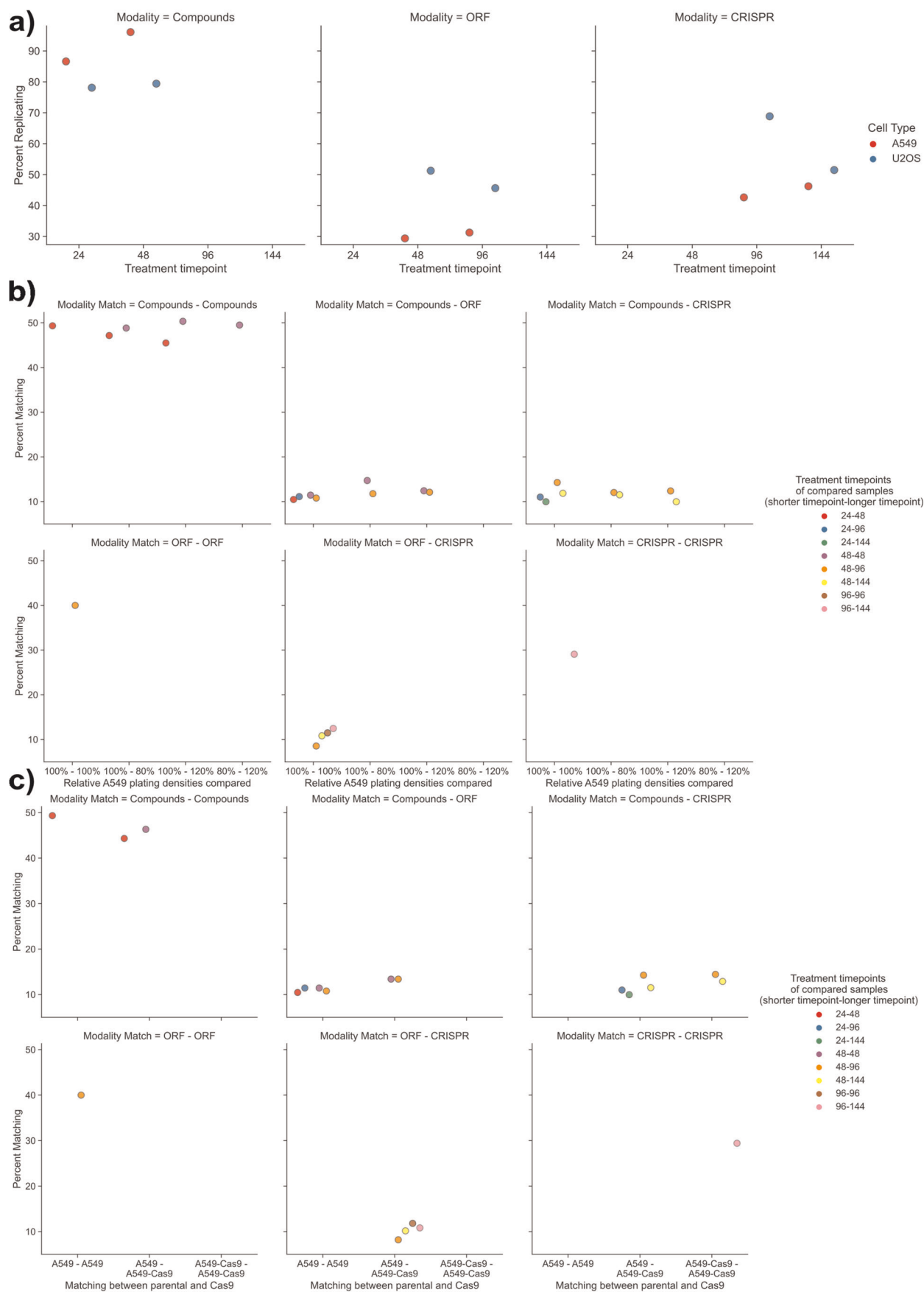
Caicedo, J. C. et al. *Mol. Biol. Cell* **33**, ar49 (2022): <https://doi.org/10.1091/mbc.E21-11-0538>

Key data used in this protocol

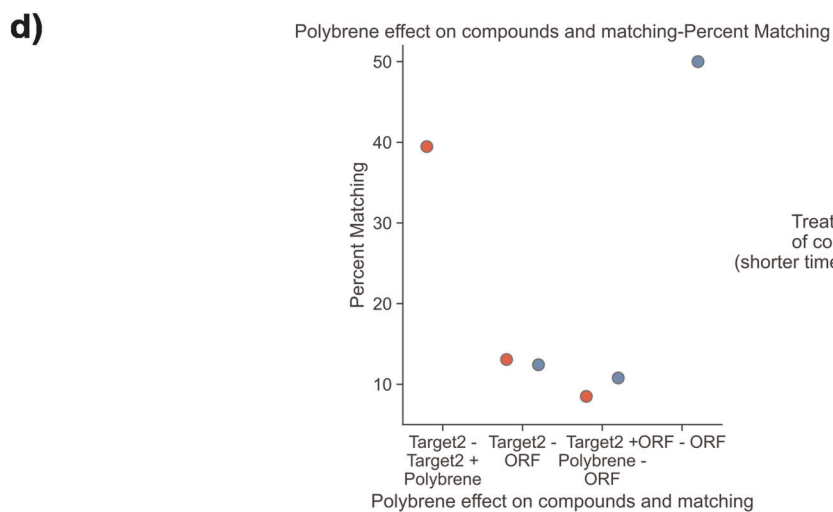
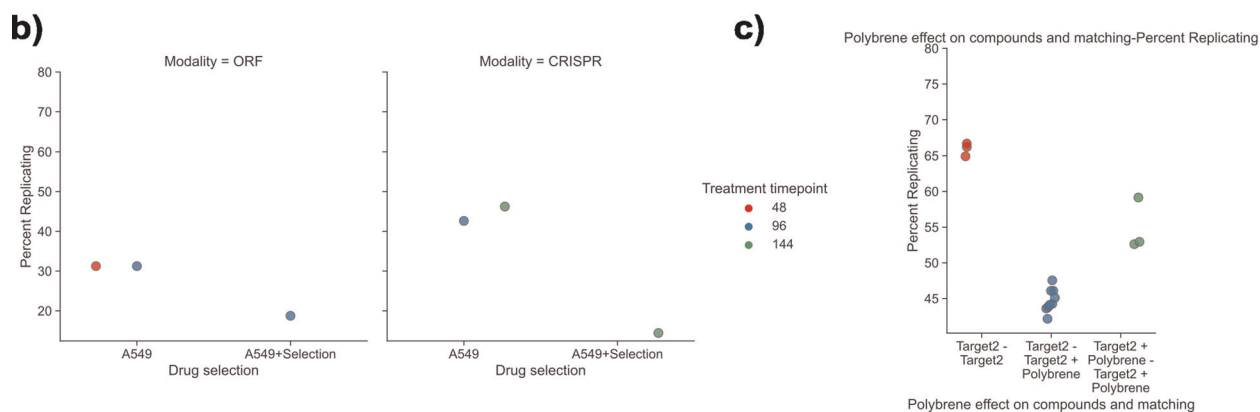
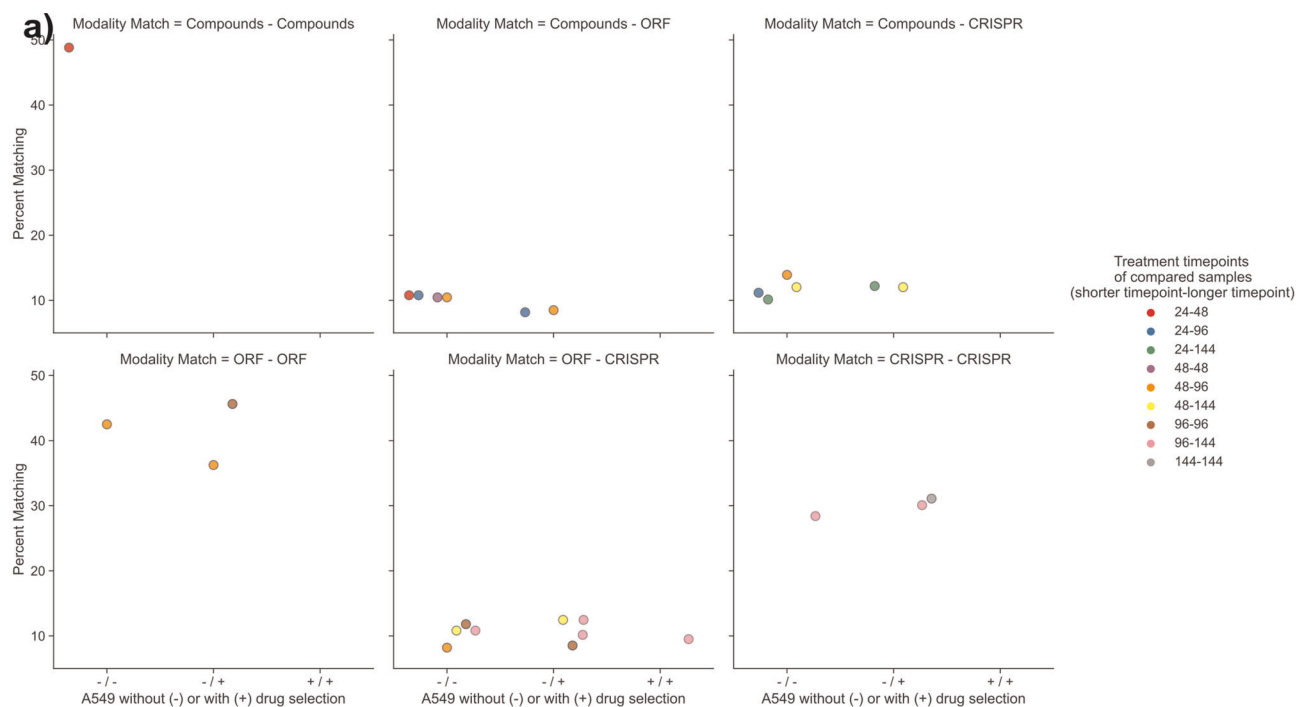
JUMP Cell Painting Consortium. Datasets (2023): <https://doi.org/10.5281/zenodo.7628768>

JUMP-MOA Plate - Compound Number																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	63	69	27	51	33	52	19	6	20	76		48	77	15	49	81	84	74	55	36	90	75	67	35
B	16	32	81	47	31	15	45	57	30	21	53	31	42	65	52	44	59		23	66	56	28	50	64
C	58	66	43	9	72	83	38	4	65	62	12	82	61	37	89	22	11	60	73	47	53	25	39	71
D	15	30	87	39	74	71	36	5	68	7	35	73	64	54	24	38		33		83	83	42	58	86
E	64	64	70	70	13	13			54	54	50	50	5	77	7	49	41	84	68	55	79	90	80	67
F	57	68	11	61	32	25		81	90	49	39	87	43	13	62	23	66	20	2	34	51	62	12	
G	66	58	9	43	83	72	4	38	62	65	82	12	3	46	14	57	29	16	32		31	85		30
H	67	89	41	35	55	40	89	26	21		22	46	88	51	34	78	63	4	65	58	38	1	76	52
I	69	63	51	27	52	33	6	19	76	20	48			61	40	89	21	11	45	73	26	53	87	39
J		90	46	22	5	45	47	36	29	67	79	37	86	8	70	56	69	6	19	10		24	78	18
K	88	2	18	1	17	44	23	56	86	28	42	10	15	3	81	14	74	29	36	32	75	31	35	
L	37	29	7	3	60	79	3	60	25	84	71	41	18	72	10	76	9	70	54	63	72	27	4	43
M	24	24	34	34			59	59	8	8	78	78	37	5	22	7	60	41	47	68	25	79	71	80
N	80	53	73	74	26		85	11	61	75	40	80	48	12	17	50	1	17	13	2	6		33	59
O	2	88	1	18	44	17	56	23	28	86	10	42	46		57	40	16	21		45	85	26	30	87
P	77		49	85	84	55	14	14	75	16		77	27	82	82	48	28	9	44	19	8	69	20	88

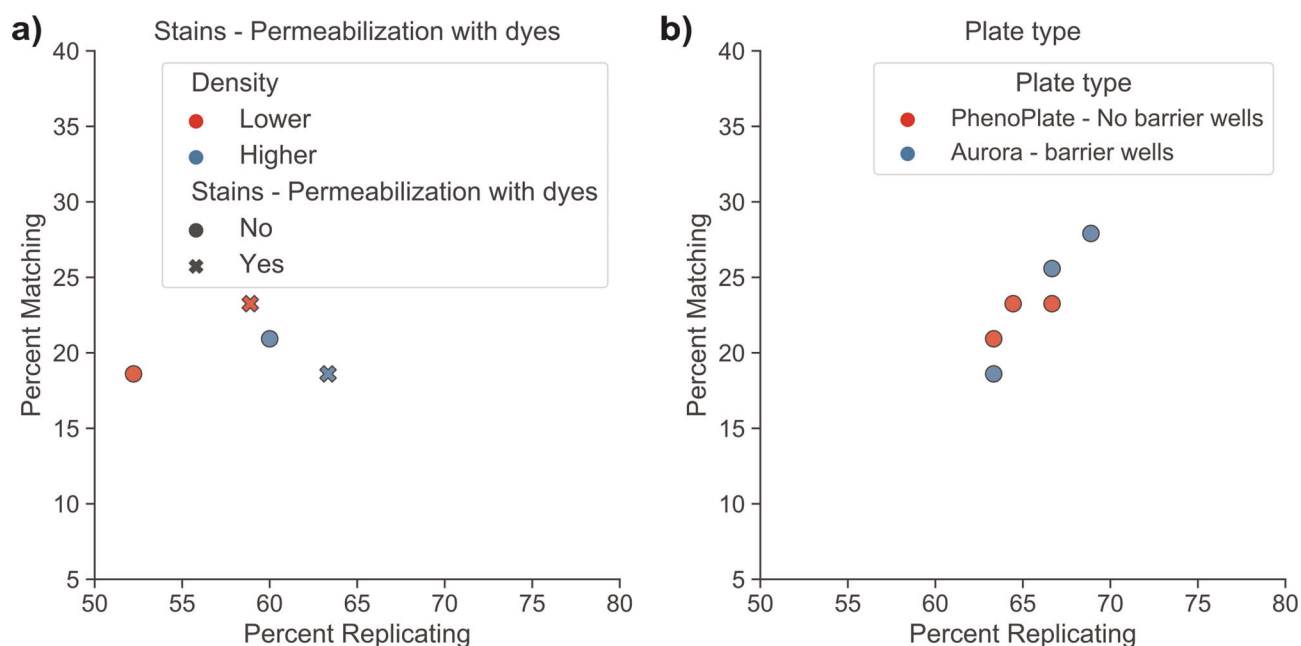
Extended Data Fig. 1 | The JUMP-MOA plate map labeled by compound. The JUMP-MOA compound plate map: unlabeled wells are DMSO only; all other wells are labeled to show the distribution of compound replicates across the entire plate. A version of this figure grouped by MOA rather than compound is available as Fig. 2a.



◀ **Extended Data Fig. 2 | Further assessment of cell type suitability for Cell Painting.** **a**, Percent replicating of each cell line and time point tested when selecting the cell type and treatment time points to be used by the consortium. A549 (red dots) has better replication at both time points of compound treatment than U2OS (blue dots) but worse performance at all time points of genetic perturbation experiments. **b**, Assessment of decreasing or increasing the number of cells plated by 20% on the ability to perform percent matching. Few consistent effects are seen when comparing across the columns present in each panel. **c**, Assessment of using parental cell lines or Cas9-expressing polyclonal lines for compound treatments. Few consistent effects are seen; the top right panel represents the desired task, the ability to match compounds added to either parental or Cas9 cells to CRISPR (which must be performed in Cas9 cells). The results in that panel do not seem extremely different either for matching 48-h compound treatment to 96 h of CRISPR treatment (compare the orange dots) or for matching 48-h compound treatment to 144 h of CRISPR treatment (compare the yellow dots). For more information including the plate(s) represented by each data point, see the Source Data file for this figure; expanded experimental details for each plate may be found in Supplementary Data File 1.

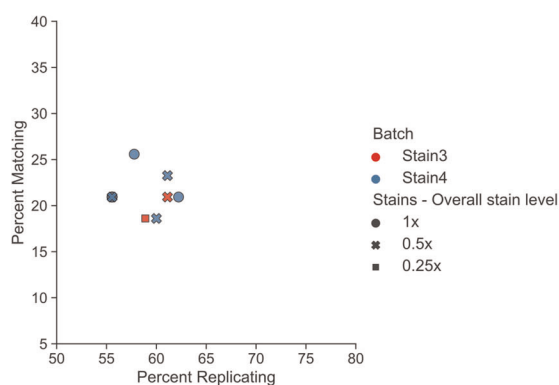


◀ **Extended Data Fig. 3 | Assessment of the effect of gene-treatment-related compounds on Cell Painting performance.** **a**, Addition of blasticidin to ORF plates or puromycin to CRISPR plates does not appear to improve cross-modality matching across modalities versus unselected plates; compare the second and third columns to the first column in the top center, top-left and bottom-middle panels. **b**, Addition of selection compounds may have a deleterious effect on percent replicating versus unselected plates, although we cannot rule out that this is due to fewer replicates for the selected plates than the unselected ones. **c**, Addition of 4 µg/ml polybrene for 24 h may produce a phenotypic effect; polybrene addition displays decreased inter-treatment cross-plate percent replicating versus intra-treatment cross-plate-percent replicating (compare the center column to the outside columns), even though both sets of plates were stained and imaged as part of the same batch. **d**, Addition of polybrene to Target2-treated cells does not improve percent matching between Target2-treated plates and ORF-treated plates; note that the Target-ORF plates came from a previous batch and were not stained and imaged in parallel to the Target2 compound plates. For more information including the plate(s) represented by each data point, see the Source Data file for this figure; expanded experimental details for each plate may be found in Supplementary Data File 1.

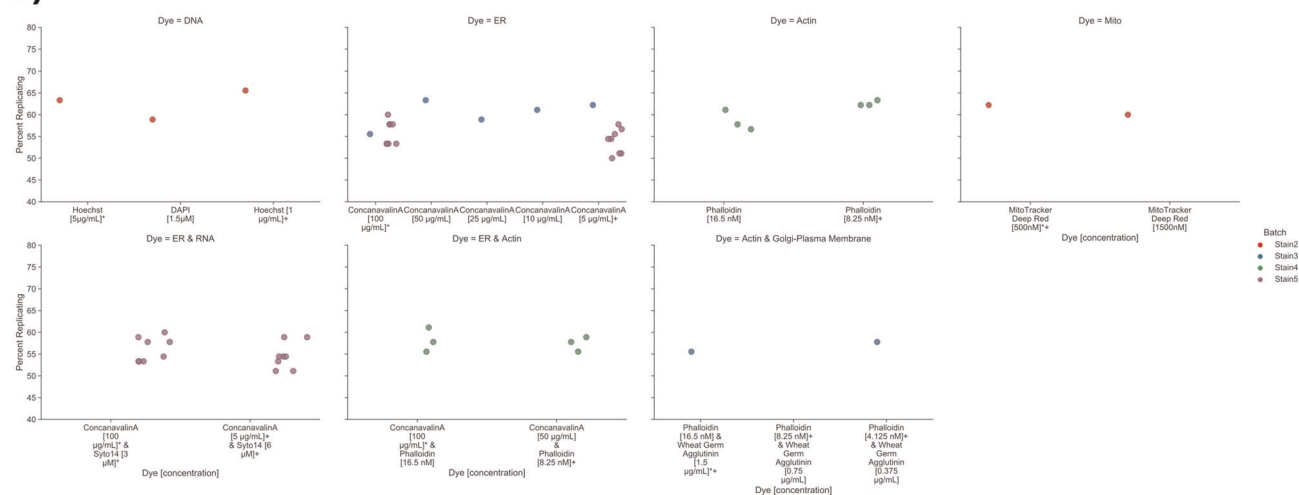


Extended Data Fig. 4 | Assessment of permeabilization timing and plate type on Cell Painting performance. **a**, Performing permeabilization at the same time as staining produces comparable percent replicating and percent matching results. **b**, PhenoPlates without barrier wells and Aurora plates with barrier wells produce comparable percent replicating and percent matching results. For more information including the plate(s) represented by each data point, see the Source Data file for this figure; expanded experimental details for each plate may be found in Supplementary Data File 1.

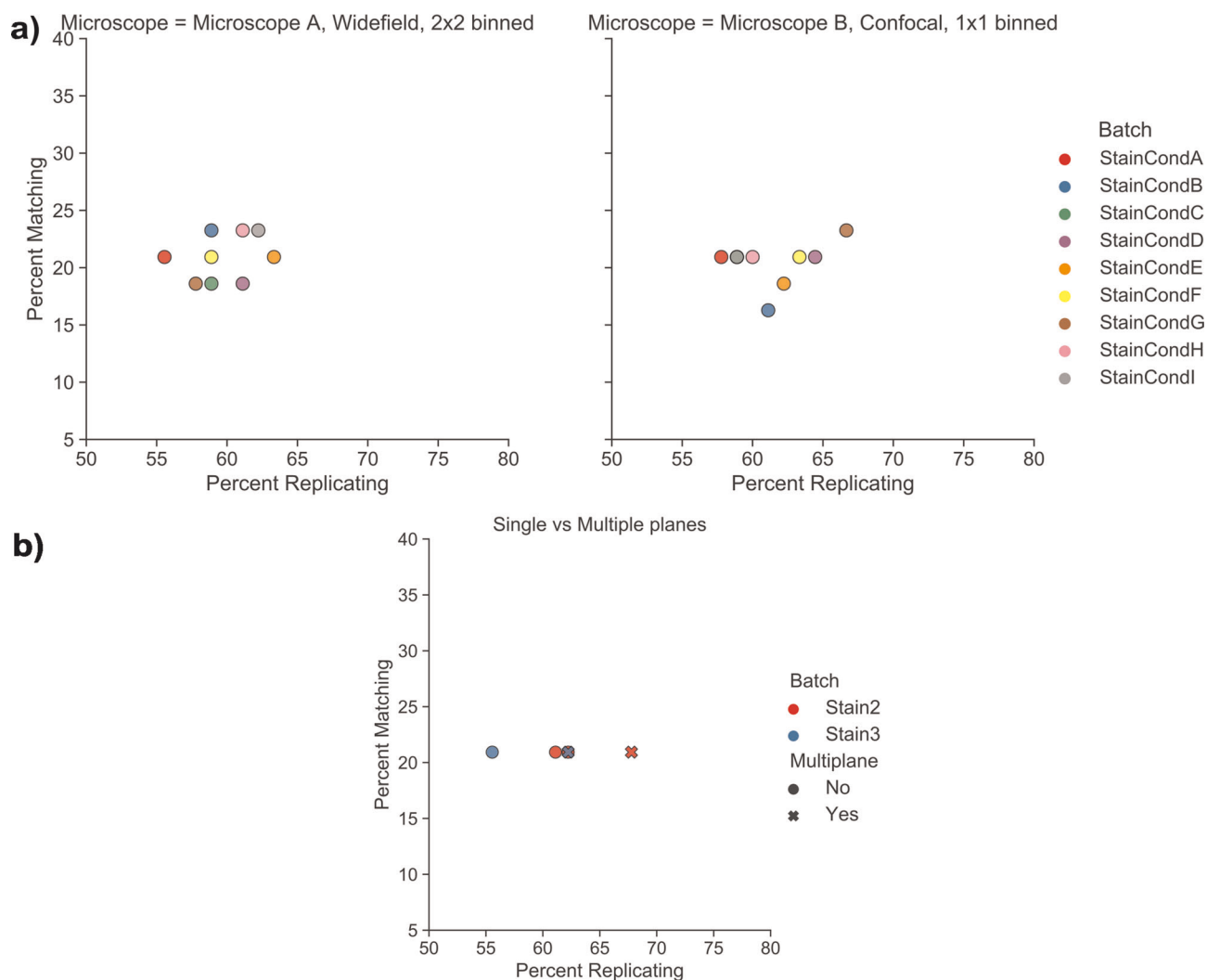
a)



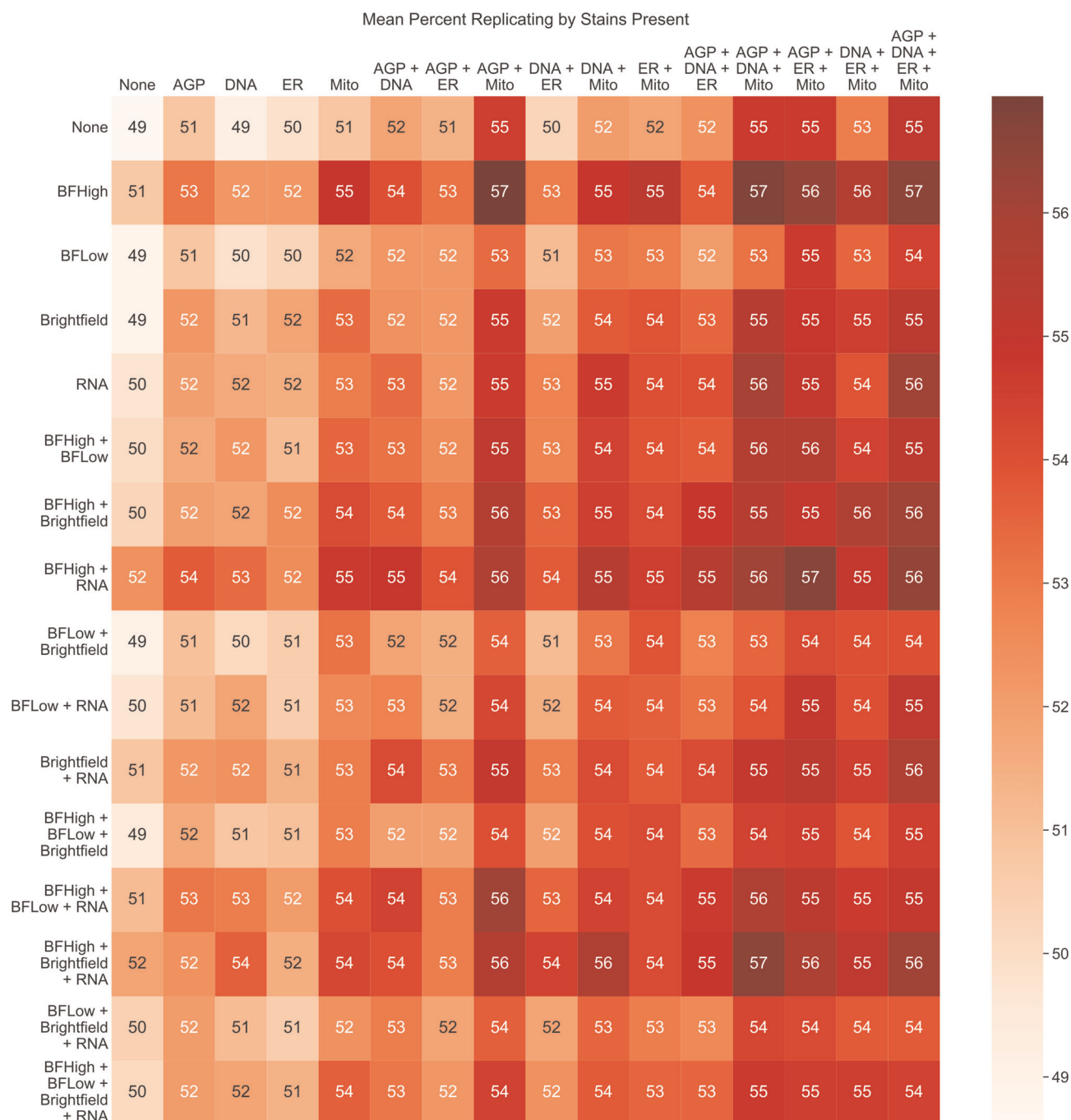
b)



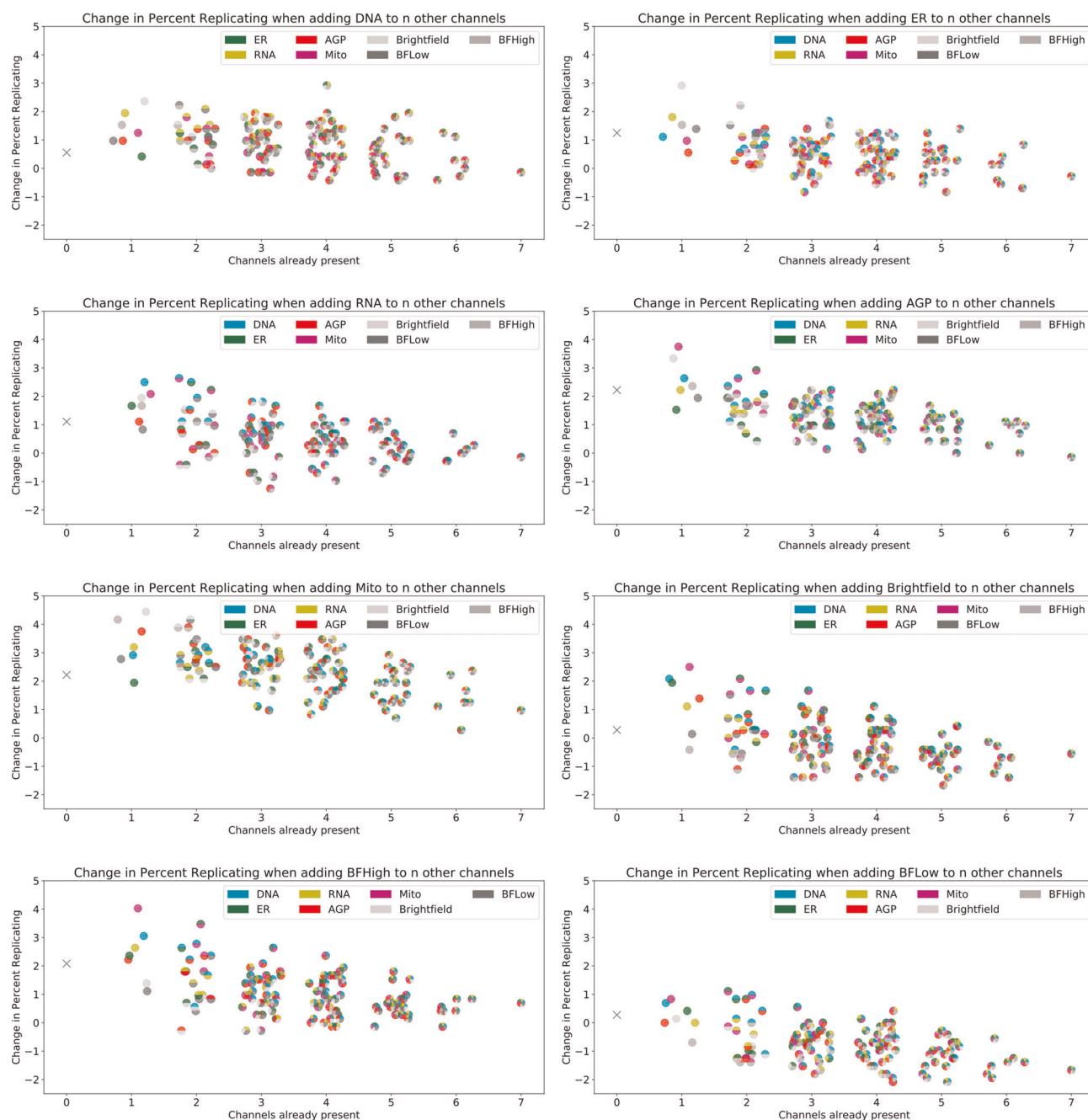
Extended Data Fig. 5 | Expanded assessment of dye concentration effects on Cell Painting performance. a, Within batches, reducing all dyes by two- or fourfold produces comparable percent replicating and comparable but perhaps slightly decreased percent matching results. **b**, All quantitatively tested stain concentration changes, broken out by the dye(s) perturbed. For more information including the plate(s) represented by each data point, see the Source Data file for this figure; expanded experimental details for each plate may be found in Supplementary Data File 1.



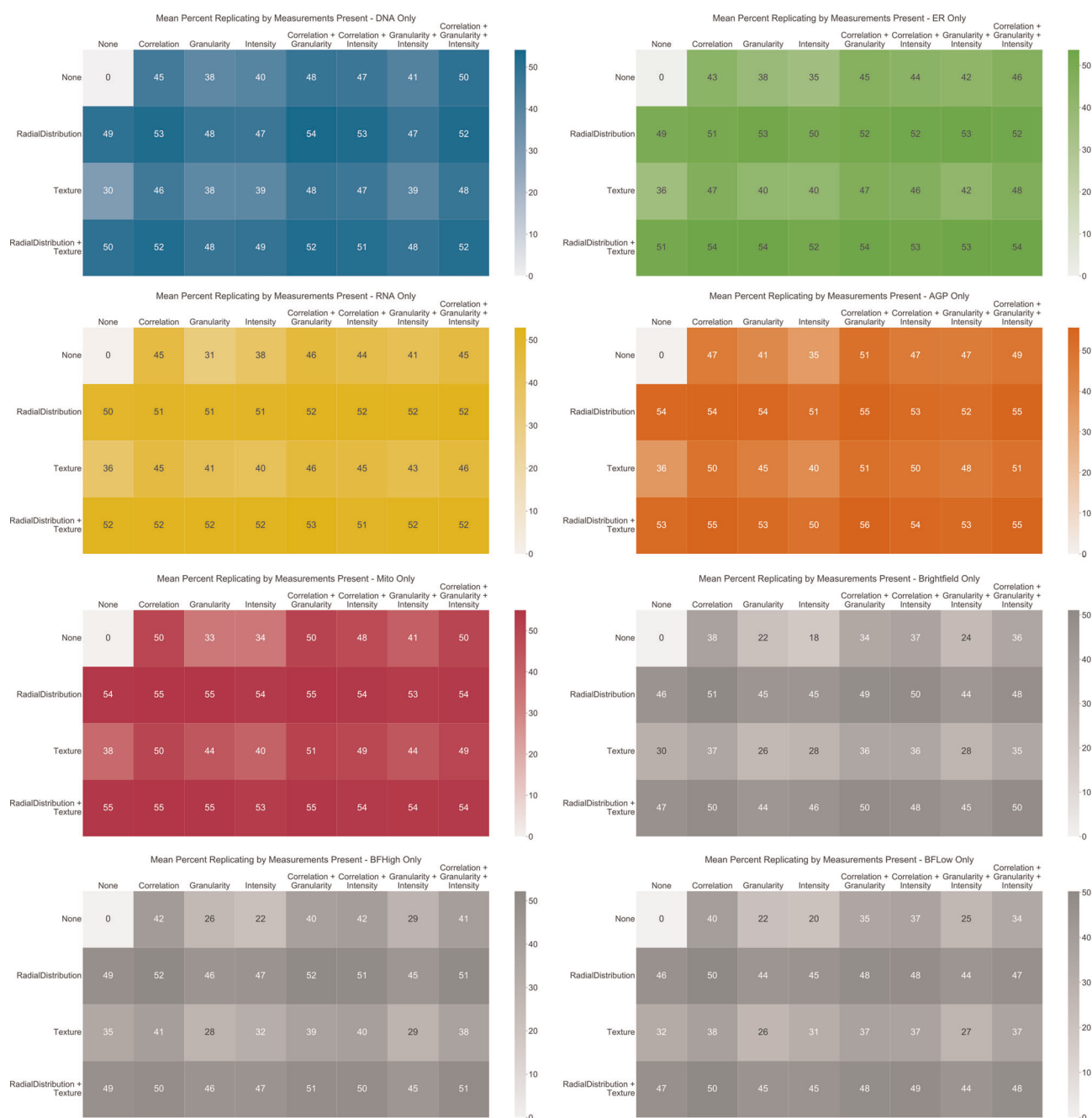
Extended Data Fig. 6 | Further assessment of imaging conditions on Cell Painting performance. **a**, Imaging of the same plates on a wide-field microscope with 2×2 binning versus a different manufacturer's microscope in confocal mode with 1×1 binning. No major differences are observed. **b**, Acquisition of multiple Z planes slightly improves percent replicating but not percent matching in two batches. For more information including the plate(s) represented by each data point, see the Source Data file for this figure; expanded experimental details for each plate may be found in Supplementary Data File 1.



Extended Data Fig. 7 | Assessment of Cell Painting performance when using various combinations of channel features. Mean percent replicating of eight JUMP-MOA plates stained with the final staining conditions after dropping out all feature names containing individual channel names before performing feature selection and calculation of percent replicating. 'None' means that only AreaShape features are present. To create a sufficiently compact data representation, the eight channels present were split four each onto the X (AGP, DNA, ER and Mito) and Y (RNA, Brightfield, BFHigh and BFLow) axes; this allows visualization of the 256 possible unique combinations. Note that these results are not the same as truly having the stains not present, because (i) a channel still may have been used in creating the initial segmentation results, and (ii) it does not account for cross-talk between stains. An alternate representation of these data is presented in Extended Data Fig. 8. For more information including the plates represented in this figure, see the Source Data file for this figure; expanded experimental details for each plate may be found in Supplementary Data File 1. AGP, actin golgi plasma membrane; BF, brightfield; Mito, mitochondria.



Extended Data Fig. 8 | Change in Cell Painting performance when adding individual channel features. Change in mean percent replicating of eight JUMP-MOA plates stained with the final staining conditions when an individual stain is present versus not; each chart shows for a particular stain when added to the non-channel-specific features plus zero or more other stains (x axis) the change in the mean percent replicating (y axis) when those features are included. The color(s) of each marker indicate which channel(s) are already present. Note that these results are not the same as truly having the stains not present, because (i) a channel still may have been used in creating the initial segmentation results, and (ii) it does not account for cross-talk between stains. The absolute percent replicating numbers are available in Extended Data Fig. 7. For more information including the plates represented in this figure, see the Source Data file for this figure; expanded experimental details for each plate may be found in Supplementary Data File 1.



Extended Data Fig. 9 | Assessment of Cell Painting performance for individual channels and feature categories. Stain-by-stain breakdown of mean percent replicating of eight JUMP-MOA plates stained with the final staining conditions after dropping out all possible combinations of features from the five stain-specific feature categories before performing feature selection and calculation of percent replicating. Unlike the analyses in Extended Data Fig. 7 and Extended Data Fig. 8, non-channel-specific feature categories (AreaShape and Neighbors) are not included here, because the goal is to assess the information contribution of each feature category in each stain. To create sufficiently compact data representation, the five categories present were split three each onto the X (Correlation, Granularity and Intensity) and two onto the Y (RadialDistribution and Texture) axes; this allows visualization of the 31 possible unique combinations. For more information including the plates represented in this figure, see the Source Data file for this figure; expanded experimental details for each plate may be found in Supplementary Data 1.