

## Abstract

### Introduction

- Manual singlet discrimination is a subjective, time-consuming process.
- Machine learning can improve singlet discrimination by utilizing higher dimensional data and automating the process, which has the potential to produce more accurate, objective, and consistent singlet discrimination.

### Methods

- 22 imaging datasets of varying cell types were used for algorithm development and testing. Images were used to establish ground truth.
- 5 feature sets and the DBSCAN algorithm were tested for precision and sensitivity.

### Results

- The LightLoss feature set demonstrated a precision of >97% and sensitivity of >88%.

### Conclusion

- DBSCAN can be used to automatically detect singlets with high precision and sensitivity.
- SingletSeeker, a free FlowJo™ 10 plugin, has been developed based on this algorithm.

## Methods

### Data Sourcing

- 22 imaging datasets (Training: 15, Testing: 7)
- Collected on BD FACSDiscover™ S8 imaging cytometers.

### Programming Environment

- Python and associated machine learning libraries were used to develop SingletSeeker algorithm.
- Scikit-Learn, Scipy, Pandas, Numpy.
- DBSCAN, Density-based Spatial Clustering in Applications with Noise

### Ground Truth

- Ground truth was established based on the images associated with each event using a semi-supervised, iterative approach.

### Feature Selection

- 5 feature sets were selected for testing to best discriminate singlets.

Feature Set	Features
LightLoss	LightLoss (Violet)-A, LightLoss (Violet)-H
Recursive Feature Elimination (RFE)	LightLoss (Violet)-A, Long Axis Moment (SSC (Imaging)), Radial Moment (FSC), Radial Moment (LightLoss (Imaging)), Radial Moment (SSC (Imaging))
Random Forest Classifier (RFC)	LightLoss (Violet)-A, LightLoss (Violet)-W, Size (FSC), FSC-A, Radial Moment (LightLoss (Imaging)), Radial Moment (SSC (Imaging))
Sequential Feature Selection (SFS)	LightLoss (Violet)-A, Short Axis Moment (LightLoss (Imaging)), SSC (Imaging)-A, LightLoss (Imaging)-A, SSC (Violet)-A
Forward Scatter (FSC)	FSC-A, FSC-H

### Manual Gating Comparison

- In addition to comparing results to the ground truth, 3 expert cytometrists and co-authors performed manual gating.

### Statistics

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

## Algorithm Development

Figure 1 – Can singlets and multiplets be separated by density?

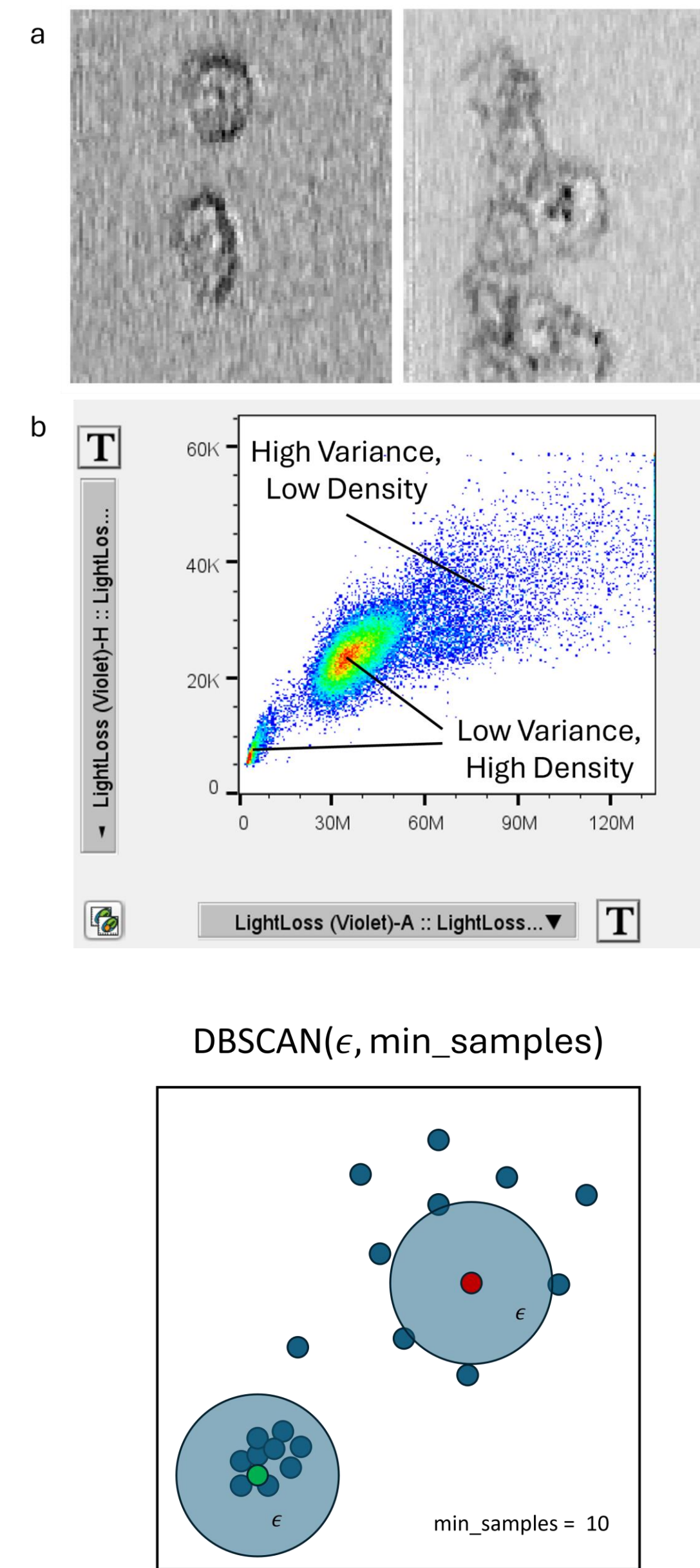


Figure 2 – Two-step process to discriminate and classify singlets

- Step 1**

  - Determine appropriate density threshold based on k-Nearest Neighbors to separate high and low density data (singlets and multiplets)
  - Run DBSCAN using selected threshold to filter out low density data (multiplets).
- Step 2**

  - Determine appropriate size threshold by finding minima in FSC/LightLoss parameters.
  - Classify singlets and debris based on position relative to minima.

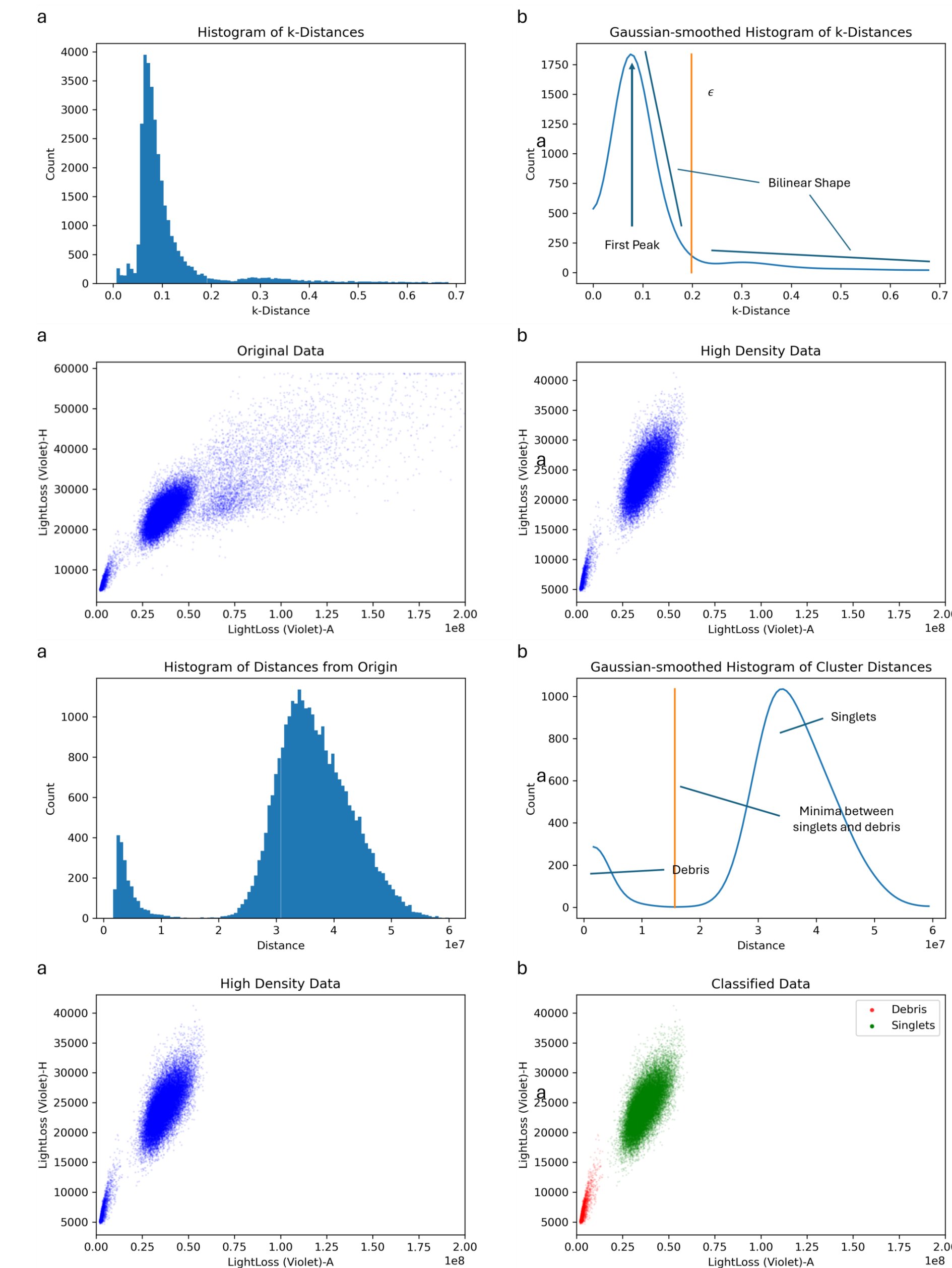
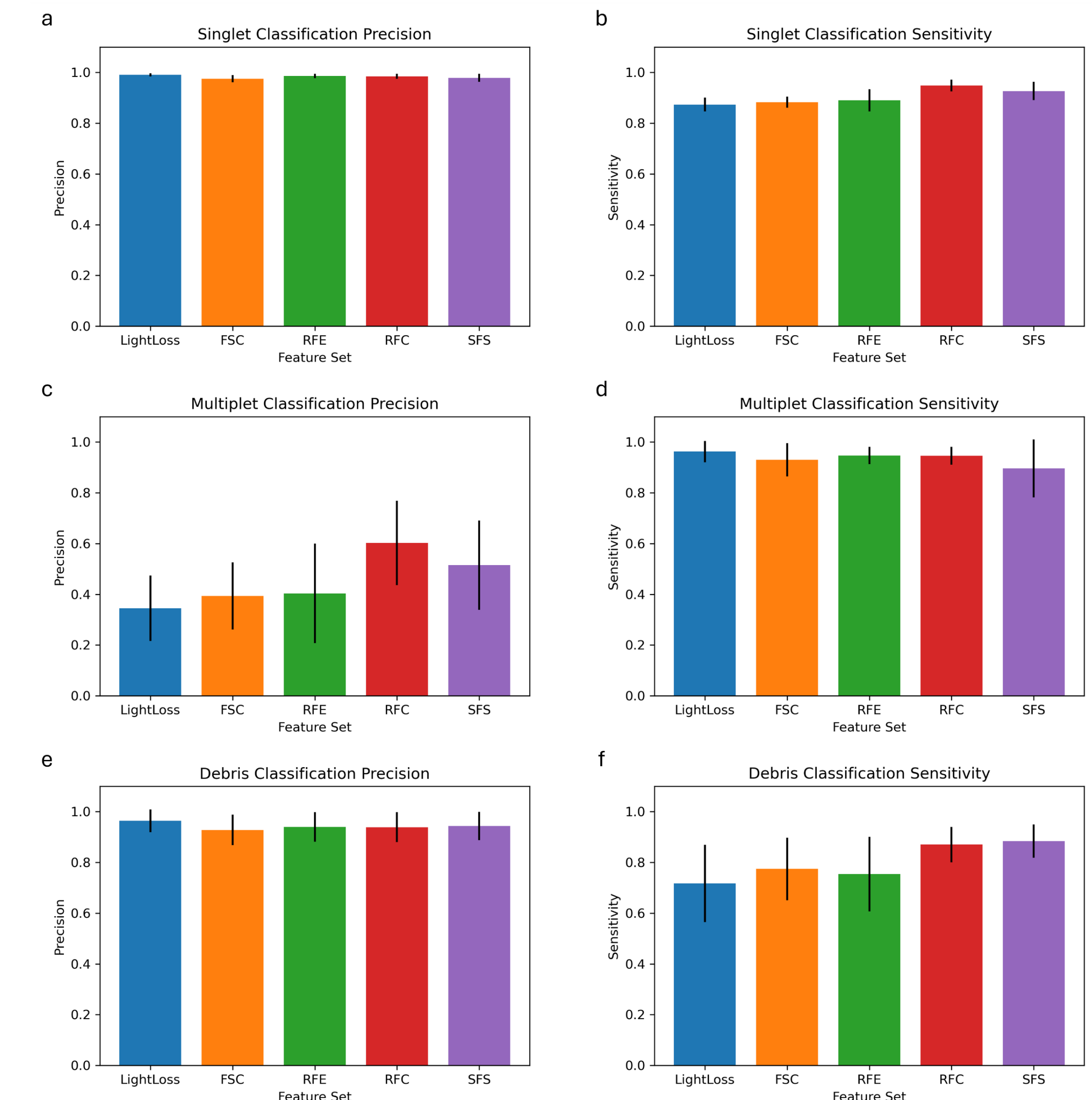
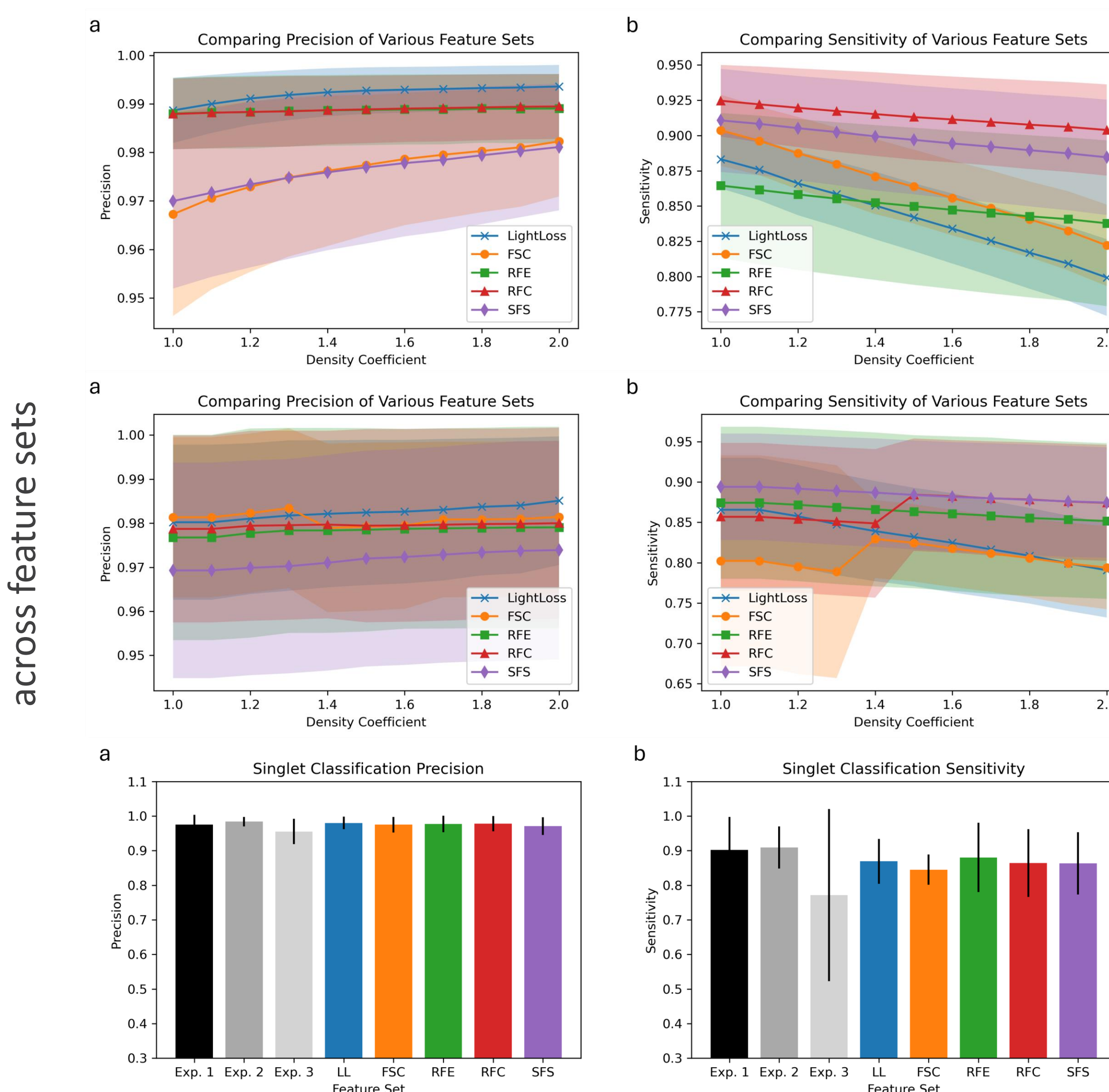


Figure 3 – Algorithm successfully classifies singlets and others based on density



## Results

Figure 4 – Feature sets provide high precision and sensitivity across feature sets



### Singlet Precision/Sensitivity – Training Data

- LightLoss provided best precision (~99%) while RFC provided best recall (~91%).

	Precision (%)		Sensitivity (%)	
LightLoss	98.87	99.36	79.93	88.32
RFE	98.80	98.91	83.78	86.45
RFC	98.79	98.95	90.39	92.47
SFS	97.00	98.11	88.46	91.07
FSC	96.73	98.23	82.10	90.37

### Singlet Precision/Sensitivity – Test Data

- LightLoss provides best precision (98%), but SFS produced the best sensitivity (~87%).

	Precision (%)		Sensitivity (%)	
LightLoss	98.02	98.51	79.06	86.56
RFE	97.68	97.91	85.16	87.42
RFC	97.87	98.00	84.87	88.42
SFS	96.93	97.39	87.45	89.40
FSC	97.89	98.34	78.89	82.93

### Comparing Manual and Automated Singlet Discrimination

- Manual gating and automated gating produce similar precision and sensitivity across all feature sets.
- Automated gating was performed

## Conclusions

### Introduction

- Singlets and multiplets can be successfully distinguished using unsupervised, density-based algorithms.
- All feature sets produce high precision with slightly lower sensitivity, though using LightLoss parameters alone yield the best precision overall, while RFE seems to provide a good balance between precision and sensitivity.
- Density coefficients are a useful user input that allows tuning of the balance between precision and sensitivity.
- The algorithm produces similar results compared to the manual gating performed by co-authors.
- As an unsupervised algorithm, this should work on any cell type as it is based on detecting differences in feature density rather than identifying features of cells themselves.
- Algorithm may facilitate more objective, consistent, and user-independent singlet gating, which may be most useful for novice cytometrists.
- The algorithm has been released as a FlowJo™ plugin, SingletSeeker, and is available for free download from FlowJo™ Exchange.