# Imaging flow cytometry

Paul Rees[1,2*], Huw D. Summers[1], Andrew Filby[3], Anne E. Carpenter[2] and Minh Doan[4]

1. Department of Biomedical Engineering, Swansea University, Bay Campus, Swansea SA1 8EN, United Kingdom.
2. Imaging Platform, Broad Institute of Harvard and MIT, Cambridge, Massachusetts MA 02142, United States of America.
3. Flow Cytometry Core Facility and Innovation, Methodology and Application Research Theme, Biosciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom.
4. Bioimaging Analytics, GlaxoSmithKline, Collegeville, PA, United States of America.

[*]email: p.rees@swansea.ac.uk

## Abstract

Imaging flow cytometry combines the high event rate nature of flow cytometry with the advantages of single cell image acquisition associated with microscopy. The measurement of large numbers of features from the resulting images provides rich datasets which have resulted in a wide range of novel biomedical applications. In this primer we discuss the typical imaging flow instrumentation, the form of data acquired and the typical analysis tools that can be applied to this data. Focusing on the first commercially available Imaging flow cytometer, the ImageStream (Luminex) we will use examples from the literature to discuss the progression of the analysis methods used in imaging flow cytometry. These methods start from the use of simple single image features and multiple channel gating strategies, followed by the design and use of custom features for phenotype classification, through to powerful machine and deep learning methods. For each of these methods, we outline the processes involved in analyzing typical datasets and provide details of example applications. Finally, we discuss the current limitations of imaging flow cytometry and the innovations and new instruments which are addressing these challenges.

# [H1] Introduction

Conventional flow cytometry is a widespread and powerful technique for the measurement of light scatter and fluorescence from cells stained with phenotypic and functional markers [1,2]. Cells are directed at high speed past laser excitation sources. Collection optics and detectors allow sampling rates of more than 10,000 cells per second, from over 30 wavelength channels. The level of fluorescence intensity measured from each channel can subsequently be used to identify cells with various phenotypes of interest, using a range of multivariate analysis tools for example FlowJo and FCS Express. Traditionally this is achieved using a series of two-dimensional scatter plots of different combinations of markers **[G]**. The user defines a polynomial region on the two-dimensional scatter plot which identifies the cells of interest, and these cells are used to generate the next scatter plot using different markers, repeating the process until all markers have been used to identify the cell phenotypes required.

Imaging flow cytometry combines the high event rate sampling of traditional flow cytometry with the acquisition of an image of each cell [3], thereby providing spatial information as well as total fluorescence intensity from each channel. For example, the ImageStream system (Luminex[4]) uses a charge-coupled device (CCD) camera with time delay integration to simultaneously acquire up to 12 images of each cell including brightfield **[G]**, darkfield **[G]** and multiple fluorescent images at rates of up to 5,000 objects per second. Time delayed integration transfers pixel information row-by-row across the detector CCD in synchrony with the cell flow velocity, enabling the high speed acquisition of focused images from low intensity objects. Time delayed integration requires a highly stable flow and precise measurement of the object transit speed using a velocity detection system which provides closed loop control to the system. An autofocus system corrects the focus of the object in the flow stream by moving the objective lens in the z direction. The acquisition of images dramatically increases the measures available for each channel, for example cell area can be measured directly and more complicated metrics such as correlation, texture and granularity give information on marker localization and cell morphology. Typically, hundreds of measures or features can then be incorporated into the gating strategy to define cell phenotypes.

Early application of imaging flow cytometry relied on the definition of simple image features from a cell's spatial information, for example the overlap of a marker's signal with the nucleus of the cell to measure nuclear translocation [5]. The rich multivariate dataset derived from the large numbers of image features has led to more powerful analyses and the application of machine and deep learning techniques to enable cell classification and functional analysis. Similarly, advances in microfluidic handling of the cells, CCD cameras and imaging modalities

have led to the prospect of significant improvements in the multi-spectral images **[G]** obtained and the speed of acquisition.

This Primer will focus on typical analyses that can be carried out using imaging flow cytometry, highlighting the advantages of the images acquired compared with traditional fluorescence flow cytometry. While new imaging flow technologies are constantly being reported, the commercially available system from Luminex [4] has been the mainstay of imaging flow studies to date and will be the main focus here. Data collection and analysis steps that can answer specific questions related to the biology of the cell are discussed using specific datasets.

Typical applications that use simple features extracted from the single cell images are described. For example, the use of spot counting to measure the uptake of nanoparticles in cells, the location of calcium in T-cells and morecomplex, user-defined features to determine the activation of eosinophils. Advanced machine and deep learning techniques can be applied to solve more advanced problems such as the classification of white blood cells and the identification of micronucleus phenotypes in cells exposed to a genotoxic compound. The limitations of imaging flow cytometry using current commercial systems are discussed and new technologies which are being developed to overcome these issues are outlined.

## [H1] Experimentation

The general experimental design of imaging flow cytometry can be considered an extension of traditional flow cytometry. The measurement of suspension cells is well suited to these instruments owing to the fluidic cell handling systems, however adherent cells can be lifted or dissociated with the correct protocols [6]. The number of cell images required depends on the application. The high event rate of imaging flow cytometry is perfectly suited to applications requiring high cell numbers, for example the identification of rare cells. This section outlines the general instrumentation setup, experimental design, sample preparation and data collection steps for the measurement of cells using imaging flow cytometry.

### [H2] Instrumentation

In general, an imaging flow cytometer enables cells suspended in a fluid to pass before an imaging system. The fluid handling system can employ a sheath fluid, as in traditional flow cytometry [1], or transport the cells in a microfluidic device [7-9]. The imaging system is usually a traditional CCD camera ,[4] but systems using radiofrequency-tagged emission fluorescence

microscopy to take advance of the sensitivity of photomultiplier detectors have also been demonstrated[10]. The ImageStream system marketed by Amnis (now part of Luminex) [11] was the first imaging flow cytometer, introduced in 2004. Cells in suspension are hydrodynamically guided into a core stream which is illuminated by an LED array and mixture of collinear and spatially separated laser lines at rates of up to 5,000 objects per second (**Fig.1a**). The standard excitation laser has a wavelength of 488 nm however the system can be expanded with up to five further lasers at 375, 405, 561, 592, and 642 nm; a higher-power 488 nm laser is also available [4]. The ImageStream MkII system doubles the original 6 channel acquisition capacity to 12 channels by using two image detection systems including filters, spectral decomposition systems and two CCD cameras. This enables the capture of images from up to 10 fluorescence channels together with brightfield and darkfield images (**Fig. 1b and 1c**). Images can be captured at 3 different magnifications 20, 40 and 60X giving a pixel resolution of 1, 0.5 and 0.3μm and native depth of focus of 8, 4 and 2.5μm respectively. The native depth can be increased using an extended depth of field [12], maintaining focus over a depth up to 16 μm, an option which is useful for spot counting applications. A high gain mode can also be introduced to adjust the gain setting, object detection thresholds and masks **[G]** to maximize the measured signal while minimally increasing the noise. This high gain mode enables the measurement of dim fluorescence markers or very small objects, such as extracellular vesicles[13] and viruses. The ImageStream system can also be outfitted with a 96-well plate reader to aim unattended acquisition, 384-well plates and larger are not supported.

A major strength of the ImageStream system is the comprehensive acquisition and analysis software for the exploration and analysis of the rich multivariate datasets. The data acquisition software (INSPIRE) enables the basic self-test, calibration and set up of the instrument. During data acquisition researchers can modify instrument operating parameters and observe the images obtained from each channel in real time. Furthermore, data can be acquired selectively, based on a gating strategy from image features to reduce the number of unwanted images in the subsequent data file.

## *[H3] Calibration*

Prior to analysis, all analytical systems should be calibrated **[Au: Edit OK?]**. The ImageStream system uses a suit of calibration tests called ASSIST. The ASSIST tests monitor excitation laser power, optical alignment, flow stream stability and focal quality. The calibration process also measures spatial registration and can correct if misaligned. During calibration, any failed tests are flagged so the user can compare the results against the accepted pass parameters. Often, rerunning the failed tests after a short wait time will rectify this, as it is quite common that after shut

down and start up, the fluidics require time to stabilise, and the lasers need to warm up. However, continued failure is a sign that there is a more serious issue and a trained service engineer may be required to diagnose the fault and find a solution.

System performance is monitored and tracked using a well characterised standard. In the case of the ImageStream system, the standards are 1µm polystyrene micro particles called speed beads. Speed beads are not fluorescently labelled and there has been some reluctance in the flow cytometry field to use them to measure or infer photonic sensitivity of the system as a flow cytometer. The ASSST tests are often supplemented using multi-level fluorescent microspheres that can be used to infer photonic sensitivity for each imaging channel [14] in molecules of equivalent soluble fluorescence (MESF). The smaller the MESF value, the more sensitive the system should be for detecting a lower number of fluorochrome molecules and associated photons per cell/particle. The quoted value for the ISXmKII is 5 MESF making it one of the most sensitive fluorescence-based flow cytometers on the market. It is also advised to use a standard biological control if available. For example, peripheral blood mononuclear cells from a healthy donor or Leukapheresis cone. In this case an antibody against the surface protein CD4 as it is highly expressed on CD4 T cells with low to intermediate expression on monocytes could be used. As such this creates a cellular control with multiple levels of signal.

### *[H3] Data compensation*

As with traditional flow cytometry, before any quantitative analysis can be performed, the data must be compensated for the spectral crosstalk between channels. The process of compensating imaging flow cytometry data is more involved given the spatial nature of the data. Essentially the spatially resolved data requires compensation at an individual pixel level [15]. Separate aliquots of sample are stained individually with each dye/marker required for the full experiment and are run through the ImageStream separately. The ImageStream Data Exploration and Analysis Software (IDEAS) can then take these individual sample results to measure the crosstalk from each marker into the empty channels and calculate a compensation matrix. The compensation matrix is multiplied by the fluorescence intensities in each pixel which removes these unwanted contributions from each channel.

## [H2] Experimental design

The first step of experimental design is to formulate the question and determine what key measurements are needed. Imaging flow cytometry tends to be more labor intensive than non-imaging flow cytometry and provides fewer fluorescence channels. If the biological question requires a spatial or morphometric read out, then imaging flow cytometry will be the preferred

technology. For example, measuring the degree or amount of FoxP3 (a protein involved in regulating the development of T-cells) in the nucleus of primary human regulatory (T-regs) cells using the cluster of differentiation 4T (CD4T) marker. The system capability impacts how many parameters can be measured per cell and which fluorochromes and dyes can be detected. The number and wavelength of available excitation lasers as well as important information about the detection channels (for example one camera versus two camera system) should be considered. There are several widely available online spectral viewers that can create a virtual machine with the right lasers and filters to aid in designing the optical setup. Next, the minimum number of biomarkers required to identify the cell type of interest from a heterogeneous population should be determined. If additional channels are available, the researcher should consider whether other parameters might be of interest and measure them simultaneously.

For example, consider a protocol previously used to identify T-regs from whole, lysed human blood [16.] To do this effectively, the sample may need to be stained with antibodies against CD45 (pan-white blood cell marker) to distinguish white blood cells from un-lysed red blood cells and debris. An antibody against CD3 should also be included to identify all T cells within the CD45-positive white blood cell (WBC) population. If focusing on regulatory CD4-positive T cells, antibodies against CD4, CD25 and CD127 (IL-7 receptor alpha chain) should be included. Each antibody would need to be tagged to a unique fluorochrome that would be compatible with the spectral setup of the system as well as each other. An advantage of imaging flow cytometry is that markers may be used in the same spectral channel if they are spatially distinct. The selection of fluorochrome to marker/target follows the same rules and approaches for conventional and full spectral fluorescence flow cytometry, where essentially, low expressed markers are assigned to bright fluorochromes and highly expressed markers to dimmer fluorochromes [17]. Using the example above, when measuring the nuclear occupancy of the FoxP3 protein, an antibody against FoxP3 tagged to a compatible fluorochrome as well as spectrally compatible nuclear dye should be selected. In all cases, the fluorescent reagents require careful titration, including the nuclear dye, because signal saturation can pose a challenge owing to the reduced dynamic range on the CCD camera (12-bit compared to 18-bit on a non-imaging flow cytometer) and the lack of control over each imaging channel (signal intensity is controlled by laser power, meaning that it can be challenging to balance a dim and bright signal for the same laser). Once reagents have been optimized, sample preparation follows the same process as with conventional flow cytometry.

## [H2] Sample preparation

Sample preparation for imaging flow cytometry is analogous, in practical terms, to any form of fluorescent antibody or dye-based technology that is used to analyze cells or particles in

suspension. Cells are prepared in a single-cell suspension and stained with surface marker antibodies, fixed, treated with intracellular antibodies and nuclear dye prior to acquisition. Nuclear dyes must be carefully titrated to ensure it does not saturate the other signals. As with conventional flow cytometry, single stained controls are required for compensation for all markers. The most noteworthy difference in sample preparation comes at the last step where it is essential to concentrate the samples in a low volume, for example 50 µl, and ideally if cell numbers allow, at a concentration of 20-30 million cells per ml (1 million cells total in 50 µl). While this may seem extreme, the imaging flow cytometer tends to run at a slower rate than conventional systems so it can take impractical amounts of time to acquire enough cells in dilute samples, particularly if looking for rarer cell types. Also, for larger volumes of a similar concentration the large number of cells measured can make the data files very large and the processing of the subsequent images becomes highly computer intensive. A highly concentrated sample of no more than 50µl volume will help to alleviate these issues, however if working with larger and sticky cell types, less concentrated samples may be preferred. Sample acquisition is relatively easy; it is often best to begin with a fully stained sample and use plots that show the raw maximum pixel **[G]** for all events in any channel and to ensure that the excitation laser powers are set to achieve maximum signal without any saturation.

## [H2] Data collection

An advantage of imaging flow cytometry is that the data output is usually in the form of single cell images because of the control afforded by channeling the cells in a sheath fluid or microfluidic device in front of the detector. The magnification is usually between 10 and 60x (image size of between 50 and 150 pixels squared) and the images may consist of as many as 10 different spectral channels. The rate at which the images are acquired depend on the flow rate of cells past the detector and the frame rate of the detector itself. Early instruments acquired 100 cells/s but new developments in opto-mechanical imaging have pushed speeds to exceed 10000 cells/s [9]. However, there is a tradeoff between the acquisition speed and the image quality, for example using time delayed integrated CCD based imaging, the binning of pixel rows required at higher flow rates reduces the spatial resolution of the images acquired.

## [H2] Defining masks

A key step in data analysis is the generation of image masks, where the cell boundaries and important intracellular regions inside the cell, such as nucleus boundaries, are identified. Most commercial instrument providers provide bespoke software which segment the individual cells

providing masks, preprocess data to exclude out of focus and cell clusters (**Fig. 2**), and generate multiple image features. Open-source software is also available.

The contrast in the brightfield image is often good enough to be able to define a cell mask but often a cytoplasm marker is used to define the extent of the cell, or a DNA marker can be used to define a nuclear mask. Simple masks can be obtained using a user defined threshold value for the marker intensity and setting pixels to 1 above that value, however a host of methods are available to automatically define the threshold value or deal with complexities such as nonuniform illumination. Once a mask has been defined it can be used to generate image features. Morphology measurements, such as area and shape measures, can be directly obtained from the masks and intensity measures are generated from the fluorescence pixel channel values where the mask pixel values are 1. Most commercial and the more recently developed imaging flow systems have associated masking software which works best with the image capture modality, for example, the INSPIRE software supplied with the ImageStream provides basic masks for each channel however these usually need to be modified for accuracy as described later.

# [H1] Results

In this section, the process of taking the basic images acquired by imaging flow cytometry to identify cell features that can be used to perform tasks such as defining cell phenotypes or measureing biological function is discussed. This will involve removing non cellular images (such as debris or calibration speed beads) or images which are out of focus. Then data analysis techniques to perform phenotypic identification and biological function assays will be outlined starting with simple gating strategies through to the use of the latest deep learning algorithms

## [H2] Cell feature generation

The process of data analysis in imaging flow cytometry follows the same procedure as for high throughput microscopy. The masks generated from the various image channels define the cell body, nucleus and any subcellular organelles to be used for analysis. The cell, nucleus and organelle shapes and size features can be extracted using typical open-source software tools or as is more often the case, researchers would use developed scripts in software languages such as Python, MATLAB and R. Commercial instruments are usually provided with software tools which generate cell features from the raw image data files and the software to analyze these features extracted, for example, using IDEAS supplied with the ImageStream system.

## [H2] Pre-processing the cell images

Before analysis of the acquired cell images can take place, the data set needs to be filtered to remove images that capture objects other than single cells or are out of focus due to the cell not coinciding with the focal plane of the lens (**Fig 2**). To identify the in-focus events, the intensity gradient along a pixel line is used (the focus building block **[G]** within IDEAS software); in a well-focused brightfield image, the cell boundaries create a sharp intensity change and hence a high intensity gradient (root mean squared) value. The in-focus sub-population can then be defined with a high-pass gating of the gradient histogram**.** The focused sub-population may be further filtered to select single cells based on object size and shape (the single cell building block). This operates on morphological features obtained from the brightfield image mask, producing a 2D scatter plot of aspect ratio **[G]** versus area. A single-cell sub-population may be defined from the dense cluster of events with high aspect ratio (tending to circular shape) and intermediate area (lying above a band of smaller objects corresponding to debris, and below higher points representing images containing multiple cells). One further pre-processing step could also involve the identifying of any dead or inactive cells providing a suitable cell marker has been included in the experimental design.

## [H2] Data analysis

Once the image features have been generated and the basic preprocessing protocols run, the image features can be used to identify phenotypes and measure biological function. Usually, the phenotype of each cell should be determined using successive gating of features. A further advantage of acquiring single cell images is the opportunity to identify subcellular locations and the spatial trafficking of proteins and signaling molecules using a combination of masks and features. Advanced multidimensional algorithms can be used for phenotype classification and functional analysis.

### *[H3] Use of image features in flow cytometry gating*

The ImageStream is provided with IDEAS software that enables the data preprocessing and the usual gating analysis associated with traditional flow cytometry. It also includes several building block tools that suggest suitable image features, masks and gating strategies for typical cell image analysis. The first step using IDEAS is to generate and implement a compensation matrix to correct for spectral spillover or crosstalk between channels using the individually stained samples. This is a familiar process for flow cytometry however in the case of imaging flow cytometry the compensation matrix is used to deconvolve the cross-channel contributions

in each single pixel. Furthermore, at instrument start-up the acquisition software corrects for the individual variations in each pixel's dark current and gain, to give a uniform photometric response for each pixel in the image. Any vertical and horizontal pixel spatial offsets are also computed, and these are corrected during acquisition. The result is a brightfield, darkfield (light scatter) and up to 10 fluorescence channel images for every event that triggers the acquisition process.

IDEAS automatically generates a segmentation mask for each channel per field of view (**Fig 1b**). This allows the user to mask the cell outline using the brightfield channel and the nucleus, for example, if a nuclear dye has been included in the experiment. Often the automated masking parameters need to be adjusted [18], for example, by changing the intensity threshold level for segmentation, a process which is critical especially when detecting subcellular organelles. Once the object masks are accurately defined, shape morphology features such as area, perimeter and aspect ratio can be measured to be used in cell gating strategies. IDEAS also allows the measurement of more complicated image features which measure the texture and granularity which considerably enhances the assay opportunities compared with traditional flow cytometry.

The ability to capture images of single cells at different wavelengths opens up many new avenues of investigation in comparison with traditional flow cytometry [19]. For example, compared with measuring just an intensity value per channel, imaging flow cytometry allows capturing morphological features for each channel, such as cell area, perimeter and shape metrics. Early applications of imaging flow cytometry exploited the use of these simple features unavailable to traditional flow cytometry for phenotype identification. For example, while a rough approximation of cell size and shape can be obtained using traditional flow cytometry using forward and side scatter, an obvious application of imaging flow cytometry would be the direct measurement of cell size and shape. Imaging flow cytometry has been used extensively to study the cell cycle control in fission yeast where a detailed measurement of cell size is critical[20] [21] [22]. Similarly, the technique has allowed the classification of the morphological phenotypes of budding yeast based on the measurement of the size of bud lengths [23]. The change in shape of the nucleus during mitosis allowed the detection of the anaphase, prophase, metaphase and telophase of the cell cycle with only a DNA marker [24].

Imaging flow cytometry can analyze subcellular structures which is far more difficult, if not impossible, with traditional flow cytometry.For example, the IDEAS analysis software allows the detection of 'spots' in the cellular image\. IDEAS v6.3 software now allows the use of connected component masks, where a channel mask can be broken down into multiple

individual component masks to label subcellular structures. All the feature measurements available for masks can then be applied to the individual components. This addition is especially useful when measuring multiple subcellular structures, for example in particle uptake studies where the intensity of individual vesicles containing nanoparticles is requird [28].

### *[H3] Use of user defined masks and features*

A strength of the IDEAS software is the flexibility it provides to take the basic features and masks and modify them to provide custom measures tailored to the application. For example, the internalization of nanoparticles by cells was quantified using an internalization score, which was derived from the correlation of the fluorescence nanoparticle pixels in the cell mask and the same cell mask that was eroded by 3μm to remove the outer membrane region [29]. A similar strategy was also used to measure the extent of the ciliary zone thickness in mature human bronchial epithelial multi-ciliated cells, as the difference in area between the cell body and the ciliated zone mask [30].

### *[H3] Machine learning analysis strategies*

Combining the large numbers of possible features which can be derived from each channel image for every cell in the population can lead to an incredibly rich dataset with the power to identify more complex phenotypes. These multivariate datasets are perfect candidates for the application for high content approaches to identifying cell phenotypes and determining cell function.

One of the first examples applying machine learning to imaging flow data identified the stages of the cell cycle including mitosis, as well as DNA content, in a completely label-free assay [31]. In the machine learning training step, Jurkat cells were stained with propidium iodide, to quantify DNA content, and a MPM2 (mitotic protein monoclonal #2) antibody, to identify mitotic cells. This enabled identifying cells in G1, S, G2 phases and the four mitotic phases–prophase, metaphase, anaphase and telophase–using traditional gating techniques. The annotated cells were used to train a network to classify the phases based on the brightfield and darkfield channels alone, without the fluorescence channels. Finally, in the prediction (or inference) step, the trained machine learning model used the label-free channels alone to classify cells into phases and predict the intensity of propidium iodide stain. This machine learning strategy has also been employed to classify human white blood cells where CD markers were used to

annotate the B, T cells, eosinophils, monocytes and granulocytes. Trained machine learning algorithms were able to identify the cell types using just the brightfield and darkfield channels [32,33].

To apply machine learning the user must extract a table of image features for each cell. IDEAS can measure large numbers of features for each channel, and these can be output for future analysis. Similarly open source software tools such as CellProfiler [34], which can extract large numbers of shape, intensity and textures features for multivariate analysis can be used [35]. The user must decide on what features are used in the analysis depending on the classification or regression task at hand. Care must be taken to remove any non-biological features that can be present such as cell number or a timestamp. The user can pre-filter the data to remove any correlated or redundant features which can confound the learning process and speed up analysis times [36,37]. Once the feature table has been extracted the user is free to choose any appropriate analysis tools, for example, MATLAB has a user-friendly machine learning toolkit, Python has extensive libraries such as Scikit-learn and also the open-source R language has been specifically designed for statistical computing. A machine learning module has recently become available within the IDEAS v6.3 software package to enable the application of machine learning techniques to the image data with no expert knowledge.

# [H1] Applications

The range of applications of imaging flow cytometry has increased over the past 15 years and this has been driven not only by new advances in the technology's hardware but also in the rapid development of computational techniques available to analyze the rich multivariate datasets acquired using these instruments. Applications have matured from using a single, simple image feature through to the deployment of advanced deep learning algorithms.

## [H2] Nanoparticle uptake

Imaging flow cytometry enables the recognition and enumeration of sub-cellular areas, such as punctate spots within a cell image. For example, the heterogeneity of fluorescent nanoparticle (quantum dots) loading into endosomes can be assessed. U2-OS Osteosarcoma cells loaded with Qtracker705 (Invitrogen) nanoparticles were excited at 488nm and images acquired using the 660-735 spectral channel. Bright spots under laser excitation corresponding to endosomal vesicles loaded with nanoparticles are clear in the fluorescent channel (**Fig 3**). Using the masks feature, these fluorescent areas can be identified using one of a number of possible masking functions such as intensity, peak or spot. A measurement feature may then be generated from the spot mask to generate a summed spot area or a spot count per cell. In this example, these represent a dose metric for the accumulated nanoparticles and would be of relevance to a nanotoxicology or nanomedicine assay. It should be noted that this process of mask generation and feature extraction relies only on the presence of distinct pixel intensity clusters that may be identified within the cell image. A fluorescence image is not therefore essential and sub-cellular morphology may be clear within the scattered light variation of the dark field channel or as dark spots in a bright field image.

The extraction of spatial metrics is not an end in itself, and the real impact of imaging flow cytometry lies in the application of post-measurement models and analysis of the data. For example, in this nanoparticle uptake data the statistical distribution of the number of nanoparticles loaded vesicles per cell is over dispersed relative to the Poisson distribution expected on a hypothesis of random particle arrival and internalization. Further study shows that this is due to cell area heterogeneity and provides predictions of the dose heterogeneity of nano-vectors [42]. The potential for probabilistic models of cell processes is realized here by the ability to extract spatial information across large populations.

## [H2] Calcium Mobilization in T Cells

Calcium acts as a ubiquitous signaling moiety in cell biology, passing on extracellular signals to drive changes in gene expression and cellular responses. In the immune system, calcium acts as a key secondary messenger downstream of the T cell receptor after recognizing foreign antigens. The measurement of calcium mobilization in T cells is often a critical assay for the characterization of cells from patients and from various transgenic mouse models, where T cell signaling is suspected to have been perturbed in some way. One of the key features of calcium signaling and mobilization is the spatial aspect, with temporal involvements and dependencies on different subcellular locations making it a very attractive model system to be measured using imaging flow cytometry.

A fluorescent dye panel that is compatible with a 4 laser, 2 camera, 12 channel imaging flow cytometer is used to identify two key intracellular organelles involved in calcium mobilization, namely the endoplasmic reticulum and the mitochondria, and secondly, to report the flux of calcium ions in these locations [43]. These dyes were carefully titrated to ensure optimal signal to noise, onward cell viability and organelle specificity. For the latter consideration, it is possible to use spatial information to confirm specificity of each organelle dye as it has been shown that excessive concentrations of such dyes will lead to a loss of specificity and a generic labeling of intracellular structures. In this case the bright detail similarity feature provides a metric for the spatial segregation of two distinct organelle dyes with low feature values (less than 1.5) representing good spatial segregation. After ~60 seconds of data collection, the sample was unloaded, a stimulus such as anti-CD3 antibodies or Calcium modulator added and then the sample reloaded to continue data acquisition. Single-stained samples were collected for compensation purposes.

Corrected and compensated data was analyzed by creating a range of masks based on the specific organelle stains and restricting the kinetic measurement of various fluorescent calcium probes to these structures versus the whole cell signal. Utilization of this approach uncovers interesting features of calcium mobilization in activated T cells, namely that mitochondria seem to be able to act as a sink for intracellular ER-derived calcium and not just from an extracellular influx; this observation is wholly dependent on the ability to obtain single cell, kinetic spatial information at a population-wide level.

## [H2] Morphology analysis of granulocytes

While traditional flow cytometry can indirectly measure granularity via the intensity of the scattered laser light, imaging flow cytometry can directly measure spatial variation in the brightfield, darkfield and fluorescence images. For example, the pronounced morphological features of eosinophils, a type of granulocyte which play a role in mediating inflammatory response can be analyzed. The granules in eosinophils are enzyme-filled vesicles and these produce high-contrast dark spots within brightfield images. When an eosinophil is activated due to an immune threat the granules increase in number and migrate to the cell surface. The procedure for assessing the sub-cellular distribution of these granules is applicable to any analysis of cell morphology based on brightfield image contrast.

Eosinophils derived from leukocytes in whole blood required gating on the high autofluorescence signal, owing to the concentration of flavin adenine dinucleotide localized within the granules[44]. The brightfield images for the eosinophil population were analyzed to determine whether the granules are located at the outer edge of the cytoplasm, next to the membrane, or more evenly distributed within the cell. The analysis is based on the creation of area masks to define general sub-regions of the cell and to define the granules (using the masks function) (**Fig 4a**). Logical mathematical operations using these masked areas can then identify degree of overlap and quantify the spatial distribution of the granule dark spots (using the features function). A mask of the brightfield channel is automatically generated by the IDEAS software; erosion of this mask isolates the inner body of the cell. Logical combination of original plus eroded masks can then produce a mask of the cell perimeter, through a Brightfield mask AND NOT Erode mask operation **[G]**: shared areas, common to both masks, are removed from the original brightfield mask. Having created location masks that define the cell interior and perimeter, the image spots corresponding to granules can be masked. This can be achieved with a number of alternative masking functions such as intensity, peak or spot, with selection determined by user preference and their relative performance when applied to the specific cell image set being analyzed. Final extraction of a measurement feature, defining the degree of membrane association of the granules, is achieved by calculating the area of a combined mask resulting from logical combination of the granule mask AND perimeter mask, which selects only those masked granule areas that lie close to the cell membrane (**Fig 4b**). A histogram of the calculated area is plotted and used to gate cells **[G]** that have a high or low degree of overlapping dark spots underneath the membrane mask. There are often alternative approaches that may be taken in a spatial analysis for example implementing a morphology-based approach using only the dark spot mask. Rather than isolating the membrane associated granules this approach seeks to classify the different spatial distributions seen across a cell population to differentiate when granules are distributed across the whole of the cell and those where they are preferentially clustered at the cell membrane. A scatter plot of these features (**Fig 5**) presents a distribution of the cell-state

extending from cells with centrally located granules on the lower left (dispersed and asymmetric pattern) to those with strong membrane association in the upper right (localized and symmetrical pattern).

## [H2] Machine learning

### *[H3] White blood cell classification*

White blood cell phenotype identification can be used to demonstrate the steps involved in applying machine learning for automated analysis. In traditional flow cytometry this is typically achieved using CD markers to label the different cell phenotypes together with forward and side scatter. However, the white blood cell phenotypes can be classified without CD markers using imaging flow cytometry and machine learning (**Fig S1**). White blood cells derived from healthy doners were stained with Fluorescein isothiocyanate (FTIC) labeled anti-bodies against the cell surface markers for monocytes (CD14), neutrophils (CD15) and lymphocytes (CD19 - B cells and CD3 – T cells). These markers together with autofluorescence were used to identify eosinophils, using traditional gating techniques on image data acquired using the ImageStream system[32]. To employ machine learning for this task the first step is to export all used features to a data text file (in this example the measurement features for darkfield, brightfield and channel 4, which contains autofluorescence images). The features together with the known phenotypes for a cell population is then used as an input to train typical machine learning algorithms[32]. This example focusses on classifying the eosinophil, neutrophil, monocyte and lymphocyte cell phenotypes. In addition to image channel features, IDEAS also exports cell object and time data columns. These biologically irrelevant metrics need to be removed from the data prior to implementation of machine learning. The ability of machine learning algorithms to correctly classify the cell phenotypes using the combined data set of all sub-populations can be assessed. In this case the combined data for the 3 channels provides a data matrix of 115 metrics for 3,168 cells. The feature data matrix may be used with any chosen machine learning software, in the form of confusion matrices **[G]**, from MATLAB's classification learner app. For illustration purposes a naive Bayes and a fine tree algorithm were chosen. Both deliver highly accurate classification and unsurprisingly the decision tree is optimum as it follows the binary signal discrimination employed in the original manual gating.

### *[H3] Deep learning*

The examples we have described so far have required measuring particular image features that are pre-defined by software. The IDEAS software, as well as open source bioimage analysis tools such as CellProfiler and ImageJ, can measure a large number of features, which can be selected by the researcher or used en-masse for machine learning. Deep learning, by contrast, has the potential to go beyond features that humans have pre-programmed into software. Deep learning algorithms (neural networks) use full images as the input to a convolutional network. When appropriately trained, the network generates the features required for the analysis applications; these features can often be more powerful than human-designed ones. Deep convolutional neural networks for image classification are well suited to small multichannel images and they require large numbers of images to train, which makes them perfect candidates for the analysis of data from imaging flow cytometers.

One of the first applications of deep learning to imaging flow data [38] trained a deep convolutional neural network to detect the different phases of the cell cycle using the pixel data of the images rather than extracting conventional image features. Other challenging applications are quite diverse. For example, a convolution neural network was trained to classify phytoplankton species and also to identify the stages of the life cycle [39]. More recently the same deep learning algorithms were used to classify large numbers of pollen species with high accuracy [40]. Furthermore, morphological and fluorescence features that were conserved at the various levels of taxonomy were determined. Similarly, deep learning was similarly used for predicting Cryptosporidium and Giardia in drinking water[41].

## *[H3] Micronuclei detection using deep learning*

A typical application of deep learning to imaging flow cytometry data is to take advantage of the large number of single cell images to classify individual phenotypes. For example, deep learning can be used to classify micronuclei events from imaging flow data. The in vitro micronucleus assay is the standard method for the assessment of possible DNA damage induced by chemical / radiative perturbation. The assay is the gold standard test of genotoxicity in the development of all chemicals and pharmaceuticals. When the nucleus divides during mitosis, chromosome fragments that fail to be incorporated into the daughter nuclei appear as micronuclei within the cell. Imaging flow cytometry has been shown to be an effective measurement tool for the micronucleus assay giving the high throughput single cell nature of the data [45 46 47]. The assay was partly automated using spot counting to find the micronuclei within the cells [48] however it was subsequently shown that this is a perfect

application for the use of deep learning to fully automate the classification of cells with micronuclei [49 50].

As with the application of deep learning to any problem, the type of neural network to be used should be determined. Several classification networks have been applied to imaging flow data including AlexNet [51,] ResNet50 [52] and VGG-16 [53], all of which have been pre trained on many thousands of annotated images. The number of layers and complexity of the network can improve classification accuracy but also increase the time required to re-train the neural network. Once the network has been selected the input layer needs to be matched to the single cell image size pixel sizes. The individual images extracted are often of different sizes and therefore they need to be cropped or padded to the network input size. Also the application will dictate which image channels will be input into the network for classification. While classification networks were the first to be applied to imaging flow data, other networks can be used. For example, a Faster region-based convolutional neural network was used to quantitatively analysis of phagocytosis in cells using imaging flow cytometry data [54].

It is also important to consider which programming language to use to implement the network. MATLAB has useful deep learning toolboxes, however Python has a host of different packages to implement convolutional neural networks including Keras, Cafe and TensorFlow. Although for non-experts, as with the addition of machine learning into the IDEAS software, likewise a deep learning module has now also been developed. This module can train using one of the popular neural networks within the user interface, removing the need for cropping and padding, making the application of deep learning easier for novice users. This tool was used recently to classify silicone oil droplets from protein particles[55], a protocol which has noteworthy application in the development of biopharmaceuticals.

To demonstrate training a deep learning neural network to classify cells with micronuclei, the publicly available dataset[50] which contains TK6 cells which exhibit mono, bi, tri and quardanuclated phenotypes together with micronuclei events (**Fig S2**) after exposure to carbendazim can be used. The human annotated dataset has both brightfield and DNA fluorescence images which have been cropped/padded to 64x64 pixels and maximum/minimum renormalised per image. As a simple example, just the DNA channel was input into the 'DeepFlow' neural network [38] developed specifically for Imaging flow cytometry data which is available in Python and MATLAB for this image size and trained on 6445 randomly selected images from each class over 30 epochs, minibatch size of 30 using the ADAM optimizer. The resulting confusion matrix (**Fig S2e**) shows the results of the trained network on 1609 test images, which gives an overall accuracy of 79.1% This can be improved

by augmenting the rarer cell classes, using the brightfield channel and increasing the number images used to train. As well as classification, the weights of the penultimate layers of the trained network can be used to visualize the performance of the network or even for regression analysis. For example, extracting the weights from the (average pooling) layer above the classification layer and using t-distributed stochastic neighbour embedding **[G]** to reduce the features to two dimensions to visualize the class prediction.

## [H1] Reproducibility and data deposition

The ImageStream system for imaging flow cytometry has an extensive calibration, self-check and initialization start-up process, resulting in excellent data reproducibility. As with all protocols that require staining or labeling cells, variable uptake of the markers or target binding can lead to problems with reproducibility in the analysis, however this is not a problem specific to imaging flow cytometry. In fact, a study to detect micronuclei events in cell conducted at three different laboratories (using different instrument settings, such as excitation laser intensities) using different DNA stains demonstrated that deep learning algorithms trained on data from one laboratory could be used to classify results from the other laboratories with high accuracy [50].

The move to open and transparent data analysis has led to authors depositing data and analysis code using platforms such as FigShare, GitHub and within supplementary information with manuscripts. The flow cytometry community has adopted a set of minimum standards required for data (MIFlowCyt)) [58] and the preferred depository, FlowRepository. While no formal standards exist for Imaging Flow Cytometry, attempts have been made to outline best practice in report results and depositing data [60] which will become more important as more data is being made available. The MIFlowCyt minimum standards for reporting results includes the details required on the experimental design, samples/specimens used, preparation, treatment and staining of samples, instrument details and the analysis applied to the data. These reporting standards apply equally to Imaging Flow Cytometry however the data analysis on the images produced is more aligned to high throughput microscopy data. While these standards are not well established in microscopy, recent attempts have been made to determine best practices for analysis[61] and reporting[62] which should also be applied to imaging flow cytometry.

## [H1] Limitations and optimizations

Imaging flow cytometry shows the value in combining the advantages of a microscope and a flow cytometer. However, the technique does have limitations, for example, in lacking capability for workflow automation, cell sorting, repeated time-lapse imaging of the same cell and 3D resolution. Nevertheless, recent advances, in the field of imaging flow cytometry itself and from other disciplines, are beginning to address these limitations.

### [H2] Automation

An imaging flow cytometry workflow involves multiple steps, in which both the laboratory procedures for data acquisition and the computational procedures for data analysis often require manual handling., Steps such as sample staining, centrifuging, washing, sample handling, instrument preparation, data capturing, event gating, triggering, data cleaning, profiling all require manual handling. For wet-lab procedures, there are currently no robotic options such as those in plate-based or slide-based high-throughput machines. Although batch processing can be used, expert-guided analysis is the norm in computational processes and thus scaling within an automated and distributed computing platform is difficult. This poses a major challenge in downstream analyses, in which over 100 unique features, typically dozens of masks for cellular objects and subcellular compartments, as well as a large collection of algorithms available for each channel, yields several thousands of combinations to identify features and populations of interest. Partial automation is available, for instance, the Luminex ImageStream system is accompanied by data acquisition software (INSPIRE) and a separate analysis suite (IDEAS). This analysis platform does provide biologist-friendly templates (wizards) to guide users through common analysis scenarios, including foundational (compensation, gating), application-specific (apoptosis, localization, internalization), and exploratory (feature finder) schemes. Moreover, there are open-source attempts to orchestrate software modules and algorithms to improve automation in analysis procedures, commonly written in Python, MATLAB [63], or R [64.]

### [H2] Sorting

Sorting is an important feature of a cytometric system, regardless of imaging capability, because it allows physical segregation of objects to isolate subpopulations of unique cell types. This can allow subsequent assays on the subpopulations, or valuable procedures such as clonal selection and expansion. Unfortunately, constructing an image-based cell sorter requires several major modernizations in highspeed image acquisition, intelligent data analysis (often machine learning-based), and microscale sorting modules. In contrast to a range of choices for

sorting flow cytometry, only a few sorting Imaging flow cytometry systems have been designed, and these are yet to become commercially available [65 66 67 68].

## [H2] Temporal resolution

In a flow-based system, once the objects flow past the imager, they are either discarded or recollected in a common container. It is not, therefore, readily feasible to enable repeated imaging of the same cell, as seen in time-lapse, slide-based microscopy. The limitation to a single snapshot of each cell also rules against implementation of 3D reconstruction approaches such as confocal sectioning. However, progress has been made in 3D cell image reconstruction using digital holography to produce tomographic flow cytometry[69]. Future development could alleviate snapshot restrictions through implementation of object tracking and unique identification using cellular barcodes [70 71]. New concepts such as the use of spatial-temporal transformations allow the use of photomultiplier detectors[72] which offer the possibility of high-speed acquisition and sorting. Likewise, the use of ultrafast quantitative phase imaging offers the prospect of high speed imaging flow cytometry which can provide label free phenotyping[73].

## [H2] Multi-object interaction

Imaging flow cytometers can capture multiple objects if they appear within the same field of view at the point of acquisition, and can therefore provide information on close-proximity, object interaction. For example, the platform has been used to identify platelet binding to white blood cells [74]. However, complex and/or long-range interactions between multiple objects would be a considerable challenge, if not impossible, to achieve.

# [H1] Outlook

New imaging flow systems are making use of new methods to flow cells past the detectors, developing completely new image capture systems and adding new functionality such as adding cell phenotype sorting. In fact, these exciting new technologies are leading the way in producing new types of assays that cannot be carried out using current technologies and will form the future commercial systems in the very near future. At their heart all these produce a large number of single cell, often multichannel images and therefore the strategies for using imaging flow cytometry data remain similar irrespective of the instrumentation and therefore the analysis examples given here will be easily adaptable to other systems.

Like many imaging systems, Imaging flow cytometry is susceptible to the triangle of imaging constraints— speed, resolution, and sensitivity—improving one parameter causes the others to suffer. These compromises become even more critical as data volumes, velocity and variety of biomedical research increase in the next 5-10 years. Even so, there are certain gaps for improvement in photonics and optics that are likely to improve Imaging flow cytometry systems. Future iterations may bring novel data acquisition and sorting technologies at higher resolution, with higher dimensions (larger 2D/3D FOV, temporal feature availability), while retaining, if not improving, the event capture rate that makes Imaging flow cytometry advantageous over other single-cell imaging platforms.

Equally important will be improvements in data analysis techniques, in which feature stability, model reproducibility, and automation should be prioritized. Even with machine learning-based assistance incorporated in today's workflows, users are still heavily taxed with many iterations of data cleaning and modeling processes, such as quality control checks, manual annotations in supervised learning, normalization of all features to a common base to offset the wide variation in feature value ranges, feature selection to alleviate the curse of dimensionality, feature ranking and combinations to optimize population separations. It would be helpful to see advanced AI methodologies incorporated into a biologist-friendly pipeline to deliver more automated, less supervised, and more reliable classifier/phenotyping models.

Given the ever-increasing levels of information to be captured from single cells, Imaging flow cytometry coupled with machine learning approaches provides a powerful platform for disease fingerprinting in clinical applications. Rare events (such as metastatic cancer cells) may be detected better than by microscopy, and disease states may be detectable that are otherwise invisible to clinicians. With sorting capability, Imaging flow cytometry would prove to be a very useful tool for clinical diagnosis and treatment monitoring, especially for hematological disorders, even without the use of biomarkers [75]. If an intelligent, label-free, sorting Imaging flow cytometry is developed, users might collect sorted cells to allow clonal selection and expansion and do so iteratively to produce an effective cell therapy. Sorting Imaging flow cytometry would excel in pooled screening campaigns, in which multitudes of gene/compound combinations can be tested in an unprecedented throughput. In pool-based format, nucleic acids, CRISPR-ed oligos, small molecules or antibodies are mixed in the microfluidic device into cellular or droplet form, then screened by image-based sorting followed by downstream omic techniques such as next-generation sequencing or proteomics. Novel readouts include combinatorial treatment responses, differential co-expression, network and pathway analyses,

to help discern complex phenotypes and regulatory programs, and subsequently prioritize candidate genes or compounds for biopharmaceutical manufacturing.

# Glossary

**Gates:** a range of bins for the histogram or a polygon for the scatter plot. This process selects cells for further analysis. The gating process can be repeated to define phenotypes which require more than two markers for identification.

**Brightfield image:** The simplest form of microscopy, where the image is formed by white light which is transmitted through the sample and then capture on a detector.

**Darkfield image:** In the context of imaging flow cytometry, the darkfield image is formed when light scattered from the cell is collected on the detector perpendicular to the excitation direction.

**Raw maximum pixel feature:** A feature in IDEAS that returns the maximum pixel value in an image acquired by the detector before any compensation. This is often used to set the laser excitation intensity to ensure that the pixel values are not saturated.

**Mask:** a binary image which defines the extent of the object in an image, the pixel values in the image are 1 inside the object perimeter and 0 elsewhere to represent the background.

**Building blocks:** suggested feature scatter plots and gating strategies to help the user with simple analysis and preprocessing tasks, such as determining in-focus cells in the IDEAS software.

**Multi-spectral images:**.An image dataset in which the same field of view is imaged in different spectral bands.

**Aspect ratio:** the ratio of the minor axis and the major axis. The major axis is the longest line that can be drawn through the shape and the minor axis is the shortest line that can be drawn through the shape at right angles to the major axis.

**AND mask operation:** The AND operator applied to two masks delivers the overlapped shared area between the masks.

**NOT mask operation:** The NOT operator is a logic operator which delivers the inverse of a mask i.e. 0s become 1s.

**Confusion matrix:** A confusion matrix is used to compare the predicted outcome of a machine learning algorithm with the known classes of the data. In general, the rows represent the number of instances of the actual class while the columns represent the number of instances of a predicted class from the algorithm, or vice versa. Therefore the diagonal elements represents the number of correct classifications and off diagonal elements can be used to assess where the algorithms is making misclassifications.

**t-distributed stochastic neighbour embedding:** An algorithm used to visualise high dimensional datasets in two or three dimensions. Nonlinear dimensional reduction of the data to the 2/3D coordinate system is used to preserve the distances between similar and dissimilar data points.

## Acknowledgements

P.R. and H.S. acknowledge the UK Engineering and Physical Sciences Research Council (EP/N013506/1) and UK Biotechnology and Biological Sciences Research Council (BB/P026818/1) for supporting this work. A.C. acknowledges the National Science Foundation (DBI 1458626) and the National Institutes of Health (R35 GM122547) for supporting this work.

## Author contributions

Introduction (H.S., A.C., M.D., P.R.); Experimentation (H.S., A.C., P.R.); Results (H.S., A.F., A.C., P.R.); Applications (H.S., A.F., A.C., P.R.); Reproducibility and data deposition (H.S., A.F., A.C., P.R.); Limitations and optimizations (H.S., A.C., M.D., P.R.); Outlook (H.S., A.C. M.D., P.R.).

## Competing interests

The authors declare no competing interests.

## Supplementary information

Supplementary information is available for this paper at https://doi.org/10.1038/s415XX-XXX-XXXX-X

# Figure legends

**Figure 1: Overview of imaging flow cytometry and images generated (**a) Diagram of the optical layout of the Imagestream flow cytometer. (b) Typical brightfield, darkfield and fluorescent images and masks obtained from the ImageStream system (c) Example of cell phenotyping using multiple CD markers using the ImageStream system

**Figure 2: Process flow employed to select in-focus, single cell images from an acquired event set.** Using a histogram of the root-mean-squared pixel values from the masked brightfield image the higher values are gated to determine in-focus cells. We note that the image which belongs in yellow bin is blurred and removed to the gating choice. Using these gated cells, a scatter plot of aspect ratio of the brightfield mask versus the area of the brightfield mask is then used to further gate a population of objects with medium area and high aspect ratio which removes cell clusters (top cell image) and speed beads (bottom image) or debris.

**Figure 3: Data analysis based on spatial information. [Au: please add a title for the figure]** Histogram of the number of nanoparticle loaded vesicles (NLV) in a cell population U2-OS cells under exposure of 1nM Qtracker705 particles for 1 hour. The distribution exhibits over-dispersion relative to a Poisson process (dotted line) with accurate representation of the data being achieved using a negative binomial distribution function (solid red line). Typical cell images together with the masks used are also show, the scale bar denotates 10μm.

**Figure 4: Spatial analysis based on area masking. [Au: please add a title for the figure]** (a) Differentiation of cell populations with membrane-associated or dispersed granules, according to mask area. A cell perimeter mask was generated using the brightfield mask AND NOT brightfield eroded mask – these are effectively the pixels that were removed by the eroding operation. The membrane associated granules are now determined by the overlap of the perimeter mask with the dark spot mask for example perimeter mask AND dark spot mask. (b) Histogram of the dark spot overlap with the perimeter mask, typical brighfield images and dark spot masks are shown for two typical histogram bins.

**Figure 5: Spatial analysis based on morphology.** Differentiation of cell populations with membrane-associated or dispersed granules, according to the morphology of their spatial distribution.

# Selected references


**Introduction to ImageStream**

Basiji, D. A., Ortyn, W. E., Liang, L., Venkatachalam, V. & Morrissey, P. Cellular Image Analysis and Imaging by Flow Cytometry. *Clinics in Laboratory Medicine* **27**, 653-670, doi:10.1016/j.cll.2007.05.008 (2007).

**Compensation of imaging flow cytometry data**

Ortyn, W. E. *et al.* Sensitivity measurement and compensation in spectral imaging. *Cytometry A* **69**, 852-862, doi:10.1002/cyto.a.20306 (2006).

**Cell cycle analysis using imaging flow cytometry using image properties**

Calvert, M. E. K., Lannigan, J. A. & Pemberton, L. F. Optimization of yeast cell cycle analysis and morphological characterization by multispectral imaging flow cytometry. *Cytometry Part A* **73A**, 825833, doi:10.1002/cyto.a.20609 (2008).

**Discussion of masks in imaging flow cytometry**

Dominical, V., Samsel, L. & McCoy, J. P., Jr. Masks in imaging flow cytometry. *Methods* **112**, 9-17, doi:10.1016/j.ymeth.2016.07.013 (2017).

**Application of machine learning to imaging flow data**

Blasi, T. *et al.* Label-free cell cycle analysis for high-throughput imaging flow cytometry. *Nat. Commun.* **7**, 10256, doi:10.1038/ncomms10256 (2016).

**Application of deep learning to imaging flow data**

Eulenberg, P. *et al.* Reconstructing cell cycle and disease progression using deep learning. *Nat. Commun.* **8**, 463, doi:10.1038/s41467-017-00623-3 (2017).

**Image based flow cytometry sorting**

Nitta, N. *et al.* Intelligent Image-Activated Cell Sorting. *Cell* **175**, 266-276.e213, doi:10.1016/j.cell.2018.08.028 (2018).

# References


1 Brown, M. & Wittwer, C. Flow Cytometry: Principles and Clinical Applications in Hematology. *Clinical Chemistry* **46**, 1221-1229, doi:10.1093/clinchem/46.8.1221 (2000).

2 De Rosa, S. C., Brenchley, J. M. & Roederer, M. Beyond six colors: a new era in flow cytometry. *Nat. Med.* **9**, 112-117, doi:10.1038/nm0103-112 (2003).

3 Basiji, D. A., Ortyn, W. E., Liang, L., Venkatachalam, V. & Morrissey, P. Cellular Image Analysis and Imaging by Flow Cytometry. *Clinics in Laboratory Medicine* **27**, 653-670, doi:10.1016/j.cll.2007.05.008 (2007).

4 Luminex. <https://www.luminexcorp.com/eu/amnis-imagestreamxmk-ii/#overview> (

5 Maguire, O., Collins, C., O'Loughlin, K., Miecznikowski, J. & Minderman, H. Quantifying nuclear p65 as a parameter for NF-κB activation: Correlation between ImageStream cytometry, microscopy, and Western blot. *Cytometry Part A* **79A**, 461-469, doi:10.1002/cyto.a.21068 (2011).

6 Kaur, M. & Esau, L. Two-step protocol for preparing adherent cells for high-throughput flow cytometry. *Biotechniques* **59**, 119-126, doi:10.2144/000114325 (2015).

7 Stavrakis, S., Holzner, G., Choo, J. & deMello, A. High-throughput microfluidic imaging flow cytometry. *Curr Opin Biotechnol* **55**, 36-43, doi:10.1016/j.copbio.2018.08.002 (2019).

8 Holzner, G. *et al.* High-throughput multiparametric imaging flow cytometry: toward diffraction-limited sub-cellular detection and monitoring of sub-cellular processes. *Cell Rep* **34**, 108824, doi:10.1016/j.celrep.2021.108824 (2021).

9 Mikami, H. *et al.* Virtual-freezing fluorescence imaging flow cytometry. *Nat Commun* **11**, 1162, doi:10.1038/s41467-020-14929-2 (2020).

10 E. D. Diebold *et al*., Digitally synthesized beat frequency multiplexing for sub-millisecond fluorescence microscopy, *Nature Photonics* **7**, 806-810 (2013).

11 George, T. C. *et al.* Distinguishing modes of cell death using the ImageStream multispectral imaging flow cytometer. *Cytometry A* **59**, 237-245, doi:10.1002/cyto.a.20048 (2004).

12 Ortyn, W. E. *et al.* Extended depth of field imaging for high speed cell analysis. *Cytometry A* **71**, 215-231, doi:10.1002/cyto.a.20370 (2007).

13 Haley Renee, P. & Raymond, K. K. Demonstration of High Gain mode in combination with Imaging Flow Cytometry for improved EV analysis. *Mian Yi Xue Za Zhi* **204**, 15 (2020).

14 Vogt, R. F., Jr., Whitfield, W. E., Henderson, L. O. & Hannon, W. H. Fluorescence intensity calibration for immunophenotyping by flow cytometry. *Methods* **21**, 289-296, doi:10.1006/meth.2000.1009 (2000).

15 Ortyn, W. E. *et al.* Sensitivity measurement and compensation in spectral imaging. *Cytometry A* **69**, 852-862, doi:10.1002/cyto.a.20306 (2006).

16 Fazekas de St Groth, B., Zhu, E., Asad, S. & Lee, L. Flow cytometric detection of human regulatory T cells. *Methods Mol. Biol.* **707**, 263-279, doi:10.1007/978-1-61737-979-6_17 (2011).

17 Holmberg-Thyden, S., Grønbæk, K., Gang, A. O., El Fassi, D. & Hadrup, S. R. A user's guide to multicolor flow cytometry panels for comprehensive immune profiling. *Anal. Biochem.* **627**, 114210, doi:10.1016/j.ab.2021.114210 (2021).

18 Dominical, V., Samsel, L. & McCoy, J. P., Jr. Masks in imaging flow cytometry. *Methods* **112**, 917, doi:10.1016/j.ymeth.2016.07.013 (2017).

19 Barteneva, N. S. & Vorobjev, I. A. *Imaging Flow Cytometry: Methods and Protocols*. (Humana Press, 2015).

20 Patterson, J. O., Swaffer, M. & Filby, A. An Imaging Flow Cytometry-based approach to analyse the fission yeast cell cycle in fixed cells. *Methods* **82**, 74-84, doi:10.1016/j.ymeth.2015.04.026 (2015).

21 Patterson, J. O., Basu, S., Rees, P. & Nurse, P. CDK control pathways integrate cell size and ploidy information to control cell division. *Elife* **10**, doi:10.7554/eLife.64592 (2021).

22 Patterson, J. O., Rees, P. & Nurse, P. Noisy Cell-Size-Correlated Expression of Cyclin B Drives Probabilistic Cell-Size Homeostasis in Fission Yeast. *Curr. Biol.* **29**, 1379-1386.e1374, doi:10.1016/j.cub.2019.03.011 (2019).
23 Calvert, M. E. K., Lannigan, J. A. & Pemberton, L. F. Optimization of yeast cell cycle analysis and morphological characterization by multispectral imaging flow cytometry. *Cytometry Part A* **73A**, 825-833, doi:10.1002/cyto.a.20609 (2008).
24 Filby, A. *et al.* An imaging flow cytometric method for measuring cell division history and molecular symmetry during mitosis. *Cytometry A* **79**, 496-506, doi:10.1002/cyto.a.21091 (2011).
25 Summers, H. D. *et al.* Statistical analysis of nanoparticle dosing in a dynamic cellular system. *Nature Nanotechnology* **6**, 170-174, doi:10.1038/nnano.2010.277 (2011).
26 Bourton, E. C. *et al.* Multispectral imaging flow cytometry reveals distinct frequencies of γ-H2AX foci induction in DNA double strand break repair defective human cell lines. *Cytometry Part A* **81A**, 130-137, doi:10.1002/cyto.a.21171 (2012).
27 Jurgielewicz, B. J., Yao, Y. & Stice, S. L. Kinetics and Specificity of HEK293T Extracellular Vesicle Uptake using Imaging Flow Cytometry. *Nanoscale Res. Lett.* **15**, 170, doi:10.1186/s11671-020-03399-6 (2020).
28 Fei, C., Lillico, D. M. E., Hall, B., Rieger, A. M. & Stafford, J. L. Connected component masking accurately identifies the ratio of phagocytosed and cell-bound particles in individual cells by imaging flow cytometry. *Cytometry A* **91**, 372-381, doi:10.1002/cyto.a.23050 (2017).
29 Vranic, S. *et al.* Deciphering the mechanisms of cellular uptake of engineered nanoparticles by accurate evaluation of internalization using imaging flow cytometry. *Part. Fibre Toxicol.* **10**, 2, doi:10.1186/1743-8977-10-2 (2013).
30 Chivukula, R. R. *et al.* A human ciliopathy reveals essential functions for NEK10 in airway mucociliary clearance. *Nat. Med.* **26**, 244-251, doi:10.1038/s41591-019-0730-x (2020).
31 Blasi, T. *et al.* Label-free cell cycle analysis for high-throughput imaging flow cytometry. *Nat. Commun.* **7**, 10256, doi:10.1038/ncomms10256 (2016).
32 Nassar, M. *et al.* Label-Free Identification of White Blood Cells Using Machine Learning. *Cytometry A* **95**, 836-842, doi:10.1002/cyto.a.23794 (2019).
33 Lippeveld, M. *et al.* Classification of Human White Blood Cells Using Machine Learning for Stain-Free Imaging Flow Cytometry. *Cytometry A* **97**, 308-319, doi:10.1002/cyto.a.23920 (2020).
34 Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100, doi:10.1186/gb-2006-7-10-r100 (2006).
35 Hennig, H. *et al.* An open-source solution for advanced imaging flow cytometry data analysis using machine learning. *Methods* **112**, 201-210, doi:10.1016/j.ymeth.2016.08.018 (2017).
36 Ding, C. & Peng, H. Minimum redundancy feature selection from microarray gene expression data. *Computational Systems Bioinformatics. CSB2003. Proceedings of the 2003 IEEE Bioinformatics Conference. CSB2003*, doi:10.1109/csb.2003.1227396.
37 Mandal, M. & Mukhopadhyay, A. An Improved Minimum Redundancy Maximum Relevance Approach for Feature Selection in Gene Expression Data. *Procedia Technology* **10**, 20-27, doi:10.1016/j.protcy.2013.12.332 (2013).
38 Eulenberg, P. *et al.* Reconstructing cell cycle and disease progression using deep learning. *Nat. Commun.* **8**, 463, doi:10.1038/s41467-017-00623-3 (2017).

39 Dunker, S., Boho, D., Wäldchen, J. & Mäder, P. Combining high-throughput imaging flow cytometry and deep learning for efficient species and life-cycle stage identification of phytoplankton. *BMC Ecol.* **18**, 51, doi:10.1186/s12898-018-0209-5 (2018).
40 Dunker, S. *et al.* Pollen analysis using multispectral imaging flow cytometry and deep learning. *New Phytol.* **229**, 593-606, doi:10.1111/nph.16882 (2021).
41 Luo, S. *et al.* Deep learning-enabled imaging flow cytometry for high-speed Cryptosporidium and Giardia detection. *Cytometry A* **99**, 1123-1133, doi:10.1002/cyto.a.24321 (2021).

42 Rees, P., Wills, J. W., Brown, M. R., Barnes, C. M. & Summers, H. D. The origin of heterogeneous nanoparticle uptake by cells. *Nat. Commun.* **10**, 2341, doi:10.1038/s41467-019-10112-4 (2019).
43 Cerveira, J., Begum, J., Di Marco Barros, R., van der Veen, A. G. & Filby, A. An imaging flow cytometry-based approach to measuring the spatiotemporal calcium mobilisation in activated T cells. *J. Immunol. Methods* **423**, 120-130, doi:10.1016/j.jim.2015.04.030 (2015).
44 Piasecka, J. *et al.* Diffusion mapping of eosinophil-activation state. *Cytometry* A **97A**, 253-258, doi: 10.1002/cyto.a.23884 (2019).
45 Rodrigues, M. A., Beaton-Green, L. A., Kutzner, B. C. & Wilkins, R. C. Automated analysis of the cytokinesis-block micronucleus assay for radiation biodosimetry using imaging flow cytometry. *Radiat. Environ. Biophys.* **53**, 273-282, doi:10.1007/s00411-014-0525-x (2014).
46 Rodrigues, M. A., Probst, C. E., Beaton-Green, L. A. & Wilkins, R. C. Optimized automated data analysis for the cytokinesis-block micronucleus assay using imaging flow cytometry for high throughput radiation biodosimetry. *Cytometry A* **89**, 653-662, doi:10.1002/cyto.a.22887 (2016).
47 Verma, J. R. *et al.* Evaluation of the automated MicroFlow® and MetaferTM platforms for high-throughput micronucleus scoring and dose response analysis in human lymphoblastoid TK6 cells. *Archives of Toxicology* **91**, 2689-2698, doi:10.1007/s00204-016-1903-8 (2017).
48 Rodrigues, M. A. Automation of the in vitro micronucleus assay using the Imagestream imaging flow cytometer. *Cytometry A* **93**, 706-726, doi:10.1002/cyto.a.23493 (2018).
49 Rodrigues, M. A. *et al.* The in vitro micronucleus assay using imaging flow cytometry and deep learning. *npj Systems Biology and Applications* **7**, doi:10.1038/s41540-021-00179-5 (2021).
50 Wills, J. W. *et al.* Inter-laboratory automation of the in vitro micronucleus assay using imaging flow cytometry and deep learning. *Arch. Toxicol.* **95**, 3101-3115, doi:10.1007/s00204-021-03113-0 (2021).
51 Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM* **60**, 84-90, doi:10.1145/3065386 (2017).
52 He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, doi:10.1109/cvpr.2016.90 (2016).
53 Karen Simonyan, A. Z. *Website*, <http://dx.doi.org/10.48550/arXiv.1409.1556> (
54 Mochalova, E. N., Kotov, I. A., Lifanov, D. A., Chakraborti, S. & Nikitin, M. P. Imaging flow cytometry data analysis using convolutional neural network for quantitative investigation of phagocytosis. *Biotechnol. Bioeng.* **119**, 626-635, doi:10.1002/bit.27986 (2022).
55 Probst, C., Zayats, A., Venkatachalam, V. & Davidson, B. Advanced Characterization of Silicone Oil Droplets in Protein Therapeutics Using Artificial Intelligence Analysis of Imaging Flow Cytometry Data. *J. Pharm. Sci.* **109**, 2996-3005, doi:10.1016/j.xphs.2020.07.008 (2020).
(
58 Spidlen, J., Breuer, K. & Brinkman, R. Preparing a Minimum Information about a Flow Cytometry Experiment (MIFlowCyt) compliant manuscript using the International Society for Advancement of Cytometry (ISAC) FCS file repository (FlowRepository.org). *Curr. Protoc. Cytom.* **Chapter 10**, Unit 10.18, doi:10.1002/0471142956.cy1018s61 (2012).

60 Filby, A. & Davies, D. Reporting imaging flow cytometry data for publication: why mask the detail? *Cytometry A* **81**, 637-642, doi:10.1002/cyto.a.22091 (2012).
61 Caicedo, J. C. *et al.* Data-analysis strategies for image-based cell profiling. *Nat Methods* **14**, 849-863, doi:10.1038/nmeth.4397 (2017).
62 Montero Llopis, P. *et al.* Best practices and tools for reporting reproducible fluorescence microscopy methods. *Nat Methods* **18**, 1463-1476, doi:10.1038/s41592-021-01156-w (2021).
63 Doan, M. *et al.* Deepometry, a framework for applying supervised and weakly supervised deep learning to imaging cytometry. *Nat. Protoc.* **16**, 3572-3595, doi:10.1038/s41596-021-00549-7 (2021).
64 Demont, Y. Tools for Imaging Flow Cytometry [R package IFC version 0.1.2]. (2021).

65 Schraivogel, D. *et al.* High-speed fluorescence image–enabled cell sorting. *Science* **375**, 315320, doi:10.1126/science.abj3013 (2022).
66 Gu, Y. *et al.* Machine Learning Based Real-Time Image-Guided Cell Sorting and Classification. *Cytometry A* **95**, 499-509, doi:10.1002/cyto.a.23764 (2019).
67 Nitta, N. *et al.* Intelligent Image-Activated Cell Sorting. *Cell* **175**, 266-276.e213, doi:10.1016/j.cell.2018.08.028 (2018).
68 Nawaz, A. A., et al., "Intelligent image-based deformation-assisted cell sorting with molecular specificity," Nature Methods 17, 595-599 (2020).
69 Merola, F., *et al.,* "Tomographic flow cytometry by digital holography," Light: Science & Applications **6**, e16241 (2017).
70 Kebschull, J. M. & Zador, A. M. Cellular barcoding: lineage tracing, screening and beyond. *Nat. Methods* **15**, 871-879, doi:10.1038/s41592-018-0185-x (2018).
71 Rees, P. *et al.* Nanoparticle vesicle encoding for imaging and tracking cell populations. *Nat. Methods* **11**, 1177-1181, doi:10.1038/nmeth.3105 (2014).
72 Han, Y. and Lo, W.-H., Imaging cells in flow cytometer using spatial-temporal transformation, *Scientific Reports* **5**, 13267 (2015).
73 K. C. M. Lee *et al.,* Quantitative phase imaging flow cytometry for ultra-large-scale single-cell biophysical phenotyping, *Cytometry* **A** 95, 510-520 (2019).
74 Ouk, C., Jayat-Vignoles, C., Donnard, M. & Feuillard, J. Both CD62 and CD162 antibodies prevent formation of CD36-dependent platelets, rosettes, and artefactual pseudoexpression of platelet markers on white blood cells: a study with ImageStream®. *Cytometry A* **79**, 477-484, doi:10.1002/cyto.a.21050 (2011).
75 Ugawa, M. *et al.* In silico-labeled ghost cytometry. doi:10.7554/eLife.67660 (2021).
76 Rosenburg, C.A., Exploring dyserythropoiesis in patients with myelodysplasticsyndrome by imaging flow cytometry and machine-learning assisted morphometrics. *Cytometry* **100**, 554–567 (2021)