The use of image-derived features to identify and exclude non-singlet events in flow cytometry sort workflows

Aaron J Middlebrook, Woodrow Lomas, Supriya Balaji Ramachandran.
BD Biosciences, San Jose, CA

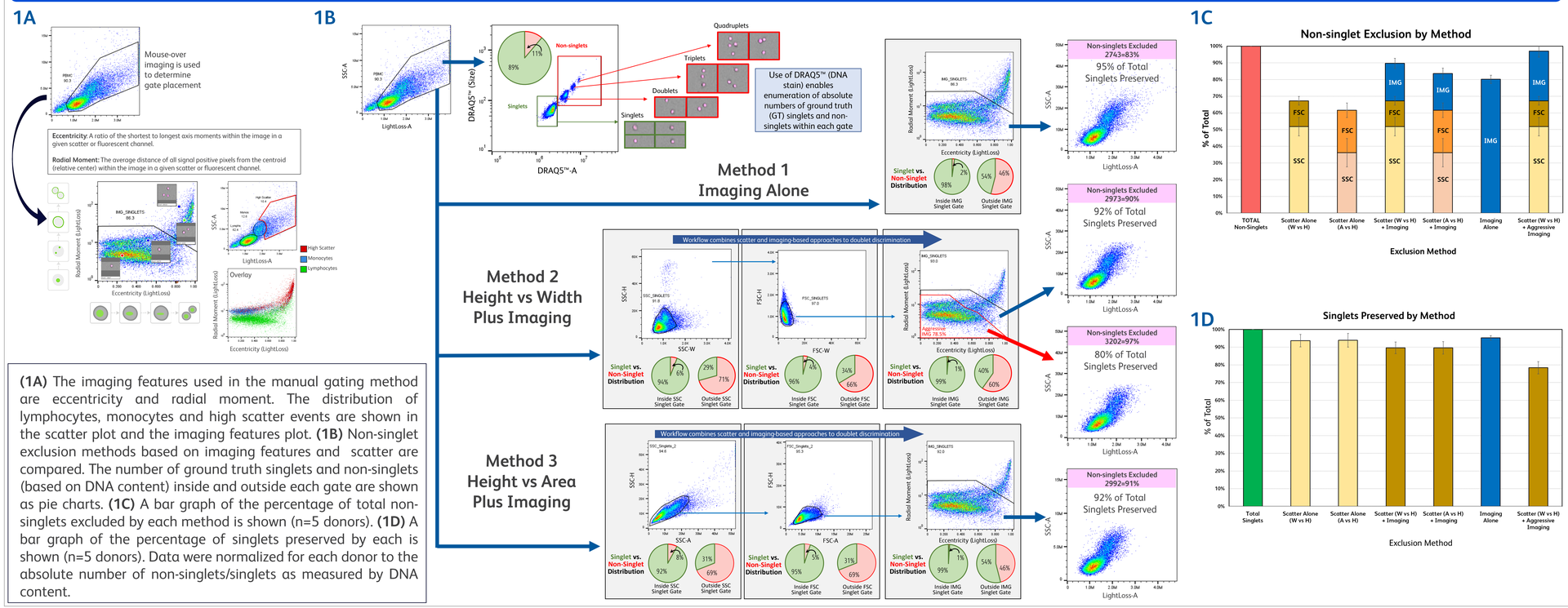
Abstract

Doublet and clump exclusion is a critical preliminary step in establishing a pure sort. Doublets can confound purity and can lead to unwanted results or unnecessary expense when performing downstream functional and or genomic analyses. Here we sought out to demonstrate the improvements in doublet discrimination that an image enabled sorter can offer.

Peripheral blood mononuclear cells (PBMCs) were prepared by Ficoll separation from five healthy donors. Sample preparations were counterstained with DRAQ5™, a membrane-permeable DNA intercalating dye that is compatible with BD CellView™ Image Technology. Singlet, doublet and clump frequency were tracked via DNA content as traditional scatter-based doublet discrimination strategies were compared with an image-based approach. Using DNA content as a ground truth for singlet vs non-singlet events, each gate along the workflow was compared. Traditional scatter-based approaches such as width vs height and area vs height were performed alongside an image-based approach using the features of eccentricity and radial moment. The image-based approach was evaluated alone and in conjunction with scatter-based methods. Singlet preservation was also evaluated as a significant loss of singlets would be an undesirable consequence of any successful doublet exclusion strategy.

While the inclusion of cellular doublets in a targeted single-cell sort has obvious consequences in that the non-target cells contaminate the sort product and can lead to misleading or undesirable results in downstream applications, we also explored the impact of non-cellular doublets. Doublets that consist of a cell and an unlysed red blood cell, large exosome, apoptotic body or platelet can also impact the results of a sort. Here we explored the inclusion and or exclusion of these events in a sort and the impact they may have on the quality of the sort product.

Results



(1A) The imaging features used in the manual gating method are eccentricity and radial moment. The distribution of lymphocytes, monocytes and high scatter events are shown in the scatter plot and the imaging features plot. **(1B)** Non-singlet exclusion methods based on imaging features and scatter are compared. The number of ground truth singlets and non-singlets (based on DNA content) inside and outside each gate are shown as pie charts. **(1C)** A bar graph of the percentage of total non-singlets excluded by each method is shown (n=5 donors). **(1D)** A bar graph of the percentage of singlets preserved by each is shown (n=5 donors). Data were normalized for each donor to the absolute number of non-singlets/singlets as measured by DNA content.

Methods

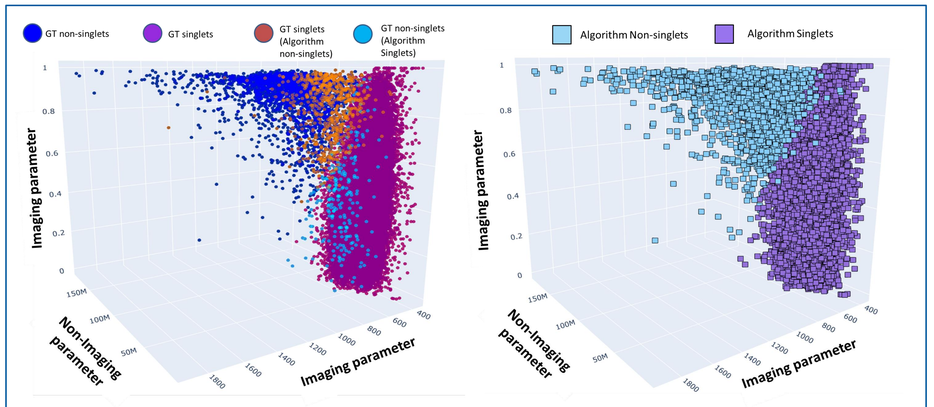
Sample prep and data acquisition
Whole blood was collected in standard EDTA sample collection tubes. Peripheral blood mononuclear cells (PBMCs) were prepared using a standard Ficoll density gradient centrifugation protocol. Blood from five healthy donors was processed in parallel. PBMCs were washed and stained with DRAQ5™ (BD Biosciences cat#564902). Cells were acquired on a BD FACSDiscover™ S8 Prototype with CellView™ Image Technology. Data was analyzed in FlowJo™ Software (BD Biosciences).

Manual gating techniques
Doublet discrimination gating based on scatter (SSC and FSC) was done without the use of imaging as we tried to capture practices employed by users on non-imaging enabled cytometers. The methods tested (Height vs. Width & Height vs. Area) were chosen based on the recommendations outlined in the article, "Guidelines for the use of flow cytometry and cell sorting in immunological studies" in the European Journal of Immunology by Cossarizza et al, 2017 (https://doi.org/10.1002/eji.201646632). Gates on all scatter plots were drawn using the Autogating tool in FlowJo™ to minimize experiential bias. The use of imaging was not limited to numerical image-derived features but also employed pop-up mouse-over imaging to guide gate placement.

Machine Learning (ML) algorithm techniques
One training dataset of 25K events was used to the find regression relationship between predictors and target (DRAQ5™ DNA stain indicator). A top ten feature list was generated and then iteratively fine tuned to arrive at a sparse 3-feature set. Underlying distribution was assumed to be a multi-modal Gaussian mixture and unsupervised clustering was performed on 375K events (10 datasets).

Results

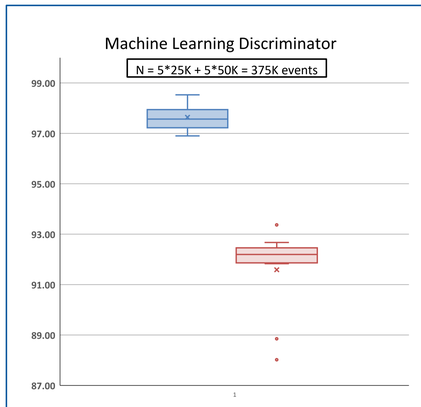
2A Clustering outcomes of One dataset with ~28K events



(2A) Ground truth (GT) is defined as true singlets and true non-singlets based on DNA content (DRAQ5™ staining).
Left: Clustering outcomes indicating GT singlets (magenta), GT non-singlets (blue), Algorithm identified non-singlets (that are true singlets) (orange) and algorithm identified singlets (that are true non-singlets) (light blue)

Right: All algorithm identified singlets (violet) and algorithm identified non-singlets (light blue)

2B Singlet inclusion and non-singlet exclusion rate (375K events), 10 datasets



(2B) Box-and-whisker plot demonstrates that median true singlet identification rate is ~97% (blue) and true doublet exclusion rate is ~92% (orange). The S.D. of singlets distribution is 0.52 and S.D. of doublets distribution is 1.72

Conclusions

- Doublet discrimination based on scatter excluded approximately 60% of doublets (H vs W and H vs A were approximately equivalent) and preserved approximately 95% of singlets. (Fig. 2)
- Doublet discrimination based on imaging features alone excluded 80% of doublets and preserved 95% of singlets. (Fig. 2)
- Combining imaging and scatter parameters, 90% of doublets were excluded and 90% of singlets were preserved. (Fig. 2)
- An aggressive imaging gate was able to exclude 97% of doublets, however only preserved 78% of true singlets. (Fig. 2)
- An ML-based algorithm can exclude 92% of true non-singlets preserve 97% of true singlets. (Fig. 3)
- A combination of imaging and non-imaging parameters is powerful for singlet vs non-singlet discrimination. (Fig. 3)
- The underlying distribution of events closely follows a multi-modal Gaussian assumption.
- Gaussian clustering is superior to K-means, Spectral, DBSCAN and BIRCH clustering methods for this application as discovered through our work (data not shown).