SLAS Discovery 28 (2023) 292–305

Contents lists available at ScienceDirect

SLAS Discovery

journal homepage: www.elsevier.com/locate/slasd

Original Research

Evolution and impact of high content imaging

Gregory P. Way [a], Heba Sailem [b], Steven Shave [c,d], Richard Kasprowicz [c], Neil O. Carragher [d,*]

[a] Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA
[b] School of Cancer and Pharmaceutical Sciences, King's College London, UK
[c] GlaxoSmithKline Medicines Research Centre, Gunnels Wood Rd, Stevenage SG1 2NY, UK
[d] Edinburgh Cancer Research, Cancer Research UK Scotland Centre, Institute of Genetics and Cancer, University of Edinburgh, UK

A B S T R A C T / O U T L I N E

The field of high content imaging has steadily evolved and expanded substantially across many industry and academic research institutions since it was first described in the early 1990's. High content imaging refers to the automated acquisition and analysis of microscopic images from a variety of biological sample types. Integration of high content imaging microscopes with multiwell plate handling robotics enables high content imaging to be performed at scale and support medium- to high-throughput screening of pharmacological, genetic and diverse environmental perturbations upon complex biological systems ranging from 2D cell cultures to 3D tissue organoids to small model organisms. In this perspective article the authors provide a collective view on the following key discussion points relevant to the evolution of high content imaging:

- Evolution and impact of high content imaging: An academic perspective
- Evolution and impact of high content imaging: An industry perspective
- Evolution of high content image analysis
- Evolution of high content data analysis pipelines towards multiparametric and phenotypic profiling applications
- The role of data integration and multiomics
- The role and evolution of image data repositories and sharing standards
- Future perspective of high content imaging hardware and software

1. Introduction

High content imaging encompasses and integrates the research disciplines of cell biology, photonics, laboratory automation and image analysis to robustly interrogate the phenotypes of individual cells, multicellular tissue samples and small model organisms at scale. The field of high content imaging (HCI), also known as high content screening (HCS), was inspired and evolved from flow cytometry and digital imaging microscopy technologies which enable multiplex labelling of biomarkers on a cell-by-cell basis [1]. In 1997, Cellomics Inc., one of the pioneers of HCI, developed the first fully integrated HCI platform (ArrayScan) for HCS applications [2]. The ArrayScan and subsequent HCI platforms from other groups provided end-to-end hardware and software solutions for automating image acquisition, image processing, image analysis, image archiving, and image visualisation (Fig. 1).

These developments revolutionised microscopic analysis of cells and supported a shift away from subjective reporting and manual quantification of observations to fully quantitative cell biology permitting an accelerated approach to new knowledge generation. Thus, similarly to advances in next generation sequencing technology, automated HCI contributes to the modern era of hypothesis-free "discovery science" complementing more traditional hypothesis-driven research paradigms. The significant efficiency gains and accelerated discovery of potential new therapeutic targets, chemical starting points and early prediction of toxicity provided by HCI was a major incentive supporting early adoption of the technology by the pharmaceutical industry. Academic groups have subsequently contributed powerful and accessible image analysis software and machine learning applications to enable deep phenotyping of cell biology (e.g. therapeutic mechanism-of-action) as well as increased biological sample complexity. Together in close partnership, academia and industry have made rapid progress in collecting and analysing HCI data.

Initial HCS was typically performed in 2-dimensional cell cultures formatted in 96- or 384-multiwell plates using platform proprietary predefined image analysis algorithms capable of extracting one to a few quantitative measurements per condition. Subsequent evolution of both commercial and general-purpose open source image analysis software packages including Definiens [3], CellProfiler [4] and Advanced Cell Classifier [5], allowed non-experts in image analysis to create sophisticated bespoke algorithms tailored towards complex phenotype

* Corresponding author.
E-mail address: n.carragher@ed.ac.uk (N.O. Carragher).

https://doi.org/10.1016/j.slasd.2023.08.009
Received 3 May 2023; Received in revised form 9 August 2023; Accepted 29 August 2023
Available online 3 September 2023

quantification. This rapid evolution of free, general purpose software provided an important alternative and complementary approach to proprietary high throughput screening technologies which emerged as the predominant drug discovery engine of the biopharmaceutical industry in the early 1990s.

As a result of sequencing the human genome and subsequent advances in understanding disease at the genetic level, the pharmaceutical industry invested heavily in target-directed high throughput screening technologies in what was perceived as a new era of rapid and efficient discovery of highly selective and potentially personalised drug candidates. While many high throughput screening campaigns and modern target-led drug discovery strategies have produced remarkable successes in delivering effective medicines, high attrition rates in late stage clinical development prevail [6]. Advances in next generation sequencing (NGS) have revealed remarkable molecular heterogeneity within and between patients and adaptation in underlying disease mechanisms for many disease types that evade treatment. These findings starkly highlight the challenges in predicting which drugs and which drug targets will translate into clinically meaningful efficacy for many complex disease areas. It has become increasingly apparent that genes and proteins function as parts of integrated signalling pathway networks which contribute to extensive compensatory capacities and plasticity in cell fate. In contrast to traditional high throughput screening assays that measure the activity of single readouts in targets of interest, HCS promises broad attainment of deeper phenotypic knowledge about the effects of therapeutics, capturing complex signals like interacting signalling pathways, transcription factor dynamics, and polypharmacology [7–9]. Thus the advent of HCS represents an important evolution in the drug discovery paradigm from reductionist target biology to more systems level understanding of cell phenotype.

Faithfully modelling human disease in medium- to high-throughput assays is perhaps the most significant challenge in drug discovery. However, recent technological breakthroughs in human induced pluripotent stem cells (iPSC), creation of genetically well-defined models of disease through CRISPR-Cas9 gene editing, derivation of primary human cells, and advances in 3-dimensional (3D) in vitro biology techniques are converging towards accurately recapitulating specific segments of disease pathology in screening formats [10] (Fig. 2). Multicellular co-culture, microphysiological systems (MPS), 3D microtissue spheroid, and organoid models have been gaining in popularity and better represent the complex tissue architecture found in vivo. Moreover, such models are particularly well suited for the latest high content imaging platforms which provide spatial resolution in X, Y, and Z dimensions [11–14] (Fig. 2). Recent advances in 3D bioprinting and miniaturisation of microfluidic devices further improve assay reproducibility and support the generation of homogeneous multicellular 2D and 3D models in standard microtiter plate screening formats [15,16] (Fig. 2).

The development of more disease-relevant models plays to the strength of academic research centers with deep understanding of disease biology and access to clinical specimens, which represent fruitful areas for academic-industry collaboration in which to maximise impact.

Analytical challenges scale alongside assays that capture increasing amounts of broad (many measurements) and deep (high number of multiparametric features) data. Early efforts to deal with data on this scale relied upon expert knowledge to craft tools for signal extraction and analysis. Advances in the application of Artificial Intelligence/Machine Learning (AI/ML) technology coupled with significant improvements in computational power have found numerous applications in the drug discovery process [17,18], which has impacted every area of the drug discovery pipeline from target identification to clinical trials [19]. This revolution has been driven by significant improvements in computational power, with consumer grade graphics cards delivering teraflops of performance coupled with the open nature of communities supporting these efforts and releasing powerful open source toolkits [20]. Consequently, AI/ML techniques are routinely applied to high content imaging data to classify cell phenotypes and predict therapeutic mechanism-of-action [21,22,23]. Artificial neural networks (ANNs) and deep learning are growing areas of interest in biological image analysis [24] and are being applied to a variety of prediction tasks. Convolutional Neural Networks (CNNs) have been particularly impactful for image analysis, enabling deep architectures [25] and techniques to be applied in unsupervised [26,27] and supervised settings for small molecule mechanism-of-action prediction [28,29].

The term "phenomics'' was first coined to describe the comprehensive study of phenotypes [30], which provides functional context to

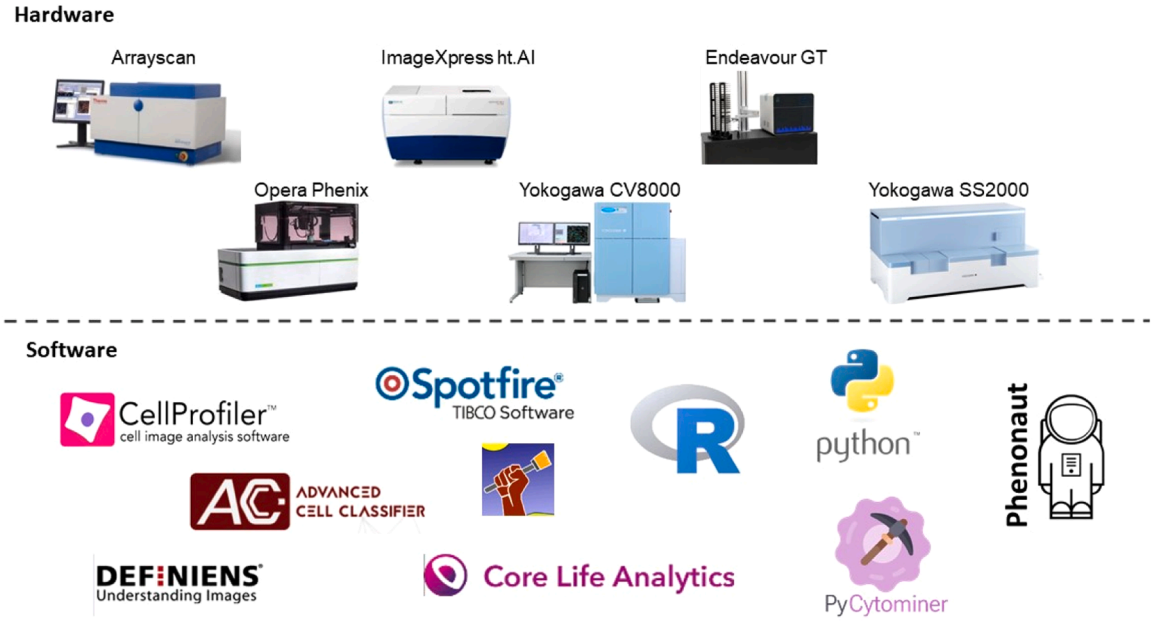

Fig. 1. Evolution of High Content Imaging hardware and software solutions, including open-source software for raw image analysis and secondary data analysis has contributed to increased adoption and variety of HCI applications. These developments include more sophisticated analysis of biological samples of increasing complexity including co-cultures and 3D models and integration of high content imaging with other single cell multiomics technologies.

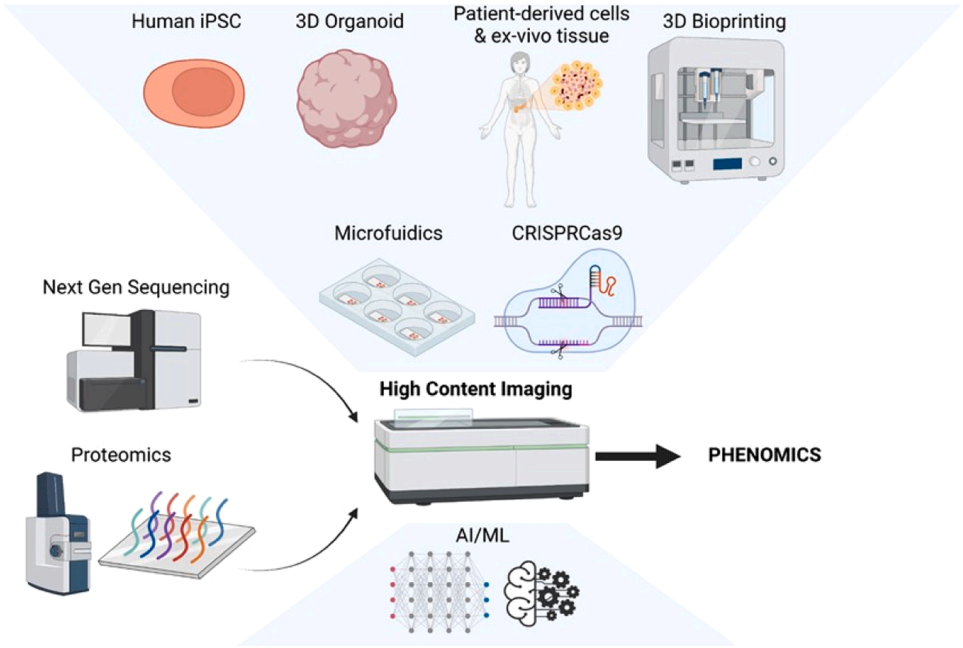

Fig. 2. Recent advances in human iPSC technology, patient-derived models, 3D bioprinting, 3D tissue organoids, CRISPRCas9 gene-editing and novel microfluidic devices are converging with the latest advances in high content imaging to produce more disease-relevant and mechanistically-informative in vitro models for drug discovery and basic research. Further integration of high content imaging data with orthogonal multiomics datasets and emerging AI/ML solutions are contributing to the new field of "phenomics".

genomic, transcriptomic and proteomics data. Integration of high content imaging data with other multiomics data types using AI/ML is required to provide a systems biology level understanding of cell phenotype and therapeutic mechanism-of-action. Multidisciplinary research collaborations across academic and industry sectors are contributing to a new era of "phenomics drug discovery" where HCI is core to the development of more disease relevant and mechanistically informative drug discovery. Below we discuss the evolution and impact of HCI from both academic and industry perspectives. We describe the significant advances in HCI analysis and the development and application of multiparametric phenotypic profiling which has revolutionized the field. We describe the important role of image data repositories and image data sharing standards to further advance the HCI field and we provide our future perspectives on the next phase of HCI hardware and software evolution.

2. Evolution and impact of high content imaging: an academic perspective

The academic sectors' considerable strengths in biology and data science are contributing significantly to the evolution of HCI. Various academic research groups have gained deep knowledge of specific areas of human disease biology from several years and even decades of intense and focused research effort. In addition, academic research centres with close links to the clinic and ready access to human volunteer and/or patient samples have provided a major contribution to the development and application of patient derived cell and tissue models for translational research. For example, development of co-culture protocols [31] have enabled production of large amounts of conditionally reprogrammed cells from accessible biopsy specimens, from both healthy tissues and tumor samples that have subsequently been applied in HCI studies [32]. Early adoption of HCI capabilities in academia pushed the boundaries of sample complexity, incorporating complex tissue samples such as coeliac tissue biopsies into primary screening assays [33]. Following many years of academic research efforts the generation and culture of 3D organoid tissues is now routine [34] and can be performed at scale for high throughput screening and HCI applications [11,12]. The adaptation of fresh human tissue samples for in vitro cell culture, ex vivo tissue slice and organoid translational research applications overcome many of the disadvantages of using transformed cell lines for drug discovery [34,35]. While primary human and patient-derived ex vivo models are of high value, the relevant tissue is, in many cases, difficult to obtain, or available only after the patient's death (e.g. heart, brain, and healthy liver). A major breakthrough in the ability to develop tissue specific cell-based disease models, including patient-derived cell assays at scale, has been achieved through the development of human induced pluripotent stem cell (iPSC) technology [36]. Protocols to derive iPSCs were first developed in academia and academia continues to be an important source of iPSC models, including specialised iPSC lines with various gene editing strategies to create genetically well-defined models of disease and "normal" gene corrected counterparts. Many human iPSC derived cell models have been adapted for high throughput and high content imaging formats [37–39]. Additionally, approaches to recapitulate disease biology with gene editing in otherwise normal tissue derived iPSC models to specifically map genotype to phenotype and simulate disease trajectories are in early stages of academic development [40]. However, maintaining human iPSC cell lines and optimising differentiation protocols at scale with high levels of consistency for screening applications requires significant resources and close adherence to standard operating procedures typically found in industry or core academic research facilities.

Data science is a rapidly emerging field in biomedical research which has been driven by the need to manage, integrate and interpret big data sets generated through advances in technology platforms such as next generation sequencing, proteomics, digital pathology and high content imaging. Academia has played a major role in developing data science by bringing together different disciplines from the numerical sciences (mathematics, statistics, computer science, bioinformatics) and other diverse fields such as astronomy and engineering to solve similar data-related problems, such as scalable data processing and data visualisation. This collaborative effort has led to many critical open source software libraries that are essential to data-driven initiatives, as well as the development of pioneering AI/ML approaches that reach across domains, and commonly into biomedical research applications. Early data science contributions to HCI from academia included the provision of open source image analysis solutions (e.g. CellProfiler [4]) that were readily compatible with automated analysis across large numbers of images. Additional early open source software developments included tools that allowed biologists to apply machine learning based classification of cell phenotypes from images such as CellClassifier [41],

Advanced Cell Classifier [5], ilastik [42] and KNIME [43]. The provision of both commercial software (e.g. Definiens) and open source tools combined with rapid advances in computing power and multiparametric imaging, facilitated faster and more reliable analyses that revealed more nuanced insights and brought forth a new era of high content imaging known as “phenotypic profiling”. We are currently living in this era, where data science and AI/ML are rapidly improving insights and all intermediate steps from experimental design to data acquisition and data processing.

Further evolution of phenotypic profiling methods in academia included development of the Cell Painting assay: a relatively low cost multiplex assay that "paints the cell" with multiple fluorescent dyes to obtain quantitative phenotypic profiles of cell morphology without the need for specific antibody labelling or genetically engineered probes [44,45]. The canonical Cell Painting assay multiplexes six fluorescent dyes, imaged in five spectral channels, to reveal eight broadly relevant cellular components or organelles. The Cell Painting assay is flexible, limited only by fluorescent channel overlap, and profiles generated are robust across a number of microscope setting changes as measured by percent replicating metrics [46]. The Cell Painting assay can be modified to specific screening conditions by, for example, swapping an existing canonical channel with a targeted dye to a specific biological entity of interest, such as adding a stain for lipid droplets in a screen for metabolic disease treatments [47]. Studies have also shown that brightfield imaging may contain just as much if not more detail than the unbiased Cell Painting stains [48,49], but this performance strength may be experiment specific [46]. In the future, we may modify the Cell Painting panel to flexibly focus on specifically known biological markers while allowing the brightfield channel to drive unbiased cell morphology analyses. In a pilot study of bioactive compounds, the Cell Painting assay detected a range of cellular phenotypes and the multiparametric phenotypic profiles were used to cluster compounds with similar annotated protein targets or chemical structure [45]. A recent study to evaluate if human genes can be functionally annotated using the Cell Painting assay demonstrated that 50% of the 220 genes tested yielded detectable morphological profiles which group into biologically meaningful gene clusters consistent with known functional annotation [50]. The JUMP-CP consortium (Joint Undertaking for Morphological Profiling-Cell Painting) led by the Broad Institute at MIT and Harvard and including several pharmaceutical industry partners aims to create a large public Cell Painting dataset of over 136,000 genetic and chemical perturbations [51]. It is anticipated that this public resource will catalyse new drug discovery programs across both academia and industry by enabling the prediction of compounds’ mode of action and toxicity, characterising disease phenotypes and uncovering new therapeutic target biology.

Academic research funding has generally been directed towards answering hypothesis-driven research questions and hypothesis-free applications have historically been dismissed by funding panels as “fishing expeditions”. However, with the abundant successes that sprung from the human genome project [52], hypothesis-free research has demonstrated significant value in basic and translational academic research and yielded a new discipline of “discovery science”. HCI assays can be designed to both test specific hypotheses and enable robust hypothesis generation with discovery science. Early adopters in academia employed HCI assays to identify compounds which control centrosome duplication [53]. Other groups combined HCI with siRNA screens to support early functional genomic screens [54,55] to reveal new biology and therapeutic targets; an approach which has now been widely adopted across academia and industry using the latest generation of arrayed CRISPR libraries [56]. In summary, academia has played a major role in the evolution of HCI capabilities and also has been a beneficiary of the overall evolution of HCI technology as a powerful tool for new knowledge creation. The academic sector is also strongly positioned to play an important role in the next evolution of HCI through continued development of hardware, software and data analysis solutions and through contributing novel biological capabilities and applications, especially applied to rare diseases. As discussed further below, these developments will benefit from strong academic-industry collaboration and partnerships.

3. Evolution and impact of high content imaging: an industry perspective

HCI is applied as a mainstay in the pharmaceutical industry across the discovery pipeline, from target identification through to candidate selection.

Two-dimensional culture multicolour imaging assays remain the assay category of primary impact in industry as gauged through assay abundance, application in progressing discovery programs, and influence on decision making. Largely this is due to simplicity of setup and direct biological relevance of the readout from the target-specific probes used. Yet even within these assays, methodological evolution is evident through increased frequency of use of primary cell, co-cultures, or iPSC-derived models. The trend to utilise more physiologically relevant cell models early in the discovery pipeline has been achieved through scalable cell factories for iPSC line generation and differentiation, in addition to simplified isolation procedures and commercial availability of primary cell sources. Indeed, this change typifies industry priorities by internalizing academic protocols to streamline methods and achieve robustness for routine, scalable implementation.

In industry, safety and efficacy profiling have been a primary beneficiary of adoption of HCI, especially when coupled with complex in vitro models. HCI in organoids in particular has shown application in screens for efficacy, lead identification, and safety [11,57]. Deploying these HCI methods aims to reduce high attrition rates of candidate therapeutics. A meta-analysis of 2003–2011 clinical phase data indicated only 10% of candidates entering clinical trials resulted in FDA approvals; with efficacy or safety cited as predominant reasons in first review response [58]. The probability of launch statistics are reflected within more recent analyses and as a sub-category, Phase II failure, in particular, has been identified as 79% attributable to safety and efficacy [59]. To provide more predictive safety assessment in early discovery, HCI has been adopted in predictive toxicology, with industry groups overwhelmingly classifying the technology as a current or near-term game-changer [60]. In particular, the use of HCI in complex models has enabled more accurate toxicological assessment in vitro, such as the use of MPS for detecting hepato- and renal-active compounds [61]. HCI also provides readouts in MPS liver models labelled with general cell health, lipid and bile canalicular markers to provide evidence of different mechanisms of hepatotoxicity [62]. The economic benefit of full integration of such assays is placed at billions of dollars per annum [62], yet it is worth recognising that safety profiling is applicable to tens-of compounds rather than hundreds, so it needs to be carefully integrated into the discovery pipeline, such as at candidate selection. For higher throughput methods that can be utilised in early discovery, spheroid systems show improved predictivity over 2D sandwich and monolayer cultures [63]; however, despite some assay systems using HCI to provide whole-spheroid intensity or volume readouts, more often than not, a simpler biochemical readout for cytotoxicity supersedes the collection of HCI data [64]. An expected revolution in the outlook of 3D HCI data use is likely to come from adoption of next-generation hardware; most notably light sheet microscopy adaptations that provide plate-based imaging facility, reduced photodamage, and improved acquisition speeds for suitable z-sampling to enable accurate measurements of 3D substructure [65,66]. The reader is referred to the Hardware section of this article for further discussion on light sheet and options available. Other HCI methods are under exploration to improve predictions of drug toxicity, which are particularly effective when augmented with omics data sources [38,67]. A notable academic-industry partnership, the Omics for Assessing Signatures for Integrated Safety (OASIS) consortium, is leading efforts in

hepatotoxicity prediction involving image profiling and exemplifies a cross-sector effort with many additional benefits including rich annotation sources of in vivo studies against anonymized compounds, assay standardization, and generation of large, well annotated image databases that can be leveraged by AI/ML methods.

The category of assays relating to image-based morphological profiling are at a stage of maturity where they are now being practically applied in industry discovery efforts. Recursion Pharmaceuticals are driving the largest-scale application of image-based profiling, having placed Cell Painting at the centre of their Phenomics strategy for hit identification and progressing five assets to clinical trials in 2022. Recursion releases large annotated image reference sets, (e.g., RxRx3), which are composed of millions of images of tens of thousands of unique perturbations. RxRx3 alongside the JUMP consortium, and OASIS provide support for innovative method development and benchmarking of computational approaches, which has had a large impact in extending utility of Cell Painting data [68] (Fig. 3).

To maximise the capacity of these large datasets as inference systems, the HCI field requires understanding of the image acquisition landmarks (i.e. perturbation replicates, sample size, plate distribution, incubation time, etc.) and analytical methods by which newly collected datasets might integrate seamlessly and demonstrate sufficient statistical similarity in order to reliably annotate perturbation cell states. An underappreciated but critical aspect when generating data to build or query these large datasets relates to careful quality control in wet-lab procedures. Indeed, for this reason industry settings are optimizing generation of high quality profiles by adopting standardised wet-lab approaches to align with the latest academic protocols (e.g., JUMP Cell Painting), full process automation, and incorporating nuisance compound sets [69,70]. Furthermore, industry is improving data quality by systematically addressing batch effects through adoption of effective quality controls including cell line stratification, plate position randomisation, cell counts, staining distribution, and image quality control measures such as focus scoring.

Image profiling exemplifies the inextricable integration of HCI within industry, as it has improved drug discovery pipelines by expediting target identification through to medicinal chemistry, providing an opportunity for in silico drug prediction, and serving as a cornerstone for AI-driven drug discovery.

4. Evolution of high content image analysis

In an image analysis experiment, a data scientist outlines, or segments, objects of interest, such as cells, in order to extract numerical descriptors suitable for downstream statistical and machine learning analyses. While there is no one-size-fits-all pipeline for all imaging datasets, we are converging on a canonical image processing pipeline (Fig. 4) [71–73]. Evolving organically, the pipeline includes image quality control, image correction, cell segmentation, cell feature extraction, and batch effect correction. After the mid-2000s, various methods have been developed to perform each step in this pipeline, but one of the most common approaches uses CellProfiler, which is a user-friendly, flexible tool that facilitates image data processing with a dynamic plugin system to incorporate and improve various pipeline steps [74]. CellProfiler allows automated image analysis and object segmentation using intensity thresholding and watershed-based methods. In addition to image segmentation, CellProfiler orchestrates the pipelines, and decoupling each of these steps has led to independent optimization and many analysis improvements.

As a first step after image acquisition, image quality control can be a manual and laborious process that is user subjective. Efforts to automatically flag cells based on poor focus and debris using machine learning and simulated data have reduced manual requirements thus increasing throughput and confidence in biological findings [75,76]. Next, image correction, which adjusts technical artefacts based on image capture, is an often overlooked step that is growing in appreciation and importance [77,78]. The most common adjustment is illumination correction (IC), which adjusts for uneven lighting induced by the microscope; most often a phenomenon called vignetting, which causes the edges of the field of view to be darker than the centre [78]. There are currently several different methods that adjust for illumination [77–79], and different microscopy approaches may require unique solutions (e.g. modelling live cell imaging for increased photobleaching over time) [79]. While increasing in importance, the field currently lacks approaches to systematically identify if illumination correction is needed, if it was successfully applied, or the extent to which it impacts biological findings. Furthermore, in multiplex imaging applications, stains can have overlapping emission wavelengths resulting in bleed through across spectral channels, which is particularly important to adjust for when measuring colocalization between structures. However, efforts to

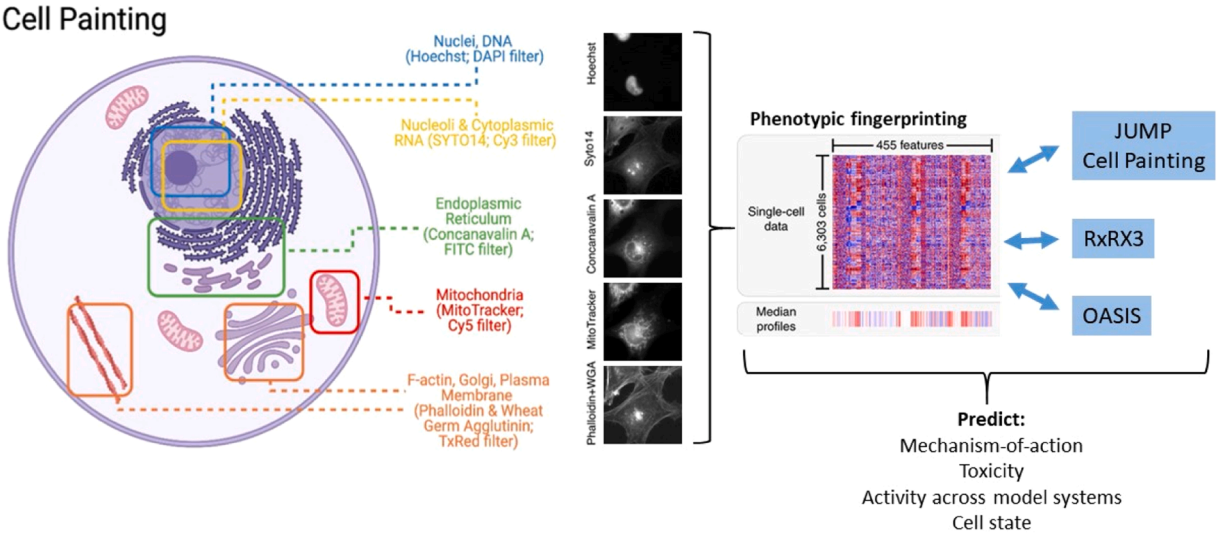

Fig. 3. The Cell Painting assay utilises a collection of fluorescent dyes to label multiple subcellular compartments, image analysis algorithms can then measure multiple features in each of these compartments to create a phenotypic fingerprint for every cell before and after compound treatment. Compound or genetic-induced phenotypic fingerprints can be interrogated by multivariate statistics or machine learning models to classify cell phenotypes and predict compound mechanism-of-action, toxicity and activity across other assays and model systems. Consortia and/or public datasets which are exploiting Cell Painting data include: JUMP-CP (https://jump-cellpainting.broadinstitute.org/); RxRX3 (https://www.rxrx.ai/) and OASIS (omics for assessing signatures for integrated safety consortia).

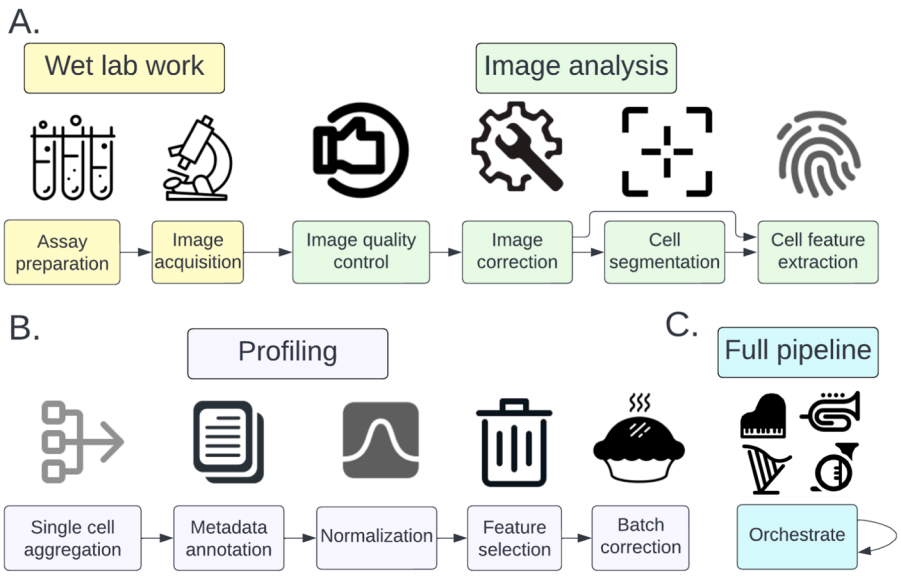

Fig. 4. Standard HCI experimental pipeline. After experimental design (A) scientists perform wet lab work to acquire high content cell images, which then requires several canonical image analysis steps. Cell segmentation is optional, but will allow single-cell profiling downstream. After image featurization, (B) scientists perform all the image-based profiling steps, to prepare data for downstream analyses. (C) This full pipeline must be orchestrated by reproducible software tools to ensure data provenance and to enable benchmarking. Both wet lab and dry lab biologists must be included in all processes from experimental design to results interpretation.

compensate for this bleed through are also in early days and require more methods, software development and statistical benchmarking [80]. Following image quality control and adjustments, data scientists extract high-dimensional cell biology features, which describe various phenotypes, cell states, and technical artefacts. There are existing tools to extract so-called hand-engineered features, which are based on classical computer vision algorithms [81–84]. There are also emerging solutions, based on deep learning, which promise to learn more informative morphology features [35,85-91]. While deep representation learning is a hot topic, it is still yet to be seen if these features will supplant the more interpretable hand-engineered features that the computer vision community has developed over decades. Additional methods, based on batch effect correction are also becoming increasingly important as data size increases, and it is unclear at what stage to perform batch correction either as an image-processing step or post feature extraction [92].

Cell segmentation is a particularly challenging and important step because of the huge variability in imaging equipment, imaging modalities, fluorescence markers, and therefore it has received a lot of research attention. It is clear that different segmentation algorithms impact how data scientists identify objects [93], but to the extent that segmentation impacts biological insights is yet to be determined. In the early days of HCI, most segmentation methods were based on manual thresholding, which is time consuming and error-prone. Localization of object centers can also be achieved using the minimum spread square loss function [94]. Today, many machine learning methods have emerged to automate segmentation of HCS data. Most segmentation models are inspired by a U-Net architecture that utilizes downsampling encoding layers followed by upsampling decoding layers to segment the image. [95] U-Net also includes skip connections between the encoder and decoder to preserve the spatial information from the input image, which improves segmentation accuracy. For example, the popular models StartDist [96], CellPose [97], DeepCell [98] are modified U-Net architectures that were trained on huge datasets and are now capable of segmenting a wide variety of image data with minimal or even no training. Furthermore, software tools such as ICY [99], QuPath [100], and ilastik [42] offer the user more flexibility to train their own algorithms. These approaches often require significant user input for challenging datasets with high confluence or heterogeneous cytoplasm, which makes generalization to other datasets difficult. However, once trained on a particular dataset they can be improved through fine-tuning. In summary, deep learning can learn the important features required for accurate cell segmentation directly from raw images and can handle heterogeneous imaging data capturing various staining and imaging modalities. It remains a rich research area with many groups proposing new approaches, both generic and cell type specific, which extends beyond high content imaging to other data modalities such as electron microscopy, pathology, and spatial transcriptomics [95,97, 101–103].

Due to the rapid advances in imaging technologies, we are able to capture different biological scales that consist of highly variable structures from organelles and molecules to organoid and vascular morphology. To date, bespoke methods are often required to segment these images [104]. However, we anticipate new deep learning models to be able to segment various types of objects. For example, the recent release of the Segment Anything Model (SAM) by Meta [105], which is not restricted to biological imaging, could be a very promising direction. For example, within a few clicks we were able to obtain very accurate segmentation of challenging tissue image data (Fig. 5).

Other emerging tools include an academic-industry partnership with the University of Wisconsin-Madison and Microsoft presenting their segmentation model Segment Everything Everywhere All At Once (SEEM) [106]. These large models based on transformer architectures offer zero-shot learning for a variety of generalized tasks, including cell segmentation, and may represent the next generation of segmentation approaches able to immediately generalize to diverse datasets. In the coming months and years, our field will continue stress testing and fine-tuning these approaches in their application to HCS segmentation.

5. Evolution of high content data analysis pipelines towards multiparametric and phenotypic profiling applications

In 2004, Perlman et al. published a landmark paper describing the ability of multiparametric high content phenotypic measurements to derive compound fingerprints, which showed that compounds with similar mechanism-of-actions induced similar cell morphologies [17]. Early examples from academia and industry explored the use of machine learning classifiers to predict mechanism-of-action of phenotypic hits by comparing the similarity of their high content phenotypic profiles with a reference library of well-annotated compounds [22,23]. Further development of high content phenotypic profiling assays combined with multivariate statistics and machine learning led several academic groups and academic-industry collaborations to further demonstrate the utility of image-based phenotypic profiling in discriminating phenotypes [22, 23,107–109]. These initial screens generated biological insights from terabytes of imaging data and were no doubt important and useful

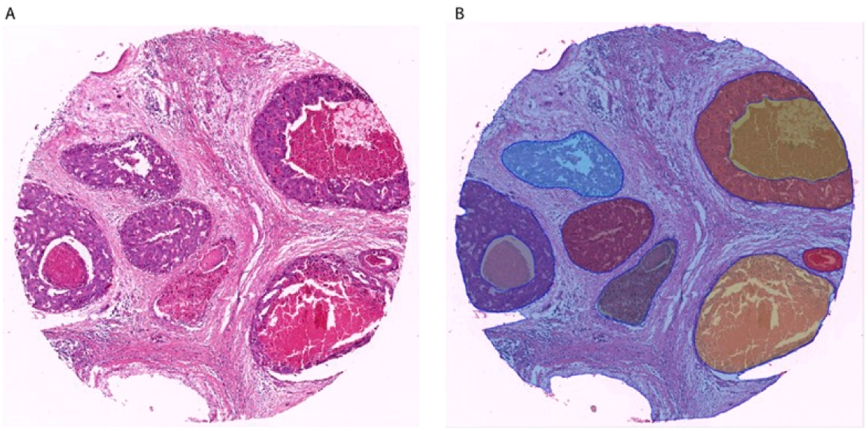

Fig. 5. The general-purpose Segment Anything Model (SAM) can accurately segment mammary ducts in Breast Ductal Carcinoma samples just with 5 clicks. (A) Represents raw image. (B) Represents SAM generated segmentation masks.

applications. However, the analysis pipelines and software infrastructure to handle this early data deluge were underdeveloped. This kicked off a scientific arms race between data collection and data analysis, and the cycle continues today as larger and larger datasets are continuously generated that outpace our ability to fully analyze them. As we are screening more and more drugs, we are also being humbled to learn that finding effective drugs with HCS is difficult, and therefore, we are pairing increased data collection with concurrent improvements in phenotypic profiling methods and software with the expectation that better infrastructure and computational approaches yield higher value screens. Novel hit calling or ranking methods that can be applied to Cell Painting profiles are of particular interest. Existing metrics, for example scalar projection, have been successfully used to rank profiles against on and off-perturbation phenotypes [110]. However, many of these measures have limitations and the field will benefit from increased collaboration with experts in mathematical disciplines to build improved approaches of similarity ranking.

Our increased ability to analyse high content data and thus derive insights from phenotypic profiling applications involve advancing method development at each step in the processing and analysis pipelines. After extracting thousands of features from each image or cell object, a data scientist must apply a bioinformatics pipeline to process these features, preparing them for downstream discovery. Much like with image analysis, this data analysis strategy has evolved organically within academic and industry labs [71]. Scientists and engineers have developed specific software tools for HCS data processing including Pycytominer [73], bioprofiling.jl [72], Phenonaut [111], and StratoMineR [112] which use either open source or closed source strategies. Canonically, the bioinformatics steps include single-cell processing to aggregate features within each well, metadata annotation, normalisation, feature selection, and consensus signature discovery (Fig. 4B). If images were not flagged in the image analysis steps, these pipelines can also filter single cells to remove incorrect segmentations, debris, out of focus cells, or other issues that may confound results. HCS routinely measures millions to billions of single cells, which makes single cell analysis extremely challenging as current data analysis infrastructure scales poorly to this many data points. Therefore, the aggregation step, while inherently losing single-cell heterogeneity information, is currently required. The aggregation step may also remove the need for additional quality control filtering as single cells will not contribute much when transforming features to their median value. Nevertheless, while the promise of microscopy is its inherent single cell nature, we will only realize this promise in HCS by developing new scalable software and methods to leverage single cell information, which are currently being developed [113].

A decade ago, building a method and demonstrating utility was good enough for high impact. While not universally the case now, it is much more impactful to release methods with usable, well documented, version controlled software. This software facilitates method application, development, and benchmarking and lives on to be further developed well beyond method publication. We have seen this success with CellProfiler [4] and FiJi [114] and now Napari [115], as software evolves much faster than print. In addition to software for methodology, we also require software for reproducible orchestration of data analysis pipelines. CellProfiler serves as an image analysis orchestration engine, but is one that requires extensive biological expertise and experience with manual parameter toggling. Pipelining software such as snakemake [116], Workflow Description Language (WDL) [117], or nextflow [118] will enable rapid pipeline development, repurposing, and benchmarking, but there are currently no existing orchestration engines tailored specifically to full HCS data analysis pipelines (Fig. 4C). Benchmarking each individual pipeline step is also incredibly challenging and requires standard benchmarks the field agrees upon. It is possible to test each individual step in isolation, but until we test how each step impacts the overall HCS end goal of quantifying phenotype, it's difficult to determine and compare performance. To date, most large-scale screens analyse their data with bespoke pipelines that are presented separately from data with questionable reproducibility. Software like AnnData [119] facilitate scalable data for specific scientific ecosystems, but the emerging paradigm is for language agnostic data types based on the Apache Software ecosystem to maximise programming language cross-compatibility and cloud computing [120].

Another important step in analysing HCS data is the ability to effectively link features to images in visualizations. Effective data visualization plays an important role when communicating results for facilitating biological interpretation. General-purpose tools, such as heatmaps or t-SNE plots, do not fully capture the structural nature of cell image data, and few tools have emerged to tackle these domain-specific challenges. For example, PhenoPlot [121] and the subsequent ShapoGraphy [122], allow generating pictorial quantitative representation of data using glyph visualisation. The tools map images to cell-like structures which enable easier interpretation (Fig. 6). The user can design their own structure in a web based app depending on the parameters they are extracting from the images. Other tools that allow data interaction in a graphical user interface include Mineotaur [123], Facetto [124], or Loon [125]. These tools link quantitative data points with raw image data to identify key trends in cell image features.

Data analysis advances developed for HCS applications are extending into cell biology studies more broadly, as the increasing volume of experiments are collecting datasets that outpace our ability to analyse everything by eye [126]. Accelerating this life cycle, publicly-available repositories dedicated to large-scale imaging data are coming on the

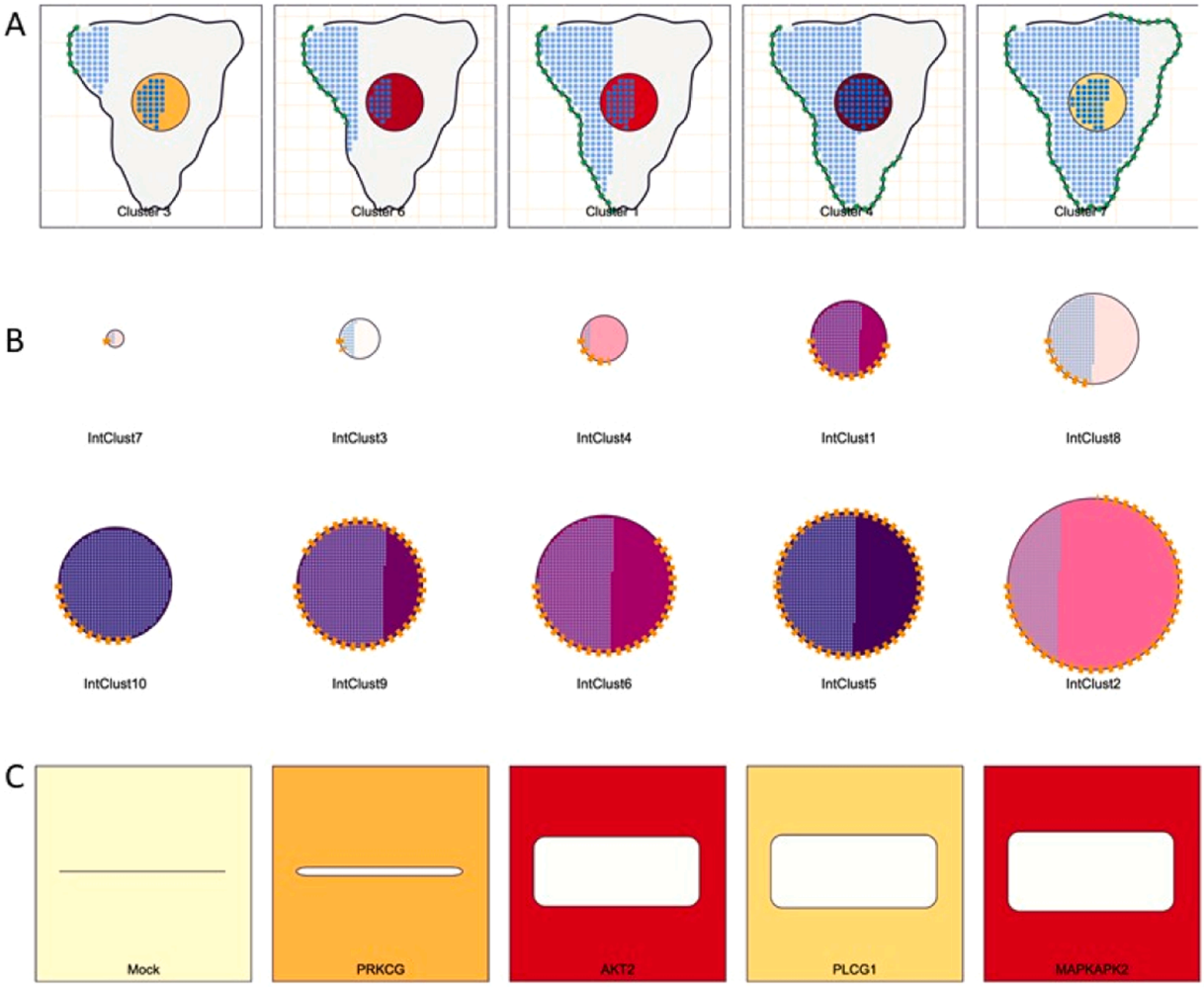

Fig. 6. Examples on data visualisation using ShapoGraphy (www.shapography.com). (A) Representation of cell signalling in the membrane, cytosol and nucleus in HeLa cells using symbols and colour. (B) Features of tumour or organoid features such as cellularity, invasion and size by mapping data to a circle-shaped object. (C) Representation of changes in wound healing assay (white rectangle) and associated number of cells based on colour.

scene (see section: *The role and evolution of image data repositories and sharing standards*), which increases the pace of method and software development and increases our agility and pace at fully leveraging HCS data to discover effective treatments.

6. The role of data integration and multiomics

The opportunity for integration of imaging and omics data is that information from integration will be greater than the sum of parts. Omics platforms have been deployed but integration remains an active area of research that will benefit from further academic-industry collaboration and the use of new frameworks for analysis. Multiomics refers to the collection and use of readouts from multiple omics technologies, aiming to capture the complete state of (macro-)molecules present in the measured biological substrate [127]. Practically, this refers to genomics, transcriptomics, proteomics, metabolomics [128], and commonly includes phenomics derived from HCI/HCS and other data sources which might provide complementary information on cell states. AI/ML literature uses the more general term "Multiview" to refer to using multiple representations or "views" of an underlying system state capturing information across different timescales, resolutions and with different batch effects and biases. It has been demonstrated multiple times that combining omics data, to derive cellular state, results in the addition of unique complementary information, enhancing performance in prediction tasks [129,130]. As earlier noted, roughly half of studied genes produce a detectable morphological change [50,131], and therefore it is often seen as the job of complementary omics to fill these blindspots. As well as providing a more complete detection of cellular responses, Joyce and Bernhard [132] document a pairwise matrix of omics views and resolve the potential biological insights achievable for all pairs including enzyme annotations, regulatory complexes, binding sites, gene regulatory networks, and functional annotations. Whilst powerful, complementary views are costly to generate in terms of time, reagents and data analysis requirements. This has typically limited multiomics to the bookends of the drug discovery pipeline. At the beginning of the pipeline, target validation makes heavy use of multiomics [133], applying multiple techniques to ensure perturbation of a given target leads to the desired therapeutic response. At the opposite end of the pipeline, candidate validation aims to collect extensive information on the most promising treatments, ensuring they are hitting the correct targets with little or no off-target effects and low toxicity [134,135]. Traditionally, the middle of the pipeline where throughput is critical, has not been augmented by the collection of multiomics data, However, this is now changing as speed of data acquisition and analysis increases, assays improve, automation increases in scale, and costs decrease. Scientists are collecting more imaging data alongside other omics technologies, which are transcending the traditional single readout high throughput screening setups. With continued development of assay technologies, increased multiomics throughout the discovery pipeline would allow many benefits, such as earlier triage of problematic

treatments, prioritisation of the most promising agents, and the overall ability to make better informed decisions during the discovery process. Thus, HCI is a prime candidate for further integration into the entire drug discovery pipeline.

Currently, science is embracing multidisciplinary applications of techniques from diverse fields, including information theory, computer science, machine learning and statistics, for integration of different omics views aiming to improve predictions over single omics/single view applications. Whilst many approaches exist, they can all be placed into three broad categories as defined by Rappoport [136]; early-, mid- and late-stage integration. Early integration combines different views in a pre-processing step isolated from any prediction task. Examples of this include simple concatenation of features across views [130], while more involved transformations align correlated features across views [137]. Mid-integration refers to techniques in which feedback from the prediction task is used to direct the transformation and joint embedding of views in a shared phenotypic space [138]. Finally, late integration covers approaches which carry out predictions on individual views and then unify predictions using a mechanism like consensus scoring [139, 140]. Another example of late integration is correlating phenotypes discovered in images to other omic data measuring comparable samples which can be applicable to single and bulk datasets [141]. Extensive reviews of multiomics integration techniques may be found in literature [136,142,143].

With such a wide choice of integration approaches available in literature, there is no clear choice of what is subjectively 'best', which cannot be determined without a comprehensive benchmark matching the omics views available to the researcher, experimental setup and prediction task. Integration techniques that work well for certain omics views are unlikely to be performant across different pairs, triplets or higher order groups. A benchmark prioritising powerful omics to pair with phenomics and a selection of prediction tasks would be a valuable asset to all within the HCI community. While compound mode of action prediction is arguably the most prioritised task in HCI and drug discovery, large public datasets are only now becoming available to benchmark AI/ML techniques applied to HCS data [51,68]. Critically for multiomics integration, benchmarking using HCI requires pairing with omics views from different sources, which will likely increase batch effects.

While a common integration task aims for better predictions, one view may also be used to predict another, with examples demonstrating the use of small molecule structure [144,145] and HCI [146,147] to predict proteomics profiles. The use of fast low cost technologies like HCI to make predictions in this manner allows application of well supported tools and databases outside of the collected omics technology ecosystem, allowing application over more of the drug discovery pipeline.

7. The role and evolution of image data repositories and sharing standards

In HCS experiments, scientists collect a large amount of data, which makes tracking both the data itself and information about the data challenging. This information about data, also known as metadata, is critical for reproducibility, especially given the high costs and low success rates of these experiments. This need for reproducibility is even more critical given the current reproducibility crisis in many fields. In our era where experiments fail to replicate more often than not [148], we must continue to report as much metadata as possible, using standardized identifiers [149]. Recently, researchers have proposed a Recommended Metadata for Biological Images (REMBI), which aims to provide metadata guidelines for diverse microscopy communities, begin discussion about standardising metadata identifiers, and promote reuse of microscopy datasets [150]. Over the past five years, more emphasis has been placed on sharing these large datasets, which has increased the importance of tracking metadata information and reproducibility. These efforts all fall under the umbrella of FAIR research, which aims to make data Findable, Accessible, Interoperable, and Reusable [151]. Ultimately, the results and utility of the initial HCS are validated downstream in the efficacy of the hits that were prioritised for follow-up characterization. Therefore, the scale and lag time to validation makes assessing experimental reproducibility especially challenging. Conversely, the reproducibility of the computational analyses is easier to track since it can be directly tested, however, this requires providing version-controlled code for the full pipeline and version-controlled computational environments.

Metadata standards and improved file types are important for the growth of data sharing and reuse, but the heterogeneity of microscopy data makes it challenging to keep track of all the different items. These include microscopy parameters, cell culturing conditions, assay materials, perturbation and other treatment details, amongst others. Emerging standards for these identifiers are improving computational analyses and machine readability to facilitate data reuse. Ten years ago, such standards did not exist, which made most data non-interoperable or required significant effort to convert metadata to the same language. Furthermore, the evolution of file types is ongoing, with TIFFs being the de facto file type shared in the past. While they contain images alongside important metadata, they can be slow to load and are not optimised for cloud computing. Emerging file types that fit the needs of academia and industry in HCI are actively being developed, such as OME-ZARR [152]. File types are also evolving for intermediate data types in table format, moving from CSV to database standards and more performant data based on Arrow (e.g., Parquet) that are now being tested and implemented. It is crucial to consider interoperability when designing new file types to ensure that they are widely adoptable by the scientific community.

Data sharing resources are now equipped to handle large high-content datasets, enabling researchers to share and access a vast amount of data with ease. One such resource is the Image Data Resource (IDR), a growing service that currently stores over 100 studies totalling about 400 TB of images of reference cell and tissue imaging data [153]. IDR can version control your data and provide a direct object identifier, allowing anyone to use and cite the data deposited. There are several other microscopy data sharing resources available including, EMPIAR [154,155], BioImage Archive [156], and Cell Image Library [157] but IDR is the primary third-party host for HCS data. More recently, the Registry of Open Data on AWS (RODA) is now storing large HCS datasets, with the JUMP Cell Painting consortium currently hosted there. There are also several in-house data sharing repositories, such as the Allen Institute for Cell Science Explorer, Recursion RxRx [158], Broad Bioimage Benchmark Collection [159], and various others operated by academic institutions. While there are many advantages to self-hosted data repositories, they tend to have less emphasis placed on metadata standards and identifiers, and enforcing compliance can be more challenging. There remains a need for further standardization and interoperability between these repositories and sharing platforms. As mentioned earlier, the heterogeneity of microscopy data makes it challenging to develop and implement standardised metadata and file types. Continued efforts to develop and improve these standards will be crucial in promoting efficient and widespread sharing of high-content microscopy data. However, with the growth of these data sharing resources, the possibility of microscopy data reuse and secondary analyses is growing, which will lead to increased use in validating experiments and testing model generalizability. In addition to sharing raw images, it is also helpful to share other high-value intermediate data types, such as illumination corrected images, embeddings, and single-cell and bulk feature extraction methods, which may be shared in other repositories. Wilson et al. describes more considerations for sharing microscopy images [154] .

The field of microscopy data sharing is still in its early stages, but there has been significant progress made over the past decade and the evolution of these repositories and standards will likely continue to be

shaped by advances in technology and changes in research practices. For example, the growing use of AI/ML in image analysis may require new approaches to data sharing, data management, and a heightened emphasis on standardization and reproducibility. One important challenge is ensuring that the data used to train machine learning models is consistent and well-documented. If the data are not standardized, then the models will not be generalizable to other datasets, leading to poor performance and limited impact. Another challenge is the interpretability of machine learning models in the context of microscopy data. While these models can often achieve impressive results, it is important to understand how they are making their predictions, especially in the context of drug discovery where a false positive or negative could have significant consequences. Efforts to develop interpretable machine learning models and standards for reporting their performance and predictions will be crucial in ensuring their utility in this field. Therefore, it will be important to continually evaluate and adapt image data repositories and sharing standards to ensure they meet the needs of the scientific community. As we continue to share and analyse high-content microscopy data, we will accelerate the pace of drug discovery and ultimately lead to faster, more cost-effective cures for a wide range of diseases.

8. Future perspective of high content imaging hardware and software

Hardware: Following a sustained period of improvements to commercial HCI instruments over several decades, accelerated technology development over the past 5 years has delivered significant advances in capability. These advances include; improved 3D imaging as exemplified by the OperaPhenix (Perkin Elemer); ImageXpress-confocal HT.ai (Molecular Devices) and CellInsight CX7 LZR Pro (ThermoFisher) platforms; unprecedented speed such as that demonstrated by endeavor GT (Araceli) and new capabilities that combine single cell imaging and sample picking for multiomics analysis provided by the Single Cellome™ System SS2000 (Yokogawa) (Fig. 1). Other advances include development of bespoke HCI platforms specifically designed for small model organisms such as Zebrafish (VAST BioimagerTM (Union Biometrica) [160].

Despite these significant advances the development and implementation of HCI infrastructure has often struggled to keep pace with the ever increasing diversity and complexity of a new generation of ever evolving and more sophisticated 3D models and microfluidic devices [161,162]. Thus, a number of HCI challenges and gaps remain including acquisition and analysis of 3D models with single cell and subcellular resolution at depth to explore the heterogeneity of complex 3D multicellular structures in both fixed samples and longitudinal monitoring in live cell assays [163]. Multiphoton microscopy is currently the most powerful technique for realising single cell fluorescence imaging and segmentation at depth, however conventional multiphoton microscopy is too slow for high throughput screening applications across sufficient sample numbers. Research efforts are underway to increase the speed of multiphoton microscopy through parallelized signal acquisition using multiple laser beams or time-gated camera detection systems, which have demonstrated proof-of-concept including in automated multiwell plate formats [164–166]. Other advances in HCI hardware development from academia include adaptation of “light sheet” microscopic imaging for multiwell plates [167]. Light sheet fluorescence microscopy (LSFM) uses a thin sheet of light to excite only fluorophores within a single planar volume in front of the objective [168]. Light sheet microscopy therefore provides true optical sectioning capability facilitating 3D imaging with reduced photobleaching and phototoxicity of the sample. Oblique Plane Microscopy (OPM) is a “fast light sheet” microscopy technique that uses a single high numerical aperture microscope objective to both illuminate a tilted plane within the specimen and to collect fluorescence from the tilted illuminated plane [169,170]. As OPM is compatible with a conventional microscope, it can be used to image conventionally mounted specimens on coverslips, tissue culture dishes or standard multiwell plates. The OPM has demonstrated its rapid multiwell plate imaging capability to image 3D responses of tumour spheroids to glucose over time and to map and quantify cell morphological plasticity in 3D [65,171]. Alternative platforms which are commercially available include the Lattice Light Sheet 7 from Zeiss which delivers an easy-to-use automated light sheet instrument suitable for multiple sample carriers including multiwell plates.

The development of multidisciplinary consortia encompassing expertise in photonics, automated microscopy, image analysis and biology to deliver open source hardware and software solutions that can be exploited by both commercial and academic institutions is well placed to accelerate the evolution and adoption of high content imaging. The MACH3CANCER (advancing Microscopy to Accelerate understanding of Complexity and Heterogeneity of 3D Cancer) is one such consortium funded by cancer Research UK (https://mach3cancer.org/). The tools and resources developed by MACH3CANCER, which will be shared with the community, include: new single cell-resolved open source high content analysis platforms (openHCA) specifically designed for 3D cell cultures and organoids. The open source approach will enable other laboratories to replicate these capabilities and ensure that the openHCA instrumentation can be easily upgraded to new functionality (with no barriers due to proprietary hardware or software) and be integrated with other HCA capabilities including commercial platforms. Another academic-industry consortium includes the Transformative Imaging for Quantitative Biology (TIQBio) partnership which aim to provide advances in high resolution imaging of 3D models in an unperturbed manner.

Software: As more complex multiparametric data analysis approaches have evolved for HCI, analysis pipelines may include techniques integrating ‘traditional’ algorithms spanning statistics, signal processing and information theory to techniques from AI/ML including classification, regression, pooling, sampling, normalisation, batch correction, imputation, transformation, and application of generative methods (see section “*Evolution of high content data analysis pipelines towards multiparametric and phenotypic profiling applications*” for more details). Closed-source proprietary tools exist and allow non-expert users to apply pipelines including such techniques like BIOVIA Dassault Systèmes’ Pipeline Pilot, Perkin Elmer’s HC profiler (powered by TIBCO Spotfire®), and HCS StratoMineR [112]. However, the significant contributions of academia to the field highlights the cutting- and often bleeding-edge nature of research being carried out within these institutions, requiring unhindered access to source code, intermediate data and the ability to migrate analysis to a variety of compute platforms unrestricted by licensing and access controls. Critically, integration of new literature techniques is rapidly enabled with the drive towards open access publishing and adherence to FAIR principles [151], resulting in source code for new techniques often being readily available in public repositories. Rapid iteration and integration of techniques is massively helped in an open-source software ecosystem, where contributions from many experts in their fields may flow into one well structured and well managed software package for unrestricted use, evaluation, and improvement by the community. Continued evolution of high content analysis includes the need to integrate HCI with multiple data types (see section “*The role of data integration and multiomics*”) [130]. Data integration workflows for multiomics data take many forms across academia and industry and efforts with limited resources can easily fall short of data integration best practices, with additional data and processing requirements dramatically increasing pipeline complexity upon combination and processing of high content imaging, proteomics, metabolomics and other omics data. Open source initiatives such as Phenonaut [111] aim to standardise multiomics and single omics workflows operating on feature data, but will only be successful in establishing standards and driving the field forward in an open, collaborative, and FAIR way with community involvement addressing new methods, benchmarking, and best practices.

9. Current gaps and recommendations

While significant advances in HCI analysis have evolved to extract multiple features and classify a broad variety of cell phenotypes in 2D cultures at single cell level, the majority of phenotypic profiling studies are performed on aggregated whole well/cell population level. This is particularly true for 3D model systems where high content phenotypic profiling at the single cell level at sufficient imaging depths remains to be fully realised. Improvements in imaging hardware for 3D resolution and software solutions for handling single cell level data are required to study the heterogeneity of cellular response and distinct cell types in more complex 2D co-culture and 3D cell models. Such developments will support the next evolution of high content phenotypic profiling applications using more complex and physiologically relevant model systems. The evolution towards open source HCI hardware could deliver new HCI instruments and upgrades at reduced costs supporting expansion of the technology beyond core screening facilities and increased adoption of robust quantitative cell biology across industry and academic research groups. The development and evolution of HCI hardware may also support a paradigm-shift in the development of screening applications beyond standard multiwell plate formats toward more bespoke and physiologically relevant 3D models and microfluidic devices.

The field of single cell technology is advancing rapidly, evolution of HCI in parallel with other single cell technologies including integrating image-based single cell phenotypic classification followed by cell picking and collection to feed into single cell transcriptomics and proteomics profiling is becoming realised with new platforms such as the Yokogawa SS2000, Beacon(R) optofluidic system, and Sartorius CellCelector platforms. However, this is an area that requires additional investment to deploy across research programs and user groups and maximise the full potential to embrace the heterogeneity in cell phenotypes within biological samples and more comprehensively explore cell state transition at an integrated phenotypic, transcriptomic and post-translational pathway level.

One of the benefits of HCI and image-based phenotypic profiling is low cost when applied at scale relative to genomic, transcriptomic, and proteomic profiling technologies. This disparity however limits multiomics integration resulting in gaps in methods for data integration and analysis. Further investment in generation of orthogonal high throughput transcriptomic and proteomic profiling technology and data sets which can be paired with appropriately matched HCI data would support benchmarking of different data integration approaches. Investments in the provision of publicly available multiomic datasets across platforms at different scales to integrate with HCI datasets is required to evolve the field of quantitative biology and HCI applications.

A concerted move away from development of bespoke image analysis pipelines which are separate from the data towards integration of language agnostic data types associated with raw image data will maximise cross comparison and quality assessment of analysis approaches, adoption across multiple research programs and institutions and thus increased scalability.

10. Conclusion

Academic and industry contributions to the field of HCI have been complementary and synergistic. Consortia activity and collaboration which provide precompetitive tools and datasets have been crucial to continued evolution of the field and include new applications and broadly adopted strategies for precise classification of cell phenotypes and prediction of biological mechanism-of-action and in vivo toxicology. These developments have contributed to significant evolution of HCI technology and applications to answer fundamental basic research questions, perform discovery science, and to improve decision making across discrete stages of the drug discovery process.

Further interdisciplinary collaboration between data science, biological assay development and HCI hardware solutions will continue to evolve the field towards delivering higher quality and more quantitative solutions which also support more disease relevant and mechanistically informative drug discovery applications. The past three decades since the inception of HCI has been witness to substantial improvements in technology and applications which have stimulated and continue to stimulate increased adoption of HCI across research institutions in academic and industry sectors. The future of HCI looks bright: undoubtedly the replacement of manual microscopic imaging and analysis with automated solutions provide a step change in increased robustness of biomolecular imaging and functional biology studies, which are less prone to bias and artefacts due to low sample throughput. This step change and the availability of a greater variety of HCI platforms and open source software and data analysis solutions will ensure the trajectory of increased HCI adoption continues. The rapid development of new technologies from the fields of 3D biology, CRISPR gene-editing, human iPSC, multiomics, single cell technologies and AI/ML are converging with HCI contributing to a new era of cell biology and drug discovery (Fig. 2). Together these developments promise to contribute significant benefits across multiple research areas including data science, cell biology, chemical biology and healthcare which in turn will support continued investment in HCI and further enhance the significant impact that HCI contributes to biomolecular and biomedical research.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by a UKRI Medical Research Council award (grant number MRC/W003996/1) to SS and NOC.

References

[1] Taylor DL. Past, present, and future of high content screening and the field of cellomics. Methods Mol Biol 2007;356:3–18.
[2] Abraham VC, Taylor DL, Haskins JR. High content screening applied to large-scale cell biology. Trends Biotechnol 2004;22:15–22.
[3] Baatz M, Arini N, Schäpe A, Binnig G, Linssen B. Object-oriented image analysis for high content screening: detailed quantification of cells and sub cellular structures with the Cellenger software. Cytometry A 2006;69:652–8.
[4] Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin DA, Chang JH, Lindquist RA, Moffat J, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol 2006;7:R100.
[5] Horvath P, Wild T, Kutay U, Csucs G. Machine learning improves the precision and robustness of high-content screens: using nonlinear multiparametric methods to analyze screening results. J Biomol Screen 2011;16:1059–67.
[6] Sun D, Gao W, Hu H, Zhou S. Why 90% of clinical drug development fails and how to improve it? Acta Pharm Sin B 2022;12:3049–62.
[7] Bakal C, Aach J, Church G, Perrimon N. Quantitative morphological signatures define local signaling networks regulating cell morphology. ScienceScience 2007;316:1753–6.
[8] Sailem H, Bousgouni V, Cooper S, Bakal C. Cross-talk between Rho and Rac GTPases drives deterministic exploration of cellular shape space and morphological heterogeneity. Open Biol 2014;4. https://doi.org/10.1098/rsob.130132.
[9] Chow YL, Singh S, Carpenter AE, Way GP. Predicting drug polypharmacology from cell morphology readouts using variational autoencoder latent space arithmetic. PLoS Comput Biol 2022;18:e1009888.
[10] Horvath P, Aulner N, Bickle M, Davies AM, Nery ED, Ebner D, Montoya MC, Östling P, Pietiäinen V, Price LS, et al. Screening out irrelevant cell-based models of disease. Nat Rev Drug Discov 2016;15:751–69.
[11] Lukonin I, Zinner M, Liberali P. Organoids in image-based phenotypic chemical screens. Exp Mol Med 2021;53:1495–502.
[12] Betge J, Rindtorff N, Sauer J, Rauscher B, Dingert C, Gaitantzi H, Herweck F, Srour-Mhanna K, Miersch T, Valentini E, et al. The drug-induced phenotypic landscape of colorectal cancer organoids. Nat Commun 2022;13:3135.

[13] Choo N, Ramm S, Luu J, Winter JM, Selth LA, Dwyer AR, Frydenberg M, Grummet J, Sandhu S, Hickey TE, et al. High-throughput imaging assay for drug screening of 3D prostate cancer organoids. SLAS Discov 2021;26:1107–24.

[14] Nguyen HTL, Kohl E, Bade J, Eng SE, Tosevska A, Shihabi AA, Hong JJ, Dry S, Boutros PC, Panossian A, et al. A rapid platform for 3D patient-derived cutaneous neurofibroma organoid establishment and screening. Biorxiv 2022. https://doi.org/10.1101/2022.11.07.515469.

[15] Kelm JM, Timmins NE, Brown CJ, Fussenegger M, Nielsen LK. Method for generation of homogeneous multicellular tumor spheroids applicable to a wide variety of cell types. Biotechnol Bioeng 2003;83:173–80.

[16] Wevers NR, van Vught R, Wilschut KJ, Nicolas A, Chiang C, Lanz HL, Trietsch SJ, Joore J, Vulto P. High-throughput compound evaluation on 3D networks of neurons and glia in a microfluidic platform. Sci Rep 2016;6:38856.

[17] Jiménez-Luna J, Grisoni F, Weskamp N, Schneider G. Artificial intelligence in drug discovery: recent advances and future perspectives. Expert Opin. Drug Discov. 2021;16:949–59.

[18] Brown N. Artificial intelligence in drug discovery. Royal Society of Chemistry; 2020.

[19] Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, Li B, Madabhushi A, Shah P, Spitzer M, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov 2019;18:463–77.

[20] Cook, S. (2013). Cuda programming: a developer's guide to parallel computing with GPUs (Newnes).

[21] Smith K, Horvath P. Active learning strategies for phenotypic profiling of high-content screens. J Biomol Screen 2014;19:685–95.

[22] Ljosa V, Caie PD, Ter Horst R, Sokolnicki KL, Jenkins EL, Daya S, Roberts ME, Jones TR, Singh S, Genovesio A, et al. Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment. J Biomol Screen 2013;18:1321–9.

[23] Caie PD, Walls RE, Ingleston-Orme A, Daya S, Houslay T, Eagle R, Roberts ME, Carragher NO. High-content phenotypic profiling of drug response signatures across distinct cancer cells. Mol Cancer Ther 2010;9:1913–26.

[24] Gupta A, Harrison PJ, Wieslander H, Pielawski N, Kartasalo K, Partel G, Solorzano L, Suveer A, Klemm AH, Spjuth O, et al. Deep learning in image cytometry: a review. Cytometry A 2019;95:366–80.

[25] Krentzel D, Shorte SL, Zimmer C. Deep learning in image-based phenotypic drug discovery. Trends Cell Biol 2023. https://doi.org/10.1016/j.tcb.2022.11.011.

[26] Janssens R, Zhang X, Kauffmann A, de Weck A, Durand EY. Fully unsupervised deep mode of action learning for phenotyping high-content cellular images. Bioinformatics 2021;37:4548–55.

[27] Godinez WJ, Hossain I, Zhang X. Unsupervised phenotypic analysis of cellular images with multi-scale convolutional neural networks. Biorxiv 2018. https://doi.org/10.1101/361410.

[28] Dürr O, Sick B. Single-cell phenotype classification using deep convolutional neural networks. J Biomol Screen 2016;21:998–1003.

[29] Godinez WJ, Hossain I, Lazic SE, Davies JW, Zhang X. A multi-scale convolutional neural network for phenotyping high-content cellular images. Bioinformatics 2017;33:2010–9.

[30] Lab Scientist to Direct Bioinformatics at CREA (2011). https://today.lbl.gov/2011/08/05/lab-scientist-to-direct-bioinformatics-at-crea/.

[31] Suprynowicz FA, Upadhyay G, Krawczyk E, Kramer SC, Hebert JD, Liu X, Yuan H, Cheluvaraju C, Clapp PW, Boucher Jr RC, et al. Conditionally reprogrammed cells represent a stem-like state of adult epithelial cells. Proc Natl Acad Sci U S A 2012;109:20035–40.

[32] Yuan H, Myers S, Wang J, Zhou D, Woo JA, Kallakury B, Ju A, Bazylewicz M, Carter YM, Albanese C, et al. Use of reprogrammed cells to identify therapy for respiratory papillomatosis. N Engl J Med 2012;367:1220–7.

[33] Cooper SEJ, Mohamed BM, Elliott L, Davies AM, Feighery CF, Kelly J, Dunne J. Adaptation of a cell-based high content screening system for the in-depth analysis of celiac biopsy tissue. Methods Mol Biol 2015;1326:67–77.

[34] Clevers H. Modeling Development and Disease with Organoids. CellCell 2016;165:1586–97.

[35] Schiff L, Migliori B, Chen Y, Carter D, Bonilla C, Hall J, Fan M, Tam E, Ahadi S, Fischbacher B, et al. Integrating deep learning and unbiased automated high-content screening to identify complex disease signatures in human fibroblasts. Nat Commun 2022;13:1590.

[36] Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. CellCell 2006;126:663–76.

[37] Huang CY, Nicholson MW, Wang JY, Ting CY, Tsai MH, Cheng YC, Liu CL, Chan DZH, Lee YC, Hsu CC, et al. Population-based high-throughput toxicity screen of human iPSC-derived cardiomyocytes and neurons. Cell Rep 2022;39:110643.

[38] Zhang CJ, Meyer SR, O'Meara MJ, Huang S, Capeling MM, Ferrer-Torres D, Childs CJ, Spence JR, Fontana RJ, Sexton JZ. A human liver organoid screening platform for DILI risk prediction. J Hepatol 2023;78:998–1006.

[39] Cheng C, Reis SA, Adams ET, Fass DM, Angus SP, Stuhlmiller TJ, Richardson J, Olafson H, Wang ET, Patnaik D, et al. High-content image-based analysis and proteomic profiling identifies Tau phosphorylation inhibitors in a human iPSC-derived glutamatergic neuronal model of tauopathy. Sci Rep 2021;11:17029.

[40] Hodis E, Torlai Triglia E, Kwon JYH, Biancalani T, Zakka LR, Parkar S, Hütter J-C, Buffoni L, Delorey TM, Phillips D, et al. Stepwise-edited, human melanoma models reveal mutations' effect on tumor and microenvironment. ScienceScience 2022;376:eabi8175.

[41] Rämö P, Sacher R, Snijder B, Begemann B, Pelkmans L. CellClassifier: supervised learning of cellular phenotypes. Bioinformatics 2009;25:3028–30.

[42] Berg S, Kutra D, Kroeger T, Straehle CN, Kausler BX, Haubold C, Schiegg M, Ales J, Beier T, Rudy M, et al. ilastik: interactive machine learning for (bio)image analysis. Nat Methods 2019;16:1226–32.

[43] Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, Ohl P, Sieb C, Thiel K, Wiswedel B. KNIME: the Konstanz Information Miner. Data analysis, machine learning and applications studies in classification, data analysis, and knowledge organization. Springer Berlin Heidelberg; 2008. p. 319–26.

[44] Gustafsdottir SM, Ljosa V, Sokolnicki KL, Anthony Wilson J, Walpita D, Kemp MM, Petri Seiler K, Carrel HA, Golub TR, Schreiber SL, et al. Multiplex cytological profiling assay to measure diverse cellular states. PLoS ONE 2013;8:e80999.

[45] Bray M-A, Singh S, Han H, Davis CT, Borgeson B, Hartland C, Kost-Alimova M, Gustafsdottir SM, Gibson CC, Carpenter AE. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat Protoc 2016;11:1757–74.

[46] Jamali N, Tromans-Coia C, Abbasi HS, Giuliano KA, Hagimoto M, Jan K, Kaneko E, Letzsch S, Schreiner A, Sexton JZ, et al. Assessing the performance of the Cell Painting assay across different imaging systems. Biorxiv 2023. https://doi.org/10.1101/2023.02.15.528711.

[47] Laber S, Strobel S, Mercader J-M, Dashti H, Ainbinder A, Honecker J, Garborcauskas G, Stirling DR, Leong A, Figueroa K, et al. Discovering cellular programs of intrinsic and extrinsic drivers of metabolic traits using LipocyteProfiler. Biorxiv 2021. https://doi.org/10.1101/2021.07.17.452050. 2021.07.17.452050.

[48] Fredin Haslum J, Lardeau C-H, Karlsson J, Turkki R, Leuchowius K-J, Smith K, Mullers E. Cell Painting-based bioactivity prediction boosts high-throughput screening hit-rates and compound diversity. Biorxiv 2023. https://doi.org/10.1101/2023.04.03.535328.

[49] Cross-Zamirski JO, Mouchet E, Williams G, Schönlieb C-B, Turkki R, Wang Y. Label-free prediction of cell painting from brightfield images. Sci Rep 2022;12:10001.

[50] Rohban MH, Singh S, Wu X, Berthet JB, Bray M-A, Shrestha Y, Varelas X, Boehm JS, Carpenter AE. Systematic morphological profiling of human gene and allele function via Cell Painting. eLifeeLife 2017;6. https://doi.org/10.7554/eLife.24060.

[51] Chandrasekaran SN, Ackerman J, Alix E, Michael Ando D, Arevalo J, Bennion M, Boisseau N, Borowa A, Boyd JD, Brino L, et al. JUMP Cell Painting dataset: morphological impact of 136,000 chemical and genetic perturbations. Biorxiv 2023. https://doi.org/10.1101/2023.03.23.534023. 2023.03.23.534023.

[52] International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. NatureNature 2004;431:931–45.

[53] Perlman ZE, Mitchison TJ, Mayer TU. High-content screening and profiling of drug activity in an automated centrosome-duplication assay. ChemBioChem 2005;6:145–51.

[54] Moffat J, Grueneberg DA, Yang X, Kim SY, Kloepfer AM, Hinkle G, Piqani B, Eisenhaure TM, Luo B, Grenier JK, et al. A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. CellCell 2006;124:1283–98.

[55] Neumann B, Walter T, Hériché J-K, Bulkescher J, Erfle H, Conrad C, Rogers P, Poser I, Held M, Liebel U, et al. Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. Nature 2010;464:721–7.

[56] Funk L, Su K-C, Ly J, Feldman D, Singh A, Moodie B, Blainey PC, Cheeseman IM. The phenotypic landscape of essential human genes. CellCell 2022;185:4634–53.e22.

[57] Keles H, Schofield CA, Rannikmae H, Edwards EE, Mohamet L. A scalable 3D high-content imaging protocol for measuring a drug induced dna damage response using immunofluorescent subnuclear γH2AX spots in patient derived ovarian cancer organoids. ACS Pharmacol Transl Sci 2023;6:12–21.

[58] Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nat Biotechnol 2014;32:40–51.

[59] Dowden H, Munro J. Trends in clinical success rates and therapeutic focus. Nat Rev Drug Discov 2019;18:495–6.

[60] Pognan F, Beilmann M, Boonen HCM, Czich A, Dear G, Hewitt P, Mow T, Oinonen T, Roth A, Steger-Hartmann T, et al. The evolving role of investigative toxicology in the pharmaceutical industry. Nat Rev Drug Discov 2023;22:317–35.

[61] Peel S, Corrigan AM, Ehrhardt B, Jang K-J, Caetano-Pinto P, Boeckeler M, Rubins JE, Kodella K, Petropolis DB, Ronxhi J, et al. Introducing an automated high content confocal imaging approach for organs-on-chips. Lab Chip 2019;19:410–21.

[62] Ewart L, Apostolou A, Briggs SA, Carman CV, Chaff JT, Heng AR, Jadalannagari S, Janardhanan J, Jang K-J, Joshipura SR, et al. Performance assessment and economic analysis of a human liver-chip for predictive toxicology. Commun Med 2022;2:154.

[63] Bell CC, Dankers ACA, Lauschke VM, Sison-Young R, Jenkins R, Rowe C, Goldring CE, Park K, Regan SL, Walker T, et al. Comparison of hepatic 2D sandwich cultures and 3D spheroids for long-term toxicity applications: a multicenter study. Toxicol Sci 2018;162:655–66.

[64] Yang S, Ooka M, Margolis RJ, Xia M. Liver three-dimensional cellular models for high-throughput chemical testing. Cell Rep Methods 2023;3:100432.

[65] Maioli V, Chennell G, Sparks H, Lana T, Kumar S, Carling D, Sardini A, Dunsby C. Time-lapse 3-D measurements of a glucose biosensor in multicellular spheroids by light sheet fluorescence microscopy in commercial 96-well plates. Sci Rep 2016;6:37777.

[66] Chatterjee K, Pratiwi FW, Wu FCM, Chen P, Chen B-C. Recent progress in light sheet microscopy for biological applications. Appl Spectrosc 2018;72:1137–69.

[67] Seal S, Carreras-Puigvert J, Trapotsi M-A, Yang H, Spjuth O, Bender A. Integrating cell morphology with gene expression and chemical structure to aid mitochondrial toxicity detection. Commun Biol 2022;5:858.

[68] Fay MM, Kraus O, Victors M, Arumugam L, Vuggumudi K, Urbanik J, Hansen K, Celik S, Cernek N, Jagannathan G, et al. RxRx3: phenomics Map of Biology. Biorxiv 2023. https://doi.org/10.1101/2023.02.07.527350. 2023.02.07.527350.

[69] Cimini BA, Chandrasekaran SN, Kost-Alimova M, Miller L, Goodale A, Fritchman B, Byrne P, Garg S, Jamali N, Logan DJ, et al. Optimizing the Cell Painting assay for image-based profiling. Biorxiv 2022. https://doi.org/10.1101/2022.07.13.499171. 2022.07.13.499171.

[70] Dahlin JL, Hua BK, Zucconi BE, Nelson Jr SD, Singh S, Carpenter AE, Shrimp JH, Lima-Fernandes E, Wawer MJ, Chung LPW, et al. Reference compounds for characterizing cellular injury in high-content cellular morphology assays. Nat Commun 2023;14:1364.

[71] Caicedo JC, Cooper S, Heigwer F, Warchal S, Qiu P, Molnar C, Vasilevich AS, Barry JD, Bansal HS, Kraus O, et al. Data-analysis strategies for image-based cell profiling. Nat Methods 2017;14:849–63.

[72] Vulliard L, Hancock J, Kamnev A, Fell CW, Ferreira da Silva J, Loizou JI, Nagy V, Dupré L, Menche J. BioProfiling.jl: profiling biological perturbations with high-content imaging in single cells and heterogeneous populations. Bioinformatics 2022;38:1692–9.

[73] Way, G., Chandrasekaran, S.N., Bornholdt, M., Fleming, S., Tsang, H., Adeboye, A., Cimini, B., Weisbart, E., Ryder, P., Stirling, D., et al. (2022). Pycytominer: data processing functions for profiling perturbations.

[74] McQuin C, Goodman A, Chernyshev V, Kamentsky L, Cimini BA, Karhohs KW, Doan M, Ding L, Rafelski SM, Thirstrup D, et al. CellProfiler 3.0: next-generation image processing for biology. PLoS Biol 2018;16:e2005970.

[75] Bray M-A, Carpenter AE. Quality control for high-throughput imaging experiments using machine learning in cellprofiler. Methods Mol Biol 2018;1683:89–112.

[76] Qiu M, Zhou B, Lo F, Cook S, Chyba J, Quackenbush D, Matzen J, Li Z, Mak PA, Chen K, et al. A cell-level quality control workflow for high-throughput image analysis. BMC Bioinformatics 2020;21:280.

[77] Smith K, Li Y, Piccinini F, Csucs G, Balazs C, Bevilacqua A, Horvath P. CIDRE: an illumination-correction method for optical microscopy. Nat Methods 2015;12:404–6.

[78] Singh S, Bray M-A, Jones TR, Carpenter AE. Pipeline for illumination correction of images for high-throughput microscopy. J Microsc 2014;256:231–6.

[79] Peng T, Thorn K, Schroeder T, Wang L, Theis FJ, Marr C, Navab N. A BaSiC tool for background and shading correction of optical microscopy images. Nat Commun 2017;8:14836.

[80] Wang Shulei, Arena ET, Eliceiri KW, Yuan Ming. Automated and robust quantification of colocalization in dual-color fluorescence microscopy: a nonparametric statistical approach. IEEE Trans Image Process 2018;27:622–36.

[81] Huang K, Murphy RF. From quantitative microscopy to automated image understanding. J Biomed Opt 2004;9:893–912.

[82] Pau G, Fuchs F, Sklyar O, Boutros M, Huber W. EBImage–an R package for image processing with applications to cellular phenotypes. Bioinformatics 2010;26:979–81.

[83] van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, Gouillart E, Yu T. Scikit-image: image processing in Python. PeerJ 2014;2:e453.

[84] Grys BT, Lo DS, Sahin N, Kraus OZ, Morris Q, Boone C, Andrews BJ. Machine learning and computer vision approaches for phenotypic profiling. J Cell Biol 2017;216:65–71.

[85] Pfaendler R, Hanimann J, Lee S, Snijder B. Self-supervised vision transformers accurately decode cellular state heterogeneity. Biorxiv 2023. https://doi.org/10.1101/2023.01.16.524226.

[86] Pratapa A, Doron M, Caicedo JC. Image-based cell phenotyping with deep learning. Curr Opin Chem Biol 2021;65:9–17.

[87] Caicedo JC, Arevalo J, Piccioni F, Bray M-A, Hartland CL, Wu X, Brooks AN, Berger AH, Boehm JS, Carpenter AE, et al. Cell painting predicts impact of lung cancer variants. Mol Biol Cell 2022;33:ar49.

[88] Moshkov N, Bornholdt M, Benoit S, Smith M, McQuin C, Goodman A, Senft RA, Han Y, Babadi M, Horvath P, et al. Learning representations for image-based profiling of perturbations. Biorxiv 2022. https://doi.org/10.1101/2022.08.12.503783.

[89] Wong DR, Logan DJ, Hariharan S, Stanton R, Kiruluta A. Deep representation learning determines drug mechanism of action from cell painting images. Biorxiv 2022. https://doi.org/10.1101/2022.11.15.516561. 2022.11.15.516561.

[90] Zaritsky A, Jamieson AR, Welf ES, Nevarez A, Cillay J, Eskiocak U, Cantarel BL, Danuser G. Interpretable deep learning uncovers cellular properties in label-free live cell images that are predictive of highly metastatic melanoma. Cell Syst 2021;12:733–47. e6.

[91] Ternes L, Dane M, Gross S, Labrie M, Mills G, Gray J, Heiser L, Chang YH. A multi-encoder variational autoencoder controls multiple transformational features in single-cell image analysis. Commun Biol 2022;5:255.

[92] Sypetkowski, M., Rezanejad, M., Saberian, S., Kraus, O., Urbanik, J., Taylor, J., Mabey, B., Victors, M., Yosinski, J., Sereshkeh, A.R., et al. (2023). RxRx1: a dataset for evaluating experimental batch correction methods.

[93] Dima AA, Elliott JT, Filliben JJ, Halter M, Peskin A, Bernal J, Kociolek M, Brady MC, Tang HC, Plant AL. Comparison of segmentation algorithms for fluorescence microscopy images of cells. Cytometry A 2011;79:545–59.

[94] Javer A, Rittscher J, Sailem HZ. DeepScratch: single-cell based topological metrics of scratch wound assays. Comput Struct Biotechnol J 2020;18:2501–9.

[95] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentation. arXiv [cs.CV].

[96] Schmidt, U., Weigert, M., Broaddus, C., and Myers, G. (2018). Cell detection with star-convex polygons. 10.1007/978-3-030-00934-2_30.

[97] Pachitariu M, Stringer C. Cellpose 2.0: how to train your own model. Nat Methods 2022;19:1634–41.

[98] Van Valen DA, Kudo T, Lane KM, Macklin DN, Quach NT, DeFelice MM, Maayan I, Tanouchi Y, Ashley EA, Covert MW. Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. PLoS Comput Biol 2016;12:e1005177.

[99] de Chaumont F, Dallongeville S, Olivo-Marin J-C. ICY: a new open-source community image processing software. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE; 2011. https://doi.org/10.1109/isbi.2011.5872395.

[100] Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, McQuaid S, Gray RT, Murray LJ, Coleman HG, et al. QuPath: open source software for digital pathology image analysis. Sci Rep 2017;7:16878.

[101] Caicedo JC, Goodman A, Karhohs KW, Cimini BA, Ackerman J, Haghighi M, Heng C, Becker T, Doan M, McQuin C, et al. Nucleus segmentation across imaging experiments: the 2018 data science bowl. Nat Methods 2019;16:1247–53.

[102] Greenwald NF, Miller G, Moen E, Kong A, Kagel A, Dougherty T, Fullaway CC, McIntosh BJ, Leow KX, Schwartz MS, et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. Nat Biotechnol 2022;40:555–65.

[103] Müller, A., Schmidt, D., Rieckert, L., Solimena, M., and Weigert, M. (2023). Organelle-specific segmentation, spatial analysis, and visualization of volume electron microscopy datasets.

[104] Sailem HZ, Al Haj Zen A. Morphological landscape of endothelial cell networks reveals a functional role of glutamate receptors in angiogenesis. Sci Rep 2020;10:13829.

[105] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., et al. (2023). Segment Anything. arXiv [cs.CV]. 10.48550/ARXIV.2304.02643.

[106] Zou, X., Yang, J., Zhang, H., Li, F., Li, L., Gao, J., and Lee, Y.J. (2023). Segment everything everywhere all at once.

[107] Tanaka M, Bateman R, Rauh D, Vaisberg E, Ramachandani S, Zhang C, Hansen KC, Burlingame AL, Trautman JK, Shokat KM, et al. An unbiased cell morphology–based screen for new, biologically active small molecules. PLoS Biol. 2005;3:e128.

[108] Gibson CC, Zhu W, Davis CT, Bowman-Kirigin JA, Chan AC, Ling J, Walker AE, Goitre L, Delle Monache S, Retta SF, et al. Strategy for identifying repurposed drugs for the treatment of cerebral cavernous malformation. CirculationCirculation 2015;131:289–99.

[109] Way GP, Kost-Alimova M, Shibue T, Harrington WF, Gill S, Piccioni F, Becker T, Shafqat-Abbasi H, Hahn WC, Carpenter AE, et al. Predicting cell health phenotypes using image-based morphology profiling. Mol Biol Cell 2021;32:995–1005.

[110] Cuccarese MF, Earnshaw BA, Heiser K, Fogelson B, Davis CT, McLean PF, Gordon HB, Skelly K-R, Weathersby FL, Rodic V, et al. Functional immune mapping with deep-learning enabled phenomics applied to immunomodulatory and COVID-19 drug discovery. Biorxiv 2020. https://doi.org/10.1101/2020.08.02.233064. 2020.08.02.233064.

[111] Shave S, Dawson JC, Athar AM, Nguyen CQ, Kasprowicz R, Carragher NO. Phenonaut: multiomics data integration for phenotypic space exploration. Bioinformatics 2023;39. https://doi.org/10.1093/bioinformatics/btad143.

[112] Omta WA, van Heesbeen RG, Pagliero RJ, van der Velden LM, Lelieveld D, Nellen M, Kramer M, Yeong M, Saeidi AM, Medema RH, et al. HC StratoMineR: a web-based tool for the rapid analysis of high-content datasets. Assay Drug Dev Technol 2016;14:439–52.

[113] Heigwer F, Scheeder C, Bageritz J, Yousefian S, Rauscher B, Laufer C, Beneyto-Calabuig S, Funk MC, Peters V, Boulougouri M, et al. A global genetic interaction network by single-cell imaging and machine learning. Cell Syst 2023. https://doi.org/10.1016/j.cels.2023.03.003.

[114] Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods 2012;9:676–82.

[115] Sofroniew, N., Lambert, T., Evans, K., Nunez-Iglesias, J., Bokota, G., Winston, P., Peña-Castellanos, G., Yamauchi, K., Bussonnier, M., Doncila Pop, D., et al. (2022). napari: a multi-dimensional image viewer for Python (Zenodo) 10.5281/ZENODO.3555620.

[116] Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, Forster J, Lee S, Twardziok SO, Kanitz A, et al. Sustainable data analysis with Snakemake. F1000Res 2021;10:33.

[117] Voss, K., Gentry, J., and Van der Auwera, G. (2017). Full-stack genomics pipelining with GATK4 + WDL + Cromwell. 10.7490/f1000research.1114631.1.

[118] Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. Nat Biotechnol 2017;35:316–9.

[119] Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 2018;19:15.

[120] Akhtar, A. (2020). Role of apache software foundation in big data projects.

[121] Sailem HZ, Sero JE, Bakal C. Visualizing cellular imaging data using PhenoPlot. Nat Commun 2015;6:5825.

[122] Khawatmi M, Steux Y, Zourob S, Sailem HZ. ShapoGraphy: a user-friendly web application for creating bespoke and intuitive visualisation of biomedical data. Front Bioinform 2022;2:788607.

[123] Antal B, Chessel A, Carazo Salas RE. Mineotaur: a tool for high-content microscopy screen sharing and visual analytics. Genome Biol 2015;16:283.
[124] Krueger R, Beyer J, Jang W-D, Kim NW, Sokolov A, Sorger PK, Pfister H. Facetto: combining unsupervised and supervised learning for hierarchical phenotype analysis in multi-channel image data. IEEE Trans Vis Comput Graph 2020;26:227–37.
[125] Lange D, Polanco E, Judson-Torres R, Zangle T, Lex A. Loon: using exemplars to visualize large-scale microscopy data. IEEE Trans Vis Comput Graph 2022;28:248–58.
[126] Driscoll MK, Zaritsky A. Data science in cell imaging. J Cell Sci 2021;134. https://doi.org/10.1242/jcs.254292.
[127] Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome Biol 2017;18:83.
[128] Bock C, Farlik M, Sheffield NC. Multi-omics of single cells: strategies and applications. Trends Biotechnol 2016;34:605–8.
[129] Cantini L, Zakeri P, Hernandez C, Naldi A, Thieffry D, Remy E, Baudot A. Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer. Nat Commun 2021;12:124.
[130] Way GP, Natoli T, Adeboye A, Litichevskiy L, Yang A, Lu X, Caicedo JC, Cimini BA, Karhohs K, Logan DJ, et al. Morphology and gene expression profiling provide complementary information for mapping cell state. Cell Syst 2022;13:911–23. e9.
[131] Nassiri I, McCall MN. Systematic exploration of cell morphological phenotypes associated with a transcriptomic query. Nucleic Acids Res 2018;46:e116.
[132] Joyce AR, Palsson BØ. The model organism as a system: integrating "omics" data sets. Nat Rev Mol Cell Biol 2006;7:198–210.
[133] Smith C. Drug target validation: hitting the target. NatureNature 2003;422. 341, 343, 345 passim.
[134] Canzler S, Schor J, Busch W, Schubert K, Rolle-Kampczyk UE, Seitz H, Kamp H, von Bergen M, Buesen R, Hackermüller J. Prospects and challenges of multi-omics data integration in toxicology. Arch Toxicol 2020;94:371–88.
[135] Nguyen N, Jennen D, Kleinjans J. Omics technologies to understand drug toxicity mechanisms. Drug Discov Today 2022;27:103348.
[136] Rappoport N, Shamir R. Multi-omic and multi-view clustering algorithms: review and cancer benchmark. Nucleic Acids Res 2019;47:1044.
[137] Rodosthenous T, Shahrezaei V, Evangelou M. Integrating multi-OMICS data through sparse canonical correlation analysis for the prediction of complex traits: a comparison study. Bioinformatics 2020;36:4616–25.
[138] Mitra S, Saha S, Hasanuzzaman M. Multi-view clustering for multi-omics data using unified embedding. Sci Rep 2020;10:13654.
[139] Brière G, Darbo É, Thébault P, Uricaru R. Consensus clustering applied to multi-omics disease subtyping. BMC Bioinformatics 2021;22:361.
[140] Lu X, Meng J, Su L, Jiang L, Wang H, Zhu J, Huang M, Cheng W, Xu L, Ruan X, et al. Multi-omics consensus ensemble refines the classification of muscle-invasive bladder cancer with stratified prognosis, tumour microenvironment and distinct sensitivity to frontline therapies. Clin Transl Med 2021;11:e601.
[141] Sailem HZ, Bakal C. Identification of clinically predictive metagenes that encode components of a network coupling cell shape to transcription by image-omics. Genome Res 2017;27:196–207.
[142] Huang S, Chaudhary K, Garmire LX. More is better: recent progress in multi-omics data integration methods. Front Genet 2017;8:84.
[143] Vahabi N, Michailidis G. Unsupervised Multi-omics data integration methods: a comprehensive review. Front Genet 2022;13:854752.
[144] Pham T-H, Qiu Y, Zeng J, Xie L, Zhang P. A deep learning framework for high-throughput mechanism-driven phenotype compound screening and its application to COVID-19 drug repurposing. Nat Mach Intell 2021;3:247–57.
[145] Nguyen, C.Q., Pertusi, D., and Branson, K.M. (2023). Molecule-morphology contrastive pretraining for transferable molecular representation. arXiv [q-bio.QM]..
[146] Rohban MH, Fuller AM, Tan C, Goldstein JT, Syangtan D, Gutnick A, DeVine A, Nijsure MP, Rigby M, Sacher JR, et al. Virtual screening for small-molecule pathway regulators by image-profile matching. Cell Syst 2022;13:724–36. e9.
[147] Mehrizi, R., Mehrjou, A., Alegro, M., Zhao, Y., Carbone, B., Fishwick, C., Vappiani, J., Bi, J., Sanford, S., Keles, H., et al. (2023). Multi-omics prediction from high-content cellular imaging with deep learning. arXiv [q-bio.QM].
[148] Begley CG, Ellis LM. Drug development: raise standards for preclinical cancer research. NatureNature 2012;483:531–3.
[149] Minding microscopy metadata (2021). Nat Methods 18, 1411.
[150] Sarkans U, Chiu W, Collinson L, Darrow MC, Ellenberg J, Grunwald D, Hériché J-K, Iudin A, Martins GG, Meehan T, et al. REMBI: recommended Metadata for Biological Images-enabling reuse of microscopy data in biology. Nat Methods 2021;18:1418–22.
[151] Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3:160018.
[152] Moore J, Basurto-Lozada D, Besson S, Bogovic J, Bragantini J, Brown EM, Burel J-M, Casas Moreno X, de Medeiros G, Diel EE, et al. OME-Zarr: a cloud-optimized bioimaging file format with international community support. Biorxiv 2023. https://doi.org/10.1101/2023.02.17.528834.
[153] Williams E, Moore J, Li SW, Rustici G, Tarkowska A, Chessel A, Leo S, Antal B, Ferguson RK, Sarkans U, et al. The image data resource: a bioimage data integration and publication platform. Nat Methods 2017;14:775–81.
[154] Wilson SL, Way GP, Bittremieux W, Armache J-P, Haendel MA, Hoffman MM. Sharing biological data: why, when, and how. FEBS Lett 2021;595:847–63.
[155] Iudin A, Korir PK, Salavert-Torres J, Kleywegt GJ, Patwardhan A. EMPIAR: a public archive for raw electron microscopy image data. Nat Methods 2016;13:387–8.
[156] Ellenberg J, Swedlow JR, Barlow M, Cook CE, Sarkans U, Patwardhan A, Brazma A, Birney E. A call for public archives for biological image data. Nat Methods 2018;15:849–54.
[157] Orloff DN, Iwasa JH, Martone ME, Ellisman MH, Kane CM. The cell: an image library-CCDB: a curated repository of microscopy data. Nucleic Acids Res 2013;41:D1241–50.
[158] Celik S, Huetter J-C, Melo-Carlos S, Lazar N, Mohan R, Tillinghast C, Biancalani T, Fay M, Earnshaw B, Haque I. Biological cartography: building and benchmarking representations of life. Biorxiv 2022. https://doi.org/10.1101/2022.12.09.519400.
[159] Ljosa V, Sokolnicki KL, Carpenter AE. Annotated high-throughput microscopy image sets for validation. Nat Methods 2012;9:637.
[160] Chang T-Y, Pardo-Martin C, Allalou A, Wählby C, Yanik MF. Fully automated cellular-resolution vertebrate screening platform with parallel animal processing. Lab Chip 2012;12:711–6.
[161] Mulholland T, McAllister M, Patek S, Flint D, Underwood M, Sim A, Edwards J, Zagnoni M. Drug screening of biopsy-derived spheroids using a self-generated microfluidic concentration gradient. Sci Rep 2018;8:14672.
[162] Kramer B, Corallo C, van den Heuvel A, Crawford J, Olivier T, Elstak E, Giordano N, Vulto P, Lanz HL, Janssen RAJ, et al. High-throughput 3D microvessel-on-a-chip model to study defective angiogenesis in systemic sclerosis. Sci Rep 2022;12:16930.
[163] Carragher N, Piccinini F, Tesei A, Trask Jr OJ, Bickle M, Horvath P. Concerns, challenges and promises of high-content analysis of 3D cellular models. Nat Rev Drug Discov 2018;17:606.
[164] Poland SP, Krstajić N, Monypenny J, Coelho S, Tyndall D, Walker RJ, Devauges V, Richardson J, Dutton N, Barber P, et al. A high speed multifocal multiphoton fluorescence lifetime imaging microscope for live-cell FRET imaging. Biomed Opt Express 2015;6:277–96.
[165] Grant DM, McGinty J, McGhee EJ, Bunney TD, Owen DM, Talbot CB, Zhang W, Kumar S, Munro I, Lanigan PM, et al. High speed optically sectioned fluorescence lifetime imaging permits study of live cell signaling events. Opt Express 2007;15:15656–73.
[166] Kumar S, Dunsby C, De Beule PAA, Owen DM, Anand U, Lanigan PMP, Benninger RKP, Davis DM, Neil MAA, Anand P, et al. Multifocal multiphoton excitation and time correlated single photon counting detection for 3-D fluorescence lifetime imaging. Opt Express 2007;15:12548–61.
[167] Ponjavic A, Ye Y, Laue E, Lee SF, Klenerman D. Sensitive light-sheet microscopy in multiwell plates using an AFM cantilever. Biomed Opt Express 2018;9:5863–80.
[168] Stelzer EHK. Light-sheet fluorescence microscopy for quantitative biology. Nat Methods 2015;12:23–6.
[169] Kumar S, Wilding D, Sikkel MB, Lyon AR, MacLeod KT, Dunsby C. High-speed 2D and 3D fluorescence microscopy of cardiac myocytes. Opt Express 2011;19:13839–47.
[170] Dunsby C. Optically sectioned imaging by oblique plane microscopy. Opt Express 2008;16:20306–16.
[171] Sparks H, Dent L, Bakal C, Behrens A, Salbreux G, Dunsby C. Dual-view oblique plane microscopy (dOPM). Biomed Opt Express 2020;11:7204–20.