nature methods

Review article

https://doi.org/10.1038/s41592-024-02528-8

Cell Painting: a decade of discovery and innovation in cellular imaging

Received: 11 April 2024

Accepted: 24 September 2024

Published online: 05 December 2024

Check for updates

Srijit Seal[1,2,8] ✉, Maria-Anna Trapotsi[3,8] ✉, Ola Spjuth[4,5], Shantanu Singh[3], Jordi Carreras-Puigvert[4,5], Nigel Greene[6], Andreas Bender[2,7] & Anne E. Carpenter[1] ✉

Modern quantitative image analysis techniques have enabled high-throughput, high-content imaging experiments. Image-based profiling leverages the rich information in images to identify similarities or differences among biological samples, rather than measuring a few features, as in high-content screening. Here, we review a decade of advancements and applications of Cell Painting, a microscopy-based cell-labeling assay aiming to capture a cell's state, introduced in 2013 to optimize and standardize image-based profiling. Cell Painting's ability to capture cellular responses to various perturbations has expanded owing to improvements in the protocol, adaptations for different perturbations, and enhanced methodologies for feature extraction, quality control, and batch-effect correction. Cell Painting is a versatile tool that has been used in various applications, alone or with other -omics data, to decipher the mechanism of action of a compound, its toxicity profile, and other biological effects. Future advances will likely involve computational and experimental techniques, new publicly available datasets, and integration with other high-content data types.

Phenotypic drug discovery (PDD) involves using a living system to identify compounds that alter the phenotype of a given disease. PDD has evolved from screening a few compounds in animals to testing millions in cell models. By contrast, target-based drug discovery (TDD) identifies compounds that interact with a pre-selected target. Although both approaches have yielded therapeutics, mounting evidence suggests that PDD yields more first-in-class medicines than does TDD[1]. Notably, many drugs approved by the US Food and Drug Administration (FDA) lack a defined molecular target, and several drugs do not work through their purported target[2]. Therefore, phenotypic strategies have gained favor precisely because they allow compounds to be explored in a target-agnostic manner, which is especially appealing for diseases that are polygenic or are associated with undruggable targets.

High-content screening (HCS) is an effective and efficient phenotypic screening strategy that uses microscopy as the readout[3]. HCS captures and measures cell phenotypes in images and can identify candidate targets (for example, when genetic perturbations are screened) and therapeutics (when small molecules are screened). At the core of HCS is cellular morphology—the visual appearance of cells, usually stained for cell structures or biomarkers—which is intricately linked to cell physiology, health, and function (Supplementary Table 1 lists some common keywords used in HCS assays).

A major development emerged in 2004, when Perlman et al. demonstrated that, instead of tailoring an image-based assay to measure a particular phenotype of interest, images can be used in a relatively unbiased way (aside from the choice of experimental conditions) to group drug treatments that have similar impacts on cell morphology[4].

[1]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [2]Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, UK. [3]Imaging and Data Analytics, Clinical Pharmacology & Safety Sciences, R&D, AstraZeneca, Cambridge, UK. [4]Department of Pharmaceutical Biosciences and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [5]Phenaros Pharmaceuticals AB, Uppsala, Sweden. [6]Imaging and Data Analytics, Clinical Pharmacology & Safety Sciences, R&D, AstraZeneca, Waltham, MA, USA. [7]STAR-UBB Institute, Babeș-Bolyai University, Cluj-Napoca, Romania. [8]These authors contributed equally: Srijit Seal, Maria-Anna Trapotsi. ✉e-mail: seal@broadinstitute.org; marianna.trapotsi1@astrazeneca.com; anne@broadinstitute.org

https://doi.org/10.1038/s41592-024-02528-8

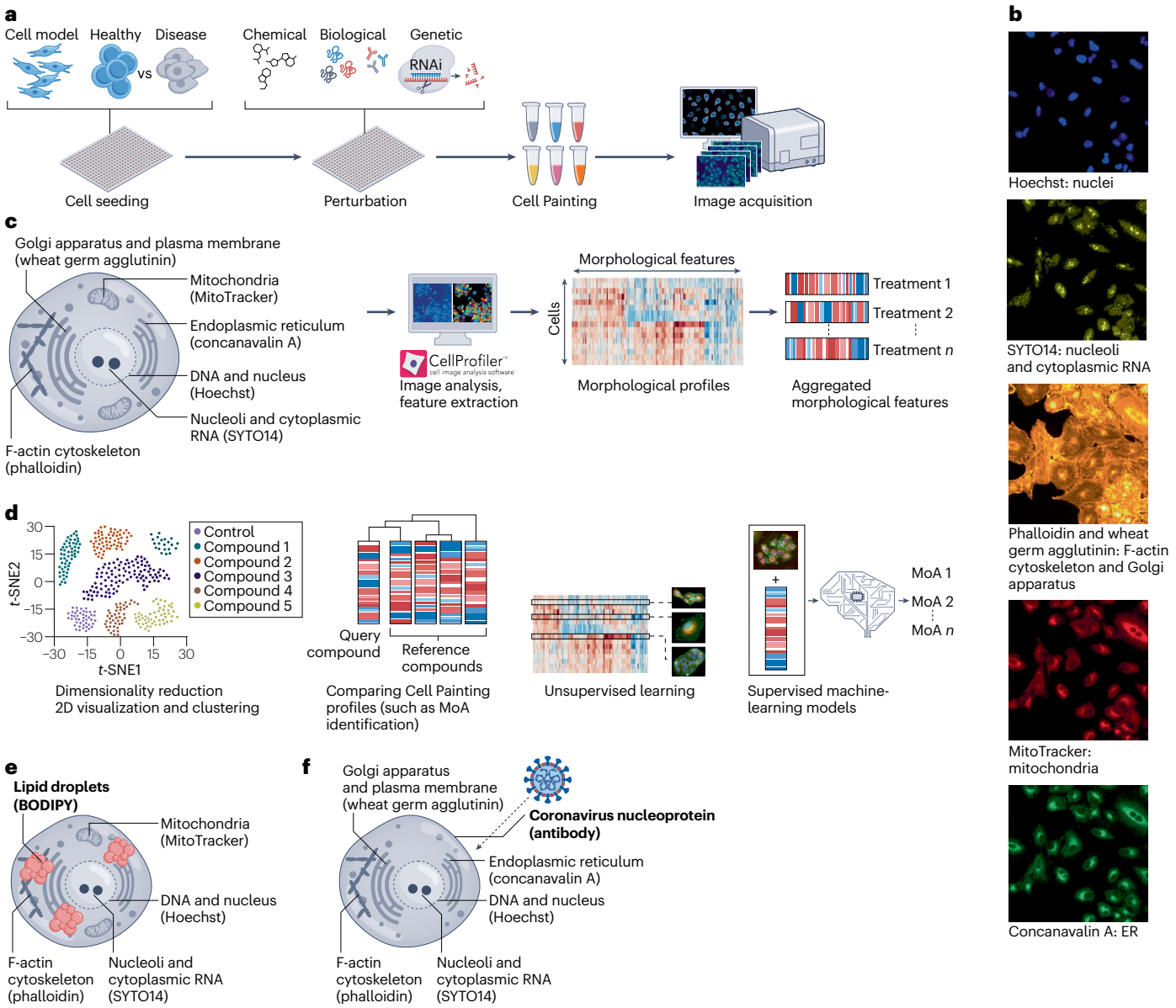

Fig. 1 | Morphological profiling using Cell Painting. a, Schematic of the Cell Painting assay; cells are incubated and perturbed and a set of six stains is applied. **b**, Images are then obtained in five channels by automated microscopy followed by nucleus and cell body segmentation. **c**, Appropriate software or deep-learning-based methods are applied to measure or calculate morphological features from the images. **d**, After feature preprocessing, downstream analysis is performed. This includes a variety of methods, including supervised and unsupervised machine learning, to better elucidate the biological effects of a compound, such as its mechanism of action or safety profile. **e**,**f**, Adaptations of the Cell Painting assay include BODIPY to mark lipid droplets in lipid-accumulating cells (**e**) and a coronavirus antibody against human coronavirus 229E (CoV-229E) viral protein (**f**).

This finding, combined with advances such as transcriptional profiling and automated sample preparation and microscopy, helped launch the field of image-based profiling and the use of image assays that maximize information content[5–7].

The most popular image-based profiling assay is Cell Painting, first described in 2013 (ref. 8). Cell Painting 'paints' the cell with many fluorescent dyes to mark major organelles or components, aiming to capture its phenotypic state and responses to perturbations (Fig. 1a). The standard dyes for Cell Painting are Hoechst 33342 (which stains DNA), concanavalin A (endoplasmic reticulum), SYTO 14 (nucleoli and cytoplasmic RNA), phalloidin (f-actin), wheat germ agglutinin (WGA) (Golgi apparatus and plasma membrane), and MitoTracker Deep Red (mitochondria). The Cell Painting assay was designed to be easy and inexpensive to implement, requiring no custom equipment beyond the proper microscope filters and relying solely on dyes, rather than antibodies. Multiplex staining is followed by processing with automated imaging pipelines (whether deep-learning-based or using classical methods, such as in the open-source CellProfiler[9]) that extract morphological profiles and standardize them against reference and control compounds (Fig. 1b). This approach yields a high-dimensional dataset for each cell and captures more than 1,000 morphological features, including size, shape, texture, and intensity. The morphological profiles are processed to apply various normalizations and batch-effect corrections and are then used for downstream analysis (Fig. 1c).

Although any image set can be used for image-based profiling, the Cell Painting assay is widely used in academic and industry research. Here, we aim to comprehensively examine the advances and impacts of Cell Painting in drug discovery and related areas over the past decade

https://doi.org/10.1038/s41592-024-02528-8

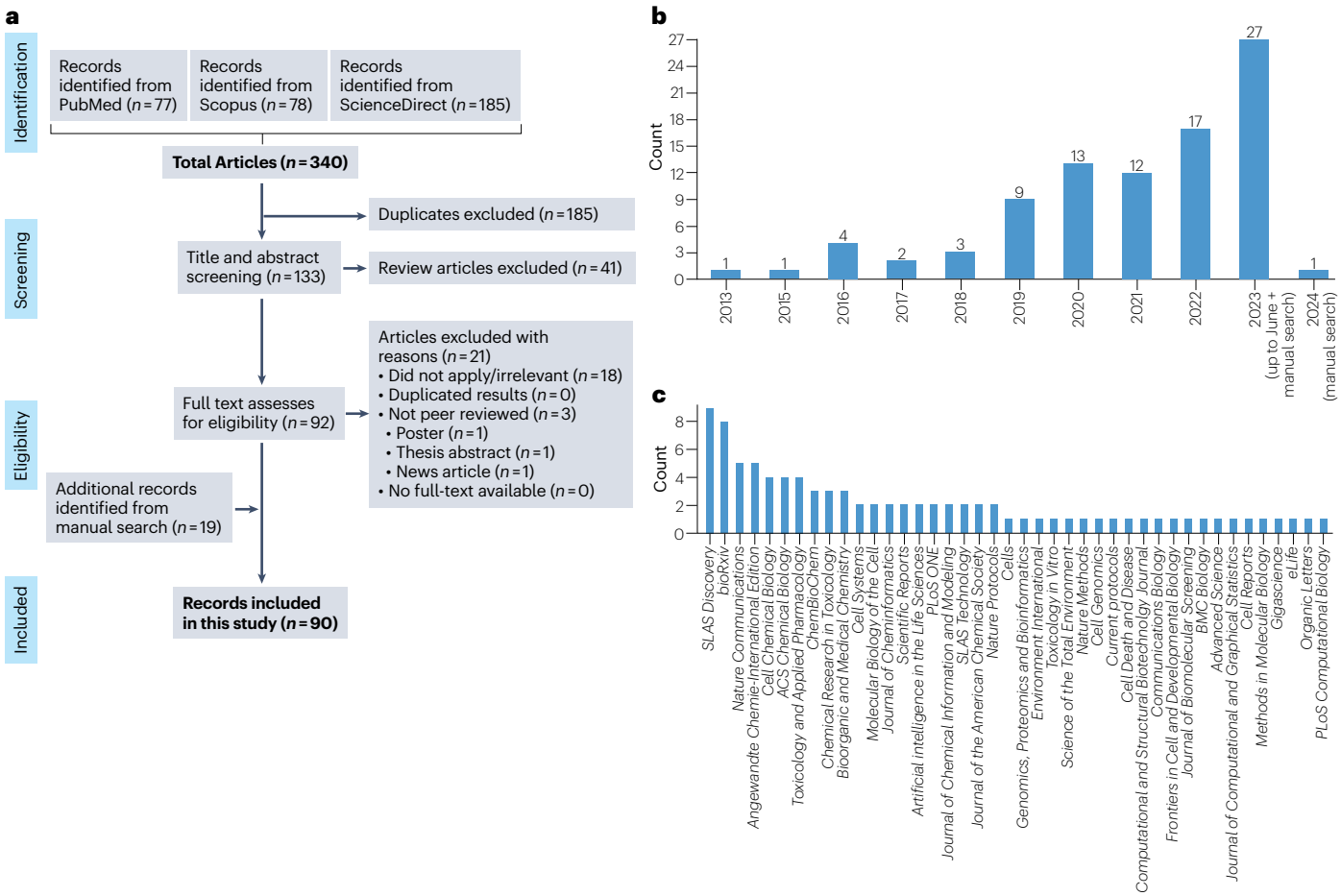

Fig. 2 | Overview of studies included in this systematic review and publication trends. a, The Preferred Reporting Items for Systematic Reviews and Meta Analyses flow chart showing the selection of the 90 studies included in this systematic review. Records from manual search included select articles published after the June 2023 cut-off date. **b**, The growth in the number of publications reviewed in this systematic review between 2013 and 2023. **c**, The journal titles in which the works were published.

(2013–2023), following a systematic review format not aiming to capture the entire field of image-based profiling. We explore how methodological advances have improved the robustness of the assay and discuss how Cell Painting has deepened our understanding of disease processes and shaped therapeutic discovery. Importantly, we discuss the integration of Cell Painting with machine learning and other -omics data. We also explore the role of Cell Painting in predictive toxicology and its significance in improving the safety and efficacy of drugs. Overall, we provide a comprehensive perspective on the impact and potential of the Cell Painting assay in drug discovery.

Systematic analysis of Cell Painting literature

Study selection

We conducted a systematic review of Cell Painting studies by retrieving 340 articles from PubMed, Scopus, and ScienceDirect (accessed June 2023) using the keyword 'Cell Painting' (in the title, abstract, subject terms, and/or keyword headings). The search was limited to articles written in English after 2012, and articles had to be peer-reviewed, with some exceptions for key preprints. Reviews, news articles, posters, thesis abstracts, and perspective papers were not included as primary research articles (these are instead listed in Supplementary Table 3 and referenced where applicable). After removing duplicates (207) and review articles (41), 92 articles underwent full-text analysis. Following further screening, 21 studies were excluded (18 were irrelevant and 3 were a poster, thesis, or news article), and a manual search added 19 relevant studies, including some published after June 2023 (listed in Supplementary Tables 2 and 3). This resulted in 90 studies for review, as shown in the preferred reporting items for systematic reviews and meta analyses flow chart (Fig. 2a), with included and excluded studies listed in Supplementary Tables 2 and 3, respectively.

Extracted data

We extracted data from Cell Painting assay publications, including authors, year, keywords, and journal title, and manually categorized the research question and major outcome. The assay's usage is increasing, with most studies published between 2021 and 2023 (Fig. 2b). *SLAS Discovery* (Society for Laboratory Automation and Screening) was the most popular journal (Fig. 2c), reflecting the assay's acceptance in the drug discovery and screening community. Other top publication choices include computational journals (for example, in cheminformatics), and journals in chemical biology and toxicology.

Advancements in Cell Painting

Assay development

The Cell Painting protocol was first developed by Gustafsdottir et al. in 2013 at the Broad Institute. It was designed to be a low-cost, high-throughput single assay that could capture many biologically relevant phenotypes[8]. As described above, six stains were selected and imaged in five channels to reveal morphological changes for eight cellular components or organelles (Fig. 1b). Gustafsdottir's publication

established the moniker 'Cell Painting'; however, an updated protocol (v2), with minor adjustments, was published in 2016 by Bray et al.[10]. A recent effort optimized the assay's cost and reproducibility, culminating in Cell Painting v3 in 2022 (ref. 11). To create the updated protocol, the JUMP-CP (Joint Undertaking for Morphological Profiling – Cell Painting; www.jump-cellpainting.broadinstitute.org) Consortium used a positive control plate of 90 compounds covering 47 diverse mechanisms of action to, for the first time in our knowledge, quantitatively optimize staining reagents, as well as experiment and imaging conditions[11]. Other studies optimized parameters, such as the duration of cell culture and image-acquisition conditions[12,13].

Cell line selection. Flat cells that rarely overlap are best for image-based assays—most cell lines meet this criterion. In general, dozens of cell lines have performed well in Cell Painting experiments, and thus the choice often depends on the goal. For example, the JUMP-CP Consortium used U2OS (osteosarcoma) cells because large-scale data existed in this cell type, and Cas9-expressing clones are available[11,13,14].

A recent study investigated the selection of optimal cell lines for image-based profiling, because different cell lines can vary in their sensitivity to specific mechanisms of action (MoAs) of compounds[15]. Cell Painting was used to profile 3,214 small molecule compounds on six cell lines: A549, OVCAR4, DU145, 786-O, HEPG2, and a fibroblast cell line derived from a person without cancer. These compounds were all annotated with information about their putative target and MoA, and included FDA-approved drugs. The cell lines were ranked on the basis of their ability to detect compound activity (termed 'phenoactivity', or 'phenotypic activity') and to predict the compound's MoA (termed 'phenotypic consistency'). Here, phenotypic activity refers to the strength of the morphological phenotypes detected by the Cell Painting assay, whereas MoA consistency describes the extent to which the compound phenocopies other compounds with the same annotated MoA. The best cell lines for detecting phenotypic activity had poor sensitivity for predicting MoA, and vice versa. This discrepancy could reflect the diverse genetic landscapes of different cell lines, which might influence the expression of targets and the cellular pathways. For example, the HEPG2 cell line's tendency to grow in highly compact colonies makes it difficult to detect alterations in cell organelles and thus blurred phenotypic distinctions between compound-treated and control groups. It should be noted that in this study, compounds were tested in the same well positions across plates from different cell lines, which can inflate phenotypic activity metrics. To avoid potential effects owing to well position, the locations of the compounds can be scrambled across plates.

Another study showed that the Cell Painting sample-preparation protocol was effective without any cell-line-specific adjustment across six biologically diverse and morphologically distinct human-derived cell lines (U2OS, MCF7, HepG2, A549, HTB-9, and ARPE-19)[16]. It was necessary to optimize image acquisition and cell-segmentation parameters only to account for differences in the size and three-dimensional (3D) shape of each cell line when cultured in monolayers. Most of the 14 tested reference chemicals showed a pronounced phenotypic effect across all cell lines, often below cytotoxic and cytostatic concentrations. However, for all but one chemical, the most sensitive features were different in each cell line. Thus, similar concentrations of a chemical altered the cellular morphology across cell types, but the specific morphological change depended on the cell type. Over the past decade, the basic Cell Painting protocol has been used on dozens more cell lines without adjustment, on the basis of our observations from the literature and personal communications.

Adaptations of the Cell Painting Assay. Adaptations of the Cell Painting assay have emerged that replace some of the original dyes with alternative fluorescent dyes to increase the spectral range and facilitate delineation of other cellular compartments and structures. For example, LipocyteProfiler (Fig. 1d) incorporates BODIPY to mark lipid droplets in lipid-accumulating cells to study metabolic disease[17]. In another study, MitoTracker was replaced with an antibody against human coronavirus 229E (CoV-229E) viral protein, introducing the opportunity to multiplex Cell Painting with specific targets (Fig. 1e)[18].

Expanding the range of perturbations. In addition to modifying the assay protocol, some studies explored the type of perturbation, going beyond small-molecule compounds. Singh et al. explored RNA interference (RNAi)-induced knockdown using the Cell Painting assay and found that morphological signatures were highly sensitive and reproducible, but there were off-target 'seed' effects of RNAi reagents that dominated the signatures[19]. These seed effects occur when a short region of the RNAi molecule, known as the seed sequence, binds non-specifically to multiple mRNAs. Other technologies include open reading frame (ORF) constructs that enable gene or protein overexpression[20] and CRISPR knockout to deplete expression[21]. A challenge with a target-agnostic assay, such as Cell Painting, is that compounds that are active in the assay can have multiple mechanisms of action, complicating the interpretation of a given bioactivity[20]. One practical solution is to include known reference perturbations; various sets of recommended control and landmark perturbations have been recently introduced by the JUMP Consortium, including two compound plates and ORF and CRISPR perturbation plates[12,14]. Dahlin et al. generated a set of Cell Painting and cellular health profiles for 218 prototypical cytotoxic and prototypical 'nuisance' compounds in U2OS cells in a concentration–response format (0.6–20 μM)[22]. Nuisance compounds, in this context, are substances that frequently show up as hits in screening assays but are ultimately considered undesirable because their effects are often non-specific, artifactual, or due to properties that interfere with the assay rather than a specific biological activity of interest. This set of compounds thus serves as a valuable resource of controls to include in image-based profiling experiments.

Optimization of microscopy imaging parameters. Although high-throughput imaging platforms have advanced over the past decade, improving speed and resolution, Tromans-Coia et al. found that various microscope imaging systems performed similarly and changing acquisition settings only minimally affected Cell Painting profile strengths[12]. Key setting alterations that improved morphological signatures included decreasing magnification, surprisingly, but only because this increases the number of cells imaged. The study provides a general set of recommendations for Cell Painting, applicable to several microscopes. It suggests that cells should be imaged at ×20 magnification across four to nine sites (fields of view), capturing approximately 2,500 cells per well, at least for the cell types considered in the study.

Extraction of morphological features from fluorescent images. Cell Painting images are often analyzed using software to extract morphological features, following the segmentation of cellular and subcellular structures. The open-source CellProfiler[9] software is one example; however, other solutions are also used, including proprietary ones (for a detailed review, see Smith et al.[23]). Cimini et al. note that although small-scale image analysis can be performed using CellProfiler on desktop computers, large-scale analysis (>1,000 images) can be computationally intensive and time-consuming and is best run on a high-performance computing cluster or cloud computing resource[11]. We have not discussed processing times for Cell Painting data in detail because technological advances quickly render estimates obsolete.

Alternative approaches to classical feature extraction have emerged, leveraging deep-learning models to recognize features directly from raw images.These include DeepProfiler[24], a tool specifically designed for morphological profiling of Cell Painting images; deep learning architectures such as convolutional neural networks (CNNs); and vision transformers such as DINO (a self-supervised learning

method)[25,26]. These approaches, in some cases, skip the single-cell segmentation and can increase the performance of Cell Painting profiles. For example, deep learning has shown an improvement of up to 29% over CellProfiler features when assessed using mean average precision (mAP) for classifying chemical perturbations[27]. Steps to further process the Cell Painting data, from morphological feature extraction to profile normalization to batch-effect correction, discussed later, are also continually improving.

Extraction of morphological features from label-free brightfield images. Replacing the information in fluorescent images with brightfield imaging enables the analysis of living cells over time, and reduces the costs, labor, and time needed to stain cells. Although brightfield imaging does not yield a clear contrast of all the cellular compartments labeled in Cell Painting, the use of deep-learning methods could potentially augment the information available in brightfield images, making this a worthwhile tradeoff.

In one study, deep-learning models were used to predict five Cell Painting fluorescent channel images from brightfield images, and CellProfiler features were calculated from the predicted images and the ground-truth images[28]. The models were trained on approximately 3,000 images (using one field of view per well from 17 batches) and then tested with 273 images. The predicted images achieved a mean Pearson correlation of 0.84 with the ground truth at the pixel level; the authors further compared extracted CellProfiler features from the ground-truth images versus predicted images obtained using brightfield microscopy. Although many morphological features extracted from the generated images showed substantial correlation with those from the ground-truth images (>0.6 correlation) and 30 features showed a correlation greater than 0.8, prediction of the features from the AGP (actin, Golgi, plasma membrane) and mitochondrial channels was more challenging. To determine whether this level of pixel-level and feature-level correlation is sufficient for biological goals, they performed a downstream analysis and investigated the ability of models to predict compounds similar to positive controls, finding eight compounds using ground truth and the label-free images, with four compounds in the intersection.

Another study tested the ability of CNN-based features extracted from brightfield images versus those extracted from fluorescence Cell Painting images to predict 10 MoA classes[29]. The features were additionally compared with CellProfiler features extracted from the Cell Painting images. Interestingly, all models showed comparable results in distinguishing the MoA of 231 compounds from 10 MoA classes[29]. Using activation maps, they determined which areas in the images were most activated for the deep-learning-based feature extractors, and found that the models focused on different cellular features depending on the image type used for training. For example, when predicting the MoA for the compound 4SC-202, the models had an accuracy of 0.89, 0.04, and 0.29 when using brightfield, fluorescent images, and CellProfiler features, respectively; the brightfield heatmap showed strong activation for small vesicles that are visible in the brightfield images but are not stained in the Cell Painting protocol. Despite the limited number and range of MoAs tested, this study suggests that applying deep learning to brightfield images holds great promise to augment or replace fluorescent stains in Cell Painting assays, saving time and money. In fact, early reports from the biotech company Recursion indicate that a transition from using Cell Painting to more commonly using brightfield imaging is occurring[30].

Feature selection for Cell Painting profiles. Not all morphological features extracted from cell images are informative. For a given task, or even for a general representation of cell phenotype, feature-selection methods are generally used to filter features and are available from virtually all data-analysis libraries (for example, www.scikit-learn.org). Pycytominer, a software package designed for analyzing Cell Painting data, incorporates feature-selection methods that reduce redundancy and increase informativeness of features[31]. Other approaches, such as AutoML (automated machine learning), enable the most informative features from Cell Painting datasets to be identified faster[32]. Using AutoML, Siegismund et al. found that a subset of only 20–30 features was sufficient to represent the most relevant information from the morphological profile and to successfully differentiate between the control class and perturbations. However, results will likely vary, depending on the endpoint being classified and the amount of data and diversity of phenotypes in the profiled dataset.

Normalization and batch correction for Cell Painting profiles. Experimental design of Cell Painting assays can substantially impact the efficacy of normalization methods attempting to mitigate technical variation, such as batch effects. For example, Janosch et al. explored the selection of features solely using dimensionality reduction methods on images from negative controls in order to discover new phenotypes based on only negative controls[33]. Typical analysis pipelines use Pycytominer to normalize data at the plate level, correcting each well either by using all wells on the plate, if they are not expected to be enriched in displaying a particular phenotype, or solely the negative control wells, if they are sufficient in number (using the RobustMAD method)[31]. Pycytominer also implements the sphering transformation (also called 'whitening'), which can be viewed as a multivariate standardization strategy[18,33]. Sphering Cell Painting profiles was found to increase the percent replicating score—a measure of reproducibility of replicates of each sample—from 24–30% to 83–84% (for compounds at 10 μM)[34], but these results are likely confounded by plate layout effects and have not been consistently high across studies.

When analyzing Cell Painting data, computing median profiles is a popular choice that averages data from multiple cells within a well[35]. This approach offers benefits like reduced data size, faster analysis, simplified interpretation, and potentially less noise, but it assumes cell homogeneity, which can obscure subtle differences between cells[36]. Although median profiles provide a summary of the cell population with a single value per feature, single-cell analysis offers heterogeneity analysis but at the cost of increased complexity, noise, and longer analysis time.

Ongoing research aims to best capture the cell population heterogeneity to improve profile performance without unwarranted increase in processing speed or noise; van Dijk et al. discuss cell heterogeneity in representations of cell populations in the Cell Painting assay[37]. Single-cell analysis can identify subpopulations, characterize heterogeneity, and inform experimental design, making it valuable for studying complex biological systems. Despite these advantages, median profiling remains the most popular approach for high-throughput image-based profiling campaigns.

As part of the JUMP Cell Painting Consortium, Arevalo et al. conducted a thorough analysis of 10 batch effect correction methods selected and adapted from a single-cell mRNA profiling benchmark study but applied to Cell Painting average profiles[38]. They used qualitative visualizations in combination with 10 metrics to assess performance on image-based profiles, focusing on batch effect reduction and preservation of biological signals. These methods were applied to JUMP Cell Painting Consortium data for five scenarios of increasing complexity: batches from within and between different laboratories, within and between different imaging equipment, and with low and high numbers of replicates. They found Harmony and Seurat RPCA noteworthy, consistently ranking among the top three methods for all tested scenarios while maintaining computational efficiency. Yet, overall, they found existing batch-correction methods' efficacy underwhelming. The proposed framework, benchmark, and metrics can be used to assess new batch-correction methods in the future. This work paves the way for improvements that enable the community to make the best use of public Cell Painting data for scientific discovery.

Table 1 | Large publicly available Cell Painting datasets covering compound or genetic perturbations

Dataset	Release date	Cell line	Number of unique perturbations (compounds or genetic)	References
Bray et al. Cell Painting dataset	January 2017	U-2 OS	30,616 compounds	117
Recursion RxRx3 dataset	January 2023	HUVEC	17,063 CRISPR–Cas9-mediated gene knockouts (most anonymized) and 1,674 compounds at 8 concentrations	118
JUMP-CP dataset	March 2023	U-2 OS	Over-expression of 12,602 genes, knockout of 7,975 genes using CRISPR–Cas9; 116,750 unique compounds	14
Periscope dataset	August 2023	HeLa and A549	Whole-genome pooled optical, >20,000 single-gene CRISPR–Cas9-mediated gene knockout screens	119

Deep-learning models are being explored for batch correction, aiming to distinguish noise from true biological signals in Cell Painting data. Yang et al. investigated a mean-teacher-based model called DeepNoise, which was tested on the RxRx1 dataset consisting of 125,510 fluorescent microscopy images from Recursion for the CellSignal competition[39]. The study found that DeepNoise achieved a multiclass accuracy of 99.60%, compared with 74.58% using plate-based normalization. The results should be interpreted with caution; the inaccessibility of the test dataset labels prevented extensive comparisons with other models and further analysis of predictions on each of the four cell types, and additional metrics such as specificity and sensitivity would be helpful. Leakage from training to test set is likely also a confounding issue. Despite these limitations, the study indicates that deep-learning methods may be more effective at learning batch-effect patterns in Cell Painting datasets than are standard normalization approaches.

Publicly available datasets

Cell Painting has been used to profile chemical compounds and genetic-perturbation libraries in numerous datasets that are publicly available. The Cell Painting Gallery provides a centralized location for these datasets[40]. Currently, there are four large public datasets (each representing tens of thousands of perturbations) for compound and/or genetic perturbations (Table 1). Figure 3a gives a distribution of the datasets used in the 90 studies reviewed in this work. Several visualization tools have been developed to explore the JUMP-CP dataset: Broad Institute (broad.io/jump-explore), Ardigen (ai.ardigen.com/jump-cp-consorcium), Spring Discovery (www.springscience.com/jump-cp), and the Max Planck Institute of Molecular Physiology (cpcse.pythonanywhere.com).

Applications of Cell Painting data

The processing and downstream analysis of Cell Painting data (Fig. 1c), often using machine-learning and statistical approaches, has enabled the identification of complex patterns and accurate predictions for many goals. Machine-learning algorithms are particularly well-explored for analyzing morphological profiles to predict the activity, safety, and toxicity of unknown compounds, both in vitro and in vivo, and to predict MoAs and targets. Supervised methods are used when labeled datasets are available (that is, the 'correct answer', or ground truth, is known for each sample), enabling the algorithms to be trained to predict specific outcomes and patterns from feature representations. Unsupervised methods are used to investigate the similarities among samples in the feature space itself (for example, to group genes or compounds) without needing labeled data. In the following sections, we discuss the applications of Cell Painting data, using machine-learning and statistical approaches, to aid drug discovery.

Predicting mechanisms of action

Cell Painting offers an unbiased view of cellular responses to compound perturbations, enabling the identification of MoAs for compounds that induce specific morphological changes (Table 2). This identification process involves comparing the phenotype of a query compound with those of 'landmark compounds' with known MoAs, or with the gene encoding the protein targeted by the compound (or other genetic perturbations in the same pathway). However, defining a compound's MoA is complex because compounds can have multiple targets with varying affinities. Additionally, protein targets, and proteins up- and downstream of the direct targets, can be differentially expressed in various cell types. As a result, binary annotations for MoAs in datasets often oversimplify reality (see Trapotsi et al. for more details)[41]. Moreover, the breadth of MoAs that can be adequately described by any profiling assay, including Cell Painting, transcriptomics, or proteomics, is limited because none can capture every possible cellular response. The applicability of Cell Painting data for each MoA must be established separately. The most commonly detected MoAs in Cell Painting readouts include microtubule modulation[42,43], DNA-damaging agents[21], mitochondrial membrane depolarisation[43–45], lysosomotropism[46], and inhibitors of the plasma membrane Na^+ pump[43], among others (Table 2). It is important to note that MoA is a broad term, and studies in this section may involve protein-target-related MoA predictions (often termed drug–target interactions) or biological-process-related MoA predictions.

The impact of stain choices and cell type on MoA identification has recently been evaluated. Cimini et al. analyzed 90 compounds comprising 47 diverse MoA classes, and found that collected data were robust to dropping individual Cell Painting channels; however, datasets were small, and specific phenotype(s) of interest may depend on compartments that were less critical for the compounds in the study[11]. Although they did not use Cell Painting, Cox et al. studied the impact of diverse fluorescence markers and cultured cell lines in morphological profiling of 1,008 approved drugs and well-characterized compounds (218 unique MoAs) at four concentrations. They used 15 reporter cell lines (three cell lineages labeled with 12 organelle and pathway markers grouped in five combinations)[47]. The best individual cell line was able to distinguish 20 of 83 MoAs (the authors considered MoAs that had at least three active compounds). The number of distinguishable MoAs increased with each cell line but quickly plateaued; ultimately, 41 of 83 MoAs were readily distinguished using data from all 15 reporter cell lines.

Computational approaches can also improve MoA classification accuracy. For example, Janosch et al. explored dimensionality-reduction methods to improve compound MoA classification[33]. They hypothesized that if a feature remains reproducible across all negative control wells, any significant changes would likely be due to a perturbation, rather than a technical variation. They benchmarked their method with a L1-norm regularization then classified MoAs using the reduced features. Removal of the noisiest parameters improved MoA classification accuracies (from 17.66% to 20.19% for the BBBC022 dataset) and showed better generalization when they trained models without seeing a class. Therefore, this method can be used to select features for downstream analysis, and the authors suggest that it could be improved by applying deep-learning methodologies.

The MoA of many compounds has been determined using Cell Painting and a 'guilt by association' strategy[48]. Autoquin, a previously uncharacterized autophagy inhibitor, was found to induce similar morphological changes in the Cell Painting assay to those caused

https://doi.org/10.1038/s41592-024-02528-8

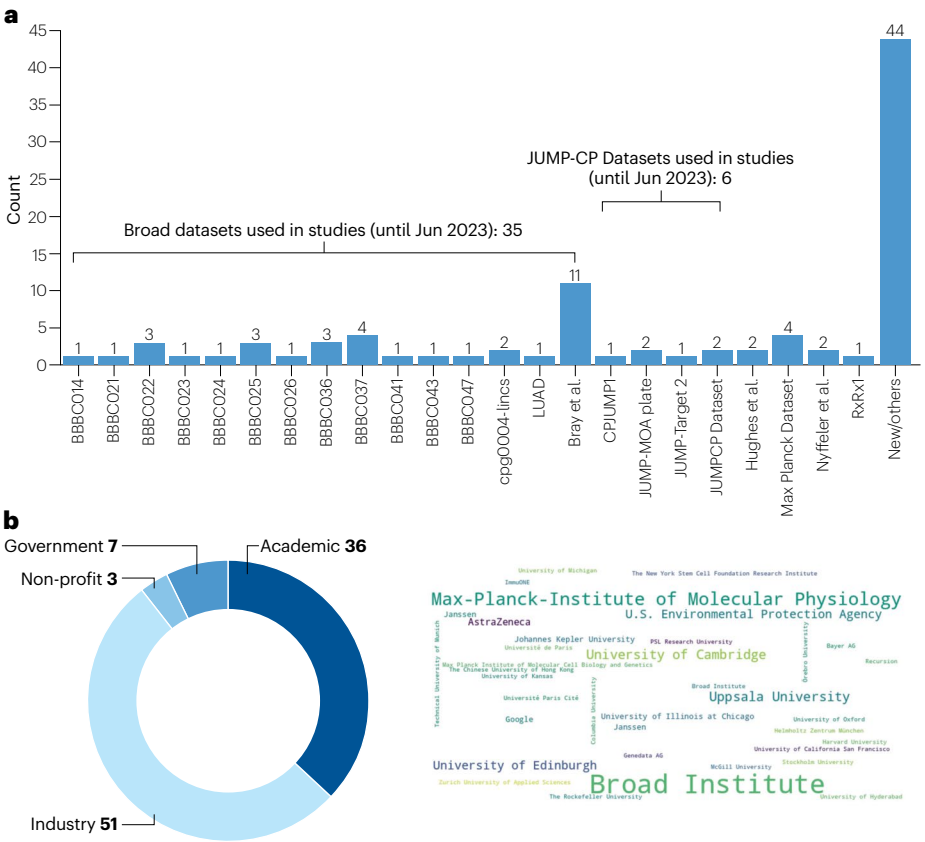

Fig. 3 | Analysis of datasets and leading institutions in studies using Cell Painting datasets. a, Analysis of the frequency of various Cell Painting datasets used in reviewed studies. Of the 90 studies reviewed in this work (some studying more than one dataset), smaller scale datasets or in-house datasets were analyzed in at least 44 studies, Broad Institute datasets in 35 studies, and the JUMP-CP dataset was used in at least six studies, despite its recent release. 'New/Others' refer to studies using datasets that were smaller in scale and/or in-house datasets that were not released publicly. **b**, Number of academic institutions, government agencies, pharmaceutical companies, and non-profit organizations that led studies evaluated in this work and/or are members of the JUMP-CP or OASIS consortium (left), with word cloud illustration (right). Further details are available in Supplementary Table 2 and Supplementary Table 4.

by oxautin-1 (ref. 49). Both autoquin and oxautin-1 have been shown to inhibit autophagy through indirect modulation of the activity of lysosomal enzymes, and autoquin has been revealed to increase Fe^{2+} levels in lysosomes. Svenningsen et al. investigated the mechanism of action of 9-methylstreptimidone, finding a high Pearson correlation ($\rho = 0.94$) with cycloheximide, a known protein-synthesis inhibitor, and confirmed its similar effects in a dose-dependent manner[50]. Other compounds have been found to modulate microtubules[42] and inhibit pyrimidine biosynthesis (inhibiting target dihydroorotate dehydrogenase)[51].

Despite these successes, predicting MoAs for cases in which reference (or 'landmark') compounds do not exist remains a significant challenge. These unclassified MoAs involve changes in cellular behavior not seen in reference compounds, making it difficult to identify and characterize these mechanisms through guilt by association. The lack of prior knowledge and annotations for unclassified MoAs hinders the development of effective machine-learning models for prediction. Integrating orthogonal approaches can provide a more comprehensive understanding of cellular behavior and alleviate this barrier. In one instance, combining Cell Painting with thermal proteome profiling enabled the discovery of diaminopyrimidine DP68 as a σ1 receptor antagonist[52]. In this case, Cell Painting revealed many potentially lysosomotropic central-nervous-system-targeting drugs that were biosimilar to diaminopyrimidine DP68, and it was challenging to discern whether the lysosomotropic effect or the σ1 receptor interaction was responsible for the phenotype[52]. Thermal proteome profiling allowed the authors to narrow it down to the σ1 receptor interaction.

Cell Painting assays have also been used to determine the MoA of dark chemical matter (DCM). DCM compounds have drug-like features but, after hundreds of assays, have revealed no biological activity, raising the question of whether they are indeed inert or whether their biological activity is very precise and just not yet discovered[53]. Pahl et al. profiled 7,700 DCM compounds with the Cell Painting assay and discovered that a remarkable 12% resulted in significant morphological changes (compared with a DMSO control)[54]. They were able to select a subset of morphological features most affected by compounds from 13 distinct bioactivty clusters. These subprofiles could then be used to identify the MoAs for new compounds by comparing them with compounds with known MoAs. Using their approach, the authors identified compounds associated with microtubule modulation, DNA synthesis, and pyrimidine biosynthesis.

Not all MoAs are desirable; in fact, some compounds are known to be frequent hits in assays with well-characterized toxic effects and are thus dubbed nuisance compounds. Dahlin et al. found that some morphological clusters of nuisance compounds were associated with cellular injury (for example, a genotoxin cluster and a tubulin poisons cluster)[22]. Other clusters represented other MoAs, including non-specific 'historical' KAT inhibitors (hKATIs). Interestingly, at higher concentrations, some compounds triggered a cellular response called a 'cytotoxicity burst' that activated multiple stress responses within the cell, rather than affecting a singular target.

https://doi.org/10.1038/s41592-024-02528-8

Table 2 | Mechanisms of action detected by Cell Painting

MoA	Biological process, direct targets	Source
Cell cycle inhibition	Cell cycle arrest in the S or G2 phase	66,120
Microtubule disruption	Microtubule organization, aurora kinase inhibitors,tubulin polymerization inhibitor	42,58,66
Cytoskeletal disruption	Actin dynamics, microtubule destabilizers	42
Protein synthesis inhibition	Protein synthesis inhibitors	66
DNA damage	Ribonucleotide reductase inhibitors, PARP inhibitors, topoisomerase inhibitors, pyrimidine biosynthesis	58,66,121
Apoptosis induction	Caspase activation, mitochondrial disruption, death receptor signaling	43,80
Autophagy	Autophagosome formation, lysosomal degradation, autophagy flux	49,84
Membrane integrity disruption	Membrane poration, lipid peroxidation	22
Mitochondrial dysfunction	Mitochondrial respiration, ATP synthesis	43,122
Oxidative stress	ROS production, antioxidant response	84
ER stress	Unfolded protein response, ER-associated degradation	84
Hormonal modulation	Hormone receptor activation, signal transduction, retinoid receptor agonists	66
Lipid metabolism inhibition	Lysosomotropism and cholesterol homeostasis regulation, HMGCR inhibition	66,121
Signal transduction inhibition	ALK tyrosine kinase receptor inhibition, src inhibitor, JAK inhibitors, AKT–PI3K–MTOR inhibitors	58,66,121
Ion channel modulation	Na^+/K^+ ATPase	121
Epigenetics	HDAC inhibitors, BET proteins	66,121
Metabolism	PPAR receptor antagonism, carbonic anhydrase inhibition HMGCR inhibition, ATPase inhibitors	58,66
Protein homeostasis	HSP inhibitors	66
Adhesion disruption	Cadherin function, integrin signaling	
Angiogenesis inhibition	VEGFR inhibitor	122
Immune modulation	Receptor antagonist, tumor necrosis factor	122
Proteolysis inhibition	Matrix metalloprotease inhibitors	58

Cell Painting has detected various MoAs across multiple pathways, as shown in publicly available studies. However, this is not an exhaustive list, and MoAs are not always annotated. As such, signals for MoAs in Cell Painting data must be established on a case-by-case basis.

MoA identification has been enhanced by the application of deep learning for image feature extraction. Instead of using pre-defined classical image features, deep-learning feature extractors are adept at extracting meaningful information directly from Cell Painting images and can therefore reduce bias and improve reproducibility by eliminating the need for feature engineering (Fig. 4).

An early study by Durr et al. developed a CNN that classified single-cell phenotypes on the basis of images generated from Cell Painting assays[55]. They trained CNNs to classify MoAs using approximately 40,000 single-cell images for 75 bioactive compounds[55]. For unbiased testing, they used 2,223 cells in a test set, from perturbations of taxol, procaine, and peruvoside; these were excluded from the training set. The CNN model, misclassifying 2.7% of all cells, performed better than CellProfiler features in a linear discriminant analysis classification model, which misclassified 5.4% of cells.

In addition to CNNs, transfer learning—in which the knowledge of a pre-trained model is transferred or fine-tuned to another model to perform a similar task—is useful, reducing the need to train models from scratch; this is being increasingly explored for cell imaging data[56]. Kensert et al. used CNN models pretrained on the 13 million natural images in the ImageNet dataset to predict 12 different MoAs across 38 compounds and 103 treatments (compound–concentration pairs)[56]. The pre-trained models achieved an accuracy between 95% and 97%; training from scratch resulted in an accuracy between 70% and 91%.

Weakly supervised deep learning can extract useful representations for unsupervised tasks by training a model on a pretext task, such as identifying replicates of biological samples from among all samples in the experiment. Moshkov et al. improved this approach by training an EfficientNet architecture on phenotypically strong compounds from diverse Cell Painting datasets, enhancing downstream analysis for matching biological compounds by 30% compared with classical features[24]. Advances in representation learning are expected to further improve results in various unsupervised Cell Painting applications beyond MoA determination. These learned representations can capture important phenotypic features and confounding factors, which can then be used to improve downstream analyses, such as predicting MoAs or matching biological compounds. Several studies have explored various approaches to representation learning using deep learning in Cell Painting assays[57,58,59].

Cell Painting in assay activity prediction

Images have proven useful in predicting compounds with desirable biological activity and identifying novel active compounds with therapeutic potential. A landmark study by Simm et al. in 2018 used images from a three-channel glucocorticoid receptor high-throughput assay to predict the activity of compounds in 535 assays spanning a wide variety of biological pathways and diseaseareas[60]. Although not using the full Cell Painting assay, the multitask models performed very well (AUC-ROC > 0.90) for 31 out of 535 assays; they validated two of these in prospective follow-up work. This study generated substantial interest in using Cell Painting for assay prediction. Hofmarcher et al. later used publicly available Cell Painting and ChEMBL data to explore the ability of CNNs trained directly using Cell Painting images to predict particular bioactivities of compounds in 209 biological assays, achieving a mean AUC-ROC of 0.73 (ref. 61). These performed better than fully connected neural networks trained using classically extracted numerical features (mean AUC-ROC = 0.68), suggesting that raw-image CNN models can capture information that might be overlooked by models trained on predefined CellProfiler features[61].

Nyffeler et al. compared various computational strategies, including multiconcentration approaches and single-concentration approaches, to identify bioactivity hits using a Cell Painting assay and showed that nine of ten approaches were highly concordant for 82% of the tested chemicals[62].

Cell Painting has also been integrated with chemical structure information to improve prediction ability. When a BMF Macau model was used with Cell Painting profiles as supplementary data, assay prediction accuracies were notably higher, with AUCs greater than 0.80 for 100 out of 224 targets. By contrast, models relying solely on chemical structural data reached AUCs greater than 0.80 for only 90 out of 224 targets[63]. In another study, models combining structural information with Cell Painting profiles, using similarities to training data, had improved performance over those that used only chemical structure information: for the combination models,

https://doi.org/10.1038/s41592-024-02528-8

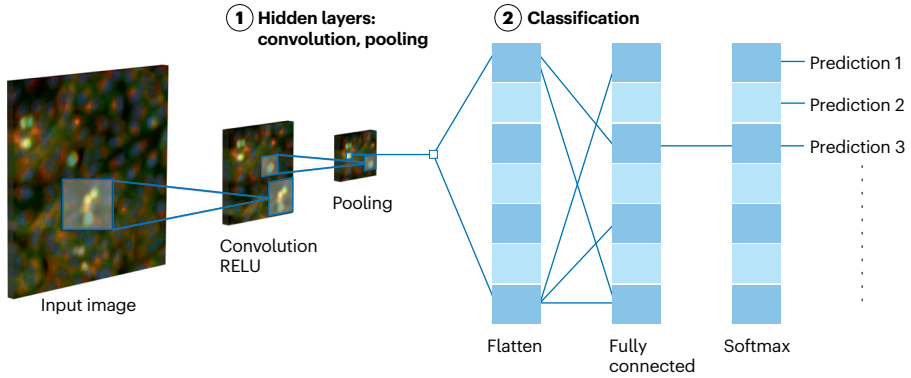

Fig. 4 | Summary of convolutional neural network analyses of Cell Painting images, one type of deep-learning network that can be used to extract image features. The input image consists of a matrix with pixel values, which can be a single cropped cell or a larger field of view. The convolution filters (smaller weight matrices) slide over the input image, detecting patterns such as edges, textures, and shapes, resulting in a feature map. An activation function (for example, ReLU) is then applied elementwise, which introduces non-linearity into the model. Pooling then reduces the spatial dimensions of the feature maps (Step 1). The final step involves extracting high-level features from the image which can then be used for model training (Step 2).

an additional 20% of assays had an AUC greater than 0.70 (79 out of 177 assays)[64].

Most recently, Sanchez-Fernandez et al. developed CLOOME, a multi-modal contrastive learning algorithm, to combine chemical structure data and Cell Painting images into a unified space. Their retrieval system correctly identified the image corresponding to a given compound with an accuracy approximately 70 times higher than that of a random baseline model; this system was also used to predict compound activity (in a similar setting as Hofmarcher et al.), and CLOOME achieved an AUC of 0.714 ± 0.20 across all prediction tasks[65]. This result indicates that the learned representations are transferable to different tasks (in this case bioactivity prediction) because no activity data were used to train the CLOOME encoders. Using images directly therefore enables unbiased insight into information contained within that image without requiring classical feature-extraction algorithms. Overall, several studies have now confirmed that adding Cell Painting image data to chemical structure information can enhance prediction ability[64,66–68].

Phenotypic profiling of bioactivity to guide chemistry

Phenotypic profiling, particularly with Cell Painting, is increasingly being used to evaluate and characterize compounds as they are synthesized, assessing the bioactivity of analogs with small changes to their structure, a process known as determining structure–activity relationship (SAR). Gerry et al. piloted this concept by synthesizing a small set of compounds and annotating their biological effects through Cell Painting[69]. They concluded that this rapid-feedback analysis could reveal unexpected phenotypes and advocated for integrating Cell Painting annotation into routine synthetic organic chemistry to accelerate medicinal chemistry.

Nelson et al. explored the use of Cell Painting-based phenotypic profiling to compare the biological activity of certain sp^3-rich compounds (carbon atoms with four single bonds), and found that two epoxy ketone diastereomers induced consistent, striking morphological changes for all doses, prompting studies to compare their morphological signatures with those of reference compounds[70]. Studies using structurally diverse, reduced flavones and their Cell Painting profiles have shown that the fraction of sp^3-hybridized atoms is not the only factor in enhanced biodiversity; stereochemistry and appendage diversity are also contributors[71,72].

More recently, biology-oriented synthesis has focused on pseudo-natural products. To this end, Christoforow et al. characterized the potential bioactivity of new classes of pyrano-furo-pyridone (PFP) pseudo-natural products[73]. They found that, among the five initial hits exhibiting bioactivity in the assay, the morphological profiles exhibited more than 70% similarity to the reference compound profiles; this then helped them to decipher their MoAs. Other studies have explored the use of target-agnostic Cell Painting to determine the phenotypic roles of novel compounds compared with reference compounds, including indocinchona alkaloids[74], a natural-product inspired flavonoid library[75]; spiroindane pyrrolidines[76]; pyrroquinoline pseudo-natural products[77]; and indofulvin pseudo-natural products[78]. Overall, phenotypic profiling enables rapid evaluation of the bioactivity, and sometimes MoA, of synthesized compounds without requiring a custom assay for each chemical series.

Predicting compound toxicity

Cell Painting assays have been used to predict multiple safety- and toxicity-related assay outcomes. For example, Way et al. found that Cell Painting data could predict aspects of cell health beyond a simple cell count, including the percentage of dead cells ($R^2 = 0.62$), number of S-phase cells ($R^2 = 0.64$), level of DNA damage in G1-phase cells ($R^2 = 0.51$), and percentage of apoptotic cells ($R^2 = 0.37$)[21]. In addition to the specific cell health phenotypes mentioned above[21,79], the Cell Painting assay could predict the outcomes of 12 cytotoxicity- and proliferation-related in vitro assays using Cell Painting profiles[80]. These models achieved an AUC of 0.71, compared with an AUC of 0.56 achieved by models using only chemical structure data (Morgan fingerprints). Trapotsi et al. successfully predicted mitochondrial toxicity with an AUC = 0.93 when using Cell Painting profiles[81]. Interestingly, this study included both small-molecule compounds and proteolysis targeting chimeras (PROTACs), which have garnered attention owing to their unique bifunctional nature and potential ability to degrade 'undruggable' targets[81]. The authors showed that Cell Painting could identify PROTAC phenotypic signatures, which were often unique as compared with the Cell Painting profiles of their individual components. Combining proxy-DILI labels with chemical and pharmacokinetic features achieved improved detection accuracy and differentiation between animal and human DILI sensitivity; it remains to be seen whether Cell Painting can be used for DILI prediction[82], which is one of the aims of the recently established OASIS consortium (www.oasisconsortium.org). Cell Painting is thus particularly useful in assessing the toxicity of new therapeutic modalities that lack the experience and best practices that have been established for small molecules.

Cell Painting is an in vitro cell-based assay, but several studies have explored its ability to predict compound perturbation effects in whole organisms. Nyffeler et al. performed an in vitro-to-in vivo extrapolation (IVIVE) of in vitro potency estimates obtained through Cell Painting[83] and compared them with effect values from in vivo mammalian

toxicity studies. Sixty-eight percent of the Cell Painting-based results were either similar to or more conservative than those of the in vivo studies, affirming the potential use of Cell Painting data in IVIVE[83]. More recently, Nyffeler et al. extended this strategy to assess whether the concentrations of environmental chemicals that induce strong phenotypic changes in the Cell Painting assay could predict unintended biological effects at the exposure at levels humans typically encounter in real-world scenarios[84]. On a larger scale, the US Environmental Protection Agency is working towards using transcriptomics and Cell Painting data in their risk assessments[85]. In the future, new sources of relevant in vivo toxicity annotations, such as hepatotoxicity[86,87] and cardiotoxicity[88], and the inclusion of pharmacokinetic information[89,90], might be used to augment Cell Painting data in predictive toxicology.

Phenotypic profiling of compound mixtures

Cell Painting has also been used to profile compound mixtures. For example, Pierozan et al. identified synergy in the responses of human breast epithelial cells to low-concentration mixtures of perfluorooctanoic acid (PFOA) and perfluorooctanesulfonic acid (PFOS), two widely used industrial chemicals[91]. Rietdijk et al. explored using Cell Painting to profile the effects of combining environmental chemicals on four cell lines; in one example, bisphenol A (BPA) did not cause significant morphological changes to cells when screened on its own, but in three out of four cell lines, it caused synergistic effects when combined with another industrial chemical[92]. Computational methods are still in development, to attempt to deconvolute the impacts of compounds on multiple target proteins in the related case of polypharmacology[93].

Using Cell Painting assays to advance disease understanding

Understanding disease biology and developing potential therapeutic interventions involves many steps, including disease modeling and biomarker discovery. The Cell Painting assay has been used to uncover functional associations between human genes and disease-associated alleles on the basis of cellular morphology when the relevant genes are perturbed or alleles are present. Rohban et al. used this approach to group 110 genes with a detectable phenotype, revealing a previously unknown interaction between the NF-κB pathway and Hippo pathways[94], and then later to identify promising compounds that match a desired phenotypic profile affecting those pathways[95]. Cancer-associated somatic variants could also be overexpressed in cells, to predict their functional impacts, including those at the single-cell level[20].

Morphological signatures can also serve as biomarkers for disease diagnosis or prognosis, or be used to monitor therapeutic responses. For example, Cell Painting assays have been used to model cancer cell morphologies to identify the distinct morphological signatures associated with esophageal adenocarcinoma and responses to selective modulators for these phenotypes[96,97].

In antiviral drug discovery, Cell Painting has been used to identify virus-induced phenotypic signatures that distinguish infected from non-infected cells[18,98,99]. Rietdijk et al. showed that treatment of infected cells with a panel of various host- and virus-targeting antivirals could reverse the morphological profile of infected host cells towards that of a non-infected cell[92]. Further, Cell Painting has been used to investigate transcription factor EB (TFEB) signaling and lysosomal dysfunction by detecting phenotypic changes in organelles in response to TFEB localization[100]. Finally, Kelley et al. used the assay to investigate drug resistance in anti-cancer therapy by identifying the morphological signature of bortezomib treatment resistance in cells[101].

Cell Painting has also been used to create disease models by leveraging natural human genetic variation. Tegtmeyer et al. aimed to investigate relationships between genetic variants and cellular morphology in induced pluripotent stem cells (iPSCs), identifying several novel associations (cell morphology quantitative trait loci)[102]. McDiarmid et al. used Cell Painting data to reveal 16 FDA-approved drugs in five mechanistic groups that could reverse morphological signatures associated with Alzheimer's disease linked to variants of the *SORL1* risk gene in neural progenitor cells[103]. Schiff et al. showed that unbiased phenotypic profiling using Cell Painting and primary fibroblasts from 91 people with Parkinson's disease and matched healthy controls could distinguish LRRK2 and sporadic Parkinson's disease lines from healthy controls[104]. Another pilot study, involving 12 healthy controls and 12 participants with the severe genetic neurological disorder spinal muscular atrophy (SMA), demonstrated that a CNN model trained on Cell Painting data from primary skin fibroblasts could distinguish disease states in cells from an unseen control–SMA pair among individuals with the disease[105]. Overall, the broad and multidimensional data generated by Cell Painting assays not only provide opportunities for new insights into complex cellular responses, but can also reveal novel therapeutic targets and strategies for drug discovery and repurposing.

Integrating Cell Painting, transcriptomics, and proteomics data

Given that Cell Painting readouts describe only one category of biological phenotype, predictive models might be improved by integrating Cell Painting data with further biological data, such as gene expression and proteomic data[106]. Nassiri and McCall compared Cell Painting data with gene expression data from the Library of Integrated Network-Based Cellular Signatures for improved insight into MoAs[107]. They used a reference database of 9,515 compounds to identify significant associations between changes in cell morphology and gene expression, revealing an interdependence that assists in inferring compound MoAs.

The relationship between Cell Painting and gene expression data was also explored in studies showing that, together, they provide complementary information for mapping cell states[34,108]. In Way et al., compounds sharing the same MoA were grouped together by Cell Painting 44% of the time and by mRNA profiles 50% of the time (across all doses); when combined, this reached 69% (ref. 34). Likewise, Haghighi et al. used morphology features from Cell Painting perturbations to predict levels of gene expression[108]. They found that cell morphology data could capture distinct biological information that was often not associated with the particular stains in the Cell Painting assay, highlighting that Cell Painting captures more information than just that for particular labeled components or some of these genes had unannotated functions. Another study combined RNA-sequencing and Cell Painting data to estimate the phenotype-altering concentration of a set of 11 mechanistically diverse compounds, and found that for 10 out of 11 compounds, both modalities could determine potency estimates within half an order of magnitude[109].

Morphological features and gene expression data were also used by Cerisier et al. to explore associations between chemicals and disease by developing a biological network combining chemical–gene–pathway–morphological perturbation and disease relationships[110]. They investigated two chemicals (amiodarone and prochlorperazine) because both showed a risk for drug-induced liver injury (DILI) in humans and thus, they assessed whether they share common information in Cell Painting and the L1000 dataset. They found a direct link between deregulated genes and cell morphology observations. Seal et al. used machine learning to predict mitochondrial toxicity by combining gene expression, Cell Painting data, and chemical structures, with detection scores of 0.40, compared with 0.25 using chemical structures alone[43]. The combination of cell morphology and gene expression at the single-cell level is an exciting research area that remains to be explored, given the higher computational demands for single-cell analysis and the technical challenges in measuring both kinds of profiles in the same assay[111].

The benefits of integrating Cell Painting data might also apply to protein profiling. One study tested 306 well-characterized compounds with established MoAs using Cell Painting and nELISA protein profiling, finding a 26.7% and 21.2% retrieval rate for MoA classes,

respectively, with an additional 33% of MoAs when combining nELISA with Cell Painting[112]. Another study combined morphological profiling with proteome analyses to reveal lysosomotropic activity leading to cholesterol homeostasis and localization for tetrahydroindolo[2,3-a] quinolizine derivative, a natural-product-inspired compound[113]. Overall, considering multiple modalities can help in elucidating chemical–phenotype observations. The integration of Cell Painting with diverse and complementary -omics modalities, such as transcriptomics and proteomics, can offer more comprehensive insights into the biological impacts of compounds.

Status check: lessons learned from the first 10 years

Over the past decade, the Cell Painting assay has become a widely used tool for drug discovery and cell biology. Cell Painting data are being used in more than 36 academic laboratories or screening centers and 51 companies worldwide, with at least 13 offering the assay as a service, providing a range of outputs including images, profiles, or matching of customer samples to an internal database of samples (Fig. 3b and Supplementary Table 4). At least four candidate therapeutics discovered using Cell Painting have entered phase 2 clinical trials (www.recursion.com/pipeline). Cell Painting uses the maximum number of channels of a typical microscope and minimizes costs and wash steps, making it amenable to screening millions of samples in a high-throughput pharmaceutical setting and testing a few samples in any academic laboratory. The assay is remarkably robust; extensive testing of different staining and imaging conditions yielded relatively similar results[11,12]. Together with image-based profiling using other assays, Cell Painting has provided insights into the complex world of cellular morphology, expanding our understanding of disease mechanisms and enhancing drug-discovery processes.

So far, Cell Painting has been most widely used for uncovering compounds' MoAs, by mapping similarity to compounds with known mechanisms. Identification of disease-associated phenotypes and prediction of assay outcomes (including various types of toxicity) are increasingly common. The potential of Cell Painting for predicting organ-level toxicity is promising and is actively being explored by the OASIS Consortium (https://oasisconsortium.org/). Less common but proven applications include characterizing newly synthesized compounds to discern structure–activity relationships, functionally annotating gene and allele impact, identifying compound mimics of a gene perturbation, designing diverse and phenotypically impactful compound libraries, and lead hopping.

The Cell Painting assay continues to evolve with advances in staining protocols, cell line diversity, and data-analysis algorithms, but its core strengths lie in striking a balance between speed, cost-effectiveness, easy implementation, and information richness, with the added potential for automation. Its affordability enables high-throughput screening, including multiple doses, varied time points, and additional cell lines per perturbation, providing a comprehensive biological perspective. Recent publications from Bayer, Pfizer and AstraZeneca, among others, have shown that Cell Painting is a promising tool in the pharmaceutical industry[58,68,44]. The potential for automation further increases its appeal, enabling large-scale phenotypic screening with minimal manual intervention.

Challenges and future directions

There are many avenues for improving image-based profiling in the future. Matching profiles across modalities—from compounds to genetic perturbations—has proven difficult so far[13]. Improving methodology for this task would accelerate several applications, such as identifying compounds with the same cellular impact as a given gene of interest. Also, although Cell Painting is widespread in matching the MoA of a query compound to known compounds, it remains to be tested whether matching compounds to genetic perturbations might aid in the much more challenging situation in which no compounds exist for a given mechanism.

Interpretation of morphological profiles is another challenge. Although sophisticated image-analysis algorithms and machine-learning methods can extract and analyze complex morphological signatures, interpreting them can be challenging, even for classical algorithms in which features are precisely defined mathematically. The BioMorph space attempts to address this by linking 827 Cell Painting features to 412 descriptive terms, on the basis of mapping to assays capturing phenotypes of cell health[79]. However, for broad utility, mapping to more assay data would be needed.

Data handling and storage requires attention—the high-content nature of the Cell Painting assay generates vast amounts of data, which can be challenging to store, manage, and share, and the data require considerable computational resources to process and analyze. Cloud-based solutions and open-source software tools can address these challenges[114], but increasing their user-friendliness would expand the use of this data type.

We see great promise in extending the Cell Painting assay to 3D cell cultures, organoids[115], tissue slices, and live cell imaging[116]. Improvements in deep-learning methods are also expected to dramatically alter Cell Painting, particularly in batch-correction methods that can extract biologically meaningful signals from technical noise[38].

The recent availability of large Cell Painting datasets stands to empower a new wave of discoveries, particularly as methods for matching new batches of data to public sets mature. Consortia serve a valuable role in creating these data sets and evaluating applications of Cell Painting data, by pooling resources and contributing expertise to experimental design. In summary, with the availability of larger datasets, increased academic and industry interest, and the potential for collaboration through consortia, the future of Cell Painting looks bright.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

References

1. Swinney, D. C. & Anthony, J. How were new medicines discovered? *Nat. Rev. Drug Discov.* **10**, 507–519 (2011).
2. Lin, A. et al. Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Sci. Transl. Med.* **11**, eaaw8412 (2019).
3. Moffat, J. G., Vincent, F., Lee, J. A., Eder, J. & Prunotto, M. Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nat. Rev. Drug Discov.* **16**, 531–543 (2017).
4. Perlman, Z. E. et al. Multidimensional drug profiling by automated microscopy. *Science* **306**, 1194–1198 (2004).
5. Schulze, C. J. et al. 'Function-first' lead discovery: mode of action profiling of natural product libraries using image-based screening. *Chem. Biol.* **20**, 285 (2013).
6. Feng, Y., Mitchison, T. J., Bender, A., Young, D. W. & Tallarico, J. A. Multi-parameter phenotypic profiling: using cellular effects to characterize small-molecule compounds. *Nat. Rev. Drug Discov.* **8**, 567–578 (2009).
7. Woehrmann, M. H. et al. Large-scale cytological profiling for functional analysis of bioactive compounds. *Mol. Biosyst.* **9**, 2604–2617 (2013).
8. Gustafsdottir, S. M. et al. Multiplex cytological profiling assay to measure diverse cellular states. *PLoS ONE* **8**, e80999 (2013).
9. Stirling, D. R. et al. CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinf.* **22**, 1–11 (2021).
10. 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).

11. Cimini, B. A. et al. Optimizing the Cell Painting assay for image-based profiling. *Nat. Protoc.* **18**, 1981–2013 (2023).
12. Tromans-Coia, C. et al. Assessing the performance of the Cell Painting assay across different imaging systems. *Cytom. Part A* **103**, 915–926 (2023).
13. Chandrasekaran, S. N. et al. Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. *Nat. Methods* **21**, 1114–1121 (2024).
14. 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).
15. Heinrich, L., Kumbier, K., Li, L., Altschuler, S. J. & Wu, L. F. Selection of optimal cell lines for high-content phenotypic screening. *ACS Chem. Biol.* **18**, 679–685 (2023).
16. Willis, C., Nyffeler, J. & Harrill, J. Phenotypic profiling of reference chemicals across biologically diverse cell types using the Cell Painting assay. *SLAS Discov.* **25**, 755–769 (2020).
17. Laber, S. et al. Discovering cellular programs of intrinsic and extrinsic drivers of metabolic traits using LipocyteProfiler. *Cell Genomics* **3**, 100346 (2023).
18. Rietdijk, J. et al. A phenomics approach for antiviral drug discovery. *BMC Biol.* **19**, 156 (2021).
19. Singh, S. et al. Morphological profiles of RNAi-induced gene knockdown are highly reproducible but dominated by seed effects. *PLoS One* **10**, e0131370 (2015).
20. Caicedo, J. C. et al. Cell Painting predicts impact of lung cancer variants. *Mol. Biol. Cell* **33**, ar49 (2022).
21. Way, G. P. et al. Predicting cell health phenotypes using image-based morphology profiling. *Mol. Biol. Cell* **32**, 995–1005 (2021).
22. Dahlin, J. L. et al. Reference compounds for characterizing cellular injury in high-content cellular morphology assays. *Nat. Commun.* **14**, 1364 (2023).
23. Smith, K. et al. Phenotypic image analysis software tools for exploring and understanding big image data from cell-based assays. *Cell Syst.* **6**, 636–653 (2018).
24. Moshkov, N. et al. Learning representations for image-based profiling of perturbations. *Nat. Commun.* **15**, 1594 (2024).
25. Caron, M. et al. Emerging Properties in Self-Supervised Vision Transformers. In *Proc.IEEE/CVF International Conference on Computer Vision* 9630–9640 (IEEE, 2021)
26. He, K. et al. Masked autoencoders are scalable vision learners. in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 15979–15988 (IEEE, 2022).
27. Kim, V., Adaloglou, N., Osterland, M., Morelli, F. M. & Zapata, P. A. M. Self-supervision advances morphological profiling by unlocking powerful image representations. Preprint at *bioRxiv* https://doi.org/10.1101/2023.04.28.538691 (2023).
28. Cross-Zamirski, J. O. et al. Label-free prediction of cell painting from brightfield images. *Sci. Rep.* **12**, 10001 (2022).
29. Harrison, P. J. et al. Evaluating the utility of brightfield image data for mechanism of action prediction. *PLoS Comput. Biol.* **19**, e1011323 (2023).
30. Belli, B. Brightfield is back: a 17th century cell imaging technique is making a comeback thanks to machine learning. https://www.recursion.com/news/brightfield-is-back-a-17th-century-cell-imaging-technique-is-making-a-comeback-thanks-to-machine-learning *Recursion* (2024).
31. Serrano, E. et al. Reproducible image-based profiling with Pycytominer. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2311.13417 (2024).
32. Siegismund, D., Fassler, M., Heyse, S. & Steigele, S. Benchmarking feature selection methods for compressing image information in high-content screening. *SLAS Technol.* **27**, 85–93 (2022).
33. Janosch, A., Kaffka, C. & Bickle, M. Unbiased phenotype detection using negative controls. *SLAS Discov.* **24**, 234–241 (2019).
34. Way, G. P. et al. Morphology and gene expression profiling provide complementary information for mapping cell state. *Cell Syst.* **13**, 911–923 (2022).
35. Caicedo, J. C. et al. Data-analysis strategies for image-based cell profiling. *Nat. Methods* **14**, 849–863 (2017).
36. Altschuler, S. J. & Wu, L. F. Cellular heterogeneity: do differences make a difference? *Cell* **141**, 559–563 (2010).
37. van Dijk, R., Arevalo, J., Babadi, M., Carpenter, A. E. & Singh, S. Capturing cell heterogeneity in representations of cell populations for image-based profiling using contrastive learning. *PLOS Comput. Biol.* **20**, e1012547 (2024).
38. Arevalo, J. et al. Evaluating batch correction methods for image-based cell profiling. *Nat. Commun.* **15**, 6516 (2024).
39. Yang, S. et al. DeepNoise: signal and noise disentanglement based on classifying fluorescent microscopy images via deep learning. *Genom. Proteom. Bioinform.* **20**, 989–1001 (2022).
40. Weisbart, E. et al. Cell Painting Gallery: an open resource for image-based profiling. *Nat. Methods* **21**, 1775–1777 (2024).
41. Trapotsi, M. A., Hosseini-Gerami, L. & Bender, A. Computational analyses of mechanism of action (MoA): data, methods and integration. *RSC Chem. Biol.* **3**, 170–200 (2022).
42. Akbarzadeh, M. et al. Morphological profiling by means of the Cell Painting assay enables identification of tubulin-targeting compounds. *Cell Chem. Biol.* **29**, 1053–1064 (2022).
43. Seal, S. et al. Integrating cell morphology with gene expression and chemical structure to aid mitochondrial toxicity detection. *Commun. Biol.* **5**, 858 (2022).
44. Herman, D. et al. Leveraging Cell Painting images to expand the applicability domain and actively improve deep learning quantitative structure–activity relationship models. *Chem. Res. Toxicol.* **36**, 1028–1036 (2023).
45. Garcia de Lomana, M., Marin Zapata, P. A. & Montanari, F. Predicting the mitochondrial toxicity of small molecules: insights from mechanistic assays and cell painting data. *Chem. Res. Toxicol.* **36**, 1107–1120 (2023).
46. Laraia, L., Robke, L. & Waldmann, H. BioactiVe compound collections: from design to target identification. *Chem* **4**, 705–730 (2018).
47. Cox, M. J. et al. Tales of 1,008 small molecules: phenomic profiling through live-cell imaging in a panel of reporter cell lines. *Sci. Rep.* **10**, 13262 (2020).
48. C. Herbert Waldmann—celebrating more than three decades in academia. *J. Med. Chem.* **66**, 15055–15060 (2023).
49. Laraia, L. et al. Image-based morphological profiling identifies a lysosomotropic, iron-sequestering autophagy inhibitor. *Angew. Chem. Int. Ed.* **59**, 5721–5729 (2020).
50. Svenningsen, E. B. & Poulsen, T. B. Establishing cell painting in a smaller chemical biology lab—a report from the frontier. *Bioorg. Med. Chem.* **27**, 2609–2615 (2019).
51. Schölermann, B. et al. Identification of dihydroorotate dehydrogenase inhibitors using the Cell Painting assay. *ChemBioChem* **23**, e202200475 (2022).
52. Wilke, J. et al. Discovery of a σ1 receptor antagonist by combination of unbiased cell painting and thermal proteome profiling. *Cell Chem. Biol.* **28**, 848–854 (2021).
53. Wassermann, A. M. et al. Dark chemical matter as a promising starting point for drug lead discovery. *Nat. Chem. Biol.* **11**, 958–966 (2015).
54. Pahl, A. et al. Illuminating dark chemical matter using the Cell Painting assay. *J. Med. Chem.* **67**, 8862–8876 (2024).
55. Dürr, O. & Sick, B. Single-cell phenotype classification using deep convolutional neural networks. *J. Biomol. Screen* **21**, 998–1003 (2016).

56. Kensert, A., Harrison, P. J. & Spjuth, O. Transfer learning with deep convolutional neural networks for classifying cellular morphological changes. *SLAS Discov.* **24**, 466–475 (2019).
57. Lafarge, M. W. et al. Capturing single-cell phenotypic variation via unsupervised representation learning. *Proc. Mach. Learn Res.* **102**, 315–325 (2019).
58. Wong, D. R. et al. Deep representation learning determines drug mechanism of action from cell painting images. *Digital Discov.* **2**, 1354–1367 (2023).
59. Liu, G., Seal, S., Arevalo, J. & Liang, Z. Learning molecular representation in a cell. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2406.12056 (2024).
60. Simm, J. et al. Repurposing high-throughput image assays enables biological activity prediction for drug discovery. *Cell Chem. Biol.* **25**, 611–618 (2018).
61. Hofmarcher, M., Rumetshofer, E., Clevert, D. A., Hochreiter, S. & Klambauer, G. Accurate prediction of biological assays with high-throughput microscopy images and convolutional networks. *J. Chem. Inf. Model.* **59**, 1163–1171 (2019).
62. Nyffeler, J. et al. Comparison of approaches for determining bioactivity hits from high-dimensional profiling data. *SLAS Discov.* **26**, 292–308 (2021).
63. Trapotsi, M. A. et al. Comparison of chemical structure and cell morphology information for multitask bioactivity predictions. *J. Chem. Inf. Model.* **61**, 1444–1456 (2021).
64. Seal, S. et al. Merging bioactivity predictions from cell morphology and chemical fingerprint models using similarity to training data. *J. Cheminform.* **15**, 56 (2023).
65. Sanchez-Fernandez, A., Rumetshofer, E., Hochreiter, S. & Klambauer, G. CLOOME: contrastive learning unlocks bioimaging databases for queries with chemical structures. *Nat. Commun.* **14**, 1–14 (2023).
66. Tian, G., Harrison, P. J., Sreenivasan, A. P., Carreras-Puigvert, J. & Spjuth, O. Combining molecular and cell painting image data for mechanism of action prediction. *Artif. Intell. Life Sci.* **3**, 100060 (2023).
67. Moshkov, N. et al. Predicting compound activity from phenotypic profiles and chemical structures. *Nat. Commun.* **14**, 1967 (2023).
68. Fredin Haslum, J. et al. Cell Painting-based bioactivity prediction boosts high-throughput screening hit-rates and compound diversity. *Nat. Commun.* **15**, 1–11 (2024).
69. Gerry, C. J. et al. Real-time biological annotation of synthetic compounds. *J. Am. Chem. Soc.* **138**, 8920–8927 (2016).
70. Nelson, S. D., Wawer, M. J. & Schreiber, S. L. Divergent synthesis and real-time biological annotation of optically active tetrahydrocyclopenta[c]pyranone derivatives. *Org. Lett.* **18**, 6280–6283 (2016).
71. Gerlach, E. M., Korkmaz, M. A., Pavlinov, I., Gao, Q. & Aldrich, L. N. Systematic diversity-oriented synthesis of reduced flavones from γ-pyrones to probe biological performance diversity. *ACS Chem. Biol.* **14**, 1536–1545 (2019).
72. Melillo, B. et al. Synergistic effects of stereochemistry and appendages on the performance diversity of a collection of synthetic compounds. *J. Am. Chem. Soc.* **140**, 11784–11790 (2018).
73. Christoforow, A. et al. Design, synthesis, and phenotypic profiling of pyrano-furo-pyridone pseudo natural products. *Angew. Chem. Int. Ed.* **58**, 14715–14723 (2019).
74. Foley, D. J. et al. Phenotyping reveals targets of a pseudo-natural-product autophagy inhibitor. *Angew. Chem. Int. Ed.* **59**, 12470–12476 (2020).
75. Hippman, R. S. et al. Multiple chemical features impact biological performance diversity of a highly active natural product-inspired library. *ChemBioChem* **21**, 3137–3145 (2020).
76. Singh, M., Garza, N., Pearson, Z., Douglas, J. & Boskovic, Z. Broad assessment of bioactivity of a collection of spiroindane pyrrolidines through 'cell painting'. *Bioorg. Med. Chem.* **28**, 115547 (2020).
77. Liu, J. et al. Design, synthesis, and biological evaluation of chemically and biologically diverse pyrroquinoline pseudo natural products. *Angew. Chem. Int. Ed.* **60**, 4648–4656 (2021).
78. Burhop, A. et al. Synthesis of indofulvin pseudo-natural products yields a new autophagy inhibitor chemotype. *Adv. Sci.* **8**, e2102042 (2021).
79. Seal, S. et al. From pixels to phenotypes: integrating image-based profiling with cell health data as BioMorph features improves interpretability. *Mol. Biol. Cell* **35**, mr2 (2024).
80. Seal, S., Yang, H., Vollmers, L. & Bender, A. Comparison of cellular morphological descriptors and molecular fingerprints for the prediction of cytotoxicity- and proliferation-related assays. *Chem. Res. Toxicol.* **34**, 422–437 (2021).
81. Trapotsi, M. A. et al. Cell morphological profiling enables high-throughput screening for proteolysis targeting chimera (PROTAC) phenotypic signature. *ACS Chem. Biol.* **17**, 1733–1744 (2022).
82. Seal, S. et al. Improved detection of drug-induced liver injury by integrating predicted in vivo and in vitro data. *Chem. Res Toxicol.* **37**, 1290–1305 (2024).
83. Nyffeler, J. et al. Bioactivity screening of environmental chemicals using imaging-based high-throughput phenotypic profiling. *Toxicol. Appl. Pharmacol.* **389**, 114876 (2020).
84. Nyffeler, J. et al. Application of Cell Painting for chemical hazard evaluation in support of screening-level chemical assessments. *Toxicol. Appl. Pharmacol.* **468**, 116513 (2023).
85. Thomas, R. S. et al. The next generation blueprint of computational toxicology at the U.S. Environmental Protection Agency. *Toxicol. Sci.* **169**, 317–332 (2019).
86. Chen, M. et al. DILIrank: the largest reference drug list ranked by the risk for developing drug-induced liver injury in humans. *Drug Discov. Today* **21**, 648–653 (2016).
87. Seal, S. et al. Improved detection of drug-induced liver injury by integrating predicted in vivo and in vitro data. *Chem. Res. Toxicol.* **37**, 1290–1305 (2024).
88. Seal, S. et al. Insights into drug cardiotoxicity from biological and chemical data: the first public classifiers for FDA drug-induced cardiotoxicity rank. *J. Chem. Inf. Model.* **64**, 1172–1186 (2024).
89. Horne, R. I. et al. Using generative modeling to endow with potency initially inert compounds with good bioavailability and low toxicity. *J. Chem. Inf. Model.* **64**, 590–596 (2024).
90. Seal, S. et al. PKSmart: an open-source computational model to predict in vivo pharmacokinetics of small molecules. Preprint at *bioRxiv* https://doi.org/10.1101/2024.02.02.578658 (2024) .
91. Pierozan, P., Kosnik, M. & Karlsson, O. High-content analysis shows synergistic effects of low perfluorooctanoic acid (PFOS) and perfluorooctane sulfonic acid (PFOA) mixture concentrations on human breast epithelial cell carcinogenesis. *Environ. Int.* **172**, 107746 (2023).
92. Rietdijk, J. et al. Morphological profiling of environmental chemicals enables efficient and untargeted exploration of combination effects. *Sci. Total Environ.* **832**, 155058 (2022).
93. Chow, Y. L., Singh, S., Carpenter, A. E. & Way, G. P. Predicting drug polypharmacology from cell morphology readouts using variational autoencoder latent space arithmetic. *PLoS Comput. Biol.* **18**, e1009888 (2022).
94. Rohban, M. H. et al. Systematic morphological profiling of human gene and allele function via cell painting. *eLife* **6**, e24060 (2017).
95. Rohban, M. H. et al. Virtual screening for small-molecule pathway regulators by image-profile matching. *Cell Syst.* **13**, 724–736 (2022).

96. Hughes, R. E., Elliott, R. J. R., Dawson, J. C. & Carragher, N. O. High-content phenotypic and pathway profiling to advance drug discovery in diseases of unmet need. *Cell Chem. Biol.* **28**, 338–355 (2021).
97. Hughes, R. E. et al. High-content phenotypic profiling in esophageal adenocarcinoma identifies selectively active pharmacological classes of drugs for repurposing and chemical starting points for novel drug discovery. *SLAS Discov.* **25**, 770–782 (2020).
98. Cuccarese, M. F. et al. Functional immune mapping with deep-learning enabled phenomics applied to immunomodulatory and COVID-19 drug discovery. Preprint at *bioRxiv* https://doi.org/10.1101/2020.08.02.233064 (2020).
99. 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).
100. Carey, K. L. et al. TFEB transcriptional responses reveal negative feedback by BHLHE40 and BHLHE41. *Cell Rep.* **33**, 108371 (2020).
101. Kelley, M. E. et al. High-content microscopy reveals a morphological signature of bortezomib resistance. *eLife.* **12**, e91362 (2023).
102. Tegtmeyer, M. et al. High-dimensional phenotyping to define the genetic basis of cellular morphology. *Nat. Commun.* **15**, 347 (2024).
103. McDiarmid, A. H. et al. Morphological profiling in human neural progenitor cells classifies hits in a pilot drug screen for Alzheimer's disease. *Brain Commun.* **6**, fcae101 (2024).
104. Schiff, L. et al. Integrating deep learning and unbiased automated high-content screening to identify complex disease signatures in human fibroblasts. *Nat. Commun.* **13**, 1590 (2022).
105. Yang, S. J. et al. Applying deep neural network analysis to high-content image-based assays. *SLAS Discov.* **24**, 829–841 (2019).
106. Liu, A., Seal, S., Yang, H. & Bender, A. Using chemical and biological data to predict drug toxicity. *SLAS Discov.* **28**, 53–64 (2023).
107. Nassiri, I. & McCall, M. N. Systematic exploration of cell morphological phenotypes associated with a transcriptomic query. *Nucleic Acids Res.* **46**, e116 (2018).
108. Haghighi, M., Caicedo, J. C., Cimini, B. A., Carpenter, A. E. & Singh, S. High-dimensional gene expression and morphology profiles of cells across 28,000 genetic and chemical perturbations. *Nat. Methods* **19**, 1550–1557 (2022).
109. Nyffeler, J. et al. Combining phenotypic profiling and targeted RNA-Seq reveals linkages between transcriptional perturbations and chemical effects on cell morphology: retinoic acid as an example. *Toxicol. Appl. Pharmacol.* **444**, 116032 (2022).
110. Cerisier, N., Dafniet, B., Badel, A. & Taboureau, O. Linking chemicals, genes and morphological perturbations to diseases. *Toxicol. Appl. Pharmacol.* **461**, 116407 (2023).
111. Camunas-Soler, J. Integrating single-cell transcriptomics with cellular phenotypes: cell morphology, Ca^{2+} imaging and electrophysiology. *Biophys. Rev.* **16**, 89–107 (2023).
112. Dagher, M. et al. nELISA: a high-throughput, high-plex platform enables quantitative profiling of the secretome. Preprint at *bioRxiv* https://doi.org/10.1101/2023.04.17.535914 (2023).
113. Schneidewind, T. et al. Combined morphological and proteome profiling reveals target-independent impairment of cholesterol homeostasis. *Cell Chem. Biol.* **28**, 1780–1794 (2021).
114. Way, G. P., Sailem, H., Shave, S., Kasprowicz, R. & Carragher, N. O. Evolution and impact of high content imaging. *SLAS Discov.* **28**, 292–305 (2023).
115. Lukonin, I., Zinner, M. & Liberali, P. Organoids in image-based phenotypic chemical screens. *Exp. Mol. Med.* **53**, 1495–1502 (2021).
116. Cottet, M. et al. Live cell painting: New nontoxic dye to probe cell physiology in high content screening. *SLAS Discov.* **29**, 100121 (2023).
117. Bray, M.-A. et al. A dataset of images and morphological profiles of 30 000 small-molecule treatments using the Cell Painting assay. *Gigascience* **6**, 1–5 (2017).
118. Fay, M. M. et al. RxRx3: phenomics map of biology. Preprint at *bioRxiv* https://doi.org/10.1101/2023.02.07.527350 (2023).
119. Ramezani, M. et al. A genome-wide atlas of human cell morphology. Preprint at *bioRxiv* https://doi.org/10.1101/2023.08.06.552164 (2023).
120. Schneidewind, T. et al. Morphological profiling identifies a common mode of action for small molecules with different targets. *ChemBioChem* **21**, 3197–3207 (2020).
121. Pahl, A. et al. Morphological subprofile analysis for bioactivity annotation of small molecules. *Cell Chem. Biol.* **30**, 839–853 (2023).
122. Lapins, M. & Spjuth, O. Evaluation of gene expression and phenotypic profiling data as quantitative descriptors for predicting drug targets and mechanisms of action. Preprint at *bioRxiv* https://doi.org/10.1101/580654 (2019).

Acknowledgements

S. Seal acknowledges funding from the Cambridge Centre for Data-Driven Discovery (C2D3) and Accelerate Programme for Scientific Discovery. A.E.C., S. Singh, and S. Seal acknowledge funding from the National Institutes of Health (R35 GM122547 to A.E.C.). O.S. acknowledges funding from the Swedish Research Council (Grants 2020-03731 and 2020-01865), FORMAS (Grant 2022-00940), Swedish Cancer Foundation (22 2412 Pj 03 H), and Horizon Europe (Grant Agreements 101057014 (PARC) and 101057442 (REMEDI4ALL)).

Author contributions

S. Seal and M.-A.T. designed and performed the systematic review on studies using Cell Painting data. S. Seal, M.-A.T., and A.E.C. wrote the manuscript with extensive discussions with all authors. All of the authors reviewed, edited, and contributed to discussions on the manuscript and approved the final version of the manuscript.

Competing interests

S. Singh and A.E.C. serve as scientific advisors for companies that use image-based profiling and Cell Painting (A.E.C.: Recursion, SyzOnc, Quiver Bioscience; S. Singh: Waypoint Bio, Dewpoint Therapeutics, DeepCell) and receive honoraria for occasional talks at pharmaceutical and biotechnology companies. J.C.P. and O.S. declare ownership in Phenaros Pharmaceuticals. M.-A.T. and N.G. were formerly employed at AstraZeneca. M.-A.T. and N.G. are currently employed at Recursion Pharmaceuticals. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41592-024-02528-8.

Correspondence and requests for materials should be addressed to Srijit Seal, Maria-Anna Trapotsi or Anne E. Carpenter.

Peer review information *Nature Methods* thanks Jeremy Jenkins and the other, anonymous, reviewer(s) for their contribution to the

peer review of this work. Primary handling editor: Rita Strack, in collaboration with the *Nature Methods* team.

nature portfolio

Corresponding author(s): Srijit Seal, Maria-Anna Trapotsi, Anne E. Carpenter

Last updated by author(s): 2024/08/22

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The literature search was conducted using three major databases: ScienceDirect, PubMed,and Scopus, and completed in June 2023
Data analysis	We used Matplotlib and Seaborn to generate plots in Python 3.11 (open source libraries)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

No new datasets were used in this study.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Not Applicable
Population characteristics	Not Applicable
Recruitment	Not Applicable
Ethics oversight	Not Applicable

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	90 studies were evaluated in a systematic review analysis
Data exclusions	Studies that used Cell Painting assays were included. Studies were excluded when they were not primarily in English, had been published before 2013 (before the Cell Painting assay was formally introduced), and had not been peer-reviewed (except in a few cases where we felt they were significant or included them during a manual search). Reviews, news articles, posters, thesis abstracts, and perspective papers were not included as primary research articles.
Replication	Not Applicable
Randomization	Not Applicable
Blinding	Not Applicable

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging