Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Applications in image-based profiling of perturbations

Juan C Caicedo[1,2], Shantanu Singh[1] and Anne E Carpenter[1]

A dramatic shift has occurred in how biologists use microscopy images. Whether experiments are small-scale or high-throughput, automatically quantifying biological properties in images is now widespread. We see yet another revolution under way: a transition towards using automated image analysis to not only identify phenotypes a biologist specifically seeks to measure ('screening') but also as an unbiased and sensitive tool to capture a wide variety of subtle features of cell (or organism) state ('profiling'). Mapping similarities among samples using image-based (morphological) profiling has tremendous potential to transform drug discovery, functional genomics, and basic biological research. Applications include target identification, lead hopping, library enrichment, functionally annotating genes/alleles, and identifying small molecule modulators of gene activity and disease-specific phenotypes.

Addresses
[1] Imaging Platform of the Broad Institute of Harvard and Massachusetts Institute of Technology, 415 Main Street, Cambridge, MA, USA
[2] Fundación Universitaria Konrad Lorenz, Bogotá, Colombia

Corresponding author: Carpenter, Anne E (anne@broadinstitute.org)

Current Opinion in Biotechnology 2016, **39**:134–142

This review comes from a themed issue on **Systems biology**

Edited by **Mark P Styczynski** and **Fabian J Theis**

For a complete overview see the Issue and the Editorial

Available online 17th April 2016

http://dx.doi.org/10.1016/j.copbio.2016.04.003

Introduction

Through shifts in both technology and culture, biology is increasingly a quantitative science. Experimental methods that capture the activity or state of multiple distinct biological processes ('multiplexed' assays) are thus increasingly valued. The quantitative increase in the number of independent measures that can be collected in a single assay has brought with it a qualitative change in experimental strategies. In fact, '*profiling*' technologies enable measuring hundreds to thousands of distinct properties from biological samples, an approach quite distinct from '*screening*', which refers to traditional, targeted experiments that seek to quantify a single process or cell function. In this paper, we draw a distinction between these two experimental designs: profiling vs. screening. Profiling aims to capture and encode as many properties of a sample as possible, while screening focuses only on capturing known properties of interest, usually just a few (see Box 1).

Whereas classical biological assays might measure a particular feature of a biological sample in response to perturbation (e.g., ATP consumption, cell size, or phosphorylation state of a single protein), profiling experiments capture a wide range of readouts and use techniques from machine learning and data mining to identify similarities and differences among the measured patterns (sample properties). Thus, typically, the identity of each measured feature is not of particular importance (as in screening experiments), but instead the discovered difference itself is the crucial readout. The particular measured features themselves become relevant only when informative similarities/differences in patterns have been identified. Profiling is a powerful approach enabling high-throughput experimentation and multiplexed readouts to generate massive amounts of mineable data for use in systems biology and drug discovery.

Microscopy, followed by image processing, is one of the few profiling methodologies suited to relatively inexpensive, large-scale experiments involving hundreds of thousands of tested samples. It is compatible with many scales of biological samples: cells, tissues, or organisms (for simplicity in this review we refer to the most common case: cells). In image-based profiling (also known as morphological profiling or cytological profiling [1]), large amounts of quantitative morphological data are extracted from microscopy images of cells to generate a profile comprised of various measures of the shape and size of various cellular compartments and the intensity, texture, and colocalization of various markers (Figure 1). The goal is to identify biologically relevant similarities and differences among samples based on these profiles using appropriate computational models (see Box 2). Profiles of biological populations can be compared to predict previously unrecognized cell states induced by different experimental perturbations of interest.

Alternate highly multiplexed assays for biological systems include the measurement of gene expression, protein levels, and metabolites [2,3]. While powerful, they tend to be low-throughput to medium-throughput (hundreds to thousands of samples per experiment) [4] and characterize the average response of a population of cells (with important exceptions: high-throughput techniques for gene-expression are emerging [5] and RNA-seq can measure mRNA at single-cell resolution albeit currently for only a few samples per experiment). Measuring the response of an arrayed panel of cell lines, for example, the NCI-60 panel, or a panel of RNAi-perturbed lines, to

Box 1 High-throughput image-based screening vs. profiling.

Screening is a distinct strategy from profiling. Although both involve large-scale (high-throughput) imaging experiments, the goals differ: in *screening*, the researcher aims to measure one or more phenotypes that are visually discernible, and choose a subset of hits for further investigation [29]. In *profiling*, a broad spectrum of measurements is captured from each sample (unguided by prior knowledge) in order to reveal important differences and similarities with other samples. Screening depends on a biologist's expertise to interrogate a particular phenomenon whereas profiling takes an unbiased approach to grouping samples, with a higher potential to capture unknown mechanisms.

Image-based *profiling* experiments remain relatively rare [30]. By far, the most common application of high-throughput imaging is *screening* large collections of small molecules in order to identify research probes and therapeutic leads with useful biological properties (often called high-content screening, HCS). High-content screening is becoming more widespread in recent years, in part due to the realization that screens based on cellular phenotypes are on average more fruitful than higher-throughput but less physiological screens on isolated protein targets [31,32]. High-throughput image-based screens involve the development of assays that measure particular morphological properties of single cells. This requires flexible software tools for extracting measurements from images and robust computational models for subsequent data analytics [33], whether a single morphological feature is the basis of the screen, or whether machine learning is used to combine multiple morphological features in order to 'score' the relevant phenotype based on expert input from biologists [34,35]. Given that image-based compound screening is now relatively routine, we refer the reader to prior comprehensive reviews [29,32,36–40].

Although somewhat less common, genetic perturbations are screened in a similar manner as small molecules. The major limiting factor is the initial construction and validation of libraries of perturbation reagents; for completed screens to date, RNA interference is used most often but, depending on the organism, alternatives include direct genetic manipulation (e.g., yeast deletion strains) and overexpression libraries. For loss-of-function screens, CRISPR-Cas9 and related technologies are an exciting prospect; relative to RNAi these are currently thought to have lesser off-target effects, thus improving the reliability of results [41–43].

each perturbation is another form of profiling [6–9] but requires a separate well for *each* measurement in the profile and is thus not generally practical for experiments with thousands of perturbations.

In contrast, high-content imaging techniques can measure hundreds of biologically meaningful features with single-cell resolution in a single assay well, and can be scaled to high-throughput assays with relative ease (Figure 2). There is therefore significant interest in devising appropriate computational techniques specifically for image-derived profiles, which come with technical challenges (Box 2). There is also great potential for combining profiles from multiple methodologies (e.g., imaging + gene expression) in the same experiment to capture a broader range of cell activities.

In this review, we aim to introduce an array of applications that can be achieved using image-based profiling, the collective potential impact of which is immense. Studies in this field are shifting from proof-of-principle to biological discovery; their collective breadth spans research in drug discovery and functional genomics. Microscopy is thus moving from a qualitative assessment tool to a powerful high-capacity quantitative modality.

We focus here on applications that involve systematically profiling large numbers of perturbations interrogated by microscopy imaging; outside this scope are other important applications such as high-throughput image-based screening (Box 1), pathology applications involving human tissue samples [10], studies of population heterogeneity [11–14], engineering extracellular microenvironments [15–17], location proteomics [18–22], and expression and architecture mapping [23–28].

Drug discovery

Identifying mechanisms of action, targets, and toxicity for small molecules

Small molecule perturbations can produce morphological changes detectable by microscopy, and these changes can reveal similarities among compounds in terms of their phenotypic impact in a cellular context. Many studies have demonstrated that morphological profiles can correctly predict the mechanism of action (plus toxicity in some cases) for blinded compounds, by grouping each unknown compound with already-annotated compounds, based on their phenotypic similarity [1,44,49,55–61]; several have made novel predictions [62–65,66••,67,68]. This builds on a foundation of earlier work that identified targets based on visual similarities, for example, the identification of the mitotic kinesin Eg5 as the target of the small molecule monastrol based on a distinctive monopolar spindle phenotype [69] and the phenotypic matching of gene-compound pairs related to cytokinesis using parallel RNA interference (RNAi) and small molecule screens [70] or suppressor/enhancer screens for an RNAi-sensitized phenotype [71]. These studies often focused on oncology/cell cycle, which is not surprising given their dramatic visual phenotypes. A more recent study on hundreds of compounds and several isogenic cell lines revealed novel gene–drug interactions, which were also mapped using image-based phenotypes [72]. As well, methods for identifying individual reporter cell lines that are most useful for grouping compounds with similar mechanisms of action have also been developed, using phenotypic image-based profiles [73]. Grouping compounds by their phenotypic effects is not only feasible for static images of cells but also for videos of complex behaviour in whole organisms; the locomotor response of zebrafish correctly predicted many small molecules' mechanism of action, some previously poorly characterized [74].

Although some studies use the term 'screening' when describing the measurement of phenotypic properties of cells, they may be referring to 'profiling' (e.g. [68,73]).

Figure 1

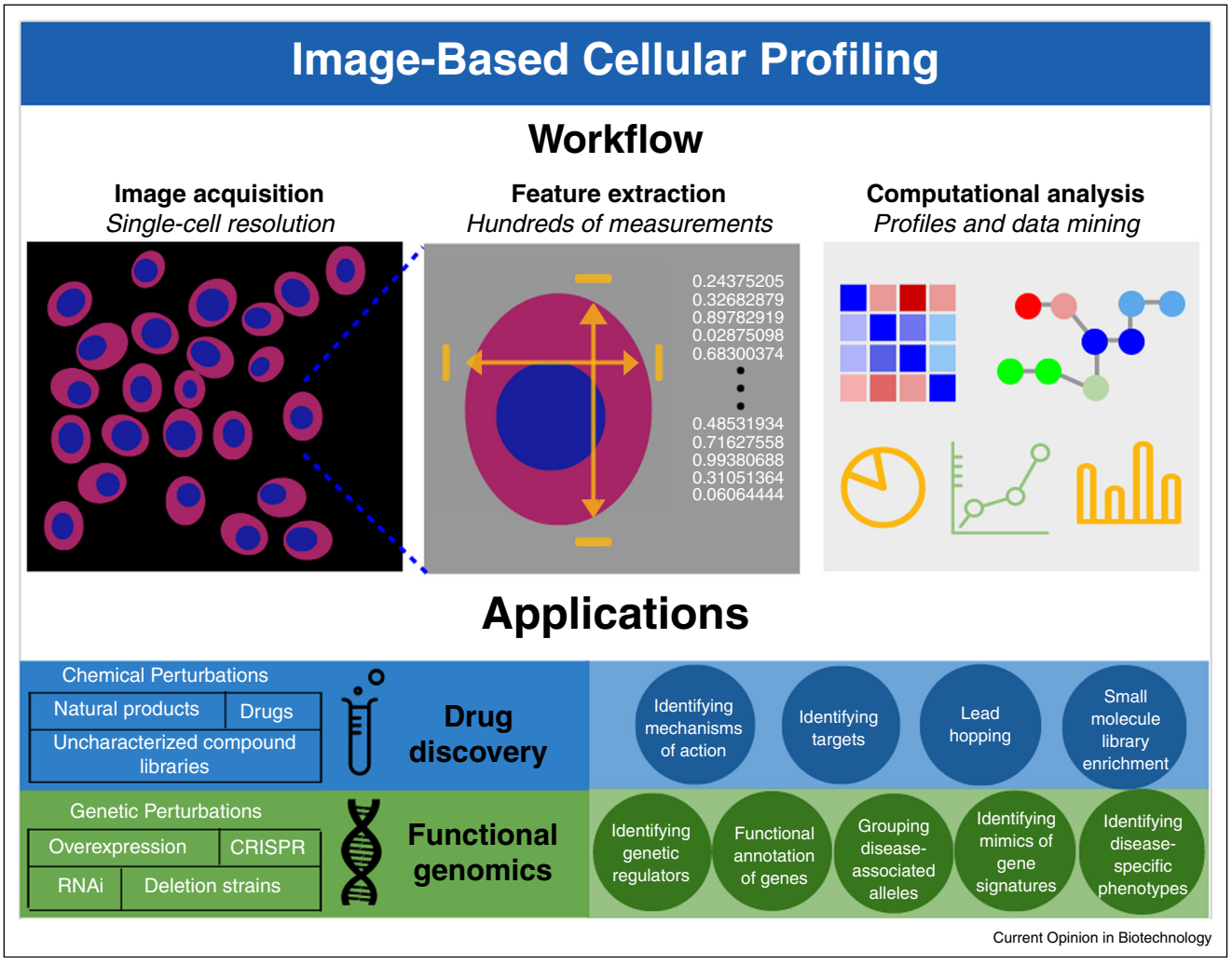


Image-based profiling workflow and applications. Perturbations are applied to a population of cells or organisms, and the resulting phenotypes are observed and captured with imaging. Hundreds of morphological measurements are extracted from images, which are then analyzed using computational models, appropriate to the target application. Those models are used to draw conclusions in a diverse range of biomedical applications.

The key distinction in these cases comes from the use of multiple features for profiling (multivariate representations of samples), and also the unbiased analysis of the experiments, while classical screening usually considers a few phenotypes and targeted analysis of a known phenomenon.

Lead hopping

Another application of similarity-matching among small molecule treatments is known as lead hopping. Although the underlying methodology is identical, the goal differs: here, a small molecule with useful phenotypic effects but undesirable structure is used to help identify other small molecules with a matching phenotypic effect but with molecular backbones better suited to medicinal chemistry. We have not yet seen a study using image-based profiling specifically for the purpose of identifying novel structures for further therapeutic development in a particular disease area, but the data type seems well-suited to this task.

Small molecule library enrichment

A typical small molecule library contains significant redundancy, that is, sets of small molecules with the same effect on a molecular target or pathway. There are likely also a large number of compounds that have no impact whatsoever on cells. The demonstrated ability of morphological profiles to group compounds with similar behaviour led to the hypothesis that smaller, more efficient small molecule libraries might be selected based on phenotypic diversity, as determined by a single profiling assay. The first study to assess this indicates that, indeed,

Box 2 Computational challenges in image-based profiling.

Morphological profiling presents many computational challenges owing to both the size and complexity of the data. One 384-well plate can generate more than 500 million single-cell measurements, and an experiment might involve hundreds of plates. How to map these raw, single-cell measurements to optimally useful perturbation profiles is an open research question. In one comparison between profiling techniques [44], methods that attempt to leverage population heterogeneity were outperformed by relatively simple population aggregation methods, highlighting the need for further research on methods for capturing heterogeneity in profiles.

Additional complications relate to the properties of the feature space: the measurements themselves are typically redundant, with many features being correlated. However, given the non-linear relationships between these features, simple correlation measures will not always be sufficient to identify these associations, thereby making feature selection non-trivial. Identifying appropriate similarity measures and dimensionality reduction methods for morphological profiles is also an open problem [45,46]. While not unique to morphological profiling per se, plate position effects (e.g., wells on the edges of a plate having different characteristics than the middle due to difference in temperature and humidity) and batch effects introduce additional confounds that make comparing profiles across plates or across experiments challenging [47]. At present, there are no standard methods for addressing these problems nor are there software packages that offer the variety of approaches that have been proposed for each step in the workflow.

Addressing many of these problems is a high priority for labs active in the field. Some of the most exciting computational developments for profiling preserve single cell data and thus take into account the increasingly well-appreciated heterogeneity of cultured cell populations [11,12,48]. For example, the Pelkmans laboratory showed that accounting for population context of a cell (e.g. whether it lies on the edge of a cell island, the surface area of contact with neighbouring cells, etc.) improved the consistency between replicate RNA interference (RNAi) screens and between siRNAs targeting the same gene [13]. The Altschuler/Wu laboratory identified cell subpopulations from images, measured the relative abundance of proteins in each of these subpopulations, and showed that grouping proteins based on this measure agreed with known functional associations [49]. The Boutros laboratory created morphological profiles of RNAi-induced knockdown of genes by first identifying cell subpopulations, which in turn were used to discover functional associations between genes [50,51]. Although methods making sophisticated use of single-cell data have mainly been used only in proof-of-principle studies so far, we anticipate further development and application to real-world problems. Image-based profiling data has always been available at single-cell resolution, making research in this area relevant to other modalities that have only recently become feasible to carry out at single-cell resolution (e.g., transcription and genomics).

Finally, the use of deep learning techniques may bring interesting benefits and solutions to some of the mentioned problems. These methods are already common practice for solving complex computer vision tasks [52], and are also starting to be applied in the bioimage informatics community [53,54]. Whether applied at the stage of segmentation, feature extraction, or classification, there is increasing interest in adopting these methodologies to push biological discoveries, and we expect to see more methods incorporating these algorithms in the near future.

morphological profiling can select enriched libraries with higher rates of activity and diverse biological performance [77••].

Functional genomics and disease phenotyping

Characterizing genetic regulators of particular biological processes

Image-based profiling has also been used in functional genomics to characterize and annotate genes; the genome is by no means fully annotated, and systematic solutions are needed. The strategy here is to up-regulate or down-regulate each gene's expression and compare the phenotypic impact, as measured in the morphological profile, to that of already-annotated genes.

The simplest such case begins with a high-throughput image-based screen (Box 1) that identifies a group of 'hit' genes that influence a particular phenotype of interest. Profiling can then be used to group those hits, based on morphological similarity using a broad spectrum of image-based phenotypes. This goes a step beyond simply cataloguing genetic regulators of a particular process: categorization based on phenotype, using the rich morphological data already available in the primary screen, can lend credibility to the involvement of entire pathways in processes and enables more efficient triaging and follow-up on individual genes.

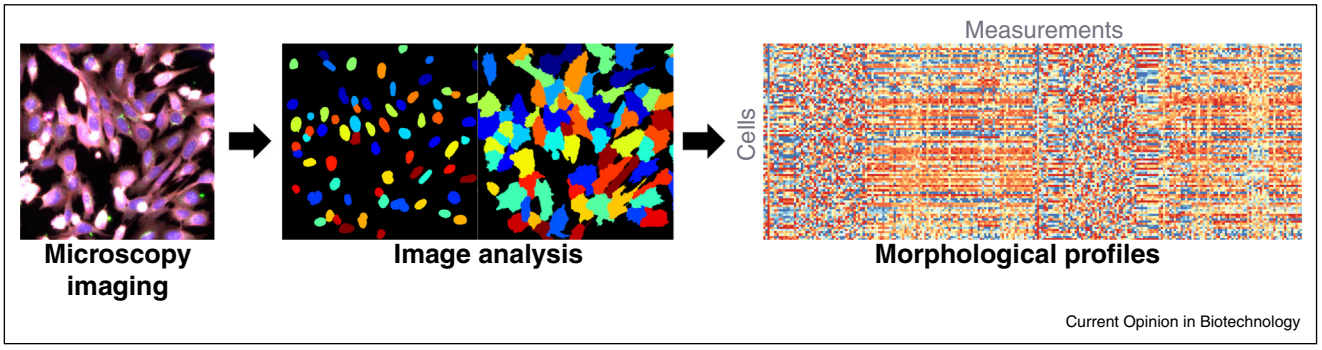
Many important cell processes have been probed using this strategy of in-depth morphological profiling of hits from a genetic perturbation screen. For example, the MitoCheck project used time-lapse microscopy to track and profile individual cells, successfully documenting and classifying hundreds of genetic regulators associated with mitosis [78]. Profiling has also been used to classify hits from a screen involving membrane-trafficking activities of endocytosis [79,80••].

Functional annotation of genes by similarity

Some recent studies take a more systems biology-oriented approach well beyond the above-described 'screen, then cluster'. Here, a large number of genes are grouped based on similarity of morphological profiles resulting from over-expression or under-expression of each gene. This approach generates hypotheses for any previously unannotated genes that closely cluster with functionally annotated genes.

For example, early work using nonessential yeast deletion mutants was able to group gene deletions into functional pathways; nearly half of the mutants yielded a discernable phenotype by imaging using a single set of three stains (cell-surface mannoprotein, actin cytoskeleton, and nuclear DNA) [81]. In the presence of a perturbation (high concentration of extracellular calcium), deletion strains clustered into functionally related groups based on changes in morphological features [82]. Synthetic genetic arrays enable assaying the morphological impact of single and double mutant yeast strains, often useful in revealing genes involved in a process [83].

Figure 2



Morphological profiling data for a single sample. A population of U-2 OS cells was treated with rapamycin, then stained and imaged according to the Cell Painting assay protocol [60,75] 144 hours later. Images were processed using the open-source software CellProfiler [76], yielding 1474 morphological measurements for all cells in the field of view. These data are displayed as a matrix: only a subset of the features are shown, for 89 cells.

In mammalian cells, the DNA damage/cell cycle functions of the DONSON gene were identified in an unbiased profiling study that grouped genes with similar loss-of-function phenotypes using RNAi [51]. Mapping genetic interactions with combinatorial RNAi is also an effective way to uncover functional relationships between genes, which can be measured with single-cell phenotypic readouts [84–86]. The largest map of directional epistatic interactions has been recently built with techniques based on large scale image analysis combined with a statistical model that reveals novel complex dependencies between genotypes and phenotypes [87••].

Most of the experiments following this strategy have used RNA interference; it is the most common method of genetic perturbation for mammalian cells. We do offer a cautionary note: we recently found that morphological profiles of RNAi-induced gene knockdown are highly sensitive and reproducible but are dominated by so-called 'seed effects', a type of off-target effect [88•]. There are workarounds to enrich on-target versus off-target effects in the context of RNAi screening where a narrow set of phenotypes are measured, but computational approaches need to be developed to enable reliable grouping of RNAi-induced multi-dimensional profiles.

Grouping disease-associated genes and alleles by functional impact

The same strategy of grouping genetic perturbations can be applied specifically to experiments involving overexpression of genes and alleles that have been linked to human disease, e.g., through genome-wide association studies. This can inform the mechanistic understanding of the disease by placing unannotated genes into pathways.

This can also, in theory, go a step further toward personalized medicine by grouping disease-causing variants based on phenotypic impact (albeit in a cell-based system). This could ultimately guide clinical treatment in cases where particular drug treatments are known to be effective only for particular alleles; a previously unobserved allele whose morphological profile is highly similar to a known allele lends a hypothesis for an effective treatment for that patient. We are not aware of any published work following this approach but expect to see examples of this strategy emerge soon.

Identifying small molecule mimics of gene signatures (and vice versa, for target identification)

In cases where a distinctive signature has been identified by morphological profiling of genetic perturbations, gene–drug connections can be made by comparison of the gene's profile to databases of small molecules that have also been morphologically profiled, in a strategy akin to the Connectivity Map, which is based on gene expression data [89]. The strategy's principle has been proved using imaging data in yeast, where the targets of four drugs with known mechanisms were re-identified by comparison of the signatures induced by those small molecules to signatures induced by yeast deletion strains [90], and where drugs impinging on a pathway show similarity to signatures of deletion strains in genes related to the same pathway [91,92,93••]. A similar approach using RNAi in mammalian cells uncovered the mechanism of action of compounds that inhibit bacterial growth via the host-pathogen interface [94••,95]. Although it is not trivial to compare multi-dimensional profiles across separate experiments using different perturbation modalities (e.g., overexpression of genes vs. treatment with small molecules), it is tremendously powerful to systematically identify small molecules

that mimic or reverse phenotypes linked to particular genetic perturbations.

Identifying disease-specific phenotypes and small molecules to revert them (signature-based screening)

The drug discovery community has increasingly recognized the effectiveness of a phenotypic approach, where a model system is sought that reflects human disease biology as faithfully as possible [31,96–99]. The physiological relevance of the assay system must be balanced against compatibility for screening large chemical libraries, or at least testing small numbers of candidate small molecules.

But what if no disease-associated phenotype is already known? A transformative approach to creating phenotypic model systems is to use either cell samples taken directly from patients with disease, or cell lines manipulated to create a genetic perturbation correlated with the disease (some aspects of this are reviewed in [100]). The important first step is identifying the phenotype of interest in an unbiased way, that is, seeking a signature of the disease state. Once the signature is known, small molecules can be screened to identify those that revert that particular signature back to a more wild-type-like state, hence the term 'signature-based screening'.

This approach was taken to first identify a cell-culture based morphological phenotype associated with loss-of-function of *CCM2*, the gene missing in patients with the hereditary stroke syndrome cerebral cavernous malformation (CCM) [101••]. The team then screened small molecules and identified those that reverted the computationally-defined image-based phenotype. Two of those drugs proved effective in animal models of the disease; of note, drugs chosen based on reverting the computationally-defined phenotype performed better in subsequent assays than drugs that reverted a human-defined phenotype for the disease, which were tested in parallel. Several other laboratories are taking this promising approach, some even beginning in animal systems. For example, 'personalized' *Drosophila* strains carrying mutations mimicking those found in a patient's tumour are being used to test therapeutic cocktails, often with a visual readout [102–104].

Conclusions

Profiling has the potential to transform many fields in biology. We expect computational advances (Box 2) to be an important force propelling image-based profiling forward. With appropriate advancements, we expect image-based profiling and analysis to be powerful tools that complement well-established -omics methods to address challenging questions in systems biology and drug discovery.

Conflict of interest

AEC has optional ownership interest in Recursion Pharmaceuticals, a biotechnology company founded in part to commercialize the findings of a study described in one of the cited works [101••].

Acknowledgement

Funding for the preparation of this article was provided by the National Science Foundation (NSF CAREER DBI 1148823 to AEC).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Perlman ZE, Slack MD, Feng Y, Mitchison TJ, Wu LF, Altschuler SJ: **Multidimensional drug profiling by automated microscopy**. *Science* 2004, **306**:1194-1198.
2. Abraham Y, Zhang X, Parker CN: **Multiparametric analysis of screening data: growing beyond the single dimension to infinity and beyond**. *J Biomol Screen* 2014, **19**:628-639.
3. Johannessen CM, Clemons PA, Wagner BK: **Integrating phenotypic small-molecule profiling and human genetics: the next phase in drug discovery**. *Trends Genet* 2015, **31**:16-23.
4. Feng Y, Mitchison TJ, Bender A, Young DW, Tallarico JA: **Multi-parameter phenotypic profiling: using cellular effects to characterize small-molecule compounds**. *Nat Rev Drug Discov* 2009, **8**:567-578.
5. Peck D, Crawford ED, Ross KN, Stegmaier K, Golub TR, Lamb J: **A method for high-throughput gene expression signature analysis**. *Genome Biol* 2006, **7**:R61.
6. Weinstein JN, Myers TG, O'Connor PM, Friend SH, Fornace AJ Jr, Kohn KW *et al.*: **An information-intensive approach to the molecular pharmacology of cancer**. *Science* 1997, **275**:343-349.
7. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S *et al.*: **The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity**. *Nature* 2012, **483**:603-607.
8. Roguev A, Talbot D, Negri GL, Shales M, Cagney G, Bandyopadhyay S *et al.*: **Quantitative genetic-interaction mapping in mammalian cells**. *Nat Methods* 2013, **10**:432-437.
9. Cowley GS, Weir BA, Vazquez F, Tamayo P, Scott JA, Rusin S *et al.*: **Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies**. *Sci Data* 2014, **1**:140035.
10. Ghaznavi F, Evans A, Madabhushi A, Feldman M: **Digital imaging in pathology: whole-slide imaging and beyond**. *Annu Rev Pathol* 2013, **8**:331-359.
11. Snijder B, Pelkmans L: **Origins of regulated cell-to-cell variability**. *Nat Rev Mol Cell Biol* 2011, **12**:119-125.
12. Altschuler SJ, Wu LF: **Cellular heterogeneity: do differences make a difference?** *Cell* 2010, **141**:559-563.
13. Snijder B, Sacher R, Rämö P, Liberali P, Mench K, Wolfrum N *et al.*: **Single-cell analysis of population context advances RNAi screening at multiple levels**. *Mol Syst Biol* 2012, **8**:579.
14. Yin Z, Sadok A, Sailem H, McCarthy A, Xia X, Li F *et al.*: **A screen for morphological complexity identifies regulators of switch-like transitions between discrete cell shapes**. *Nat Cell Biol* 2013, **15**:860-871.
15. Unadkat HV, Hulsman M, Cornelissen K, Papenburg BJ, Truckenmüller RK, Carpenter AE *et al.*: **An algorithm-based topographical biomaterials library to instruct cell fate**. *Proc Natl Acad Sci U S A* 2011, **108**:16565-16570.
16. Dalby MJ, Gadegaard N, Tare R, Andar A, Riehle MO, Herzyk P *et al.*: **The control of human mesenchymal cell differentiation using nanoscale symmetry and disorder**. *Nat Mater* 2007, **6**:997-1003.

17. LaBarge MA, Nelson CM, Villadsen R, Fridriksdottir A, Ruth JR, Stampfer MR *et al.*: **Human mammary progenitor cell fate decisions are products of interactions with combinatorial microenvironments**. *Integr Biol* 2009, **1**:70-79.

18. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A *et al.*: **Proteomics. Tissue-based map of the human proteome**. *Science* 2015, **347**:1260419.

19. Buck TE, Li J, Rohde GK, Murphy RF: **Toward the virtual cell: automated approaches to building models of subcellular organization learned from microscopy images**. *BioEssays* 2012, **34**:791-799.

20. Handfield L-F, Chong YT, Simmons J, Andrews BJ, Moses AM: **Unsupervised clustering of subcellular protein expression patterns in high-throughput microscopy images reveals protein complexes and functional relationships between proteins**. *PLoS Comput Biol* 2013, **9**:e1003085.

21. Chong YT, Koh JLY, Friesen H, Duffy SK, Duffy K, Cox MJ *et al.*: **Yeast proteome dynamics from single cell imaging and automated analysis**. *Cell* 2015, **161**:1413-1424.

22. Koh JLY, Chong YT, Friesen H, Moses A, Boone C, Andrews BJ *et al.*: **CYCLoPs: a comprehensive database constructed from automated analysis of protein abundance and subcellular localization patterns in *Saccharomyces cerevisiae***. *G3* 2015, **5**:1223-1232.

23. Kasthuri N, Hayworth KJ, Berger DR, Schalek RL, Conchello JA, Knowles-Barley S *et al.*: **Saturated reconstruction of a volume of neocortex**. *Cell* 2015, **162**:648-661.

24. Oh SW, Harris JA, Ng L, Winslow B, Cain N, Mihalas S *et al.*: **A mesoscale connectome of the mouse brain**. *Nature* 2014, **508**:207-214.

25. Ahrens MB, Orger MB, Robson DN, Li JM, Keller PJ: **Whole-brain functional imaging at cellular resolution using light-sheet microscopy**. *Nat Methods* 2013, **10**:413-420.

26. Swierczek NA, Giles AC, Rankin CH, Kerr RA: **High-throughput behavioral analysis in *C. elegans***. *Nat Methods* 2011, **8**:592-598.

27. Chiang A-S, Lin C-Y, Chuang C-C, Chang H-M, Hsieh C-H, Yeh C-W *et al.*: **Three-dimensional reconstruction of brain-wide wiring networks in *Drosophila* at single-cell resolution**. *Curr Biol* 2011, **21**:1-11.

28. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS *et al.*: **The genetic landscape of a cell**. *Science* 2010, **327**:425-431.

29. Boutros M, Heigwer F, Laufer C: **Microscopy-based high-content screening**. *Cell* 2015, **163**:1314-1325.

30. Singh S, Carpenter AE, Genovesio A: **Increasing the content of high-content screening: an overview**. *J Biomol Screen* 2014, **19**:640-650.

31. Swinney DC, Anthony J: **How were new medicines discovered?** *Nat Rev Drug Discov* 2011, **10**:507-519.

32. Eggert US: **The why and how of phenotypic small-molecule screens**. *Nat Chem Biol* 2013, **9**:206-209.

33. Eliceiri KW, Berthold MR, Goldberg IG, Ibáñez L, Manjunath BS, Martone ME *et al.*: **Biological imaging software tools**. *Nat Methods* 2012, **9**:697-710.

34. Sommer C, Gerlich DW: **Machine learning in cell biology – teaching computers to recognize phenotypes**. *J Cell Sci* 2013, **126**:1-11.

35. Jones TR, Carpenter AE, Lamprecht MR, Moffat J, Silver SJ, Grenier JK *et al.*: **Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning**. *Proc Natl Acad Sci U S A* 2009, **106**:1826-1831.

36. Megason SG, Fraser SE: **Imaging in systems biology**. *Cell* 2007, **130**:784-795.

37. Carpenter AE: **Image-based chemical screening**. *Nat Chem Biol* 2007, **3**:461-465.

38. Eggert US, Mitchison TJ: **Small molecule screening by imaging**. *Curr Opin Chem Biol* 2006, **10**:232-237.

39. Bickle M: **The beautiful cell: high-content screening in drug discovery**. *Anal Bioanal Chem* 2010, **398**:219-226.

40. Antony PMA, Trefois C, Stojanovic A, Baumuratov AS, Kozak K: **Light microscopy applications in systems biology: opportunities and challenges**. *Cell Commun Signal* 2013, **11**:24.

41. Taylor J, Woodcock S: **A perspective on the future of high-throughput RNAi screening: will CRISPR cut out the competition or can RNAi help guide the way?** *J Biomol Screen* 2015, **20**:1040-1051.

42. Shalem O, Sanjana NE, Zhang F: **High-throughput functional genomics using CRISPR-Cas9**. *Nat Rev Genet* 2015, **16**:299-311.

43. Mohr SE, Smith JA, Shamu CE, Neumüller RA, Perrimon N: **RNAi screening comes of age: improved techniques and complementary approaches**. *Nat Rev Mol Cell Biol* 2014, **15**:591-600.

44. Ljosa V, Caie PD, Ter Horst R, Sokolnicki KL, Jenkins EL, Daya S *et al.*: **Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment**. *J Biomol Screen* 2013, **18**:1321-1329.

45. Reisen F, Zhang X, Gabriel D, Selzer P: **Benchmarking of multivariate similarity measures for high-content screening fingerprints in phenotypic drug discovery**. *J Biomol Screen* 2013, **18**:1284-1297.

46. Kümmel A, Selzer P, Beibel M, Gubler H, Parker CN, Gabriel D: **Comparison of multivariate data analysis strategies for high-content screening**. *J Biomol Screen* 2011, **16**:338-347.

47. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE *et al.*: **Tackling the widespread and critical impact of batch effects in high-throughput data**. *Nat Rev Genet* 2010, **11**:733-739.

48. Ng AYJ, Rajapakse JC, Welsch RE, Matsudaira PT, Horodincu V, Evans JG: **A cell profiling framework for modeling drug responses from HCS imaging**. *J Biomol Screen* 2010, **15**:858-868.

49. Loo L-H, Lin H-J, Steininger RJ 3rd, Wang Y, Wu LF, Altschuler SJ: **An approach for extensibly profiling the molecular states of cellular subpopulations**. *Nat Methods* 2009, **6**:759-765.

50. Volz HC, Heigwer F, Wuest T, Galach M, Utikal J, Katus HA *et al.*: **Single-cell phenotyping of human induced pluripotent stem cells by high-throughput imaging**. *bioRxiv* 2015, **5**:026955 http://dx.doi.org/10.1101/026955.

51. Fuchs F, Pau G, Kranz D, Sklyar O, Budjan C, Steinbrink S *et al.*: **Clustering phenotype populations by genome-wide RNAi and multiparametric imaging**. *Mol Syst Biol* 2010, **6**:370.

52. LeCun Y, Bengio Y, Hinton G: **Deep learning**. *Nature* 2015, **521**:436-444.

53. Kandaswamy C, Silva LM, Alexandre LA, Santos JM: **High-content analysis of breast cancer using single-cell deep transfer learning**. *J Biomol Screen* 2016, **21**:252-259.

54. Kraus OZ, Frey BJ: **Computer vision for high content screening**. *Crit Rev Biochem Mol Biol* 2016, **51**:102-109.

55. MacDonald ML, Lamerdin J, Owens S, Keon BH, Bilter GK, Shang Z *et al.*: **Identifying off-target effects and hidden phenotypes of drugs in human cells**. *Nat Chem Biol* 2006, **2**:329-337.

56. Loo L-H, Wu LF, Altschuler SJ: **Image-based multivariate profiling of drug responses from single cells**. *Nat Methods* 2007, **4**:445-453.

57. Young DW, Bender A, Hoyt J, McWhinnie E, Chirn G-W, Tao CY *et al.*: **Integrating high-content screening and ligand-target prediction to identify mechanism of action**. *Nat Chem Biol* 2008, **4**:59-68.

58. Slack MD, Martinez ED, Wu LF, Altschuler SJ: **Characterizing heterogeneous cellular responses to perturbations**. *Proc Natl Acad Sci U S A* 2008, **105**:19306-19311.

59. Low J, Chakravartty A, Blosser W, Dowless M, Chalfant C, Bragger P *et al.*: **Phenotypic fingerprinting of small molecule cell cycle kinase inhibitors for drug discovery**. *Curr Chem Genomics* 2009, **3**:13-21.

60. Gustafsdottir SM, Ljosa V, Sokolnicki KL, Anthony Wilson J, Walpita D, Kemp MM *et al.*: **Multiplex cytological profiling assay to measure diverse cellular states**. *PLOS ONE* 2013, **8**:e80999.

61. Pardo-Martin C, Allalou A, Medina J, Eimon PM, Wählby C, Fatih Yanik M: **High-throughput hyperdimensional vertebrate phenotyping**. *Nat Commun* 2013, **4**:1467.

62. Tanaka M, Bateman R, Rauh D, Vaisberg E, Ramachandani S, Zhang C *et al.*: **An unbiased cell morphology-based screen for new, biologically active small molecules**. *PLoS Biol* 2005, **3**:e128.

63. Futamura Y, Kawatani M, Kazami S, Tanaka K, Muroi M, Shimizu T *et al.*: **Morphobase, an encyclopedic cell morphology database, and its use for drug target identification**. *Chem Biol* 2012, **19**:1620-1630.

64. Woehrmann MH, Bray WM, Durbin JK, Nisam SC, Michael AK, Glassey E *et al.*: **Large-scale cytological profiling for functional analysis of bioactive compounds**. *Mol Biosyst* 2013, **9**:2604-2617.

65. Schulze CJ, Bray WM, Woerhmann MH, Stuart J, Lokey RS, Linington RG: **Function-first lead discovery: mode of action profiling of natural product libraries using image-based screening**. *Chem Biol* 2013, **20**:285-295.

66. •• Ochoa JL, Bray WM, Lokey RS, Linington RG: **Phenotype-guided natural products discovery using cytological profiling**. *J Nat Prod* 2015, **78**:2242-2248.
This study is the latest in a series from the Lokey/Linington laboratories that clusters members of a 5304 member natural product library based on the similarity of their image-based profiles in mammalian cell culture, using two parallel image-based assays with three stains each. Here, they identify novel constituents that cluster with known antimitotic drugs.

67. Kurita KL, Glassey E, Linington RG: **Integration of high-content screening and untargeted metabolomics for comprehensive functional annotation of natural product libraries**. *Proc Natl Acad Sci U S A* 2015 http://dx.doi.org/10.1073/pnas.1507743112.

68. Reisen F, Sauty de Chalon A, Pfeifer M, Zhang X, Gabriel D, Selzer P: **Linking phenotypes and modes of action through high-content screen fingerprints**. *Assay Drug Dev Technol* 2015, **13**:415-427.

69. Mayer TU, Kapoor TM, Haggarty SJ, King RW, Schreiber SL, Mitchison TJ: **Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen**. *Science* 1999, **286**:971-974.

70. Eggert US, Kiger AA, Richter C, Perlman ZE, Perrimon N, Mitchison TJ *et al.*: **Parallel chemical genetic and genome-wide RNAi screens identify cytokinesis inhibitors and targets**. *PLoS Biol* 2004, **2**:e379.

71. Castoreno AB, Smurnyy Y, Torres AD, Vokes MS, Jones TR, Carpenter AE *et al.*: **Small molecules discovered in a pathway screen target the Rho pathway in cytokinesis**. *Nat Chem Biol* 2010, **6**:457-463.

72. Breinig M, Klein FA, Huber W, Boutros M: **A chemical-genetic interaction map of small molecules using high-throughput imaging in cancer cells**. *Mol Syst Biol* 2015, **11**:846.

73. Kang J, Hsu C-H, Wu Q, Liu S, Coster AD, Posner BA *et al.*: **Improving drug discovery with high-content phenotypic screens by systematic selection of reporter cell lines**. *Nat Biotechnol* 2016, **34**:70-77.

74. Rihel J, Prober DA, Arvanites A, Lam K, Zimmerman S, Jang S *et al.*: **Zebrafish behavioral profiling links drugs to biological targets and rest/wake regulation**. *Science* 2010, **327**:348-351.

75. Bray M-A, Singh S, Han H, Davis CT, Borgeson B, Hartland C *et al.*: **Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes**. *Nature Protocols* 2016. in press.

76. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O *et al.*: **CellProfiler: image analysis software for identifying and quantifying cell phenotypes**. *Genome Biol* 2006, **7**:R100.

77. •• Wawer MJ, Li K, Gustafsdottir SM, Ljosa V, Bodycombe NE, Marton MA *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**:10911-10916.
This was the first large-scale imaging experiment to carry out image-based profiling using the Cell Painting assay [75]. The Wawer study, which involved our laboratory, found that morphological profiling provided sufficient information to enrich a library of small molecules for those that are bioactive in cell culture and that yield diverse phenotypic effects. The study also produced high-throughput gene expression profile data for ~20,000 of the small molecules, which enabled some comparisons between the two profiling modalities.

78. Neumann B, Walter T, Hériché J-K, Bulkescher J, Erfle H, Conrad C *et al.*: **Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes**. *Nature* 2010, **464**:721-727.

79. Collinet C, Stöter M, Bradshaw CR, Samusik N, Rink JC, Kenski D *et al.*: **Systems survey of endocytosis by multiparametric image analysis**. *Nature* 2010, **464**:243-249.

80. •• Liberali P, Snijder B, Pelkmans L: **A hierarchical map of regulatory genetic interactions in membrane trafficking**. *Cell* 2014, **157**:1473-1487.
The Pelkmans laboratory generated RNAi-induced knockdown profiles across two cell lines and 13 different assays for endocytic pathways and organelles. Based on this data, they inferred the regulatory interactions among different pathways related to endocytosis.

81. Ohya Y, Sese J, Yukawa M, Sano F, Nakatani Y, Saito TL *et al.*: **High-dimensional and large-scale phenotyping of yeast mutants**. *Proc Natl Acad Sci U S A* 2005, **102**:19015-19020.

82. Ohnuki S, Nogami S, Kanai H, Hirata D, Nakatani Y, Morishita S *et al.*: **Diversity of Ca^{2+}-induced morphology revealed by morphological phenotyping of Ca^{2+}-sensitive mutants of *Saccharomyces cerevisiae***. *Eukaryot Cell* 2007, **6**:817-830.

83. Vizeacoumar FJ, van Dyk N, Vizeacoumar FS, Cheung V, Li J, Sydorskyy Y *et al.*: **Integrating high-throughput genetic interaction mapping and high-content screening to explore yeast spindle morphogenesis**. *J Cell Biol* 2010, **188**:69-81.

84. Horn T, Sandmann T, Fischer B, Axelsson E, Huber W, Boutros M: **Mapping of signaling networks through synthetic genetic interaction analysis by RNAi**. *Nat Methods* 2011, **8**:341-346.

85. Laufer C, Fischer B, Billmann M, Huber W, Boutros M: **Mapping genetic interactions in human cancer cells with RNAi and multiparametric phenotyping**. *Nat Methods* 2013, **10**:427-431.

86. Laufer C, Fischer B, Huber W, Boutros M: **Measuring genetic interactions in human cells by RNAi and imaging**. *Nat Protoc* 2014, **9**:2341-2353.

87. •• Fischer B, Sandmann T, Horn T, Billmann M, Chaudhary V, Huber W *et al.*: **A map of directional genetic interactions in a metazoan cell**. *eLife* 2015, **4** http://dx.doi.org/10.7554/eLife.05464.
This is the latest from series of papers from the Boutros laboratory on using morphological profiles to infer genetic interactions. The paper builds on previous work by not only identifying the interactions as being alleviating or aggravating, but also predicting their directionality.

88. • Singh S, Wu X, Ljosa V, Bray M-A, Piccioni F, Root DE *et al.*: **Morphological profiles of RNAi-induced gene knockdown are highly reproducible but dominated by seed effects**. *PLOS ONE* 2015, **10**:e0131370.
This study from our laboratory showed that morphological profiling using the Cell Painting assay is highly sensitive and reproducible. Profiles induced by a given RNA interference reagent are almost always unique, but unfortunately pairs of RNAi reagents targeting the same gene tend to have very different profiles. A significant component of this off-target effect was shown to be due to seed region matches in the RNA interference sequence.

89. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ *et al.*: **The Connectivity Map: using gene-expression**

signatures to connect small molecules, genes, and disease. *Science* 2006, **313**:1929-1935.

90. Ohnuki S, Oka S, Nogami S, Ohya Y: **High-content, image-based screening for drug targets in yeast**. *PLoS ONE* 2010, **5**:e10177.

91. Iwaki A, Ohnuki S, Suga Y, Izawa S, Ohya Y: **Vanillin inhibits translation and induces messenger ribonucleoprotein (mRNP) granule formation in *Saccharomyces cerevisiae*: application and validation of high-content, image-based profiling**. *PLOS ONE* 2013, **8**:e61748.

92. Okada H, Ohnuki S, Roncero C, Konopka JB, Ohya Y: **Distinct roles of cell wall biogenesis in yeast morphogenesis as revealed by multivariate analysis of high-dimensional morphometric data**. *Mol Biol Cell* 2014, **25**:222-233.

93. •• Piotrowski JS, Okada H, Lu F, Li SC, Hinchman L, Ranjan A *et al.*: **Plant-derived antifungal agent poacic acid targets β-1,3-glucan**. *Proc Natl Acad Sci* 2015, **112**:E1490-E1497.
This study is the latest in a series from the Ohya laboratory using the strategy of profiling a small molecule's effects on yeast cells' morphology, then searching for a phenotypic match among a large set of yeast deletion strains. In this instance, the cellular target of poacic acid, a plant-derived antifungal agent, was identified as beta-1,3-glucan.

94. •• Sundaramurthy V, Barsacchi R, Samusik N, Marsico G, Gilleron J, Kalaidzidis I *et al.*: **Integration of chemical and RNAi multiparametric profiles identifies triggers of intracellular mycobacterial killing**. *Cell Host Microbe* 2013, **13**:129-142.
The Zerial laboratory used morphological profiling to identify groups of compounds that inhibited bacterial growth without directly killing the bacteria. By querying against RNAi-induced knockdown profiles of genes involved in endocytosis, the authors posited possible mechanism of action of compounds.

95. Sundaramurthy V, Barsacchi R, Chernykh M, Stöter M, Tomschke N, Bickle M *et al.*: **Deducing the mechanism of action of compounds identified in phenotypic screens by integrating their multiparametric profiles with a reference genetic screen**. *Nat Protoc* 2014, **9**:474-490.

96. Swinney DC: **The contribution of mechanistic understanding to phenotypic screening for first-in-class medicines**. *J Biomol Screen* 2013, **18**:1186-1192.

97. Zheng W, Thorne N, McKew JC: **Phenotypic screens as a renewed approach for drug discovery**. *Drug Discov Today* 2013, **18**:1067-1073.

98. Lee JA, Berg EL: **Neoclassic drug discovery: the case for lead generation using phenotypic and functional approaches**. *J Biomol Screen* 2013, **18**:1143-1155.

99. Solmesky LJ, Weil M: **Personalized drug discovery: HCA approach optimized for rare diseases at Tel Aviv University**. *Comb Chem High Throughput Screen* 2014, **17**:253-255.

100. Ross NT, Wilson CJ: **In vitro clinical trials: the future of cell-based profiling**. *Front Pharmacol* 2014, **5**:121.

101. •• Gibson CC, Zhu W, Davis CT, Bowman-Kirigin JA, Chan AC, Ling J *et al.*: **Strategy for identifying repurposed drugs for the treatment of cerebral cavernous malformation**. *Circulation* 2015, **131**:289-299.
Unbiased image-based profiling was used to identify morphological phenotypes associated with a monogenic rare disease studied by the Li laboratory, using RNA interference to mimic the disease state in cultured cells. Then, drugs were screened to identify those that could reverse the cell-based phenotype, which were then confirmed effective in a mouse model of the disease.

102. Kasai Y, Cagan R: **Drosophila as a tool for personalized medicine: a primer**. *Pers Med* 2010, **7**:621-632.

103. Pandey UB, Nichols CD: **Human disease models in *Drosophila melanogaster* and the role of the fly in therapeutic drug discovery**. *Pharmacol Rev* 2011, **63**:411-436.

104. Dar AC, Das TK, Shokat KM, Cagan RL: **Chemical genetic discovery of targets and anti-targets for cancer polypharmacology**. *Nature* 2012, **486**:80-84.