Trends in Biotechnology

CellPress REVIEWS

Forum

Diagnostic Potential of Imaging Flow Cytometry

Minh Doan,[1] Ivan Vorobjev,[2,3] Paul Rees,[1,4] Andrew Filby,[5] Olaf Wolkenhauer,[6,7] Anne E. Goldfeld,[8] Judy Lieberman,[8] Natasha Barteneva,[8] Anne E. Carpenter,[1] and Holger Hennig[1,6,*]

Imaging flow cytometry (IFC) captures multichannel images of hundreds of thousands of single cells within minutes. IFC is seeing a paradigm shift from low- to high-information-content analysis, driven partly by deep learning algorithms. We predict a wealth of applications with potential translation into clinical practice.

Imaging Flow Cytometry

Imaging flow cytometry (IFC) combines the high-throughput, multiparameter capabilities of conventional flow cytometry with morphological and spatial information, all at single-cell resolution. Multichannel digital images of hundreds of thousands of individual cells can be captured within minutes (Figure 1), and include several fluorescence channels as well as bright field (transmitted light) and dark field (scattered light). The throughput of IFC means that it is especially well suited to the analysis of rare cell types such as circulating tumor cells (which are cancer cells that escaped from a primary tumor and circulate in the bloodstream) [1] and transition states, such as cell cycle phases (mitosis) [2].

By extracting information from these digital images (Figure 1, middle panel), IFC can quantify multiple properties of constituents of interest (including proteins, nucleic acids, glycolipids) in multiple subcellular compartments (nucleus, mitochondria, etc.). The rich information makes IFC ideal for high-content analysis, as well as machine learning, raising the possibility to profile complex cell phenotypes, identify rare cells and transition states, and, importantly, discover useful targets for disease diagnosis, personalized medicine, and drug development. Here, we discuss significant recent developments in the IFC field and a perspective on where IFC could be adopted as a diagnostic tool in clinical practice.

Advances in Instrumentation

Conventional Flow Cytometry

Conventional (zero-spatial-resolution) flow cytometry, that is, without imaging, saw its first commercial instruments in the 1970s. Thanks to continuous standardization and improvement, flow cytometry is now routinely used as a diagnostic instrument for health disorders, especially hematologic diseases [3]. Although conventional flow cytometry is considered high throughput because it analyzes up to 100 000 cells per second, it is considered to be low in information content because typically only a single feature (integrated intensity) is measured per fluorescence marker. The current trend is to increase the number of parameters that can be simultaneously measured by developing instruments with more lasers and detectors in combination with new fluorochromes that can be used in concert with one another. Mass cytometry can measure in excess of 40 markers simultaneously using antibodies tagged with rare earth metals [4]. This platform significantly increases the number of parameters measured beyond what is currently achievable with conventional flow cytometry and has driven the adoption of machine learning techniques when analyzing such multidimensional data, however, it is still limited to intensity-based features.

Imaging Flow Cytometry

Equipped with 20×, 40×, or 60× objectives and up to two charge-coupled device cameras, IFC allows thousands of morphological and spatial properties to be measured for each individual cell. These include bright field, dark field, and up to ten fluorescent channels (Figure 1) [5]. Similar to its flow cytometry-based siblings, IFC is well-suited to image non-adherent or dissociated cells, key for many clinical applications such as analyses of bodily fluids like blood, whose structures can be distorted (smeared) by placement onto a slide.

Advances in Data Analysis

Low-Content Analysis

IFC suffers from a 'content gap': although the images are high in information content, containing rich morphological and spatial information (even in a single bright-field channel), data analyses often have low information content, that is, analyses are based on only very few selected features, which are often manually identified by applying binary gates on cell populations of interest. These approaches are highly subjective, require significant user interaction, and only utilize a few morphological features instead of the hundreds that are inherently present in the data, for example, only using cell size to denote cell growth when there are likely tens or hundreds of other features that could describe and resolve the biological differences in a more reliable and powerful way. Such analyses make use of far less information than is present in the images. Yet, even with these limitations, some IFC applications are already heading toward the clinic [6], such as the diagnostic assessment of acute leukemia [7].

High-Content Analysis and Deep Learning

IFC data are now beginning to benefit from significant advancements in high-content analysis from the world of

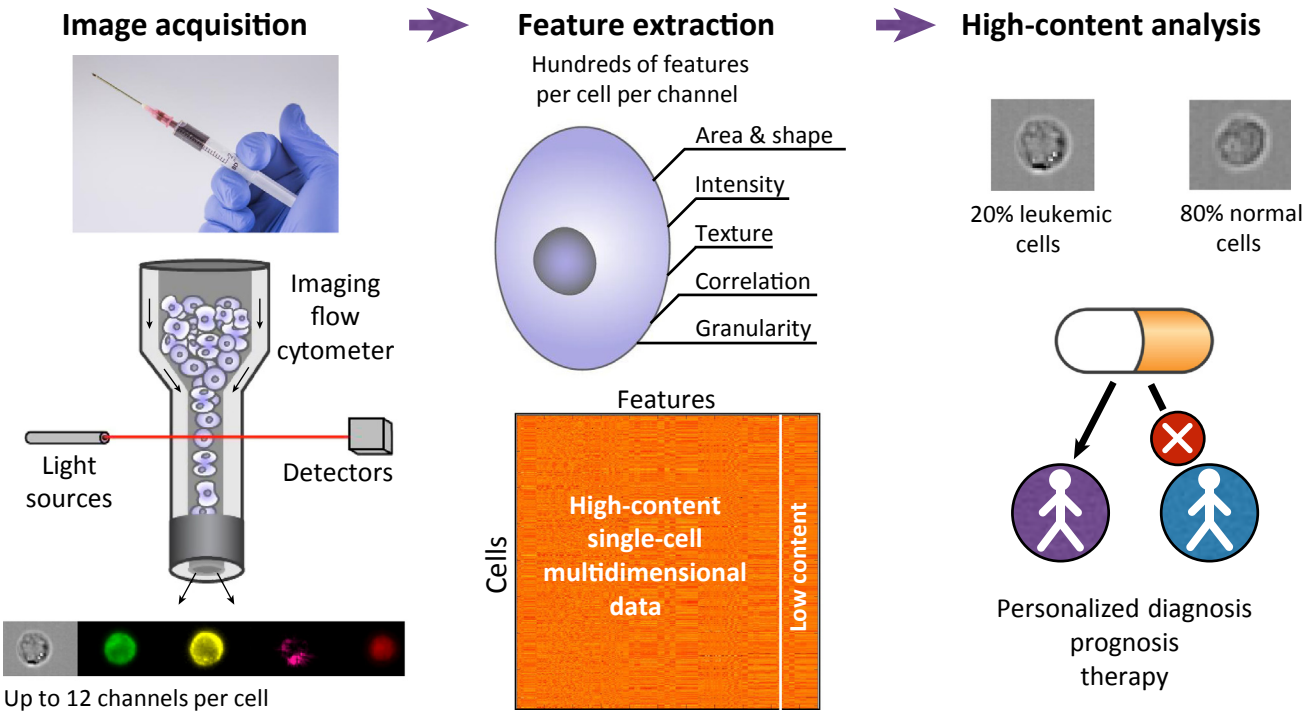

Figure 1. Imaging Flow Cytometry Acquires Images of Single Cells in High Throughput. Typical throughput is up to 5000 cells/s, but modifications to the instruments can increase this to 100 000 cells/s [14]. The images of each cell are captured by charge-coupled device detectors as the cell flows past the light sources (left panel) and morphological features are extracted (middle panel). For instance, patient blood could be analyzed to distinguish leukemic from normal cells or monitor *in vitro* or *in vivo* response to therapeutic intervention (right panel). High-content analysis could be used for personalized diagnosis, prognosis, and therapy.

high-throughput microscopy [8]. High-content analysis might unveil disease states hidden to the eye of even specialists, leading to entirely new diagnostic capabilities. In image-based profiling, once single cells are identified and segmented, thousands of quantitative metrics can be extracted, including shape, intensity, texture, and object relationships. The general strategy is to 'measure everything first, then ask questions later'. A morphological profile is created, which is like a fingerprint of each cell. Then the similarities (correlations) between profiles can be compared to define cell subpopulations or identify disease-specific phenotypes [9]. This profiling strategy might be particularly useful when applying IFC for clinical phenotyping, where the cell changes present in a chronic disease often spread throughout a spectrum of phenotypes, from pluripotent precursors to nascent immature cells to terminally differentiated cells, each with distinct morphologies and associated functions.

However, advanced IFC analysis pipelines often involve multistep workflows challenging to those who are not computational experts. Momentum to enable a broader group of biomedical researchers and clinicians to carry out complex analyses is growing; for example, a user-friendly open-source high-content IFC protocol based on machine learning is now available [2,10].

Deep learning is revolutionizing computer vision across many domains, including computational biology [11]. In particular, deep convolutional neural networks have proven to be very powerful (Box 1). The nature of data obtained using IFC is perfectly suited for deep learning for several reasons. Deep learning requires a large number of examples to train the network and IFC can quickly produce millions of single-cell images. In addition, deep learning operates at the pixel level and does not depend on prerequisite preprocessing or object segmentation, which is often prone to errors. Deep learning has shown success in identifying colon cancer cells [5] and reconstructing cell cycle and disease progression [12]. However, major challenges for applying deep learning to IFC in a clinical setting are the development of user-friendly workflows, satisfying heavy computational requirements, evaluation of diagnostic accuracy, and approval for clinical use. The Food and Drug Administration (FDA) approved the first clinical deep learning application

Box 1. Image Processing and Deep Learning

Bioimaging research has recently made great strides thanks to deep learning, a subclass of machine learning. A deep learning network consists of an input layer, a number of hidden layers, and an output layer. The larger the number of hidden layers, the deeper the network. Among several deep learning methods, convolutional neural networks have shown the most impressive results in object detection [11].

Convolutional neural networks operate on the raw pixels of each image where the network learns to enhance complex abstractions in images, such as eyes, faces, or in the case of IFC, a cellular phenotype or signature of disease, while suppressing irrelevant information in the images (see Figure I). By rastering a window over the input image, the feature maps are created (by a mathematical operation called 'convolution'). The feature maps represent how the input image looks when seen through the convolutional filter. Pooling refers to downsampling the feature maps to reduce the number of pixels while keeping the relevant information [11]. The output layer is, the classification probability which assigns a class to each input image with a certain probability, for example, Class 1 and 2 could correspond to leukemic and normal cells, respectively.

In contrast to deep learning, in conventional image analysis, the images are first preprocessed, then cellular objects are identified (segmented) by analysis software, followed by the extraction of hundreds of human-engineered features per channel and object (middle panel in Figure 1), including shape, intensity, and texture [8,9]. While classical machine learning techniques rely on human-engineered features, deep learning filters can have a much higher level of abstraction and complexity.

Recent open-source frameworks, such as TensorFlow, Keras, PyTorch and CAFFE, provide convenient and effective means to adopt powerful deep learning architectures into bioimaging research. With a wide range of practical applications emerging in various industries, we expect that such frameworks will become increasingly stable and robust, and thereby suitable for clinical applications.

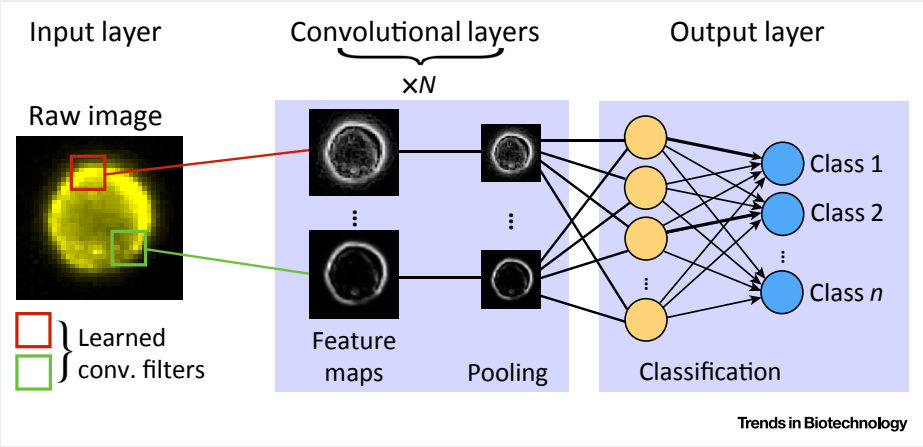

Figure I. Illustration of a Convolutional Neural Network.

for health care in 2017 (www.forbes.com/sites/bernardmarr/2017/01/20/first-fda-approval-for-clinical-cloud-based-deep-learning-in-healthcare).

Opportunities Ahead

Many potential clinical uses of IFC are conceivable, for instance a differential diagnosis of acute and chronic lymphocytic/myeloid leukemia, possibly using fewer biomarkers, or ideally unstained and unmanipulated blood samples. Reducing the number of required, descriptive biomarkers would greatly simplify the laboratory sample preparation and help preserve the intact 'nativeness' of samples, which are often fragile in hematological diseases. Recently, IFC was shown to deliver integrated leukemia diagnostics in one test [7].

Another potential clinical use of IFC is to analyze bodily fluids for rare cells, for example, the typically small number of leukemic cells that remain in the patient during treatment, known as minimal residual disease (Figure 1, right panel) [6]. There is substantial interest in liquid biopsy, the analysis of circulating tumor cells, which are extremely rare [12]. Liquid biopsy might detect cancer at an early stage, circulating metastatic or drug-resistant neoplastic cells, clotting abnormalities of platelet microparticles, or fetal abnormalities. A major advantage of the liquid biopsy is that it can be carried out in a simple, noninvasive way. One reason it took the liquid biopsy so long to develop is that circulating tumor cells are found in low concentrations in the bloodstream [12]. IFC in combination with machine learning has the potential to identify a

single tumor cell out of millions of cells with unprecedented accuracy. Blood specimens can then be sampled in serial fashion (Figure 1, right panel), providing a more comprehensive monitoring of a patient's cancer than can be obtained through traditional methods, as was demonstrated for hepatocellular carcinoma [1]. Liquid biopsies could also be analyzed by imaging technologies other than IFC: Recently, multiplex protein detection on circulating tumor cells using imaging mass cytometry has been demonstrated [13].

Despite this promise, IFC is currently primarily used in research rather than clinical practice. We see the data analysis as the primary hurdle: it is often prone to variation, manual tuning, and interpretation. These issues might be overcome with machine learning approaches. As well, there is a need for standardization of IFC, which should include standard operating procedures and standardized quality control of hardware performance. Although a common practice for conventional flow cytometry, this has not yet been implemented as such in IFC.

User-friendly, robust, and standardized workflows that can facilitate machine learning, especially deep learning, will accelerate the paradigm shift from low- to high-content analysis in IFC. Furthermore, cloud computing can overcome the computational infrastructure hurdles. These developments are key for practical IFC applications to reach the clinic, fueling the applicability of IFC as a diagnostic, prognostic, and therapeutic tool.

Acknowledgments

This work was supported in part by the US National Science Foundation/UK Biotechnology and Biological Sciences Research Council under a joint grant NSF DBI 1458626 and BB/N005163 (A.E.C. and P.R.).

[1]Imaging Platform at the Broad Institute of Harvard and MIT, 415 Main Street, Cambridge, MA 02142, USA
[2]School of Science and Technology, Nazarbayev University, Kababnbay batyr avenue 53, 010000, Astana, Kazakhstan
[3]A.N. Belozersky Institute of Physico-Chemical Biology/ Biological Faculty, M.V. Lomonosov Moscow State University, 119991, Moscow, Russian Federation
[4]College of Engineering, Swansea University, Singleton Park, Swansea SA2 8PP, UK
[5]Flow Cytometry Core Facility, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, UK
[6]Department of Systems Biology and Bioinformatics, University of Rostock, 18051 Rostock, Germany
[7]Stellenbosch Institute for Advanced Study (STIAS), Stellenbosch, 7600, South Africa
[8]Program in Cellular and Molecular Medicine, Boston Children's Hospital and Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA

*Correspondence: holgerh@broadinstitute.org (H. Hennig).

https://doi.org/10.1016/j.tibtech.2017.12.008

References

1. Ogle, L.F. *et al.* (2016) Imagestream detection and characterisation of circulating tumour cells – a liquid biopsy for hepatocellular carcinoma? *J. Hepatol.* 65, 305–313
2. Blasi, T. *et al.* (2016) Label-free cell cycle analysis for high-throughput imaging flow cytometry. *Nat. Commun.* 7, 10256
3. Betters, D.M. (2015) Use of flow cytometry in clinical practice. *J. Adv. Pract. Oncol.* 6, 435–440
4. Giesen, C. *et al.* (2014) Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* 11, 417–422
5. Han, Y. *et al.* (2016) Review: imaging technologies for flow cytometry. *Lab Chip* 16, 4639–4647
6. Barteneva, N.S. and Vorobjev, I.A. (2016) *Imaging Flow Cytometry: Methods and Protocols,* Springer
7. Grimwade, L.F. *et al.* (2017) Applications of imaging flow cytometry in the diagnostic assessment of acute leukaemia. *Methods* 112, 39–45
8. Zanella, F. *et al.* (2010) High content screening: seeing is believing. *Trends Biotechnol.* 28, 237–245
9. Caicedo, J.C. *et al.* (2017) Data-analysis strategies for image-based cell profiling. *Nat. Methods* 14, 849–863
10. Hennig, H. *et al.* (2017) An open-source solution for advanced imaging flow cytometry data analysis using machine learning. *Methods* 112, 201–210
11. Angermueller, C. *et al.* (2016) Deep learning for computational biology. *Mol. Syst. Biol.* 12, 878
12. Friedrich, M.J. (2017) Going with the flow: the promise and challenge of liquid biopsies. *JAMA* 318, 1095–1097
13. Gerdtsson, E. *et al.* (2018) Multiplex protein detection on circulating tumor cells from liquid biopsies using imaging mass cytometry. *Converg. Sci. Phys. Oncol.* 4, 015002
14. Chen, C.L. *et al.* (2016) Deep learning in label-free cell classification. *Sci. Rep.* 6, 21471