

UC San Diego

UC San Diego Previously Published Works

Title

Machine Learning Based Real-Time Image-Guided Cell Sorting and Classification.

Permalink

<https://escholarship.org/uc/item/1q22f180>

Journal

Cytometry Part B: Clinical Cytometry, 95(5)

Authors

Gu, Yi

Zhang, Alex

Han, Yuanyuan

et al.

Publication Date

2019-05-01

DOI

10.1002/cyto.a.23764

Peer reviewed



Published in final edited form as:

Cytometry A. 2019 May ; 95(5): 499–509. doi:10.1002/cyto.a.23764.

Machine Learning Based Real-Time Image-Guided Cell Sorting and Classification

Yi Gu^a, Alex Ce Zhang^a, Yuanyuan Han^a, Jie Li^b, Clark Chen^b, and Yu-Hwa Lo^{a,1}

^aDepartment of Electrical and Computer Engineering, University of California, San Diego, La Jolla, California 92093-0407, USA

^bDepartment of Neurosurgery, University of Minnesota, Minneapolis, D429 Mayo Memorial Building, 420 Delaware Street S.E., MMC 96MN 55455, USA

Abstract

Cell classification based on phenotypical, spatial, and genetic information greatly advances our understanding of the physiology and pathology of biological systems. Technologies derived from next generation sequencing (NGS) and fluorescent activated cell sorting (FACS) are cornerstones for cell- and genomic-based assays supporting cell classification and mapping. However, there exists a deficiency in technology space to rapidly isolate cells based on high content image information. Fluorescence-activated cell sorting can only resolve cell-to-cell variation in fluorescence and optical scattering. Utilizing microfluidics, photonics, computation microscopy, real-time image processing and machine learning, we demonstrate an image-guided cell sorting and classification system possessing the high throughput of flow cytometer and high information content of microscopy. We demonstrate the utility of this technology in cell sorting based on 1) nuclear localization of glucocorticoid receptors, 2) particle binding to the cell membrane, and 3) DNA damage induced γ -H2AX foci.

INTRODUCTION

There are a far greater number of cell types than people realized in the past, and classifying cells from healthy and diseased tissues in much finer detail than before can bring significant insight in biology and medicine. While sequencing of single cells becomes the technology cornerstone for cell classification, selection of these single cells for genomic analyses rely on fluorescence activated cell sorting (FACS) systems [1,2]. A small biological sample can contain millions of cells, hence analyzing even as many as 100,000 single cells represent only a very small percentage of cells in the sample. Thus intelligent selection of this small percentage of cells for downstream analysis is critical to efficient and accurate cell classification. However, today's cell selection techniques are purely based on fluorescent biomarkers and/or light scattering intensity, without resorting to high content image information that has the most distinctive power to support smart and logical selection of cells, especially those rare cells and cells without known or unique biomarkers.

¹To whom correspondence should be addressed. ylo@ucsd.edu.

Using machine learning and other innovative techniques, we demonstrate an image-guided flow cytometer cell sorter. The availability of flow cytometers with the capability to classify and isolate cells guided by high-content cell images is enabling and transformative[3]. It provides a new paradigm to allow researchers and clinicians to isolate cells using multiple user-defined characteristics encoded by both fluorescent signals and morphological and spatial features.

Examples of applications include isolation of cells based on organelle translocation, cell cycle, detection and counting of phagocytosed particles, and protein co-localization, to name a few.[4–7] Some specific applications include translocation of glucocorticoid receptor (GR) from cytosol to nucleus under dexamethasone treatment[8], glucocorticoid receptor and sequential p53 activation by drug mediated apoptosis[9], and translocation of protein kinase C (PKC) from cytosol to membrane in the context of oncogenesis[10]. β -arrestin-GFP is often used to measure the internalization (inactivation) of g-protein coupled receptors (GPCRs) as β -arrestin-GFP moves from cytosol to membrane. The ~800 GPCRs include the opioid receptors (heroin, morphine, pain pills), the dopamine receptors (cocaine, methamphetamine, addiction/reward), and hundreds of others, many awaiting discovery or “adoption” of ligands. Other specific application examples include immunology studies of B-cell or T-cell responses to various drug treatments, asymmetric B-cell division in the germinal center reaction [11,12], the erythroblast enucleation process, signaling and cytoskeletal requirements in erythroblast enucleation [13,14], uptake and internalization of exosomes by various cancer cells, response of infected cells to drugs, use of antibody-drug conjugates for tracking drugs in/outside sub-cellular compartments, and locating antigens, enzymes or other molecules [15–19].

The reported machine learning based real-time image-guided cell sorting and classification technology possesses the high throughput of flow cytometer and high information content of microscopy, being able to isolate cells according to their imaging features at 1000X faster rate than laser microdissection[20] and single-cell aspiration[21]. We have applied a microfluidic platform and a spatial-temporal transformation method [22–24] to acquire cell images in real time with extremely simple hardware. We also developed a methodology of user-interface (UI) to generate sorting criteria by supervised machine learning, as described next.

After hundreds of cells pass through the imaging flow cytometer, the software generates a distribution of cell parameters, as well as several categories of cell images based on the built-in image processing and statistical classification algorithms. Users then apply point-and-click selection of desired cell images for the basis of gating the cells for sorting from the sample. After collecting an additional number of cells based on user’s instructions, the software displays both the conventional flow cytometer parameters (i.e. fluorescence intensity) and a new set of image/morphology related parameters (e.g. nucleus size, cell area, circularity, fluorescence patterns, etc.), as well as the representative cell images of the cells. This iterative feedback process gives users the chance to confirm their original choice criteria and to modify the “gating”. Based on the displayed image and conventional data feedback, users may adjust the gating criteria. These criteria can be “ratio of fluorescence area over the total cell area”, “variations of fluorescence intensity profile over the cell”, “size of

nucleus”, or numerous other choices utilizing the spatial features of the cells. As a result, the image-guided cell selection process becomes a user-interface (UI) and user-experience (UX) interactive process with machine learning occurring in the background to present users with representative images of cell classes that most closely match the user needs and even suggest features possibly overlooked by users. As a result, users are given unprecedented intuitive visual assistance and insight to enhance their studies.

To demonstrate the above concepts and enabling capabilities, we report three experiments in this paper: (1) sorting of pEGFP-GR plasmids translocated HEK-297T Human Embryonic Kidney Cells based on intracellular protein distribution, (2) sorting of MDCK Madin-Darby Canine Kidney Epithelial Cells based on the number of particles bound to cell membrane, and (3) sorting of Human Glioblastoma Cells by extent of radiation-induced DNA damage. In all these experiments, we have set the flow speed of cells at 8cm/s in a microfluidic device, and used a throughput of 100 cells per second, limited by the processor speed of our current electronic (FPGA) hardware which can be upgraded by at least 5 fold with GPUs.

METHODS AND MATERIALS

System design of machine learning based real-time image-guided cell sorter

As shown in Figure 1 (a), the image-guided cell sorting and classification system consists of: (1) an imaging optical system with a spatially coded optical filter to perform spatial-temporal transformation, (2) a real-time image processing and feature extraction module, (3) an off-line post processing module for construction of cell images for human vision and generation of cell classification criteria, and (4) a microfluidic chip integrated with an on-chip piezoelectric (PZT) cell sorting actuator.

The optical filter encodes the fluorescent or light scattering signal of a cell into a temporal photocurrent waveform in the output of a photomultiplier tube (PMT) detector. Through a mathematical transformation described in [22,25], the 1D time domain signal is transformed into a 2D cell image. Due to the simplicity of the transformation algorithm, real-time signal processing can be implemented to extract image features of each cell to allow cell sorting based on these image features.

Machine learning is needed to generate and adjust image features to guide cell sorting [26–28]. To start, training samples are flowed through the system to produce a set of training data. Off-line processing is employed to construct high resolution, cell images to interface with users whose inputs will aid the selection and adaptation of cell classification criteria for real-time sorting, a method of supervised machine learning. During the real-time cell sorting experiments, real-time processing module reconstructs cell images, extracts image features and makes sorting decisions based on the off-line trained sorting criteria. When a decision is made to sort a cell, a voltage pulse is applied to the on-chip PZT actuator, which instantaneously bends the bimorph PZT disk to deflect the cell away from the central flow into the sorting channel. For a proof-of-concept prototype, we have used a field-programmable-gate-array (FPGA) platform to implement real-time image processing which produces a latency of a few milliseconds. Higher performance GPU processors can reduce the processing time by 100 folds or more to microseconds.

Optical imaging setup

In the optical imaging system, suspended single cells flow in a microfluidic channel made of soft-molded polydimethylsiloxane (PDMS) bonded to a glass substrate. Sheath flow is used to hydro-dynamically focus the travelling cells to the center of the microfluidic channel. At the interrogation zone, each flowing cell is illuminated simultaneously by a 500mW 455nm LED (Thorlabs) and a 100mW 488nm laser (iBeam-SMART, Toptica) to generate bright field and fluorescent images. The output beam of 488nm laser is collimated, focused and expanded to illuminate a 100 μ m (x-direction) by 250 μ m (y-direction) area. The LED light is collimated and focused at the laser illumination area. Both the fluorescent emissions and the transmitted bright field signal are detected by PMTs (H9307-02, Hamamatsu). To accommodate the geometry of the microfluidic device, the laser beam is introduced to the optical interrogation area by a 52-degree miniature dichroic mirror positioned in front of a 50X objective lens (NA=0.55, working distance=13 mm, Mituyoyo). The LED is placed at the opposite side of the channel and the light is focused to the same position as the laser beam. The spatially coded optical filter is inserted at the image plane in the detection path. The pattern of the filter is shown in Figure 1 (b). With the spatial filter, fluorescence/scattering from different parts of the cell will pass different slits at different times. As a result, the waveform of the fluorescent/scattering signal from the PMT consists of a sequence of patterns separated in time domain, and each section of the signal in the time domain corresponds to the fluorescent/scattering signal generated by each particular segment of the cell. After the light intensity profile over each slit is received, the cell image of the entire cell can be constructed by splicing all the profile together. The image resolution in x- (transverse to the flow) direction is primarily determined by the number of slits on the spatial filter, and the resolution in y- (flow) direction is mainly determined by the sampling rate and cell flow speed. In our system, the raw image resolution is 2 μ m in x-direction and 0.4 μ m in y-direction.

Dichroic mirrors are used to route the desired emission bands to their respective PMTs.

Real-Time image processing

The real-time image processing module is implemented in a field-programmable-gate-array (FPGA) platform (National Instrument cRIO-9039). The processing module performs the functions of cell detection and image processing, to be discussed next.

Both the bright field and fluorescent images of a cell are reconstructed from respective PMT readouts. Each image covers a field-of-view of 20X20 μ m² with 50X50 pixels, matched to the finest spatial resolution achievable by the system. All relevant image features are calculated and compared against the cell selection (gating) criteria to make sorting decision.

Cell detection algorithm

The cell detection function determines whether a cell exists within a certain time interval, and subsequently instructs the system whether to store and transmit the signal over this time interval. The system records the outputs from PMTs in the First-in First-out (FIFO) data structure over a chosen length of time. Each time a new set of PMT readout enters the FIFO, cell detection algorithm is activated to determine whether there is a cell within the optical

system's field of view. As soon as the system detects the presence of cell within the data set, the system processes the data immediately to construct images and extract image features. Otherwise, the system continues to examine the next set of PMT readout.

Since fluorescent and bright-field images are generated simultaneously, only one fluorescent signal is used in cell detection algorithm. To shorten the processing time, we calculate the *Brightness* of signal by integrating the fluorescent intensity stored in the FIFO. The cell detection algorithm is shown in Supplementary Figure 1.

First, the *Brightness* is compared to a preset threshold defined as *Threshold1*. If the *Brightness* is greater than *Threshold1*, the system assumes a cell is entering the field of view. The calculated *Brightness* enters a FIFO data structure named $FIFO_{Brightness}$. The length of $FIFO_{Brightness}$ is 100.

Second, the time derivative of *Brightness* is evaluated to check if the cell is within the field of view. The time derivative of *Brightness* can be represented by the quantity of "difference", defined as $(\max_{Brightness} - \min_{Brightness}) / \max_{Brightness}$, where $\max_{Brightness}$ and $\min_{Brightness}$ are the maximum and minimum elements in $FIFO_{Brightness}$. If the "difference" is smaller than a preset *Threshold2* (e.g. *Threshold2* = 5%), the system considers the cell is within the field of view, and the function of real-time image processing is activated. The values of $\max_{Brightness}$ and $\min_{Brightness}$ are updated each time a new element enters $FIFO_{Brightness}$. The algorithm to calculate $\max_{Brightness}$ and $\min_{Brightness}$ is shown in Supplementary Figure 2.

Third, *Brightness* is compared to *Threshold1* again as time goes on. If the *Brightness* falls below *Threshold1* at some point of time, the algorithm determines the cell has left the field of view.

Brightness is calculated from the PMT readout stored in the FIFO data structure. The FIFO for fluorescent signal is referred as $FIFO_{PMT}$. Each time when a new PMT readout enters $FIFO_{PMT}$, *Brightness* value is updated by including the latest PMT readout and excluding the dequeued element in $FIFO_{PMT}$. *Brightness* is calculated by equation 1.

$$Brightness[t_n] = Brightness[t_{n-1}] + Voltage_{PMT} - FIFO_{PMT}[dequeue] \quad (1)$$

where $Brightness[t_n]$ is the updated *Brightness*, $Brightness[t_{n-1}]$ is *Brightness* of last time step, $Voltage_{PMT}$ is the incoming PMT readout, which is the enqueued element of $FIFO_{PMT}$, and $FIFO_{PMT}[dequeue]$ is the dequeued element in $FIFO_{PMT}$.

Image processing algorithm

For all applications, we essentially follow the same flow for real time image processing, involving denoising, image resizing, contour definition, area calculation, feature enumeration, etc. Some of the processes can run in parallel to simultaneously extract

multiple image derived features pertinent to image-guided sorting. In the following, we depict the specific processes applicable to each specific experiment.

Real time image processing algorithm for protein translocation experiment.

The image processing algorithm is illustrated in Supplementary Figure 3. The image processing algorithm includes the following steps: (1) Denoise PMT signals with a 10th-order Hamming low-pass filter. (2) Reconstruct both bright-field and fluorescent images from PMT signals. Since both bright-field and fluorescent signals are generated by the same slit although from different light sources, they are well synchronized. Thus the image reconstruction algorithm only needs to be launched once for both bright-field and fluorescent images. (3) Resize the images from 10×50 pixels (due to asymmetric resolution in raw images) to 50×50 pixels. (4) Detect contours of cell images by first converting the grayscale images to binary images, then eliminating spurious noise with open filter, and finally applying the contour detection algorithm to the binary images. (5) Extract all image-derived parameters.

Examples of image-derived parameters that can be extracted from each cell in real time are shown in Supplementary Table 1. As discussed later in this paper, these image-derived parameters are evaluated with Receiver Operating Characteristics (ROC). The 3 parameters receiving the highest ROC score are used as default parameters for real-time image-guided sorting. The latency of real-time processing algorithm is 5.8ms/cell with the current FPGA system, and the latency can be reduced by over 100 times with high performance GPU (e.g. NVIDIA QUADRO P6000).

Real time image reconstruction and speed detection.—An optical spatial filter consisting of 10 slits is placed at the image plane of the signal. With a 50X (or 20X if desired for extended focal depth) objective lens, the image projected onto the optical spatial filter is 50X (20X) times greater than the object in the microfluidic channel. For each cell travelling in the microfluidic channel, its PMT readout produces 10 peaks, each corresponding to the cell's fluorescent or bright-field transmitted signal passing one of 10 slits on the spatial filter. At a cell travel speed of 8 cm/s and at 200 kSamples/s, each peak consists of 50 sampling points. To reconstruct cell images, two factors need to be considered. First, since cells do not travel at a perfectly uniform speed, the actual number of sampling points for each of 10 peaks may vary slightly. Second, since both cell travelling speed and cell position within the 20 μm by 20 μm image area can vary, the starting time point of the PMT readout also varies. In the image reconstruction algorithm, we refer “m” to be the starting point of PMT readout, “n” to be the number of sampled points in each peak. Based on cell speed variations, n ranges from 46 to 51. This leads to a range of m from 0 to 519–10n. The algorithm sweeps m and n to assure the best combination of (m,n) to reconstruct the cell image. Summation of intensities at the starting point of each peak is calculated for every (m,n) combination. The combination yielding the smallest sum is the right (m,n) which we use to reconstruct the image. Travelling speed of the cell can be detected based on the calculated value “n”. The algorithm is shown in equation 2 and Supplementary Figure 4. In Supplementary Figure 4, The black “*” are starting points for each peak found by the reconstruction algorithm.

$$(m, n) = \left\{ (m, n) \left| \min \sum_{i=0}^{10} \sum_{m=0}^{519-10n} FIFO_{PMT1}[m+i \times n] \right. \right\} \quad (2)$$

After image reconstruction, grayscale cell images are resized to 50×50 pixel images by linear interpolation. Then resized grayscale images are converted to binary images based on the preset intensity threshold. The conversion is described as follows:

$$binary_image(i, j) = \begin{cases} 1, & \text{if } grayscale_image(i, j) > threshold \\ 0, & \text{if } grayscale_image(i, j) < threshold \end{cases} \quad (3)$$

where (i, j) refers to the pixel located at row i and column j .

In the step of open filter, we use a 3×3 pixel neighborhood in our image processing algorithm.

Real time image contour detection.—As shown in Supplementary Figure 5, in the contour finding algorithm all the pixels in a binary cell image are scanned. For those pixels of non-zero value, the algorithm checks all eight pixels surrounding the center pixel in a 3×3 matrix. If the number of non-zero neighboring pixels is between 1 and 7, then this pixel is determined to be on the cell contour. Otherwise, the pixel is either inside or outside the contour. The criteria can be described in (4):

$$image_contour(i, j) = \begin{cases} 1, & \text{if } 0 < n < 8 \ \& \ binary_image(i, j) = 1 \\ 0, & \text{Otherwise} \end{cases} \quad (4)$$

where n is the number of non-zeros pixels surrounding the center pixel of the 3×3 matrix.

Real time image processing algorithm for sorting MDCK cells according to the number of particles on cell surfaces.

The image processing algorithm for sorting MDCK Madin-Darby Canine Kidney Epithelial Cells based on the number of beads bonded to the cells is essentially the same as the previous cases except a top-hat filter with 7X7 pixels is used to extract features of small particles. The entire algorithm takes about 6 ms with current hardware. The flow chart of real-time image processing algorithm is shown in Supplementary Figure 6.

Real time image processing algorithm for sorting human glioblastoma cells by the extent of radiation induced DNA damage.

The image processing algorithm is illustrated in Supplementary Figure 7. The same top-hat filter with 7X7 pixels used before is also applied to remove background in gamma-h2ax image. Also the same algorithm is used to convert both the GFP image and the background removed gamma-h2ax image into binary images, and extract all image-derived parameters. The latency of the algorithm is about 6.7ms/cell.

Cell sample preparation

Translocating pEGFP-GR plasmids to HEK-297T human embryonic kidney cells.—First GR-GFP plasmid DNA is obtained from bacterial culture. Then two plates of HEK 293T cells are cultured and transfected with GR-GFP. After transfection, cells are cultured for 2~3 days. Then one plate of cells is treated with dexamethasone. The dexamethasone treatment is supposed to cause migration of pEGFP-GR protein from cytoplasm to nucleus. For untreated cells, pEGFP-GR protein stays in cytoplasm.

Bond fluorescent beads to MDCK cells.—MDCK cells are cultured in a 10cm diameter dish. Then add 100 μ L solution of 1 μ m diameter fluorescent beads to the culture dish and keep the cell culture overnight. In the final step, cells are fixed and stained with Carboxyfluorescein succinimidyl ester (CFSE).

Irradiation and antibody staining of Human Glioblastoma Cells.—To induce DNA double-strand breaks (DSB), GFP labeled Human Glioblastoma Cells (GBM-CCC-001) are treated with 6Gy irradiation. The treated cells are washed once with phosphate buffered saline (PBS) and fixed with 1% paraformaldehyde 30 minutes post irradiation. The fixed cells are washed with PBS twice. Then 70% ethanol is added to the cells and the cells are incubated on ice for 1 hour. After ethanol treatment, cells are washed with PBS twice and incubated in 1% TritonX-100 at room temperature for 10 minutes. Then cells are washed with PBS once and incubated in 5% Bovine Serum Albumin (BSA) in PBS for 30 minutes at room temperature on shaker. Then cells are washed with PBS once and incubated in Anti-phospho-Histone H2A.X (Ser139) Antibody, clone JBW301 at 1:300 dilution on ice on shaker for 1 hour. After the primary antibody treatment, cells are washed twice with 5% BSA and incubate in PerCP/Cy5.5 anti-mouse IgG1 Antibody at 1:100 dilution on ice on shaker for 1 hour. At last, the stained cells are washed twice with 5% BSA and resuspended in 1:3 diluted stabilize fixative buffer in MilliQ water.

RESULTS

Sorting cells by spatial distribution of specific protein.

Spatial distribution of certain protein or organelles such as lysosomes and mitochondria carry important biological information. Image guided cell sorter is, to our knowledge, the only tool that can capture cells of sufficient quantity and purity based on such information. Here we demonstrate such functionality using pEGFP-GR plasmids translocated HEK-297T cells and un-translocated HEK-297T cells. pEGFP-GR expresses the eGFP protein fused to the N-terminal end of the glucocorticoid receptor.

Nuclear import and export of glucocorticoid receptor are important cellular processes related to numerous cancers, chronic inflammatory diseases and developmental disorders [29–31]. Protein translocation does not necessarily change the overall fluorescent or light scattering intensity of cells. Hence conventional flow cytometer is unable to distinguish translocated cells from un-translocated cells.

In the experiment, HEK-293T cells are transfected with GR-GFP and separated into 2 plates. One plate of cells is treated with dexamethasone that causes migration of GR-GFP protein

from cytoplasm to nucleus. The other plate of cells is untreated so the GR-GFP protein stays in cytoplasm. The example microscope cell images are shown in Figure. 2(a). The mixture of both types of cells are flown through the system and imaged, and the interested subgroup can be isolated based on the real-time captured cell images.

For supervised machine learning, treated and untreated cells are flown through the system and all image-related parameters from their fluorescent and bright field images are obtained. These extracted image parameters are used to generate gating criteria for real-time image guided cell sorting.

Figure 2(b) shows typical reconstructed cell images with bright-field image defining the cell boundary and the fluorescent image delineating the spatial distribution of GR-GFP protein. All image-derived parameters are also extracted to generate criteria for real-time image-based cell sorting. In this example, the image-derived parameters include fluorescent area, bright-field area, perimeters, shape factors (i.e. area/perimeter ratio), etc.

Next all image-derived parameters are ranked by their Receiver Operating Characteristics (ROC) score, which quantifies each parameter's ability to distinguish translocated cells from un-translocated cells. The three highest ranked parameters are selected for image-guided sorting parameters by default unless users enter extra inputs for the system to adopt different sorting criteria. In this experiment, the top 3 parameters are the area ratio between fluorescent and bright field signals, perimeter ratio of fluorescent and bright field signals, and total fluorescent area.

Next a 3D hyperplane separating two cell populations based on the selected top 3 parameters is formed by Support Vector Machine (SVM). This hyperplane (Figure 2 (c)) defines the criteria for real-time image-guided cell sorting. 300 un-translocated and 300 translocated cells were used as SVM training set. The accuracy of classifying the translocated cells using the SVM hyperplane was obtained by 4-fold cross-validation. Based on the 4-fold cross-validation, the classification accuracy of translocated cells was $97\% \pm 3\%$.

To test the above methodology, we isolated translocated cells from a 50:50 mixture of translocated and un-translocated cells at a concentration of 200,000 cells/mL in phosphate buffered saline (PBS). After sorting, a fluorescence microscope was used to image the sorted cells. From 171 fluorescent microscope images, we calculated a sorting purity of 89% based on the above image-guided criteria.

Sorting cells according to particle binding on cell membrane

To show the capabilities of isolating cells based on surface markers, we sort MDCK cells based on the number of fluorescent particles bonded to the cell membrane. Fluorescent polystyrene beads (1 μ m diameter) functionalized with carboxylic groups can be adsorbed to almost any membrane proteins. By adjusting the concentration of the beads and cells in the mixture, the test system can produce a large variety of the number of beads bonded to each cell. Here we use a 20X objective lens to increase the depth of focus, and the dimension of optical spatial filter is adjusted accordingly. The fluorescent signals of cells (520nm wavelength) and beads (645nm wavelength) are detected by 2 PMTs. The image processing

algorithm is similar to the previous cases, including generation of fluorescent images (50×50 pixels) of the cells and the beads, noise suppression with a digital filter, finding image contours by converting gray scaled images into binary images with defined thresholds, and extraction of image-derived features (See more details in Methods section). Examples of the reconstructed images are shown in Figure 3(a) from which one can unambiguously enumerate the number of particles on the cell surface. Figure 3(b) is the histogram of normalized fluorescence area from the beads. One can see a strong correlation between the fluorescent area and the number of beads. The hyperplane generated by SVM separating cell populations bonded with different number of beads is shown in Figure 3(c). The top 3 parameters used to generate the 3D hyperplane are normalized fluorescent area from the beads, normalized perimeter of the beads and the net intensity within the central area of beads. The “net intensity within central area of beads” is different from the overall fluorescent intensity of beads because it includes signals only from the region of the highest intensity in each spot, thus minimizing the effects of background noise, blurring, and color bleeding in the signal of conventional FACS systems. 430 cells bonded with 1 bead, 327 cells bonded with 2 beads, 185 cells bonded with 3 beads, and 145 cells bonded with 4 beads were used as SVM training set. The accuracy of classifying the cells bonded with 2 or more than 2 beads using the SVM hyperplane was obtained by 4-fold cross-validation. Based on the 4-fold cross-validation, the classification accuracy of cells bonded with 2 or more beads was 100%.

To test the above methodology, we isolated cells bonded with 2 or more beads. After sorting, a fluorescence microscope was used to verify the sorting purity. From 122 fluorescent microscope images, we calculated a sorting purity of 92%.

Sorting cells by the extent of radiation damage.

Next we report experiment of sorting human glioblastoma cells by the extent of radiation induced DNA damage. The number of gamma-h2ax foci is a key parameter measuring the number of DNA double-strand breaks (DSB) induced by cytotoxic agents including ionizing radiation.[32,33] Currently, fluorescent microscopy equipped with high throughput image processing is used to count gamma-h2ax foci number as a measure of DNA damage.[34] However, besides the lower than desired throughput for microscopy, no technique can efficiently isolate cells according to cell's resistance to radiation or cytotoxic damage to support downstream molecular analysis over the targeted cell groups. The throughput of laser microdissection is much too low to support such studies. Image-guided cell sorter fills the technology gap by using machine learning and real time image processing to sort cells based on the foci count of gamma-h2ax that directly delineates the sections of broken DNAs by ionizing radiation.

GFP transfected human glioblastoma cells (GBM-CCC-001) are treated with 6Gy irradiation, and then the cells are fixed by paraformaldehyde. The fixed cells are stained with primary antibody (mouse-anti-gH2AX) and secondary antibody conjugated with fluorophore (PerCP-Cy5.5). The GFP signal delineates the area of the entire cell and the distribution of fluorescent signals from gamma-h2ax represents the fragments of double strand DNAs broken by radiation.

As stated previously, the spatial resolution of the raw image of our system is $2\mu\text{m}$ by $0.4\mu\text{m}$. Hence image processing is required to enhance the effective resolution to resolve the gamma-h2ax foci smaller than $1\mu\text{m}$ diameter. The typical images of the fluorescent spots from gamma-h2ax foci are shown in Figure 4 (a). Here off-line compressive sampling algorithm is used to produce higher spatial resolution ($0.4\mu\text{m} \times 0.4\mu\text{m}$) images from the lower resolution ($2\mu\text{m} \times 0.4\mu\text{m}$) raw images.[35,36] Such off-line processed higher resolution images are only displayed to users, but not used for real time image-derived parameter extraction.

The compressive sampling reconstruction can be described by equation 5, where $I_{measure}$ is the background removed image, $f(x,y)$ is the spatial mask, and $\psi(x,y)$ is the point spread function of objective lens (NA=0.55) approximated by a Gaussian function with $0.35\mu\text{m}$ root-mean-square (RMS) width, and I_{Foci} is the gamma-h2ax fluorescence distribution. I_{Foci} is solved using ℓ_1 -Regularized Least Squares Solver.[37] Using the solved I_{Foci} , the image $I_{Objective}$ formed by objective lens (NA=0.55) is calculated by equation 6. The image $I_{Objective}$ has higher resolution compared to the raw image $I_{measure}$, which has $2\mu\text{m}$ by $0.4\mu\text{m}$ asymmetric resolution equivalent to a 10×50 pixel image. The example cell images reconstructed by compressive sampling algorithm are shown in Figure 4 (b).

$$I_{measure} = f(x,y) \otimes \psi(x,y) \otimes I_{Foci} \quad (5)$$

$$I_{Objective} = \psi(x,y) \otimes I_{Foci} \quad (6)$$

Although compressive sampling algorithms reconstruct images with high resolution, it is not used for image-guided sorting in our experiment because of the limited processing speed of our current hardware. In the experiment reported here we still use the lower resolution raw images in Figure 4 (a) to guide cell sorting. Two image-derived parameters extracted from the real-time reconstructed images are used to estimate the foci count: the total perimeter of gamma-h2ax images and the net gamma-h2ax intensity within the central area of each focus area. The latter is different from the overall fluorescent intensity of gamma-h2ax because it includes signals only from the region of highest intensity in each focus area, thus minimizing the effects of background noise, blurring, and color bleeding in the signal of conventional FACS systems. Figure 5 (a–c) shows how conventional intensity-based sorting is compared against image-guided sorting with the above image-derived parameters.

To obtain the ground truth for each cell being investigated, we apply off-line processing to resolve high-resolution images of 1800 cells, and then find how the total fluorescent intensity and real-time image-derived parameters are related to the ground truth images. The actual foci count is derived from the ground truth images, and Poisson regression is used to predict the foci count based on total fluorescent intensity and real-time image-derived parameters. The scatter plot of predicted foci count and actual foci count are shown in

Figure 5 (a) and (b), and the respective R-squared values are calculated. Using total fluorescent intensity to predict the foci count, the R-squared value is found to be 0.632 ± 0.029 . In contrast, using real-time image-derived parameters to predict foci count, the R-squared value is increased to 0.881 ± 0.008 , significantly greater than intensity-based sorting. Figure 5 (d) and (e) shows histograms of foci count predicted by total gamma-h2ax fluorescent intensity and predicted by image derived parameters. Here bin1 includes cells with 16~17 foci and bin2 includes cells with 25~31 foci. Consistent with the conclusion from Figure 5(a-c), Figure 5 (d) and (e) indicates that intensity based sorting between bin1 and bin2 has greater overlap (i.e. ambiguity) than image based sorting. The result can be more quantitatively represented from the Receiver Operating Characteristic (ROC) analysis in Figure 5 (f). The ROC analysis indicates superior performance of image-guided sorting compared to conventional intensity-based sorting. Also notably in Figure 5 (c), there are about 1% “outliers” in the scatter plot of total gamma-h2ax fluorescent intensity versus foci count, indicating the presence of a small population of cells with exceptionally high gamma-h2ax intensity but not particularly large foci count. If cells of such properties are to be studied, one can use sorting criteria that combine intensity and image-derived parameters to isolate such rare population. On the other hand, using only intensity for sorting criteria as in conventional FACS, we would misclassify those high-intensity but moderate foci count cells, presenting the risk of missing vital biological information. This is just another example of the unique capability image-guided cell sorting can bring to the research community.

To quantitatively evaluate the performance of image-guided sorting and intensity-based sorting, we have compared sorting purity and yield by isolating cells with more than 23 foci using both techniques. By shifting the gating threshold, we estimated the sorting accuracy versus sorting yield. As shown in Figure 6 (b), image-guided sorting with $2\mu\text{m} \times 0.4\mu\text{m}$ spatial resolution has consistently shown superior performance than conventional intensity-based sorting.

DISCUSSION

We demonstrate a machine learning based real-time image-guided cell sorting and classification system that enables: (1) generation of bright-field and fluorescent images of single cells in real time, (2) generation of image-derived gating criteria with machine learning, (3) generation and scoring of image-guided sorting parameters adaptive to user inputs, (4) cell sorting based on image-guided “gating”.

Importantly, all these additional features are established with essentially the same hardware as conventional FACS, allowing cost effective realization of the system. For the same token, the system can be easily expanded to include more parameters/colors by leveraging the multi-parameter FACS systems available today, and the throughput of the system can be enhanced with higher speed electronics.

For purposes of functionality demonstration, we report three experiments representing applications enabled by the system: (1) sorting of pEGFP-GR plasmids translocated HEK-297T human embryonic kidney cells, (2) sorting of MDCK cells based on the number of beads bonded to the cell surface, (3) sorting of human glioblastoma cells by the extent of

radiation induced DNA damage. Of course many other applications such as isolating cells by subcellular phenotype, particle internalization, protein co-localization, etc. can be conceived and inspired by the system.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENT

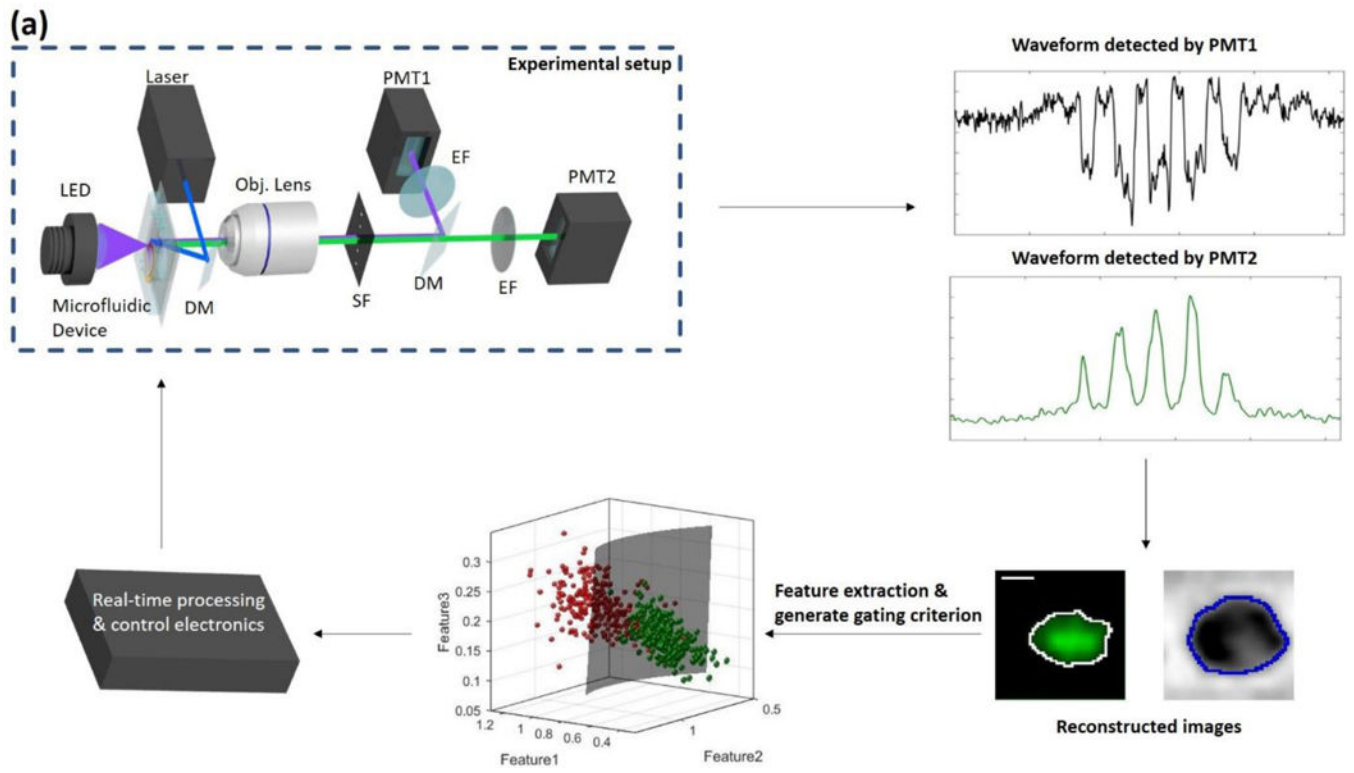
This work was performed in part at the San Diego Nanotechnology Infrastructure (SDNI) of UCSD, a member of the National Nanotechnology Coordinated Infrastructure (NNCI), which is supported by the National Science Foundation (Grant ECCS-1542148). Research reported in this publication was supported by the National Institutes of Health under award number R44DA042636. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Yuhwa Lo has an equity interest in NanoCollect Biomedical, Inc. as a co-founder and a member of the company's Scientific Advisory Board. NanoCollect may potentially benefit from the results of this research.

Reference

- [1]. Herzenberg LA, Parks D, Sahaf B, Perez O, Roederer M, & Herzenberg LA (2002). The history and future of the fluorescence activated cell sorter and flow cytometry: a view from Stanford. *Clinical chemistry*, 48(10), 1819–1827. [PubMed: 12324512]
- [2]. Cho SH, Chen CH, Tsai FS, Godin JM, & Lo YH (2010). Human mammalian cell sorting using a highly integrated micro-fabricated fluorescence-activated cell sorter (μ FACS). *Lab on a Chip*, 10(12), 1567–1573. [PubMed: 20379604]
- [3]. Nitta N, Sugimura T, Isozaki A, Mikami H, Hiraki K, Sakuma S, ... & Fukuzawa H (2018). Intelligent image-activated cell sorting. *Cell*, 175(1), 266–276. [PubMed: 30166209]
- [4]. Feiguin F, Ferreira A, Kosik KS, & Caceres A (1994). Kinesin-mediated organelle translocation revealed by specific cellular manipulations. *The Journal of Cell Biology*, 127(4), 1021–1039. [PubMed: 7962067]
- [5]. Juan G, Hernando E, & Cordon-Cardo C (2002). Separation of live cells in different phases of the cell cycle for gene expression analysis. *Cytometry Part A*, 49(4), 170–175.
- [6]. Sbarra AJ, & Karnovsky ML (1959). The biochemical basis of phagocytosis. *J biol chem*, 234, 1355–1362. [PubMed: 13654378]
- [7]. Bolte S, & Cordelieres FP (2006). A guided tour into subcellular colocalization analysis in light microscopy. *Journal of microscopy*, 224(3), 213–232. [PubMed: 17210054]
- [8]. Htun H, Barsony J, Renyi I, Gould DL, & Hager GL (1996). Visualization of glucocorticoid receptor translocation and intranuclear organization in living cells with a green fluorescent protein chimera. *Proceedings of the National Academy of Sciences*, 93(10), 4845–4850.
- [9]. Li H, Qian W, Weng X, Wu Z, Li H, Zhuang Q, Feng B, & Bian Y (2012). Glucocorticoid receptor and sequential P53 activation by dexamethasone mediates apoptosis and cell cycle arrest of osteoblastic MC3T3-E1 cells. *PLoS One*, 7(6), e37030. [PubMed: 22719835]
- [10]. Newton AC (1995). Protein kinase C: structure, function, and regulation. *Journal of Biological Chemistry*, 270(48), 28495–28498. [PubMed: 7499357]
- [11]. Hawkins ED, Oliaro J, Kallies A, Belz GT, Filby A, Hogan T, ... & Seddon B (2013). Regulation of asymmetric cell division and polarity by Scribble is not required for humoral immunity. *Nature communications*, 4, 1801.
- [12]. Barnett BE, Ciocca ML, Goenka R, Barnett LG, Wu J, Laufer TM, ... & Reiner SL (2012). Asymmetric B cell division in the germinal center reaction. *Science*, 335(6066), 342–344. [PubMed: 22174128]
- [13]. Yoshida H, Kawane K, Koike M, Mori Y, Uchiyama Y, & Nagata S (2005). Phosphatidylserine-dependent engulfment by macrophages of nuclei from erythroid precursor cells. *Nature*, 437(7059), 754. [PubMed: 16193055]

- [14]. Konstantinidis DG, Pushkaran S, Johnson JF, Cancelas JA, Manganaris S, Harris CE, ... & Kalfa TA (2012). Signaling and cytoskeletal requirements in erythroblast enucleation. *Blood*, 119(25), 6118–6127. [PubMed: 22461493]
- [15]. Franzen CA, Simms PE, Van Huis AF, Foreman KE, Kuo PC, & Gupta GN (2014). Characterization of uptake and internalization of exosomes by bladder cancer cells. *BioMed research international*, 2014.
- [16]. Vallhov H, Gutzeit C, Johansson SM, Nagy N, Paul M, Li Q, ... & Gabrielsson S (2011). Exosomes containing glycoprotein 350 released by EBV-transformed B cells selectively target B cells through CD21 and block EBV infection in vitro. *The Journal of Immunology*, 186(1), 73–82. [PubMed: 21106852]
- [17]. Nichols LA, Adang LA, & Kedes DH (2011). Rapamycin blocks production of KSHV/HHV8: insights into the anti-tumor activity of an immunosuppressant drug. *PloS one*, 6(1), e14535. [PubMed: 21264294]
- [18]. Michel ML, Pang DJ, Haque SF, Potocnik AJ, Pennington DJ, & Hayday AC (2012). Interleukin 7 (IL-7) selectively promotes mouse and human IL-17-producing $\gamma\delta$ cells. *Proceedings of the National Academy of Sciences*, 109(43), 17549–17554.
- [19]. Imai T, Kato Y, Kajiwara C, Mizukami S, Ishige I, Ichiyangi T, ... & Udono H (2011). Heat shock protein 90 (HSP90) contributes to cytosolic translocation of extracellular antigen for cross-presentation by dendritic cells. *Proceedings of the National Academy of Sciences*, 108(39), 16363–16368.
- [20]. Datta S, Malhotra L, Dickerson R, Chaffee S, Sen CK, & Roy S (2015). Laser capture microdissection: Big data from small samples. *Histology and histopathology*, 30(11), 1255. [PubMed: 25892148]
- [21]. Hochmuth RM (2000). Micropipette aspiration of living cells. *Journal of biomechanics*, 33(1), 15–22. [PubMed: 10609514]
- [22]. Han Y, & Lo YH (2015). Imaging cells in flow cytometer using spatial-temporal transformation. *Scientific reports*, 5, 13267. [PubMed: 26281956]
- [23]. Wu TF, Mei Z, Pion-Tonachini L, Zhao C, Qiao W, Arianpour A, & Lo YH (2011). An optical-coding method to measure particle distribution in microfluidic devices. *AIP advances*, 1(2), 022155.
- [24]. Zhang AC, Gu Y, Han Y, Mei Z, Chiu YJ, Geng L, ... & Lo YH (2016). Computational cell analysis for label-free detection of cell properties in a microfluidic laminar flow. *Analyst*, 141(13), 4142–4150. [PubMed: 27163941]
- [25]. Han Y, Gu Y, Zhang AC, & Lo YH (2016). imaging technologies for flow cytometry. *Lab on a Chip*, 16(24), 4639–4647. [PubMed: 27830849]
- [26]. Kotsiantis SB, Zaharakis I, & Pintelas P (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160, 3–24.
- [27]. Chen CL, Mahjoubfar A, Tai LC, Blaby IK, Huang A, Niazi KR, & Jalali B (2016). Deep learning in label-free cell classification. *Scientific reports*, 6, 21471. [PubMed: 26975219]
- [28]. Blasi T, Hennig H, Summers HD, Theis FJ, Cerveira J, Patterson JO, ... & Rees P (2016). Label-free cell cycle analysis for high-throughput imaging flow cytometry. *Nature communications*, 7, 10256.
- [29]. McLane LM, & Corbett AH (2009). Nuclear localization signals and human disease. *IUBMB life*, 61(7), 697–706. [PubMed: 19514019]
- [30]. Usmani OS, Ito K, Manechotesuwan K, Ito M, Johnson M, Barnes PJ, & Adcock IM (2005). Glucocorticoid receptor nuclear translocation in airway cells after inhaled combination therapy. *American journal of respiratory and critical care medicine*, 172(6), 704–712. [PubMed: 15860753]
- [31]. Vandevyver S, Dejager L, & Libert C (2012). On the trail of the glucocorticoid receptor: into the nucleus and back. *Traffic*, 13(3), 364–374. [PubMed: 21951602]
- [32]. Kuo LJ, & Yang LX (2008). γ -H2AX-a novel biomarker for DNA double-strand breaks. *In vivo*, 22(3), 305–309. [PubMed: 18610740]

- [33]. Sharma A, Singh K, & Almasan A (2012). Histone H2AX phosphorylation: a marker for DNA damage. In DNA repair protocols (pp. 613–626). Humana Press, Totowa, NJ.
- [34]. Lapytsko A, Kollarovic G, Ivanova L, Studencka M, & Schaber J (2015). FoCo: a simple and robust quantification algorithm of nuclear foci. BMC bioinformatics, 16(1), 392. [PubMed: 26589438]
- [35]. Donoho DL (2006). Compressed sensing. IEEE Transactions on information theory, 52(4), 1289–1306.
- [36]. Coskun AF, Sencan I, Su TW, & Ozcan A (2010). Lensless wide-field fluorescent imaging on a chip using compressive decoding of sparse objects. Optics express, 18(10), 10510–10523. [PubMed: 20588904]
- [37]. Kim SJ, Koh K, Lustig M, Boyd S, & Gorinevsky D (2007). An Interior-Point Method for Large-Scale ℓ_1 -Regularized Least Squares. IEEE journal of selected topics in signal processing, 1(4), 606–617.



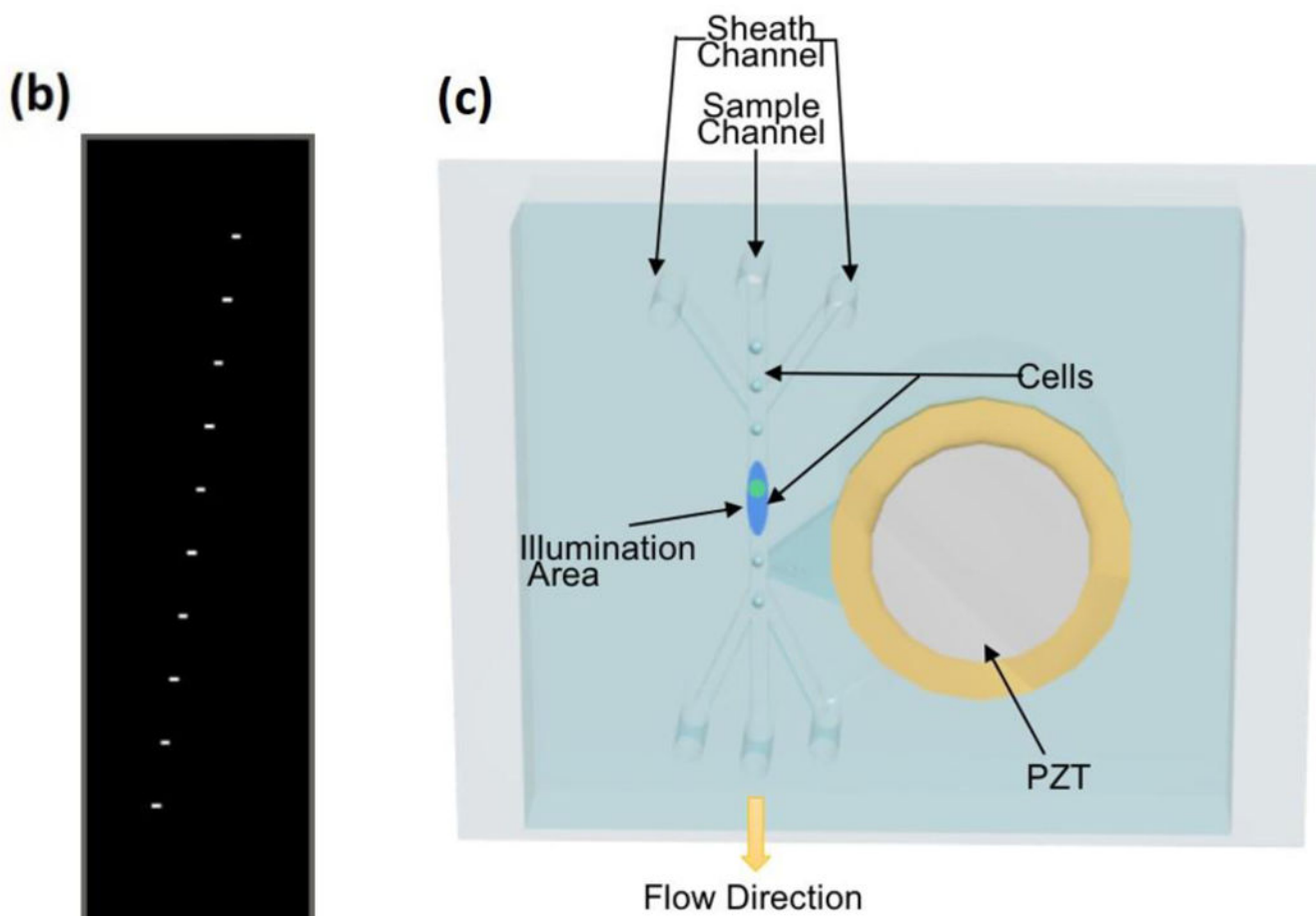


Figure 1. The Machine Learning Based Real-Time Image-Guided Cell Sorting and Classification system. (a) Schematic diagram of the image-guided cell sorting system. (Scale bar is $5\mu\text{m}$). Bright field and fluorescence cell images are at first encoded into time domain waveforms and detected by PMTs. Cell images are then reconstructed from time domain waveforms. Next, image features are extracted and image-based gating criteria are generated. Finally, cell images are processed and sorting decision is made in real-time based on the image-guided gating criteria with supervised machine learning. DM, dichroic mirror; SF, spatial filter; EF, emission filter; PMT, photomultiplier tube. (b) Design of optical spatial filter having ten $100\ \mu\text{m}$ by $50\ \mu\text{m}$ slits positioned apart. (c) Microfluidic device with an on-chip piezoelectric PZT actuator to deflect selected cells in the microfluidic channel for image-guided cell sorting.

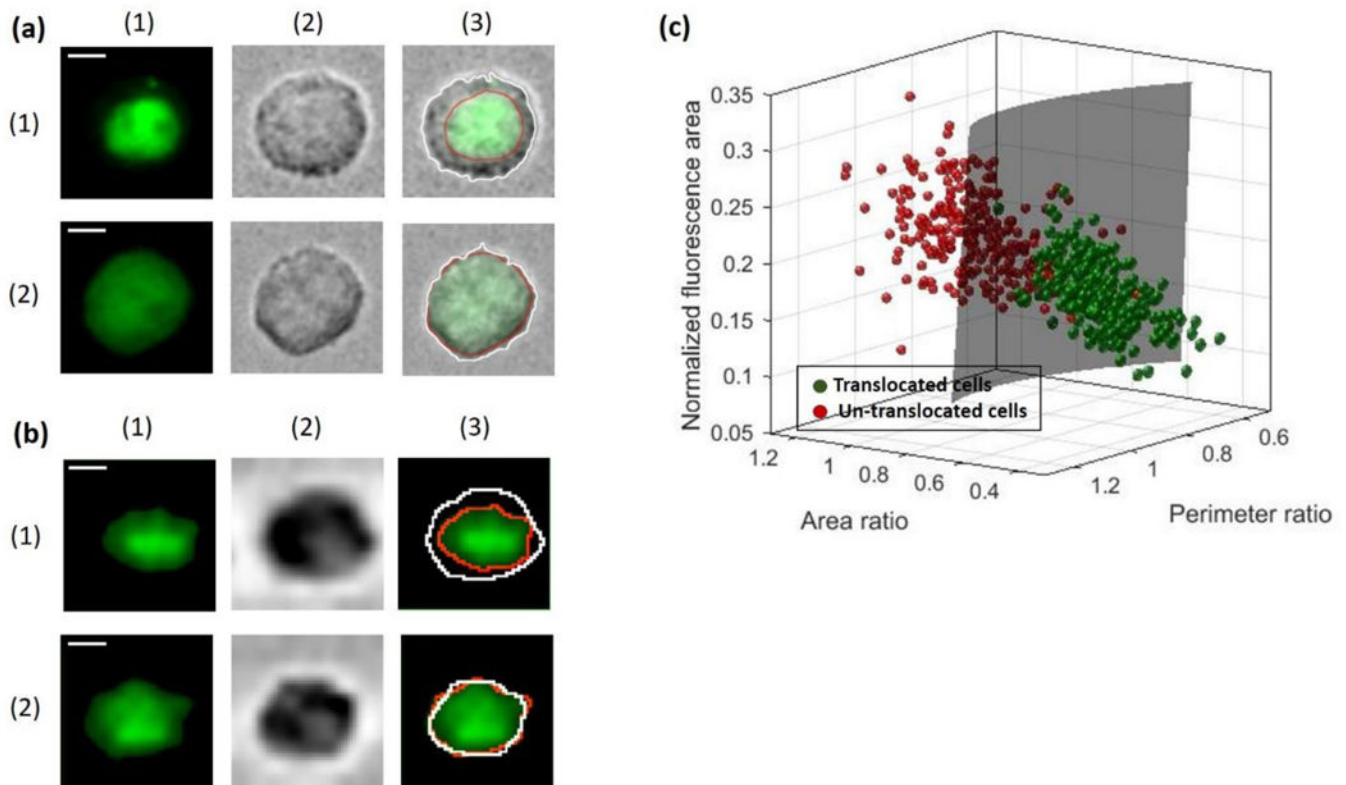


Figure 2.

Sorting cells by spatial distribution of specific protein. (a) Example of microscope cell images. Row(1) shows translocated cells, and row(2) shows un-translocated cells. Column(1) shows fluorescent images, column(2) shows bright-field images, and column(3) shows overlaid images with their respective contours defined by the computer-generated red and white curves. (b) Example cell images generated by our system. Row(1) shows translocated cells, and row(2) shows un-translocated cells. Column(1) shows fluorescent images, column(2) shows bright-field images, and column(3) shows fluorescent images overlaid with bright-field images with their respective contours defined by the computer-generated red and white curves. (c) Hyperplane formed by SVM. A $5\mu\text{m}$ scale bar is shown in each row of micrographs.

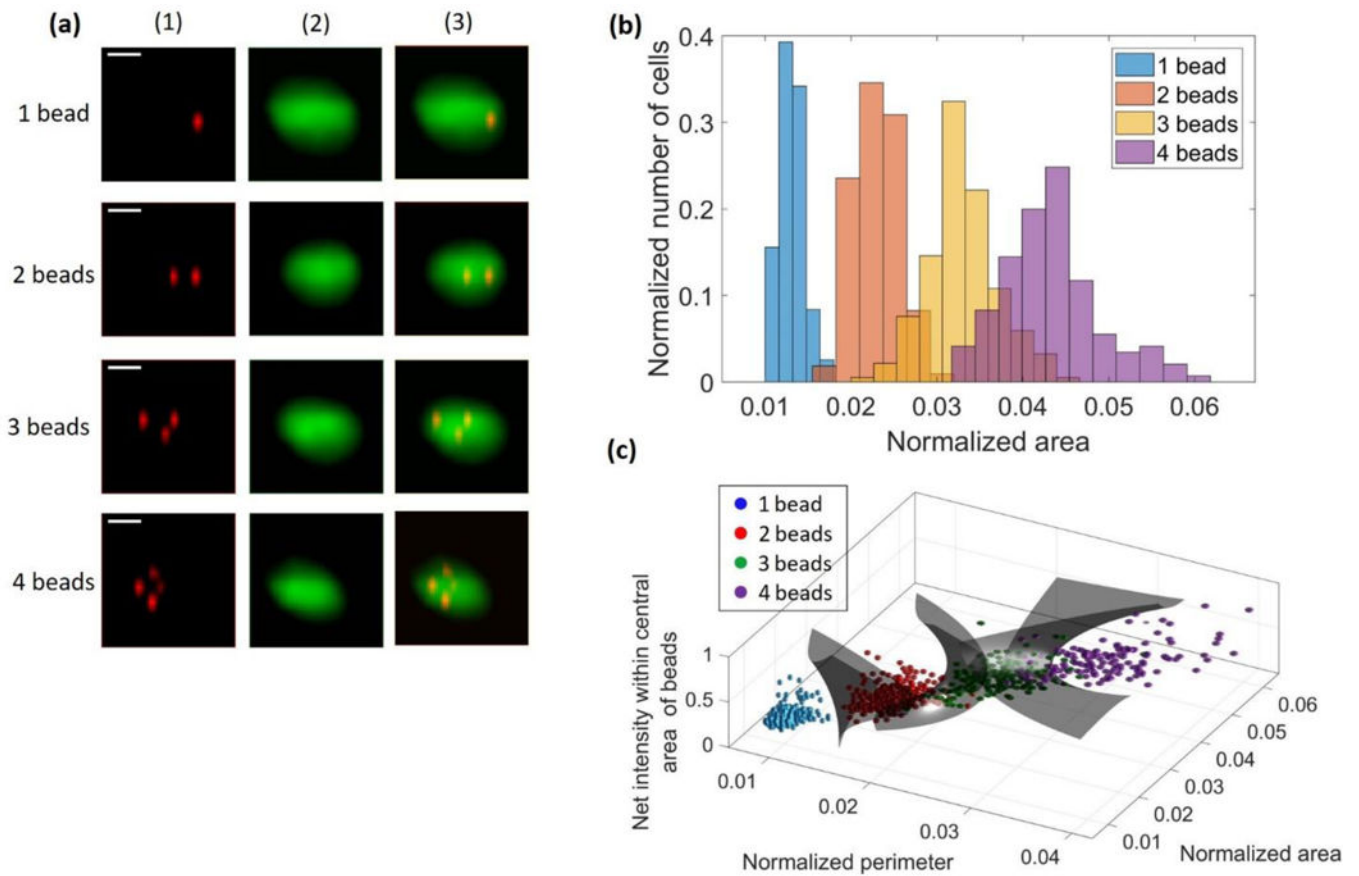


Figure 3. Sorting cells according to particle binding on cell membrane. (a) Example cell images generated by our system. Column(1) shows fluorescent images at 645nm, column(2) shows fluorescent images at 520nm, and column(3) shows overlaid images. (b) Histogram for normalized fluorescent area from beads. (c) Hyperplane formed by SVM. A $5\mu\text{m}$ scale bar is shown in each row of micrographs.

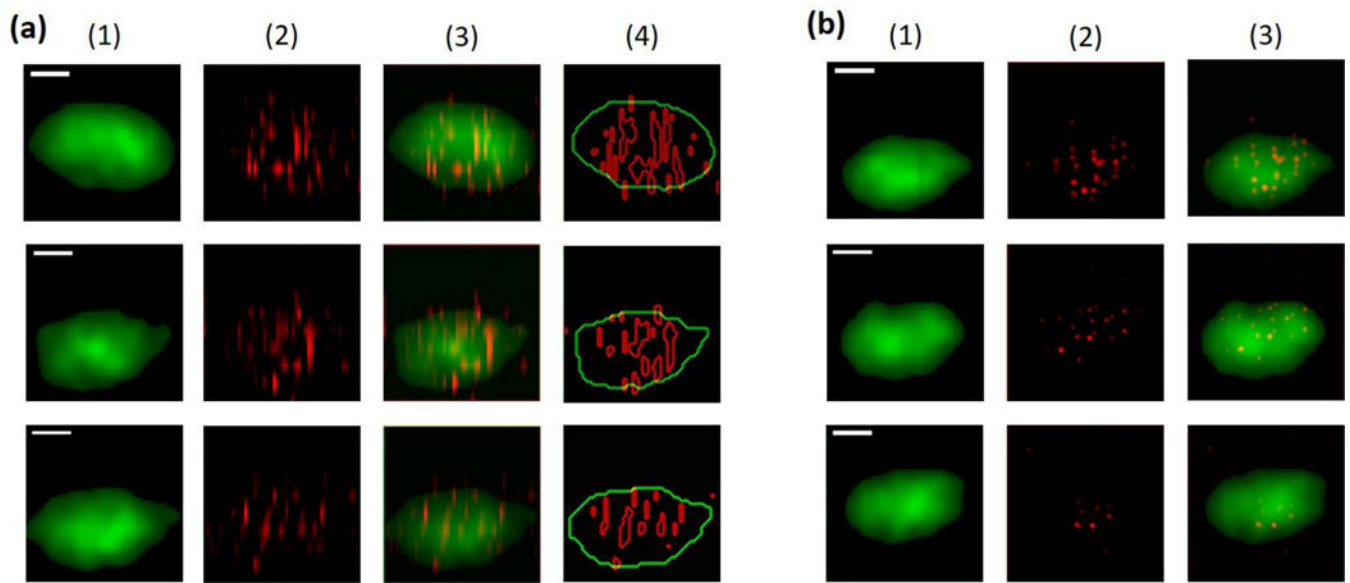


Figure 4. Example images of irradiated cells. (a) Example cell images generated by real-time algorithm. Column(1) shows GFP images, column(2) shows gamma-h2ax images with background removed, column(3) shows overlaid images, and column(4) shows image contours, with green contour for GFP image and red contours for gamma-h2ax images. (b) Example cell images generated by off-line processing for human vision. Column(1) shows GFP images, column(2) shows gamma-h2ax images, and column(3) shows overlaid images. A 5µm scale bar is shown in each row of micrographs.

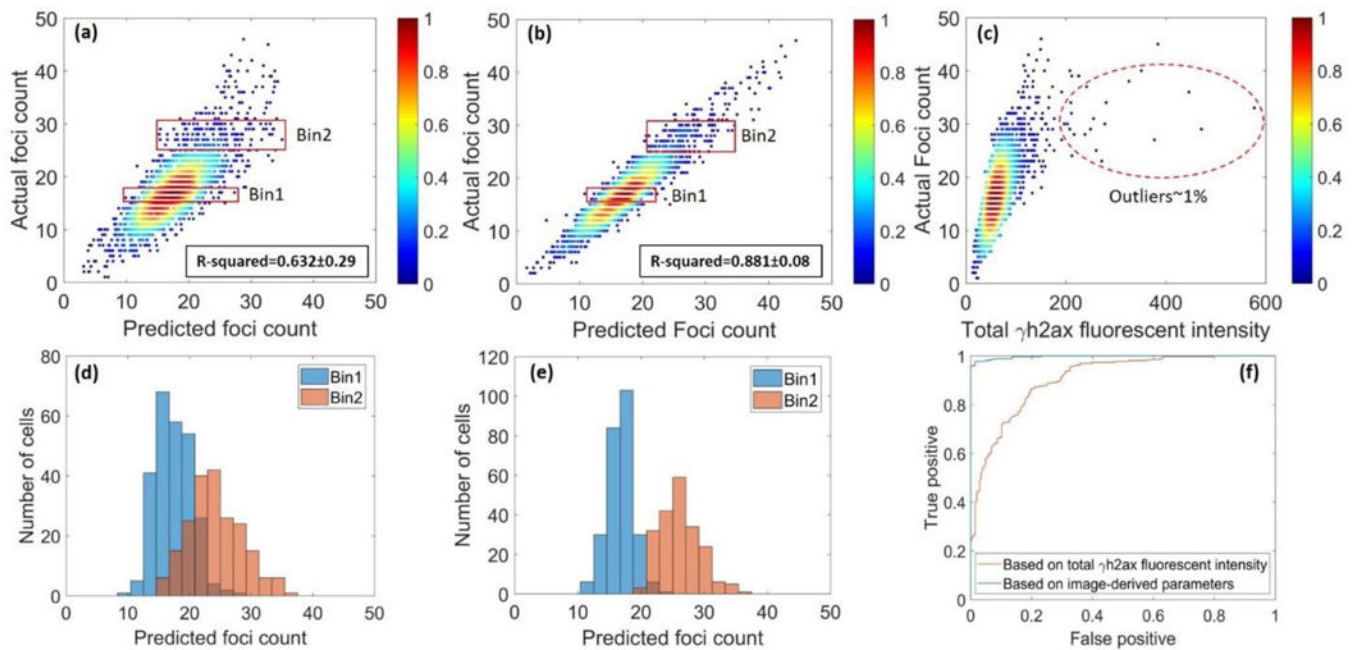


Figure 5.

Estimation of foci count based on image-derived parameters and total fluorescent intensity.

(a) Scatter plot of predicted foci count based on total gamma-h2ax intensity versus actual foci count (b) Scatter plot of predicted foci count based on image-derived parameters versus actual foci count (c) Scatter plot of total gamma-h2ax intensity versus actual foci count including outliers (d) Histogram of predicted foci count based on total gamma-h2ax intensity of bin1 and bin2 (e) Histogram of predicted foci count based on image-derived parameters of bin1 and bin2 (f) Receiving Operating Characteristic (ROC) analysis of bin1 and bin2, showing superior performance of image-guided sorting than conventional intensity-based sorting.

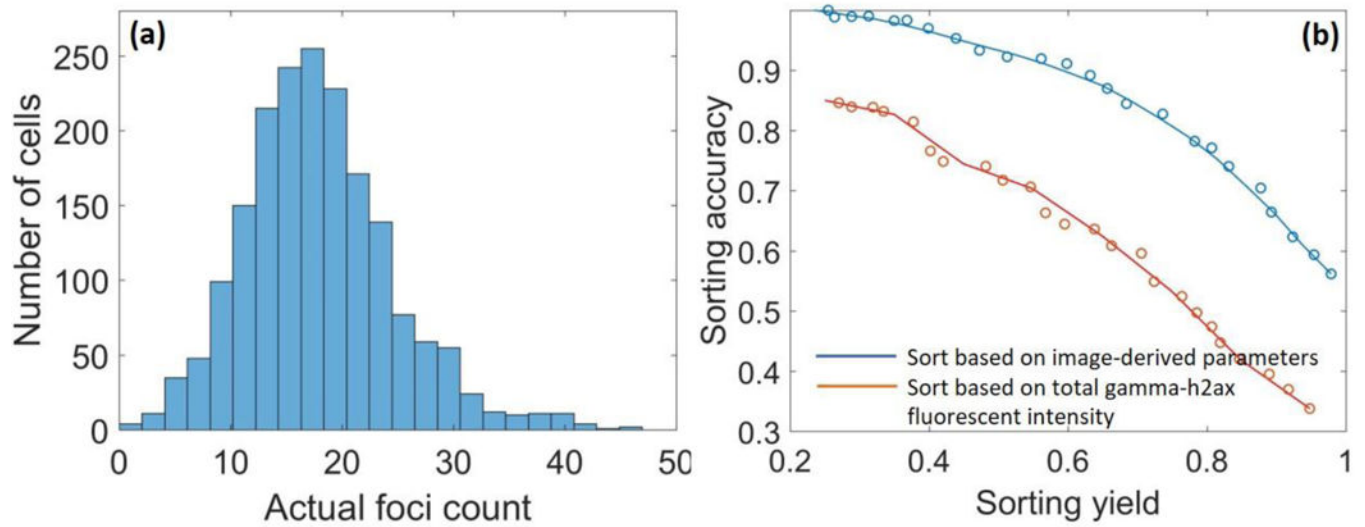


Figure 6. Sorting purity and yield. (a) Histogram of actual foci count for Gamma-ray irradiated cells. (b) Sorting yield versus sorting accuracy for isolation of cells with greater than 23 foci using image-guided and intensity-based methods.