Published in final edited form as:
Nat Methods. 2025 February ; 22(2): 254–268. doi:10.1038/s41592-024-02528-8.

Cell Painting: A Decade of Discovery and Innovation in Cellular Imaging

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

[1]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States

[2]Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW, Cambridge, United Kingdom

[3]Imaging and Data Analytics, Clinical Pharmacology & Safety Sciences, R&D, AstraZeneca, 1 Francis Crick Avenue, Cambridge, CB2 0AA, United Kingdom

[4]Department of Pharmaceutical Biosciences and Science for Life Laboratory, Uppsala University, Box 591, SE-75124, Uppsala, Sweden

[5]Imaging and Data Analytics, Clinical Pharmacology & Safety Sciences, R&D, AstraZeneca, 35 Gatehouse Drive, Waltham, MA 02451, USA

[6]STAR-UBB Institute, Babeş-Bolyai University, Cluj-Napoca, Romania

Abstract

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 strategy 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 due to improvements in the protocol, adaptations

---

[*]Corresponding Authors: seal@broadinstitute.org, marianna.trapotsi1@astrazeneca.com, anne@broadinstitute.org.

Author Contributions

S. Seal and M.A. Trapotsi designed and performed the systematic review on studies using Cell Painting data. S. Seal and M.A. Trapotsi and A. Carpenter 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.

[‡]These authors contributed equally.

Code availability

No code was used in this study.

Competing Interests Declaration

The authors declare the following competing financial interest(s): 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. The remaining authors declare no competing interests.

Supplementary information

The Supplementary Datasets are available as Supplementary Table 1: Definitions of Essential Terms in Image-Based Profiling (XLSX); Supplementary Table 2: 90 studies included in this study (XLSX); Supplementary Table 3: 65 studies excluded from this study (XLSX); Supplementary Table 4. Academic Institutions, Government Agencies, Pharmaceutical Companies, Non-Profits who led studies evaluated in this work and/or are members of the JUMPCP and OASIS consortiums (XLSX)

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.

Editor’s summary

This Review synthesizes the literature from over ten years of Cell Painting for image-based profiling and highlights how advances in this technology enable new biological discovery regarding cellular phenotypes and cell responses to perturbations.

INTRODUCTION

Phenotypic drug discovery (PDD) identifies compounds that alter a given disease phenotype in a living system. PDD has evolved from screening a few compounds in animals to testing millions in cell models. In contrast, target-based drug discovery (TDD) identifies compounds that interact with a pre-selected target. While both approaches have yielded therapeutics, mounting evidence suggests that PDD yields more first-in-class medicines than TDD.[1] Notably, many FDA-approved drugs lack a defined molecular target, and several drugs do not work via 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 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 might be used in an relatively unbiased way (besides the choice of experimental conditions) to group drug treatments that have similar impacts on cell morphology.[4] This finding, combined with other 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[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 (Figure 1a). The standard dyes for Cell Painting are Hoechst 33342 (DNA), concanavalin A (endoplasmic reticulum), SYTO 14 (nucleoli and cytoplasmic RNA), phalloidin (f-actin), wheat germ agglutinin (WGA) (Golgi apparatus and plasma membrane), and Mito Tracker 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 (Figure 1b). This approach yields a high-dimensional dataset for each cell and captures over a thousand 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 (Figure 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 (2013–2023), following a systematic review format. 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 title, abstract, or subject terms/keyword headings; limited to articles written in English after 2012; peer-reviewed with some exceptions for key pre-prints). 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 irrelevant and 3 poster/thesis/news articles), and a manual search added 19 relevant studies, including some published after June 2023 (listed in Supplementary Tables S2 and S3). This resulted in 90 studies for review, as shown in the Preferred Reporting Items for Systematic Reviews and Meta Analyses (PRISMA) flow chart (Figure 2a), with included and excluded studies listed in Supplementary Tables S2 and S3, respectively.

Extracted Data

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

ADVANCEMENTS IN CELL PAINTING

1. 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 single assay capable of capturing many biologically relevant phenotypes with high throughput.[8] As described above, six stains were selected and imaged in five channels to reveal morphological changes for eight cellular components or organelles (Figure 1b). Gustafsdottir's publication established the moniker "Cell Painting"; however, an updated protocol (v2) published in 2016 by Bray et al., while also making minor adjustments.[10] A recent effort optimized the assay's cost and reproducibility, culminating in Cell Painting v3 in 2022.[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, *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]

1.1 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 been used and performed well for Cell Painting experiments, and thus the selection 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] 3,214 small molecule compounds were profiled with Cell Painting on six different cell lines: A549, OVCAR4, DU145, 786-O, HEPG2 and a non-cancer patient-derived fibroblast cell line. These compounds were all annotated, having information about their putative target and MoA, and included FDA approved drugs. The cell lines were ranked based on their ability to detect compound activity (termed "phenoactivity", or "phenotypic activity") and to predict the compound's MoA (termed "phenosimilarity", or "phenotypic consistency"). Here, compound activity refers to the strength of the morphological phenotypes detected by the Cell Painting assay, whereas MoA describes the extent to which the compound phenocopies other compounds with the same annotated MoA. The cell lines that were best for detecting phenotypic activity had poor sensitivity for predicting MoA and *vice versa*. This discrepancy may reflect the diverse genetic landscapes of different cell lines, which may influence the expression of targets and the cellular pathways. For example, 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. To avoid potential well position effects, 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 only necessary to optimize image acquisition and cell segmentation parameters to account for differences in the size and 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 different 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 of additional cell lines without adjustment, based on our observations from the literature and personal communications.

1.2 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 (Figure 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 (Figure 1e).[18]

1.3 Improvements in the Choice 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 RNAi-induced knockdown using the Cell Painting assay and found that the morphological signatures are highly sensitive and reproducible but there were off-target ‘seed’ effects of RNAi reagents that dominated the signatures. 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/protein overexpression[19] and CRISPR knockout to deplete expression.[20] A challenge with a target agnostic assay, such as Cell Painting, is that compounds active in the assay can act by multiple mechanisms, which complicates the interpretation of a given bioactivity[19]. 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).[21] ‘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 nonspecific, 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.

1.4 Development of Microscopy Imaging—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, suggesting that cells should be imaged at 20x 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.

2.1 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.[22]). Cimini et al. note that while small-scale image analysis can be performed using CellProfiler on desktop computers, large-scale analysis (>1000 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, particularly because technological advances quickly render estimates obsolete.

Some alternative approaches to classical feature extraction have emerged that leverage deep learning models to recognize features directly from raw images, such as DeepProfiler[23], Vision Transformers such as DINO (a self-supervised learning method)[24,25], and Convolutional Neural Networks (CNNs). 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 up to a 29% improvement over CellProfiler features when assessed using mean average precision (mAP) for classifying chemical perturbations.[26] Steps to further process the Cell Painting data, from morphological feature extraction to profile normalization to batch effect correction discussed in Section 3.1 and 3.2, are also continuously improving.

2.2 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 of staining cells. Although brightfield imaging does not yield a clear contrast of all 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.[27] 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 from the predicted images from brightfield. 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, the features from the AGP (actin, Golgi, plasma membrane) and mitochondrial channels were more challenging to predict. To determine if 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 a sensitivity of 62.5% and specificity of 98.0%.

Another study tested the ability of CNN-based features extracted from brightfield images versus from fluorescence Cell Painting images to predict 10 MoA classes.[28] The features were additionally compared to 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.[28] 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 deep learning applied to brightfield images holds great promise to augment or replace fluorescent stains in Cell Painting assays, saving time and cost. In fact, early reports from the techbio company Recursion indicate a transition from Cell Painting to brightfield imaging.[29]

3.1 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 (e.g. 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.[30] Other approaches, such as AutoML (automated machine learning), enable the most informative features from Cell Painting datasets to be identified faster.[31] Siegismund et al. found using AutoML that a subset of only 20–30 features was sufficient to represent the most relevant information from the morphological signature and 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 the diversity of phenotypes in the profiled dataset.

3.2 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 negative controls only.[32] 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).[30] Pycytominer also implements the sphering transformation (also called “whitening”), which can be viewed as a multivariate standardization strategy.[18,32] 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 10uM)[33], but these results are likely confounded by plate layout effects and have not been consistently high across studies.

When analyzing Cell Painting data, well plate averaging (or median profiles) is a popular choice that averages data from multiple cells within a well.[34,35] This approach offers benefits like reduced data size, faster analysis, simplified interpretation, and potentially less noise, but it assumes cell homogeneity, which may obscure subtle differences between cells.[36] While well plate averaging provides 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; the reader is referred to van Dijk et al. for a discussion of cell heterogeneity in representations of cell populations in the Cell Painting assay.[37] Overall, understanding population heterogeneity through single-cell analysis can identify subpopulations, characterize heterogeneity, and inform experimental design, making it valuable for studying complex biological systems. Despite these advantages, well plate averaging 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 well-averaged image profiles.[38] They used qualitative visualizations in combination with ten 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.

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 to 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 compared to standard normalization approaches.

4. Publicly Available Datasets

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

5. Applications of Cell Painting Data

The processing and downstream analysis of Cell Painting data (Figure 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 labelled data. In the following sections we discuss the different applications of Cell Painting data, using machine learning and statistical approaches, to aid drug discovery.

5.1 Predicting Mechanisms of Action

The Cell Painting assay offers an unbiased view of cellular responses to compound perturbations, enabling the identification of MoA for compounds that induce relatively specific morphological changes (see Table 2). This identification process involves comparing the phenotype of a query compound to those of ‘landmark compounds’ with known MoAs, or to 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, as compounds can have multiple targets with varying affinities. Additionally, protein targets, and proteins up and downstream of the direct targets, may be altered in various cell types. As a result, binary annotations for MoAs in datasets often oversimplify the reality (see Trapotsi et al. for more details).[43] 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[44,45], DNA damaging agents[20], mitochondrial membrane depolarisation[45–47], lysosomotropism[48], and inhibitors of the plasma membrane Na+ pump[45], among others (Table 2). It is important to note that ‘mechanism of action’ (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 was recently 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 not using 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).[58] The best individual cell line was able to distinguish 20 out of 83 MoAs (authors considered MoAs that had at least three active compounds). The number of distinguishable MoAs increased with each additional cell line but quickly plateaued; ultimately 41 out 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.[32] 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 dimensionality reduction method (L1-norm) and 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 using a “guilt-by-association” strategy.[59] Autoquin, a previously uncharacterized autophagy inhibitor, was found to induce similar morphological changes in the Cell Painting assay to that of oxautin-1.[56] Both autoquin and oxautin-1 were shown to inhibit autophagy via an indirect modulation of the activity of lysosomal enzymes, and autoquin was 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.[60] Other compounds have been found to modulate microtubules[44] and inhibit pyrimidine biosynthesis (inhibiting target dihydroorotate dehydrogenase).[61]

Despite these successes, predicting MoAs where 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 by “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 $\sigma 1$ receptor antagonist.[62] Here, Cell Painting revealed many potentially lysosomotropic CNS-targeting drugs as biosimilar to Diaminopyrimidine DP68, and it was challenging to discern whether the lysosomotropic effect or the $\sigma 1$ receptor interaction was responsible for the phenotype.[62] Thermal proteome profiling allowed the authors to narrow it down to $\sigma 1$ 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 instead their biological activity is very precise and just not yet discovered.[63] 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 to a DMSO control).[64] They could select a subset of morphological features most affected by compounds with 12 distinct. 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 (e.g., a genotoxin cluster and a tubulin poisons cluster).[21] Other clusters represented other MoAs, including nonspecific ‘historical’ KAT inhibitors (hKATIs). Interestingly, some compounds triggered cellular responses called a “cytotoxicity burst” at higher concentrations that activated multiple stress responses within the cell rather than affecting a singular target.

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 be used to classify cellular phenotypes (Figure 4).

An early study by Durr et al. developed a CNN that classified single-cell phenotypes based on images generated from Cell Painting assays.[65] They trained CNNs to classify MoAs using approximately 40,000 single-cell images for 75 bioactive compounds.[65] For unbiased testing, they used 2223 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%.

In addition to CNNs, transfer learning—where the knowledge of a pre-trained model (e.g., image-based knowledge) is transferred or fine-tuned to another model to perform a similar task (e.g. Cell Painting data)—is useful, reducing the need to train models from scratch; this is being increasingly explored for cell imaging data.[66] 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).[66] The pre-trained models achieved an accuracy between 95–97% compared to training from scratch which achieved an accuracy between 70–91%.

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 to classical features.[23] 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.[34,51,67]

5.2 Cell Painting in Assay Activity Prediction—Images have proven useful in predicting compounds showing desirable biological activity and identifying novel active compounds with therapeutic potential. A landmark study by Simm et al. (2018) investigated using 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 disease areas.[68] 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. then 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.[69] Because these performed better than fully connected neural networks (FNNs) trained using classically extracted numerical features (mean AUC-ROC=0.68), this provided evidence that the raw-image CNN models could capture information that may be overlooked by models trained on predefined CellProfiler features.[69]

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 out of ten approaches were highly concordant for 82% of the tested chemicals.[70]

Cell Painting has also been combined with chemical structure information to enhance prediction ability. Assay prediction accuracies were better when using a BMF Macau model with Cell Painting profiles as side information (100 out of 224 targets with AUC>0.80) compared to using only chemical structural data as side information (90 out of 224 with AUC>0.80).[71] In another study, models that combined structural information and Cell Painting profiles, using similarities to training data, with an additional 20% assays with AUC > 0.70 (79 out of 177 assays) compared to models that only used chemical structure information.[72]

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 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.[73] 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.[50,72,74,75]

5.3 Phenotypic Profiling of Bioactivity to Guide Chemistry—Phenotypic profiling, particularly with the Cell Painting assay, 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.[76] 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 (carbon atoms with four single bonds) and found that two epoxy ketone diastereomers synthesized caused striking induced consistent morphological changes for all doses, prompting studies to compare their morphological signatures to reference compounds.[77] 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 for enhanced biodiversity, but stereochemistry and appendage diversity are also contributors.[78,79]

More recently, biology-oriented synthesis has focused on pseudo-natural products, and to this end, Christoforow et al. characterized the potential bioactivity of novel classes of pyrano-furo-pyridone (PFP) pseudo-natural products.[80] 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 to reference compounds, including indocinchona alkaloids[81], a natural-product inspired flavonoid library[82]; spiroindane pyrrolidines[83]; pyrroquinoline pseudo-natural products[84]; and indofulvin pseudo-natural products[85]. Overall, phenotypic profiling enables rapid evaluation of the bioactivity, and sometimes MoA, of synthesized compounds without requiring a custom assay for each chemical series.

5.4 Predicting Compound Toxicity—Cell Painting assays have been used to predict multiple safety/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 G_1-phase cells (R^2=0.51), and percentage of apoptotic cells (R^2=0.37).[20] In addition to the specific cell health phenotypes mentioned above[20,86], the Cell Painting assay was able to predict the outcomes of 12 cytotoxicity- and proliferation-related *in vitro* assays using Cell Painting profiles.[54] 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[87]. Interestingly, this study included both small molecule compounds and PROteolysis Targeting Chimeras (PROTACs), which have garnered attention due to their unique bifunctional nature and potential ability to degrade ‘undruggable’ targets.[87] Trapotsi et al. showed that Cell Painting could identify PROTAC phenotypic signatures, which were often unique as compared to 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 and it remains to be seen if Cell Painting can be used for DILI prediction, which is one of the aims of the recently established OASIS consortium (www.oasisconsortium.org).[88] Cell Painting is thus particularly useful to assess the toxicity of new therapeutic modalities that lack the experience and best practices established for small molecules.

5.5 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 (PFOS) and perfluorooctane sulfonic acid (PFOA), two widely used industrial chemicals.[89] 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.[90] Computational methods are still developing to attempt to deconvolute the impacts of compounds on multiple target proteins in the related case of polypharmacology.[91]

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[92] and compared them to effect values from *in vivo* mammalian toxicity studies. 68% of the Cell Painting-based results were either similar to or more conservative than the *in vivo* studies, affirming the potential to use Cell Painting data for IVIVE.[92] More recently, Nyffeler et al. extended the strategy to predict human exposure. This study revealed that the majority of toxicants that exhibited activity in the Cell Painting assay are unlikely to be bioactive or cause unintended effects at typical human exposure concentrations.[55] On a larger scale, the U.S. Environmental Protection Agency (EPA) is working towards using transcriptomics and Cell Painting data in their risk assessments.[93] In the future, new sources of relevant *in vivo* toxicity annotations such as hepatotoxicity[94,95] and cardiotoxicity[96], and the inclusion of pharmacokinetic (PK) information[97,98], might be used to augment Cell Painting data in predictive toxicology.

5.6 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 functionally associate human genes and disease-associated alleles based on the morphology of cells when those genes are perturbed, or alleles are present. Rohban et al. used this approach to group 110 genes with a detectable phenotype out of 220 ORFs, revealing a previously unknown interaction between the NF-κB pathway and Hippo pathways[99], and then later to identify promising compounds that match a desired phenotypic profile impinging on those pathways.[100] Cancer-associated somatic variants could also be overexpressed in cells, to predict their functional impacts, including at the single-cell level.[19]

Morphological signatures can also serve as biomarkers for disease diagnosis or prognosis, or 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.[101,102]

In antiviral drug discovery, Cell Painting has been used to identify virus-induced phenotypic signatures that distinguish infected from non-infected cells.[18,103,104] 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.[90] Cell Painting has also been used to investigate transcription factor EB (TFEB) signaling and lysosomal dysfunction by detecting phenotypic changes in organelles in response to TFEB localization.[105] 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.[106]

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" - cmQTLs).[107] McDiarmid et al. used Cell Painting data to reveal 16 FDA-approved drugs from five mechanistic groups that were able to reverse morphological signatures associated with Alzheimer disease based on risk gene SORL1 variants in neural progenitor cells.[108] Schiff et al. showed that an unbiased phenotypic profiling using Cell Painting and primary fibroblasts from 91 Parkinson's disease patients and matched healthy controls, was able to distinguish LRRK2 and sporadic Parkinson's disease lines from healthy controls.[109] Another pilot study, involving 12 healthy controls and 12 subjects 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 patients.[110] Overall, the broad and multidimensional data generated by Cell Painting assays not only provides opportunities for new insights into complex cellular responses but can also reveal novel therapeutic targets and strategies for drug discovery and repurposing.

5.7 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.[111] Nassiri and McCall compared Cell Painting with LINCS (Library of Integrated Network-Based Cellular Signatures) gene expression data for improved insight into MoAs.[112] They used a reference database of 9,515 compounds to identify compounds with similar gene expression changes, followed by 'cell morphology enrichment analysis'. The enrichment analysis involved identifying significant associations between changes in cell morphology and gene expression, and then modeling these associations using machine learning and hence generating a dataset of the associations between genes and each morphological feature, revealing a novel interdependence between gene expression and cell morphologies 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.[33,113] 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), but when combined, attained 69%.[33] Likewise Haghighi et al. used morphology features from Cell Painting genetic perturbations to predict the genes.[113] They found that cell morphology data could capture distinct biological information often not associated with the particular stains in the Cell Painting assay, highlighting that Cell Painting captures more information than just for particular labelled components or some of these genes had unannotated functions. Another study combined RNA-Seq 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.[114]

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.[115] They investigated two chemicals (amiodarone and prochlorperazine) because both showed a risk for drug induced liver injury (DILI) in humans and thus, they assessed if they share common information in Cell Painting and L1000 dataset. They found a direct relation between deregulated genes and cell morphology observations. Their study demonstrated that some compounds shared similar genes, pathways, and morphological profiles. 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 to 0.25 using chemical structures alone.[45] The combination of cell morphology and gene expression at a single-cell level is also an exciting research area that remains to be explored, given the higher computational demands for single-cell analysis and the technical challenges to measuring both kinds of profiles in the same assay.[116]

The benefits of integrating Cell Painting data may also apply to protein profiling. One study tested 306 well-characterized compounds with established MoAs using Cell Painting and nELISA protein profiling, finding a 21.2% and 26.7% retrieval rate for MoA classes, with an additional 33% of MOAs when combining nELISA with Cell Painting.[117] 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.[118] Overall, considering multiple modalities together 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 TEN 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 over 36 academic laboratories or screening centers and 51 industries worldwide, with at least 13 offering the assay as a service, offering a range of outputs including images, profiles, or matching of customer samples to an internal database of samples (Figure 3b, 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 maximal 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. Identifying disease-associated phenotypes and predicting 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 have shown that Cell Painting is a promising tool in the pharmaceutical industry[51,75,119]. 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 where no compounds exist for a given mechanism.

Interpretation of morphological profiles is another challenge. While sophisticated image analysis algorithms and machine learning methods can extract and analyze complex morphological signatures, interpreting these computational/statistical signatures can be challenging, even for classical algorithms where features are precisely defined mathematically. The BioMorph space attempts to address this by linking 827 Cell Painting features to 412 descriptive terms, based on mapping to assays capturing phenotypes of cell health[120]. 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 requires considerable computational resource to process and analyze. Cloud-based solutions and open-source software tools can address these challenges[121], 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[122], tissue slices, and live cell imaging.[123] 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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)). Figure 1 was created using BioRender.

Data availability

No new datasets were used in this study.

REFERENCES

1. Swinney DC & Anthony J How were new medicines discovered? Nature Reviews Drug Discovery 2011 10:7 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, (2019).
3. Moffat JG, Vincent F, Lee JA, Eder J & Prunotto M Opportunities and challenges in phenotypic drug discovery: an industry perspective. Nature Reviews Drug Discovery 2017 16:8 16, 531–543 (2017).
4. Perlman ZE et al. Multidimensional drug profiling by automated microscopy. Science (1979) 306, 1194–1198 (2004).
5. Schulze CJ et al. ‘Function-first’ Lead Discovery: Mode of Action Profiling of Natural Product Libraries Using Image-Based Screening. Chem Biol 20, 285 (2013). [PubMed: 23438757]
6. Feng Y, Mitchison TJ, Bender A, Young DW & Tallarico JA Multi-parameter phenotypic profiling: using cellular effects to characterize small-molecule compounds. Nature Reviews Drug Discovery 2009 8:7 8, 567–578 (2009).
7. Woehrmann MH et al. Large-scale cytological profiling for functional analysis of bioactive compounds. Mol Biosyst 9, 2604–2617 (2013). [PubMed: 24056581]
8. Gustafsdottir SM et al. Multiplex cytological profiling assay to measure diverse cellular states. PLoS One 8, (2013).
9. Stirling DR et al. CellProfiler 4: improvements in speed, utility and usability. BMC Bioinformatics 22, 1–11 (2021). [PubMed: 33388027]
10. Bray MA et al. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat Protoc 11, 1757–1774 (2016). [PubMed: 27560178]

11. Cimini BA et al. Optimizing the Cell Painting assay for image-based profiling. Nat Protoc 18, 1981–2013 (2023). [PubMed: 37344608]
12. Tromans-Coia C et al. Assessing the performance of the Cell Painting assay across different imaging systems. Cytometry Part A 103, 915–926 (2023).
13. Chandrasekaran SN et al. Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. Nature Methods 2024 21:6 21, 1114–1121 (2024).
14. Chandrasekaran SN et al. JUMP Cell Painting dataset: morphological impact of 136,000 chemical and genetic perturbations. bioRxiv 2023.03.23.534023 (2023) doi:10.1101/2023.03.23.534023.
15. Heinrich L, Kumbier K, Li L, Altschuler SJ & Wu LF Selection of Optimal Cell Lines for High-Content Phenotypic Screening. ACS Chem Biol 18, 679–685 (2023). [PubMed: 36920184]
16. Willis C, Nyffeler J & Harrill J Phenotypic Profiling of Reference Chemicals across Biologically Diverse Cell Types Using the Cell Painting Assay. SLAS Discovery 25, 755–769 (2020). [PubMed: 32546035]
17. Laber S et al. Discovering cellular programs of intrinsic and extrinsic drivers of metabolic traits using LipocyteProfiler. Cell Genomics 3, 100346 (2023). [PubMed: 37492099]
18. Rietdijk J et al. A phenomics approach for antiviral drug discovery. BMC Biol 19, (2021).
19. Caicedo JC et al. Cell Painting predicts impact of lung cancer variants. Mol Biol Cell 33, (2022).
20. Way GP et al. Predicting cell health phenotypes using image-based morphology profiling. Mol Biol Cell 32, 995–1005 (2021). [PubMed: 33534641]
21. Dahlin JL et al. Reference compounds for characterizing cellular injury in high-content cellular morphology assays. Nature Communications 2023 14:1 14, 1–16 (2023).
22. 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). [PubMed: 29953863]
23. Moshkov N et al. Learning representations for image-based profiling of perturbations. Nature Communications 2024 15:1 15, 1–17 (2024).
24. Caron M et al. Emerging Properties in Self-Supervised Vision Transformers. Proceedings of the IEEE International Conference on Computer Vision 9630–9640 (2021) doi:10.1109/ICCV48922.2021.00951.
25. Kraus O et al. Masked Autoencoders are Scalable Learners of Cellular Morphology. (2023).
26. Kim V, Adaloglou N, Osterland M, Morelli FM & Zapata PAM Self-supervision advances morphological profiling by unlocking powerful image representations. bioRxiv 2023.04.28.538691 (2023) doi:10.1101/2023.04.28.538691.
27. Cross-Zamirski JO et al. Label-free prediction of cell painting from brightfield images. Sci Rep 12, 10001 (2022). [PubMed: 35705591]
28. Harrison PJ et al. Evaluating the utility of brightfield image data for mechanism of action prediction. PLoS Comput Biol 19, e1011323 (2023). [PubMed: 37490493]
29. Scaling Biology: Chris Gibson, Co-Founder and CEO of Recursion Pharmaceuticals. https://www.decodingbio.com/p/scaling-biology-chris-gibson-co-founder.
30. Serrano E et al. Reproducible image-based profiling with Pycytominer. ArXiv (2024).
31. 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). [PubMed: 35058213]
32. Janosch A, Kaffka C & Bickle M Unbiased Phenotype Detection Using Negative Controls. SLAS Discovery 24, 234–241 (2019). [PubMed: 30616488]
33. Way GP et al. Morphology and gene expression profiling provide complementary information for mapping cell state. Cell Syst 13, 911–923.e9 (2022). [PubMed: 36395727]
34. Lafarge MW et al. Capturing Single-Cell Phenotypic Variation via Unsupervised Representation Learning. Proc Mach Learn Res 102, 315–325 (2019).
35. Caicedo JC et al. Data-analysis strategies for image-based cell profiling. Nat Methods 14, 849–863 (2017). [PubMed: 28858338]
36. Altschuler SJ & Wu LF Cellular heterogeneity: do differences make a difference? Cell 141, 559–563 (2010). [PubMed: 20478246]

37. Dijk R. van, Arevalo J, Babadi M, Carpenter AE & Singh S Capturing cell heterogeneity in representations of cell populations for image-based profiling using contrastive learning. bioRxiv 2023.11.14.567038 (2023) doi:10.1101/2023.11.14.567038.

38. Arevalo J et al. Evaluating batch correction methods for image-based cell profiling. Nature Communications 2024 15:1 15, 1–12 (2024).

39. Fay MM et al. RxRx3: Phenomics Map of Biology. bioRxiv 2023.02.07.527350 (2023) doi:10.1101/2023.02.07.527350.

40. Weisbart E et al. Cell Painting Gallery: an open resource for image-based profiling. (2024) doi:10.37921/977328pjvbca.

41. 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).

42. Ramezani M et al. A genome-wide atlas of human cell morphology. bioRxiv 2023.08.06.552164 (2023) doi:10.1101/2023.08.06.552164.

43. Trapotsi MA, Hosseini-Gerami L & Bender A Computational analyses of mechanism of action (MoA): data, methods and integration. RSC Chem Biol 3, 170–200 (2022). [PubMed: 35360890]

44. 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.e3 (2022). [PubMed: 34968420]

45. Seal S et al. Integrating cell morphology with gene expression and chemical structure to aid mitochondrial toxicity detection. Commun Biol 5, 858 (2022). [PubMed: 35999457]

46. 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). [PubMed: 37327474]

47. Garcia de Lomana M, Marin Zapata PA & Montanari F Predicting the Mitochondrial Toxicity of Small Molecules: Insights from Mechanistic Assays and Cell Painting Data. Chem Res Toxicol 36, 1107–1120 (2023). [PubMed: 37409673]

48. Laraia L, Robke L & Waldmann H Bioactive Compound Collections: From Design to Target Identification. Chem 4, 705–730 (2018).

49. Schneidewind T et al. Morphological Profiling Identifies a Common Mode of Action for Small Molecules with Different Targets. ChemBioChem 21, 3197–3207 (2020). [PubMed: 32618075]

50. Tian G, Harrison PJ, Sreenivasan AP, Carreras-Puigvert J & Spjuth O Combining molecular and cell painting image data for mechanism of action prediction. Artificial Intelligence in the Life Sciences 3, 100060 (2023).

51. Wong DR et al. Deep representation learning determines drug mechanism of action from cell painting images. Digital Discovery 2, 1354–1367 (2023).

52. 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.e3 (2022). [PubMed: 34968420]

53. Pahl A et al. Morphological subprofile analysis for bioactivity annotation of small molecules. Cell Chem Biol 30, 839–853.e7 (2023). [PubMed: 37385259]

54. 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). [PubMed: 33522793]

55. Nyffeler J et al. Application of Cell painting for chemical hazard evaluation in support of screening-level chemical assessments. Toxicol Appl Pharmacol 468, (2023).

56. Laraia L et al. Image-Based Morphological Profiling Identifies a Lysosomotropic, Iron-Sequestering Autophagy Inhibitor. Angewandte Chemie - International Edition 59, 5721–5729 (2020). [PubMed: 31769920]

57. Lapins M & Spjuth O Evaluation of Gene Expression and Phenotypic Profiling Data as Quantitative Descriptors for Predicting Drug Targets and Mechanisms of Action. bioRxiv 580654 (2019) doi:10.1101/580654.

58. Cox MJ et al. Tales of 1,008 small molecules: phenomic profiling through live-cell imaging in a panel of reporter cell lines. Sci Rep 10, 1–14 (2020). [PubMed: 31913322]

59. Biology V. in C. Herbert Waldmann—Celebrating More than Three Decades in Academia. J Med Chem 66, 15055–15060 (2023). [PubMed: 37933865]

60. Svenningsen EB & Poulsen TB Establishing cell painting in a smaller chemical biology lab – A report from the frontier. Bioorg Med Chem 27, 2609–2615 (2019). [PubMed: 30935791]

61. Schölermann B et al. Identification of Dihydroorotate Dehydrogenase Inhibitors Using the Cell Painting Assay. ChemBioChem 23, (2022).

62. 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.e5 (2021). [PubMed: 33567254]

63. Wassermann AM et al. Dark chemical matter as a promising starting point for drug lead discovery. Nature Chemical Biology 2015 11:12 11, 958–966 (2015).

64. Pahl A et al. Illuminating Dark Chemical Matter Using the Cell Painting Assay. J Med Chem 67, 8862–8876 (2024). [PubMed: 38687818]

65. Dürr O & Sick B Single-cell phenotype classification using deep convolutional neural networks. J Biomol Screen 21, 998–1003 (2016). [PubMed: 26950929]

66. Kensert A, Harrison PJ & Spjuth O Transfer Learning with Deep Convolutional Neural Networks for Classifying Cellular Morphological Changes. SLAS Discovery 24, 466–475 (2019). [PubMed: 30641024]

67. Liu G, Seal S, Arevalo J & Liang Z Learning Molecular Representation in a Cell. ArXiv 1, 1–21.

68. Simm J et al. Repurposing High-Throughput Image Assays Enables Biological Activity Prediction for Drug Discovery. Cell Chem Biol 25, 611–618.e3 (2018). [PubMed: 29503208]

69. Hofmarcher M, Rumetshofer E, Clevert DA, 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). [PubMed: 30840449]

70. Nyffeler J et al. Comparison of Approaches for Determining Bioactivity Hits from High-Dimensional Profiling Data. SLAS Discovery 26, 292–308 (2021). [PubMed: 32862757]

71. Trapotsi MA et al. Comparison of Chemical Structure and Cell Morphology Information for Multitask Bioactivity Predictions. J Chem Inf Model 61, 1444–1456 (2021). [PubMed: 33661004]

72. Seal S et al. Merging bioactivity predictions from cell morphology and chemical fingerprint models using similarity to training data. J Cheminform 15, 56 (2023). [PubMed: 37268960]

73. Sanchez-Fernandez A, Rumetshofer E, Hochreiter S & Klambauer G CLOOME: contrastive learning unlocks bioimaging databases for queries with chemical structures. Nature Communications 2023 14:1 14, 1–14 (2023).

74. Moshkov N et al. Predicting compound activity from phenotypic profiles and chemical structures. Nat Commun 14, (2023).

75. Fredin Haslum J et al. Cell Painting-based bioactivity prediction boosts high-throughput screening hit-rates and compound diversity. Nature Communications 2024 15:1 15, 1–11 (2024).

76. Gerry CJ et al. Real-Time Biological Annotation of Synthetic Compounds. J Am Chem Soc 138, 8920–8927 (2016). [PubMed: 27398798]

77. Nelson SD, Wawer MJ & Schreiber SL Divergent Synthesis and Real-Time Biological Annotation of Optically Active Tetrahydrocyclopenta[c]pyranone Derivatives. Org Lett 18, 6280–6283 (2016). [PubMed: 27978655]

78. Gerlach EM, Korkmaz MA, Pavlinov I, Gao Q & Aldrich LN Systematic Diversity-Oriented Synthesis of Reduced Flavones from γ-Pyrones to Probe Biological Performance Diversity. ACS Chem Biol 14, 1536–1545 (2019). [PubMed: 31184855]

79. 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). [PubMed: 30133283]

80. Christoforow A et al. Design, Synthesis, and Phenotypic Profiling of Pyrano-Furo-Pyridone Pseudo Natural Products. Angewandte Chemie - International Edition 58, 14715–14723 (2019). [PubMed: 31339620]

81. Foley DJ et al. Phenotyping Reveals Targets of a Pseudo-Natural-Product Autophagy Inhibitor. Angewandte Chemie - International Edition 59, 12470–12476 (2020). [PubMed: 32108411]

82. Hippman RS et al. Multiple Chemical Features Impact Biological Performance Diversity of a Highly Active Natural Product-Inspired Library. ChemBioChem 21, 3137–3145 (2020). [PubMed: 32558167]

83. 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). [PubMed: 32546297]

84. Liu J et al. Design, Synthesis, and Biological Evaluation of Chemically and Biologically Diverse Pyrroquinoline Pseudo Natural Products. Angewandte Chemie - International Edition 60, 4648–4656 (2021). [PubMed: 33200868]

85. Burhop A et al. Synthesis of Indofulvin Pseudo-Natural Products Yields a New Autophagy Inhibitor Chemotype. Advanced Science 8, e2102042 (2021). [PubMed: 34346568]

86. 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).

87. Trapotsi MA et al. Cell Morphological Profiling Enables High-Throughput Screening for PROteolysis TArgeting Chimera (PROTAC) Phenotypic Signature. ACS Chem Biol 17, 1733–1744 (2022). [PubMed: 35793809]

88. Seal S et al. Improved Detection of Drug-Induced Liver Injury by Integrating Predicted In Vivo and In Vitro Data. Chem Res Toxicol (2024) doi:10.1021/ACS.CHEMRESTOX.4C00015/SUPPL_FILE/TX4C00015_SI_002.XLSX.

89. 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). [PubMed: 36731186]

90. Rietdijk J et al. Morphological profiling of environmental chemicals enables efficient and untargeted exploration of combination effects. Science of the Total Environment 832, 155058 (2022). [PubMed: 35390365]

91. Chow YL, Singh S, Carpenter AE & Way GP Predicting drug polypharmacology from cell morphology readouts using variational autoencoder latent space arithmetic. PLoS Comput Biol 18, e1009888 (2022). [PubMed: 35213530]

92. Nyffeler J et al. Bioactivity screening of environmental chemicals using imaging-based high-throughput phenotypic profiling. Toxicol Appl Pharmacol 389, (2020).

93. Thomas RS et al. The Next Generation Blueprint of Computational Toxicology at the U.S. Environmental Protection Agency. Toxicological Sciences 169, 317–332 (2019). [PubMed: 30835285]

94. 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). [PubMed: 26948801]

95. Seal S et al. Improved Detection of Drug-Induced Liver Injury by Integrating Predicted In Vivo and In Vitro Data. Chem Res Toxicol (2024) doi:10.1021/ACS.CHEMRESTOX.4C00015.

96. 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). [PubMed: 38300851]

97. Horne RI et al. Using Generative Modeling to Endow with Potency Initially Inert Compounds with Good Bioavailability and Low Toxicity. J Chem Inf Model (2024) doi:10.1021/ACS.JCIM.3C01777.

98. Seal S et al. PKSmart: An Open-Source Computational Model to Predict in vivo Pharmacokinetics of Small Molecules. bioRxiv 2024.02.02.578658 (2024) doi:10.1101/2024.02.02.578658.

99. Rohban MH et al. Systematic morphological profiling of human gene and allele function via cell painting. Elife 6, e24060 (2017). [PubMed: 28315521]

100. Rohban MH et al. Virtual screening for small-molecule pathway regulators by image-profile matching. Cell Syst 13, 724–736.e9 (2022). [PubMed: 36057257]

101. Hughes RE, Elliott RJR, Dawson JC & Carragher NO High-content phenotypic and pathway profiling to advance drug discovery in diseases of unmet need. Cell Chem Biol 28, 338–355 (2021). [PubMed: 33740435]

102. Hughes RE 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 Discovery 25, 770–782 (2020). [PubMed: 32441181]

103. Cuccarese MF et al. Functional immune mapping with deep-learning enabled phenomics applied to immunomodulatory and COVID-19 drug discovery. bioRxiv 2020.08.02.233064 (2020) doi:10.1101/2020.08.02.233064.

104. 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. bioRxiv 2020.04.21.054387 (2020) doi:10.1101/2020.04.21.054387.

105. Carey KL et al. TFEB Transcriptional Responses Reveal Negative Feedback by BHLHE40 and BHLHE41. Cell Rep 33, 108371 (2020). [PubMed: 33176151]

106. Kelley ME et al. High-content microscopy reveals a morphological signature of bortezomib resistance. Elife 12, (2023).

107. Tegtmeyer M et al. High-dimensional phenotyping to define the genetic basis of cellular morphology. Nat Commun 15, 1–12 (2024). [PubMed: 38169466]

108. McDiarmid AH et al. Morphological profiling in human neural progenitor cells classifies hits in a pilot drug screen for Alzheimer's disease. Brain Commun 6, (2024).

109. Schiff L et al. Integrating deep learning and unbiased automated high-content screening to identify complex disease signatures in human fibroblasts. Nat Commun 13, (2022).

110. Yang SJ et al. Applying Deep Neural Network Analysis to High-Content Image-Based Assays. SLAS Discovery 24, 829–841 (2019). [PubMed: 31284814]

111. Liu A, Seal S, Yang H & Bender A Using chemical and biological data to predict drug toxicity. SLAS Discovery 28, 53–64 (2023). [PubMed: 36639032]

112. Nassiri I & McCall MN Systematic exploration of cell morphological phenotypes associated with a transcriptomic query. Nucleic Acids Res 46, (2018).

113. Haghighi M, Caicedo JC, Cimini BA, Carpenter AE & Singh S High-dimensional gene expression and morphology profiles of cells across 28,000 genetic and chemical perturbations. Nat Methods 19, 1550–1557 (2022). [PubMed: 36344834]

114. 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). [PubMed: 35483669]

115. Cerisier N, Dafniet B, Badel A & Taboureau O Linking chemicals, genes and morphological perturbations to diseases. Toxicol Appl Pharmacol 461, (2023).

116. Camunas-Soler J Integrating single-cell transcriptomics with cellular phenotypes: cell morphology, Ca2+ imaging and electrophysiology. Biophysical Reviews 2023 16:1 16, 89–107 (2023).

117. Dagher M et al. nELISA: A high-throughput, high-plex platform enables quantitative profiling of the secretome. bioRxiv, 10.1101/2023.04.17.535914 (2023).

118. Schneidewind T et al. Combined morphological and proteome profiling reveals target-independent impairment of cholesterol homeostasis. Cell Chem Biol 28, 1780–1794.e5 (2021). [PubMed: 34214450]

119. 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). [PubMed: 37327474]

120. 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, (2024).

121. Way GP, Sailem H, Shave S, Kasprowicz R & Carragher NO Evolution and impact of high content imaging. SLAS Discovery 28, 292–305 (2023). [PubMed: 37666456]

122. Lukonin I, Zinner M & Liberali P Organoids in image-based phenotypic chemical screens. Experimental & Molecular Medicine 2021 53:10 53, 1495–1502 (2021).

123. Cottet M et al. Live cell painting: New nontoxic dye to probe cell physiology in high content screening. SLAS Discovery (2023) doi:10.1016/J.SLASD.2023.10.005.

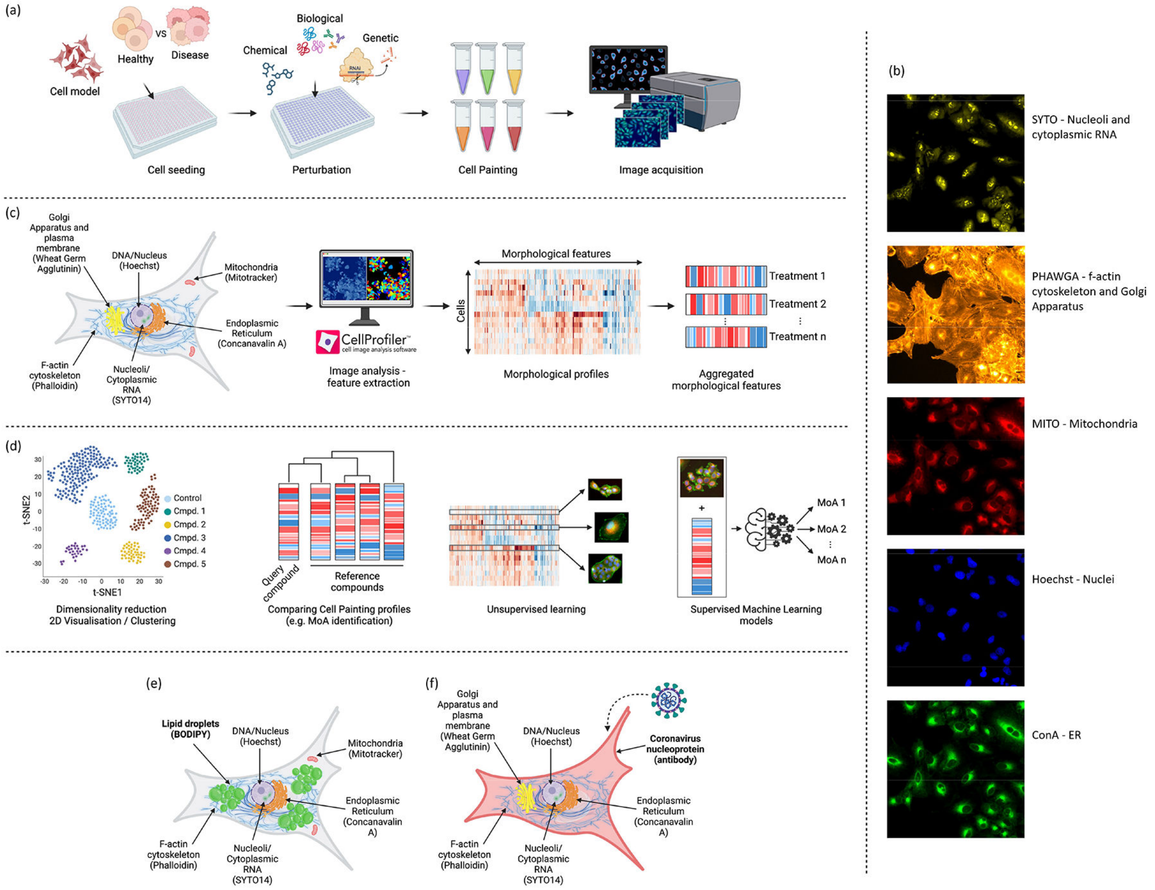

Figure 1.

Morphological profiling using the Cell Painting assay. (a) Schematic representation of 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 pre-processing, 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. Adaptations of the Cell Painting assay include (e) BODIPY to mark lipid droplets in lipid-accumulating cells and (f) a coronavirus antibody against human coronavirus 229E (CoV-229E) viral protein.

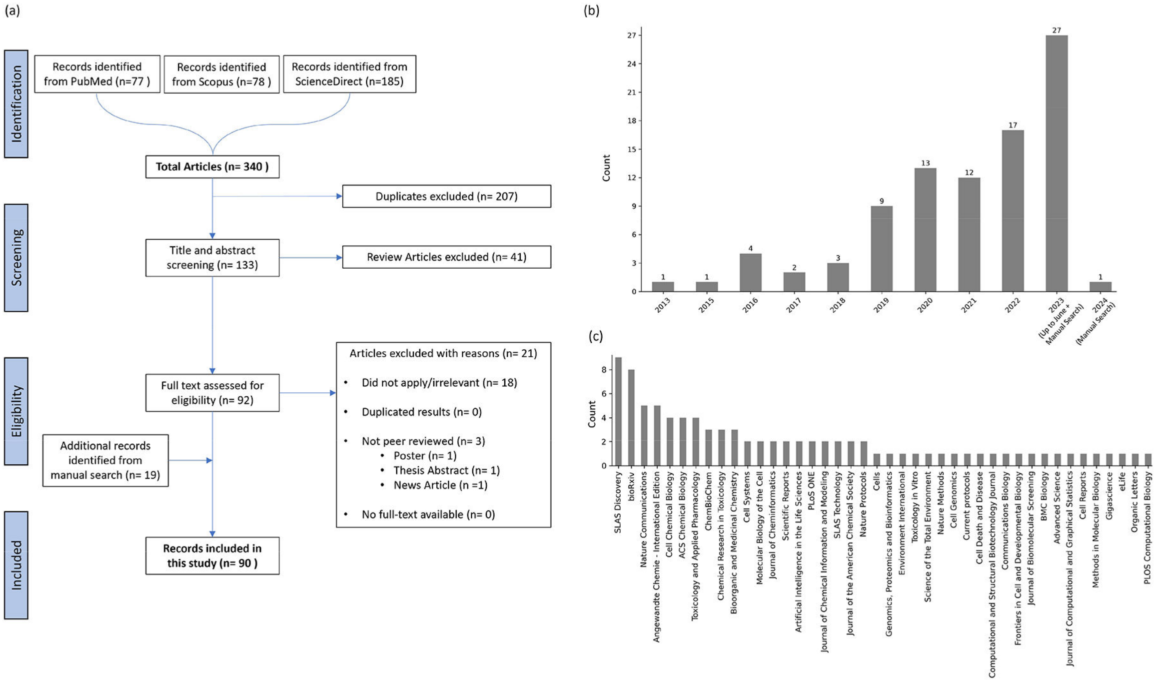

Figure 2.

(a) The Preferred Reporting Items for Systematic Reviews and Meta Analyses (PRISMA) flow chart diagramming 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 publications reviewed in this systematic review between 2013 and 2023, and (c) the journals where the publications were published.

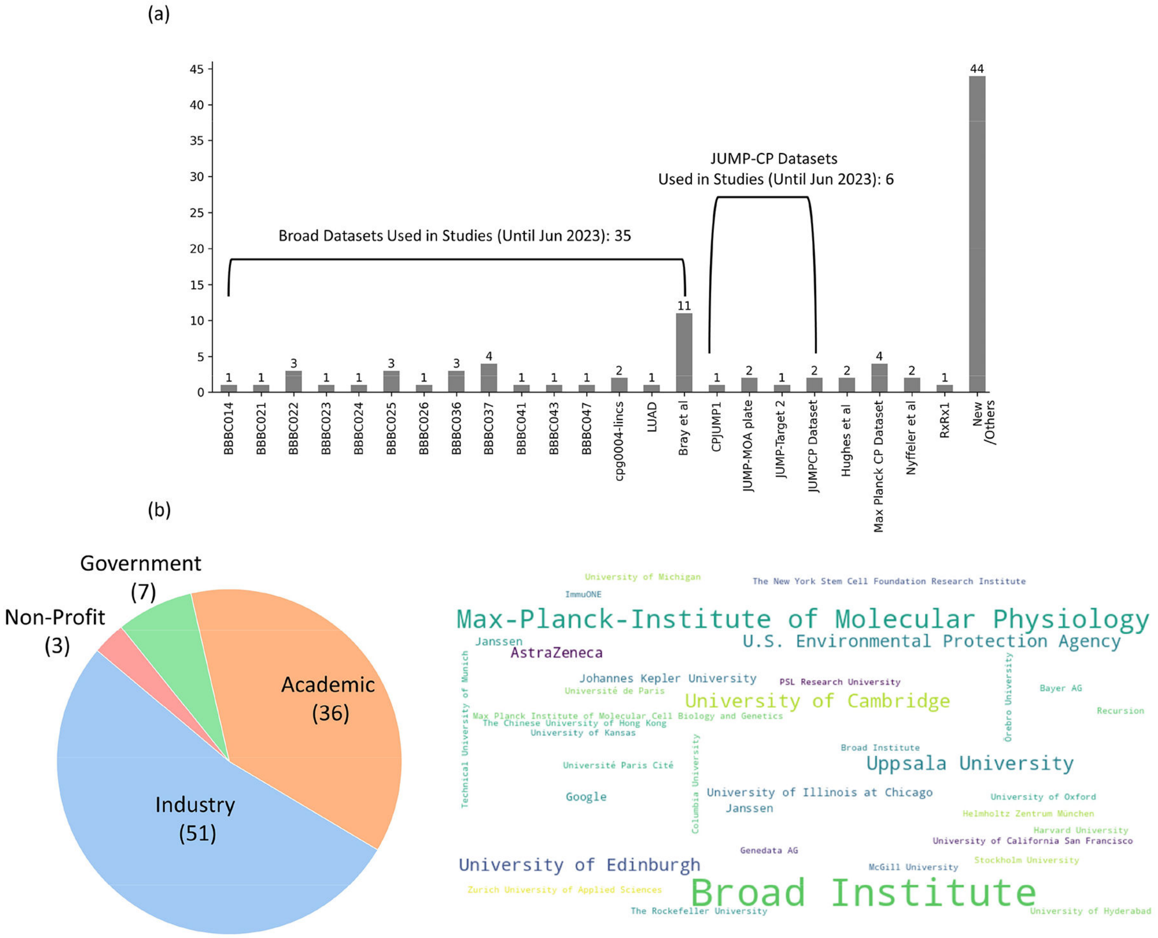

Figure 3.

(a) Analysis 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) Academic Institutions, Government Agencies, Pharmaceutical Companies, Non-Profits who led studies evaluated in this work and/or are members of the JUMP-CP and OASIS consortium. Further details see Supplementary Table 2 and Supplementary Table 4.

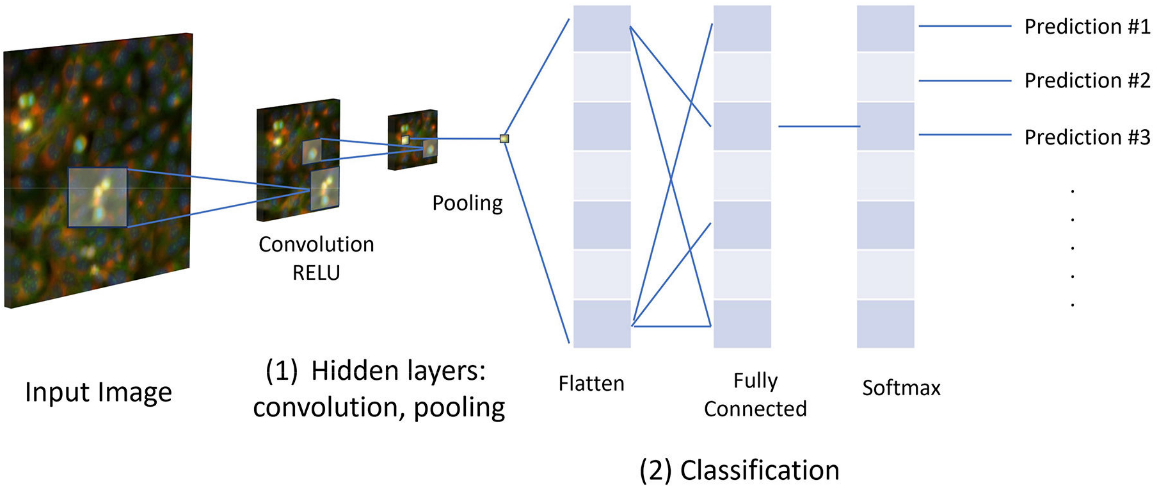

Figure 4.

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 (e.g., 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).

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	U2OS	30,616 compounds	[41]
Recursion RxRx3 dataset	January 2023	HUVEC	17,063 CRISPR/Cas9-mediated gene knockouts (most anonymized) and 1,674 compounds at 8 concentrations	[39]
JUMP-CP dataset	March 2023	U2OS	Over-expression of 12,602 genes, knockout of 7,975 genes using CRISPR-Cas9, 116,750 unique compounds	[14]
Periscope Dataset (cpg0021)	August 2023	HeLa and A549	Whole-genome pooled optical, >20,000 single-gene CRISPR/Cas9-mediated gene knockout screens	[42]

Table 2.

Cell Painting has detected various Mechanisms of Action (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	Biological Process/Direct Targets	Source
Cell Cycle Inhibition	Cell-cycle arrest in the S or G2 phase	49,50
Microtubule Disruption	Microtubule organization, Aurora kinase inhibitors, Tubulin polymerization inhibitor	50–52
Cytoskeletal Disruption	Actin dynamics, Microtubule destabilizers	52
Protein Synthesis Inhibition	Protein synthesis inhibitors	50
DNA Damage	Ribonucleotide reductase inhibitors, PARP inhibitors, Topoisomerase inhibitors, Pyrimidine Biosynthesis	50,51,53
Apoptosis Induction	Caspase activation, Mitochondrial disruption, Death receptor signalling	45,54
Autophagy	Autophagosome formation, Lysosomal degradation, Autophagy flux	55,56
Membrane Integrity Disruption	Membrane poration, Lipid peroxidation	21
Mitochondrial Dysfunction	Mitochondrial respiration, ATP synthesis	45,57
Oxidative Stress	ROS production, Antioxidant response	55
ER Stress	Unfolded Protein Response (UPR), ER-associated degradation (ERAD)	55
Hormonal Modulation	Hormone receptor activation, Signal transduction, Retinoid receptor agonists	50
Lipid Metabolism Inhibition	Lysosomotropism/cholesterol homeostasis regulation, HMGCR inhibition	50,53
Signal Transduction Inhibition	ALK tyrosine kinase receptor inhibition, src inhibitor, JAK inhibitors, AKT/PI3K/MTOR inhibitors	50,51,53
Ion Channel Modulation	Na+/K+ ATPase	53
Epigenetics	HDAC inhibitors, BET proteins	50,53
Metabolism	PPAR receptor antagonism, Carbonic anhydrase inhibition HMGCR inhibition, ATPase inhibitors	50,51
Protein Homeostasis	HSP inhibitors	50
Adhesion Disruption	Cadherin function, Integrin signalling	
Angiogenesis Inhibition	VEGFR inhibitor	57
Immune Modulation	receptor antagonist, TNF	57
Proteolysis Inhibition	Matrix metalloprotease inhibitors (MMP inhibitors)	51