Harnessing the power of microscopy images to accelerate drug discovery: what are the possibilities?

Justin Boyd , Myles Fennell & Anne Carpenter

Published online: 21 Mar 2020.

Submit your article to this journal

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

EDITORIAL

Check for updates

Harnessing the power of microscopy images to accelerate drug discovery: what are the possibilities?

Justin Boyd [a], Myles Fennell[b] and Anne Carpenter [c]

aInternal Medicines Research Unit, Pfizer Inc., Cambridge, MA, USA; bNeuroscience and Platform Biology, Arvinas, New Haven, CT, USA; cImaging Platform, Broad Institute of MIT and Harvard, Cambridge, MA, USA

1. Introduction

We've all heard it or said ourselves: *a picture is worth a thousand words*. Today, with the improvements in both capability and user-friendliness of imaging and image analysis tools, it is not hyperbolic to say that a picture is worth more than a thousand data points. This is true for all spatial scales of imaging and especially true for microscopy, when it comes to drug discovery. The potential to provide numerous quantitative data measurements in combination with scalability makes cellular imaging particularly valuable for drug hunters. Imaging and image analysis have provided key robust means to quantify complex phenotypes associated with disease and to assess the efficacy of treatment conditions at a cellular level [1–3]. Moreover, image-based profiling provides additional value to drug discovery by enabling researchers to generate comprehensive phenotypic signatures, comparable to other profiling technologies, such as genomics and proteomics, but at much lower cost. Advances in high dimensional data analytics and the adoption of artificial intelligence (AI) techniques promise to yield even further improvements in image-based profiling.

2. High-content screening and image-based profiling

High-content screening combines automated microscopy with automated image analysis and is a common phenotypic drug discovery strategy. Discrete cellular features, defined by the researcher, are quantified through measuring segmented cellular features and used to characterize disease-associated phenotypes [4]. HCS takes a pre-defined approach to identify features that differentiate cellular systems in a predictive way based upon specific feature changes. In other words, HCS focuses on quantifying single cellular processes or functions in the context of a disease. Historically, only a few user-defined features – fewer than 6, and in most cases 1–2 – have been used to differentiate treatment conditions [5]. Robust phenotypes associated with perturbations (e.g. nuclear translocation) can be quantified with few measurements, making screening tractable. By limiting the measurements to discrete features proximal to the biology of interest, researchers can quickly and effectively identify conditions (e.g. compounds or genetic perturbations) which provide the desired effect. Moreover, this approach is amenable to determining the potency

and efficacy of compounds for structure-activity relationship (SAR) determination.

Image-based profiling takes an automated unbiased approach that relies on measuring as many features as possible to generate context-dependent signatures and relies on high dimensional data analytical techniques in order to cluster experimental conditions. A key difference between the two approaches centers on the contribution of individual features: for HCS, specific features define the phenotype; for profiling, the signature based on hundreds to thousands of feature measurements defines the phenotype. In image-based profiling, readouts become data signatures representing the functional consequences of treatments and can be exquisitely sensitive and unbiased [6]. This method can be applied in cases where markers relating to specific biological activity are used (e.g. reporter genes, antibody labeling, or conjugated ligand) and in cases where cell compartment dyes are used, such as those in Cell Painting [7].

3. Target identification

With the advent of efficient cell engineering, access to patient cells, and better methods for growing cells in biologically relevant contexts, our ability to explore disease-associated phenotypes in vitro (and ultimately screen small molecule therapeutics at scale) continues to improve [8]. Genomics and proteomics are commonly used to characterize the functional consequences associated with disease, especially in the context of assessing risk association from GWAS studies. These techniques are powerful tools for characterizing disease targets, but they lack spatial context. Using cellular imaging, genes contributing to disease biology can be identified through functional genomic studies in intact cells and tissue. This modality complements the other omics approaches with the advantage of spatial context. Additionally, imaging provides a signature which can be used at scale for efficient screening. Imaging has been used to identify genes that regulate cell cycle progression and define them into distinct functional clusters by labeling and measuring nuclear features [9]. In an siRNA study, authors used image-based profiling based on CD4-associated phenotypes defined by 15 features to identify 45 host genes as potentially novel targets in HIV [10]. Other examples include studies characterizing the contribution of genes to cellular functions using image-based profiling [11].

4. Small molecule screening

The most common application of automated imaging and image analysis for drug discovery is high-content screening for small molecules. For over two decades, this approach has become a staple strategy for phenotypic drug discovery across diverse applications [2,3,12]. The vast majority of HCS relies on a few discrete features to define hits [5]. Conventional screens focus on single primary feature measurements – how bright, how many, what size, etc. – relating to the proximal biology being targeted for defining hits, which often requires identifying a cell-based biomarker associated with desired disease-related outcomes. Cell count is often used as a proxy for toxicity and for normalizing primary signals. The potential for more content to define the functional consequences of a small molecule is inherent to the technology (hence 'high-content screening'), yet rarely leveraged. In a seminal study, Perlman et al. [13] presented the strategy of profiling compounds using automated imaging and analysis. The authors describe how 11 probes were combined to generate 93 image-based measurements for profiling the activity of compounds. Although the study was limited to 100 compounds, it demonstrated the power of image-based profiling to cluster compounds by mechanism. Developments in the 16 years since that study have increased the number of feature measurements and developed improved computational tools for dealing with high dimensional data to cluster compounds [14]. For example, the latest version of CellProfiler, a common image analysis tool for HCS and image-based profiling, typically measures over 2000 feature measurements per imaged cell, enabling a tremendous amount of content contributing to morphological profiles.

Strategies taking advantage of the enormous information content in images typically use machine learning to define the signatures for screening small molecules. In a study screening known drugs in a model of cerebral cavernous malformation, researchers leveraged image-based profiling and machine learning to identify hits [15]. One of these has progressed to the clinic, highlighting the effectiveness of image-based profiling to identify drugs. Additional value of image-based small molecule screens may be found in repurposing existing image sets: Simm et al. [16] demonstrated that images from a previous screen could be reanalyzed to generate phenotypic fingerprints using machine learning to predict biological activity in assays spanning diverse disease areas. Such approaches could enable researchers to apply legacy data to select small compound sets with a high probability of hits for cases where biologically relevant assays may not be scalable for high-throughput screening. Challenges remain with the use of imaging for compound profiling, particularly when using to guide SAR against molecular targets. If a metric such as IC50 is required to measure potency, it is often difficult to interpret image signatures across a range of concentrations rather than hit identification. Image-based profiling captures numerous mechanisms in any given context. Additionally, individual features are likely to respond to dose/time differently, contributing to the complexity of quantifying pharmacological effects. Perhaps distilling a signature down to specific features around which SAR is performed can be a tractable approach. As compounds are optimized, image-based profiling can then be used to assess hit-to-leads more comprehensively.

5. Compound annotation

Many institutions and companies have accumulated large sets of images and phenotypic measurements from HCS and profiling experiments. It is becoming clear that these data may be used to annotate compound libraries. Cellular imaging is one of the few profiling technologies that can be run at scale and thus, together with other types of compound annotations, the rich phenotypic profiles from imaging can help to deconvolve mechanisms and/or be used to identify putative liabilities (e.g. toxicity). This provides higher resolution for clustering compounds with those that produce similar phenotypes. Such clusters can be used to predict the mechanism of action or reveal potential toxic effects. Since pioneering studies more than a decade ago [17–19], some pharmaceutical companies routinely cluster hit compounds based on morphological profiles to reveal common underlying mechanisms. Because the clusters are based on morphological phenotypes, compound annotation is not limited to structure groups. Both compounds sharing structural similarity and compounds targeting the same pathway can be clustered based upon the biological activity using morphological measurements. Taking advantage of image-based profiling, researchers have been able to identify novel mechanisms, annotate compounds for diverse biological activity, and characterize newly synthesized compounds. For example, researchers used profiling to characterize library sets with predicted phenotypes – known mechanisms – and unpredicted phenotypes – novel mechanisms [20]. In another study, researchers demonstrated that image-based profiling using Cell Painting was able to enrich compounds for more biological diversity compared to the conventional approach of using compound structure diversity to sample biological space [21]. Image-based profiling using Cell Painting was also used to instruct the synthesis strategy of novel compounds [22,23].

6. Conclusion

Within every cellular image, there are combinations of features that can be used to cluster treatment conditions. As automated imaging and image analysis have improved, cellular imaging has become a viable means to identify new targets, screen for compounds, and define the mechanism of action of molecules. Moreover, automated imaging and analysis can be used to annotate libraries based upon morphological profiling, and reanalyzing images can provide a method to repurpose data generated from previous imaging screens. With these advances, cellular imaging and image analysis are already beginning to improve the efficiency of phenotypic drug discovery.

7. Expert opinion

HCS and image-based profiling are viable strategies to drive drug discovery. High-throughput microscopy assays are now robust; image acquisition is not rate limiting; image analysis excels at feature extraction and measurement; and advanced analytics provide paths for meaningful use of data. While HCS is currently a popular modality for phenotypic screening, image-based profiling has been gaining attention as a means of clustering compounds and identifying disease phenotypes and targets. However, the intense data load and computational needs, as well as

hesitation to consider imaging a true quantitative data type, have historically been a challenge for image-based profiling, contributing to slow adoption into mainstream applications of drug discovery.

Image-based profiling so far represents only a small fraction of the overall image-based experimentation in both academic and pharmaceutical settings; conventional high-content screening makes up the majority of high-throughput microscopy experiments. The computational expertise and infrastructure for profiling is a relatively easily solved problem. We believe overcoming the inertia of the current drug discovery enterprise will play an important role as well. For example, drug discovery is typically organized by disease area whereas many applications of image-based profiling span multiple. Typically, confidence in the biological rationale of a target continues to trump data-driven discovery. Although each new computational approach to drug discovery ought to be assessed individually based on its merits and actual actionable information content, it can be difficult to overcome a bias or anxiety stemming from prior predictive computational approaches. In addition, the current processes and infrastructure of drug discovery favor adopting hypotheses with a narrow focus in order to streamline processes and to define stage gates for actionable steps. Most hit-to-lead efforts continue to take iterative steps to optimize compounds based on efficacy (Emax) and potency (EC50) against simple phenotypes. Rather than relying on one experiment with complex profiles for clustering compounds, multiplexing HCS with orthogonal readouts has been the choice for triaging hits from genetic or compound screens. Such a strategy is illustrated in a screen using a *Mycobacterium tuberculosis* transposon mutant library for virulence factors by combining HCS readouts with a cytokine panel [24]. Targeted phenotypes with simplistic readouts continue to define the mainstream high-content screens in drug discovery over profiling [5], even in cases where biologically relevant model systems are limited in scale. For example, Berg et al. [25] report success in using a simple high-content measurement of fluid droplet morphology as the central criterion for screening in a cystic fibrosis model. In both cases above, it would be interesting to test whether image-based profiling could have provided a stronger ability to differentiate hits. Moreover, would more chemical diversity have been achieved if the researchers had used previous high-content image sets to predict activity? The latter approach may provide a low-risk option for companies interested in deploying profiling in a low-commitment way: it can impact multiple programs across disease areas by repurposing legacy data, requiring only computational resources.

An area that could help propel image-based profiling to mainstream drug discovery is in 3D cell-based assays. Advances in 3D biology are areas where additional considerations need to be taken in order to accommodate the opportunity for utilizing more biologically relevant systems. There is an increasing interest in 3D biology applications for drug discovery, and imaging is a critical modality to assess disease and treatment associated phenotypes. Moreover, image-based profiling is likely the most effective means to extract meaningful insight due to the inherent complexity and limited scalability of these systems, compared to conventional monolayer cultures [26].

As AI analytics continue to be developed and applied to image-based experiments, it is likely that profile-based screens will gain more traction. We are witnessing this trend with companies like Recursion Pharma and insitro, two AI-driven startups where image-based phenotypic profiles are a key component of their drug discovery engine. Whether building a drug discovery entity from the ground up or augmenting an existing major R&D enterprise, there is growing interest in adopting AI strategies throughout the entire pipeline of drug discovery to inform omics, chemistry, biology, and clinical data [27]. Images offer a low-risk point of entry into AI, with many successfully proven applications.

Lessons learned from other profiling approaches will influence image-based profiling, in terms of the applications attempted, the computational strategies employed, and perhaps most important of all: the social/cultural acceptance of images as a useful data type in drug discovery. Interest groups and pre-competitive consortia in this area have recently formed and can provide feedback on successes and challenges associated with extracting more from images to drive drug discovery. During the 2019 annual meeting, the Society for Biomolecular Imaging and Informatics hosted a colloquium to discuss image-based profiling and specifically the utility of Cell Painting in drug discovery. In 2020, additional efforts between industry and academic partners will be underway to establish best practices for this field and build the framework for public reference databases. Such efforts will catalyze the expansion of imaging as a tool for drug discovery.

Funding

Declaration of interest

JD Boyd is an employee of Pfizer Inc while M Fennell is an employee of Arvinas. Furthermore, AE Carpenter holds options in Recursion Pharma. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Reviewer disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

ORCID

Justin Boyd 🄳 http://orcid.org/0000-0002-5197-4724
Anne Carpenter 🄳 http://orcid.org/0000-0003-1555-8261

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Mattiazzi Usaj M, Styles EB, Verster AJ, et al. High-content screening for quantitative cell biology. Trends Cell Biol. 2016;26(8):598–611.
2. Grys BT, Lo DS, Sahin N, et al. Machine learning and computer vision approaches for phenotypic profiling. J Cell Biol. 2017;216 (1):65–71.
3. Scheeder C, Heigwer F, Boutros M. Machine learning and image-based profiling in drug discovery. Curr Opin Syst Biol. 2018;10:43–52.

4. Dorval T, Chanrion B, Cattin ME, et al. Filling the drug discovery gap: is high-content screening the missing link? Curr Opin Pharmacol. 2018;42:40–45.

   • **This paper is a current overview of high content screening and its applications in the context of drug discovery. Included in this review and opinion piece are recommended considerations to be taken for choosing high content as an appropriate application for a project.**

5. Singh S, Carpenter AE, Genovesio A. Increasing the content of high-content screening: an overview. J Biomol Screen. 2014;19 (5):640–650.

6. Caicedo JC, Singh S, Carpenter AE. Applications in image-based profiling of perturbations. Curr Opin Biotechnol. 2016;39:134–142.

7. Bray MA, Singh S, Han H, et al. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat Protoc. 2016;11(9):1757–1774.

   • **This reference is the canonical protocol for Cell Painting. It is a detail methods paper for performing Cell Painting on adherent monolayer cells. Troubleshooting notes are included in this manuscipt, making it easy for researchers to apply Cell Painting.**

8. Pegoraro G, Misteli T. High-throughput imaging for the discovery of cellular mechanisms of disease. Trends Genet. 2017;33(9):604–615.

9. Mukherji M, Bell R, Supekova L, et al. Genome-wide functional analysis of human cell-cycle regulators. Proc Natl Acad Sci U S A. 2006;103(40):14819–14824.

10. Genovesio A, Kwon YJ, Windisch MP, et al. Automated genome-wide visual profiling of cellular proteins involved in HIV infection. J Biomol Screen. 2011;16(9):945–958.

11. Rohban MH, Singh S, Wu X, et al. Systematic morphological profiling of human gene and allele function via cell painting. Elife. 2017;6. DOI:10.7554/eLife.24060.

12. Thomas N. High-content screening: a decade of evolution. J Biomol Screen. 2010;15(1):1–9.

13. Perlman ZE, Slack MD, Feng Y, et al. Multidimensional drug profiling by automated microscopy. Science. 2004;306(5699):1194–1198.

14. Caicedo JC, Cooper S, Heigwer F, et al. Data-analysis strategies for image-based cell profiling. Nat Methods. 2017;14(9):849–863.

   • **This methods paper describes the requirements for image based profiling. Important steps from image acquisition, to image processing, to image analysis and data analytics including data visualization are described.**

15. Gibson CC, Zhu W, Davis CT, et al. Strategy for identifying repurposed drugs for the treatment of cerebral cavernous malformation. Circulation. 2015;131(3):289–299.

   •• **This manuscript represents an example of image based profiling combined with machine learning to quickly identify compounds (in this case, FDA approved drugs) that revert a disease associated phenotype back to healthy. Importantly, this paper outlines the process and platform that Recursion Pharma uses for drug discovery, which has led to compounds to be tested in the clinic.**

16. Simm J, Klambauer G, Arany A, et al. Repurposing high-throughput image assays enables biological activity prediction for drug discovery. Cell Chem Biol. 2018;25(5):611–618e3.

   •• **This manuscipt describes the application of machine learning to predict biological activity of compounds in assays from a single image dataset from a high content screen. Notably, the field has discussed reusing imaging data since the beginning of high content screening as a means of expanding data value. In this work, reanalyzing previous data using ML showed that value is not limited to the screen, itself, but also to predicting the outcomes of other screens.**

17. Loo LH, Lin HJ, Steininger RJ 3rd, et al. An approach for extensibly profiling the molecular states of cellular subpopulations. Nat Methods. 2009;6(10):759–765.

18. Young DW, Bender A, Hoyt J, et al. Integrating high-content screening and ligand-target prediction to identify mechanism of action. Nat Chem Biol. 2008;4(1):59–68.

19. MacDonald ML, Lamerdin J, Owens S, et al. Identifying off-target effects and hidden phenotypes of drugs in human cells. Nat Chem Biol. 2006;2(6):329–337.

20. Kummel A, Selzer P, Siebert D, et al. Differentiation and visualization of diverse cellular phenotypic responses in primary high-content screening. J Biomol Screen. 2012;17(6):843–849.

21. Wawer MJ, Li K, Gustafsdottir SM, et al. Toward performance-diverse small-molecule libraries for cell-based phenotypic screening using multiplexed high-dimensional profiling. Proc Natl Acad Sci U S A. 2014;111(30):10911–10916.

   • **This manuscripts describes an application of image-based profiling and gene expression profiling for selecting compound diversity sets based on biological activity. Additionally, compounds profiles predicted MOA by clustering based upon phenotypes.**

22. Gerry CJ, Hua BK, Wawer MJ, et al. Real-time biological annotation of synthetic compounds. J Am Chem Soc. 2016;138(28):8920–8927.

23. Zimmermann S, Akbarzadeh M, Otte F, et al. A Scaffold-Diversity synthesis of biologically intriguing cyclic sulfonamides. Chemistry. 2019;25(68):15498–15503.

24. Barczak AK, Avraham R, Singh S, et al. Systematic, multiparametric analysis of Mycobacterium tuberculosis intracellular infection offers insight into coordinated virulence. PLoS Pathog. 2017;13(5):e1006363.

25. Berg A, Hallowell S, Tibbetts M, et al. High-throughput surface liquid absorption and secretion assays to identify F508del CFTR correctors using patient primary airway epithelial cultures. SLAS Discov. 2019;24(7):724–737.

26. Booij TH, Price LS, Danen EHJ. 3D cell-based assays for drug screens: challenges in imaging, image analysis, and high-content analysis. SLAS Discov. 2019;24(6):615–627.

27. Zhavoronkov A. Artificial intelligence for drug discovery, biomarker development, and generation of novel chemistry. Mol Pharm. 2018;15(10):4311–4313.