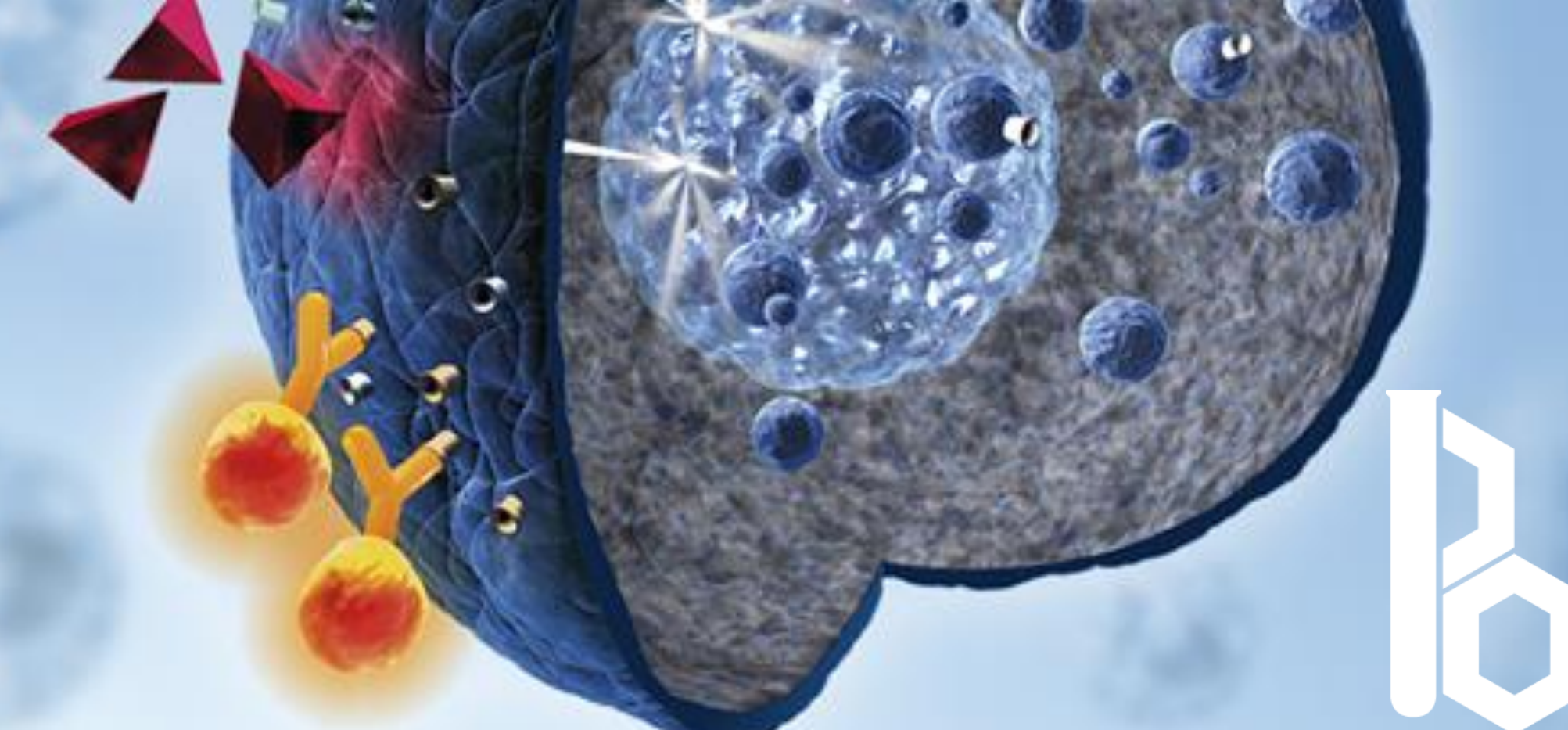


AUTOMATED GATING STRATEGIES FOR BASOPHIL ACTIVATION TEST: A COMPARISON TO MANUAL DATA ANALYSIS

Christina U. Bauer^{1,2}, Dominik Vogt¹, Anna Melone¹, Tonino Gerber¹, Thomas Schuster¹, Christian-Benedikt Gerhold¹, Michael A. Gerspach^{1,2}

¹ BÜHLMANN Laboratories AG, Schönenbuch, Switzerland

² to whom correspondence shall be addressed: cba@buhmannlabs.ch, mg@buhmannlabs.ch

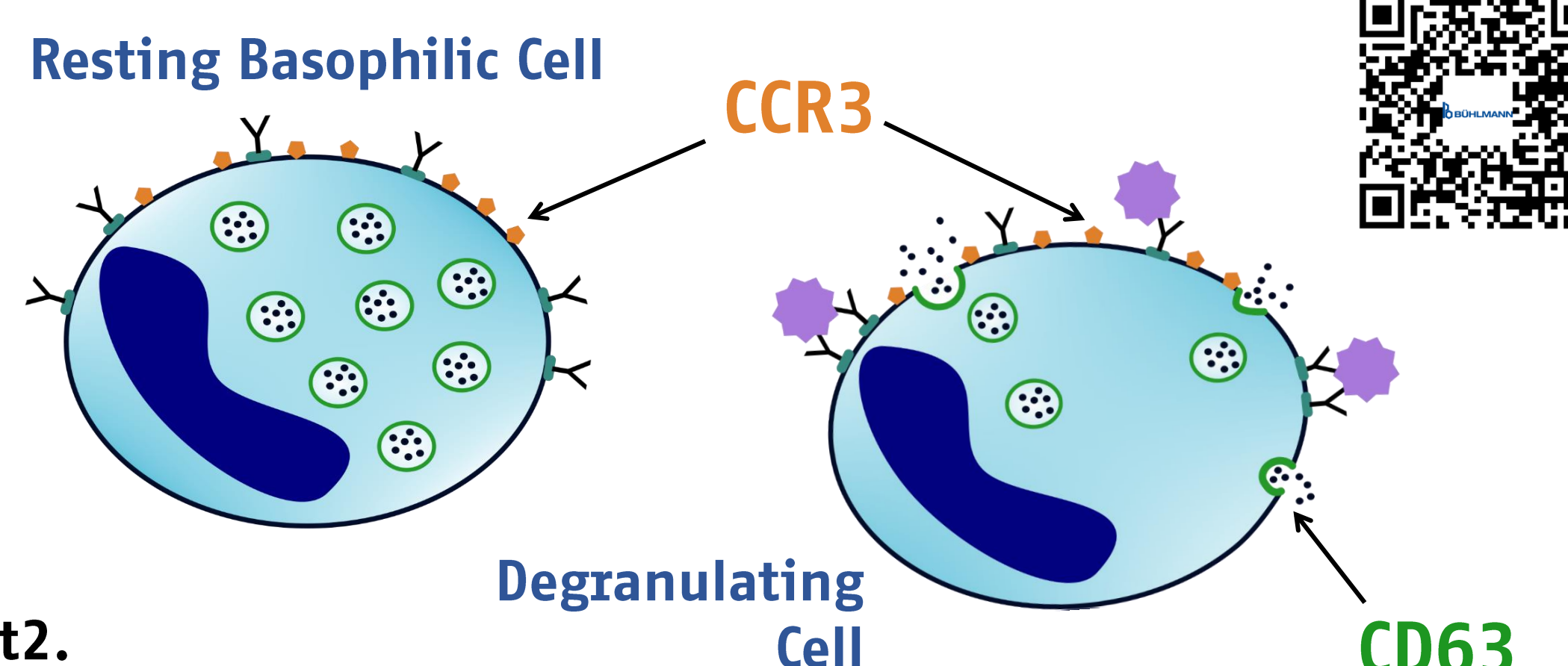


Introduction: The Flow CAST[®] as a Standardized BAT

The basophil activation test (BAT) is a unique allergy diagnostic test, in which live blood cells of the patient are directly exposed to different allergens in a test tube. If an allergic reaction is triggered, basophils become activated and degranulate. In BÜHLMANN's Flow CAST[®] BAT, this activation is quantified as the percentage of CD63+ basophils using flow cytometry, as CD63 is exposed on the cell surface upon degranulation. For each allergen, there is a diagnostically relevant cut-off value, ranging from 5 to 15 % CD63+ cells.

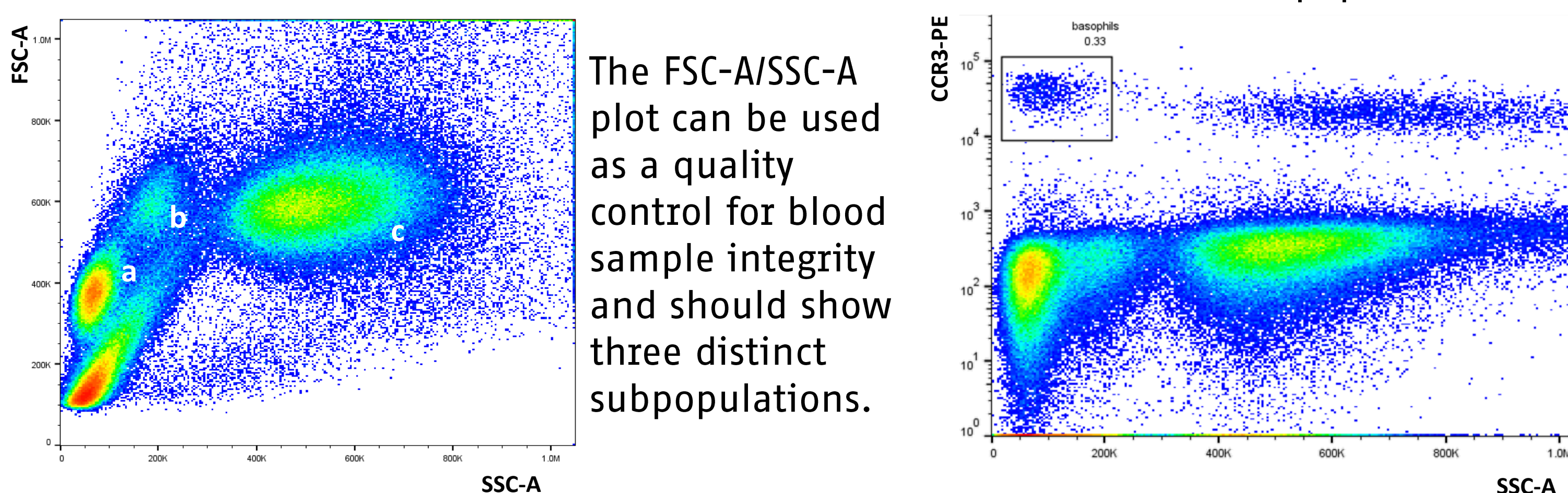
The purpose of this study is to evaluate different algorithms for automated data analysis of the BAT.

Computations have been performed with R, version 4.2.3, using the packages flowCore, randomForest, and ggplot2.

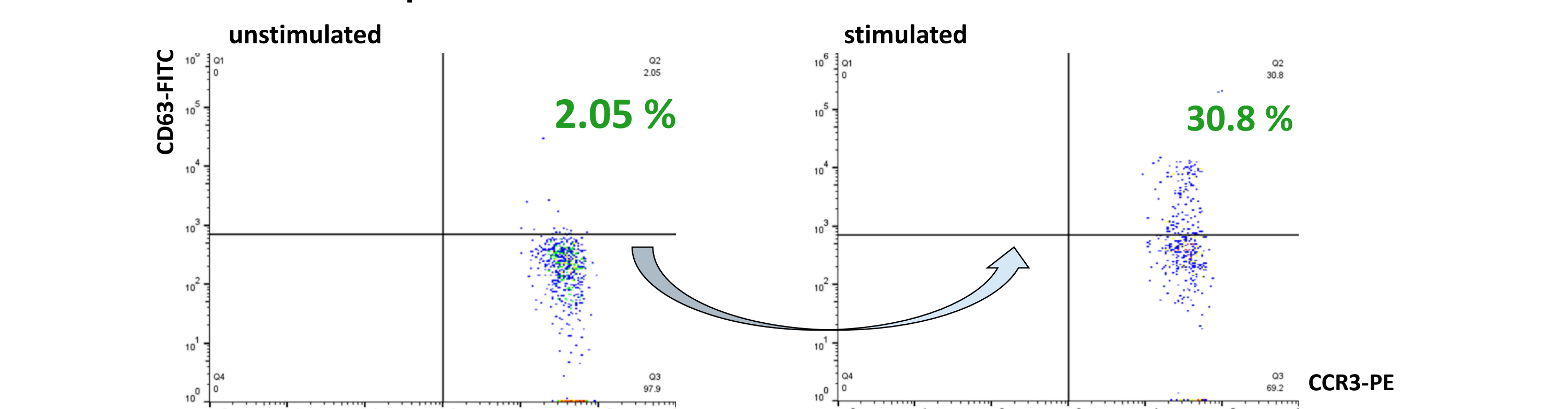


Manual Gating

(1) Basophil identification



(2) Measure basophil activation



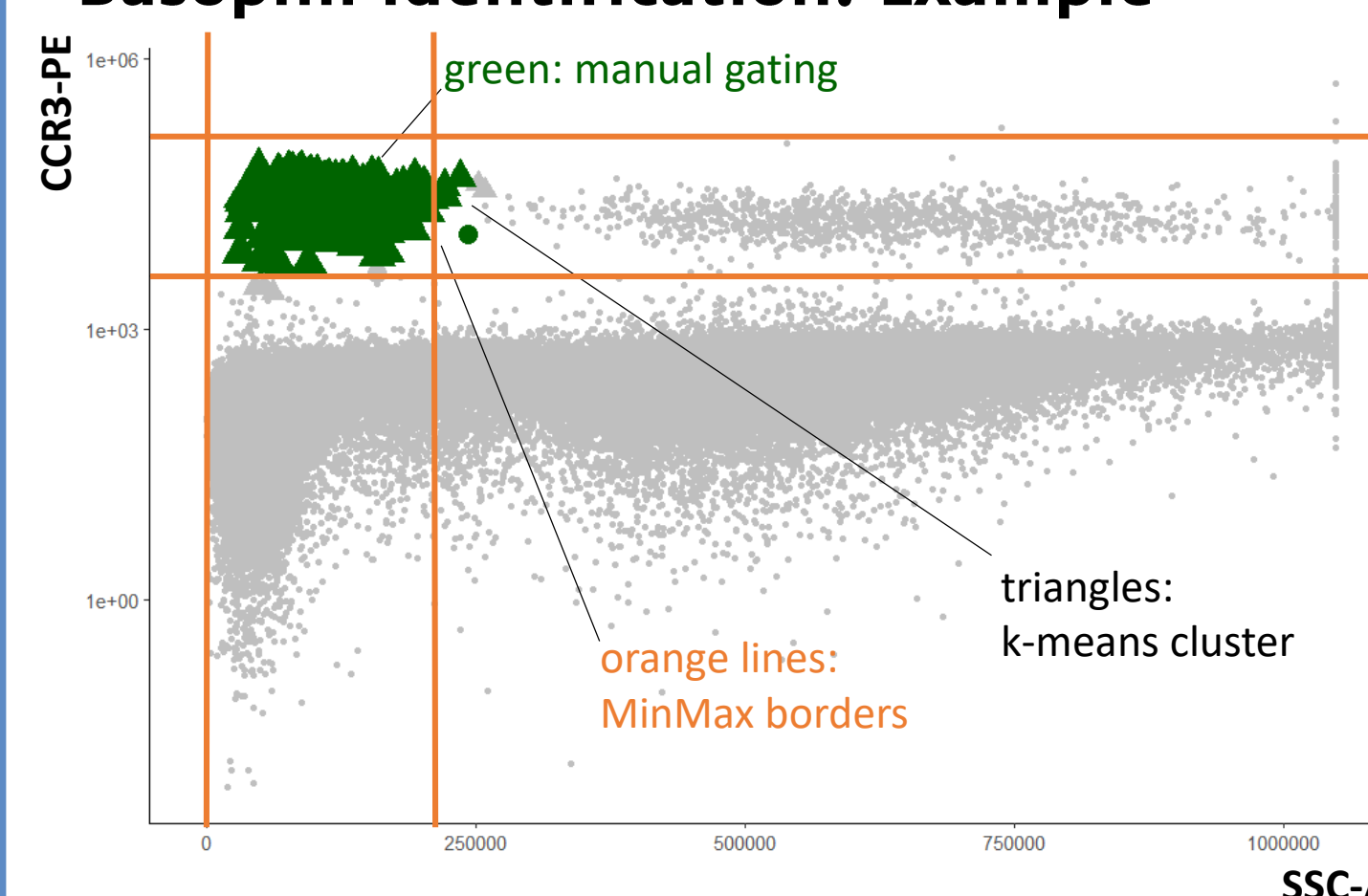
Comparison

The two hard-coded gating algorithms (MinMax and k-means) are compared to manual gating based on a dataset with 442 unstimulated and 442 stimulated BAT experiments.

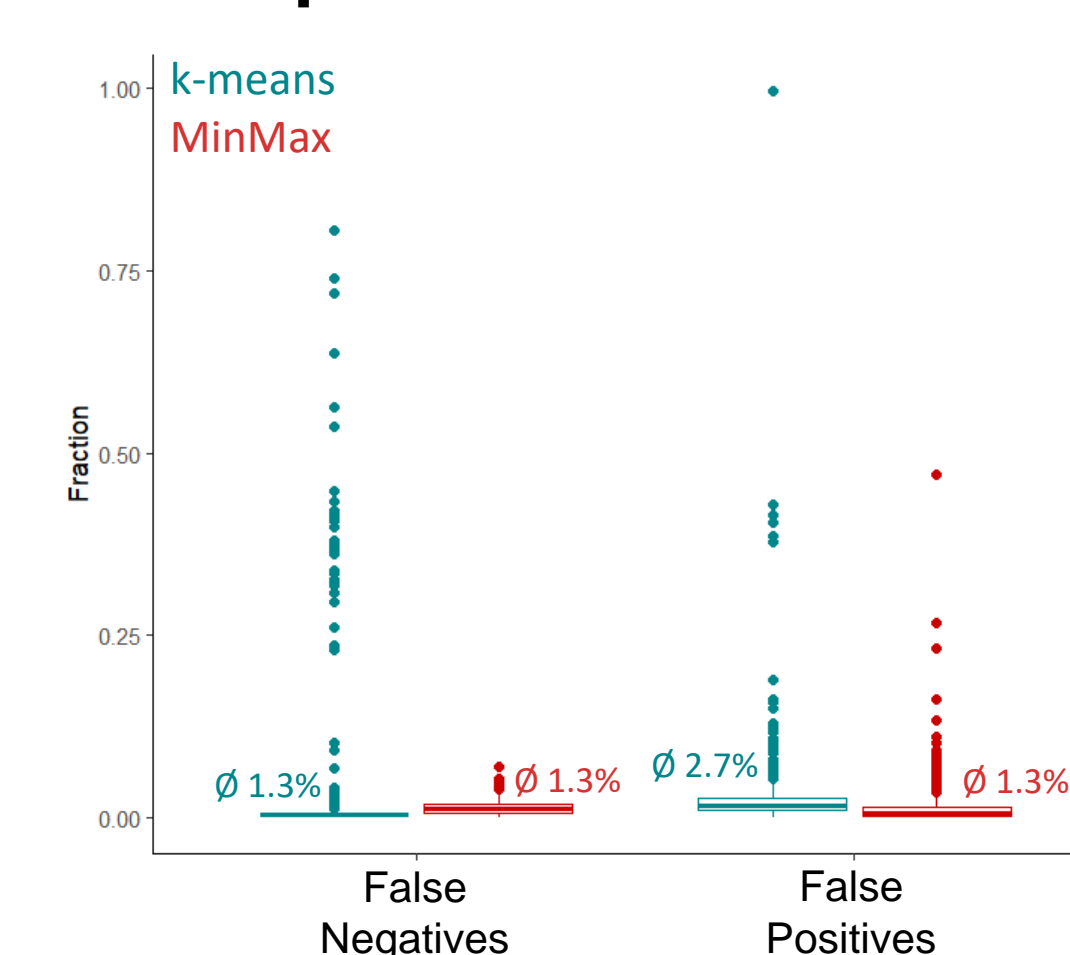
	Manual	k-means	MinMax
runtime	~ 7.4 h	~ 1 h	~ 45 min

Note: The majority of the runtime for the automated data analyses is needed for loading and opening of the fcs files.

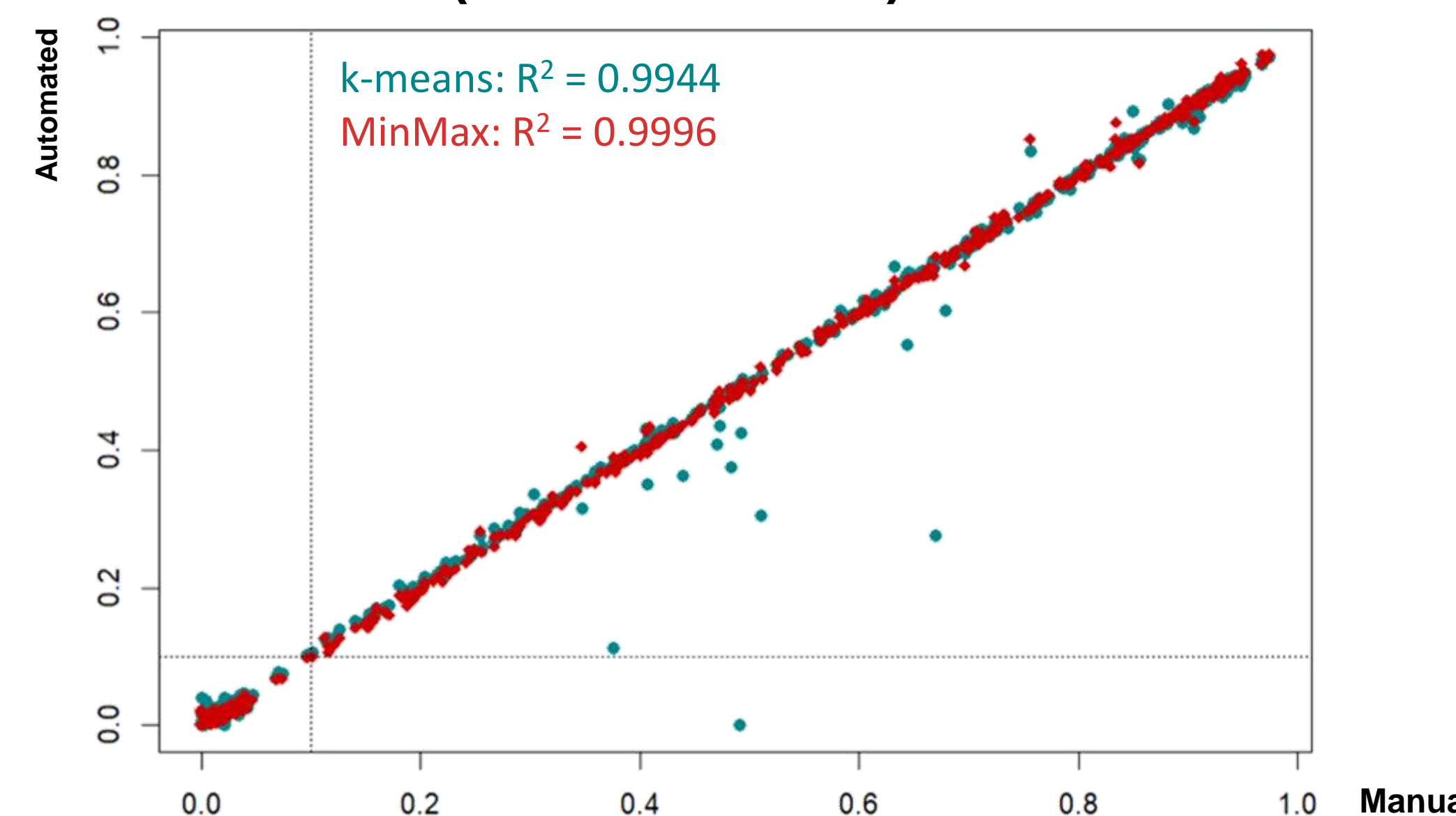
Basophil identification: Example



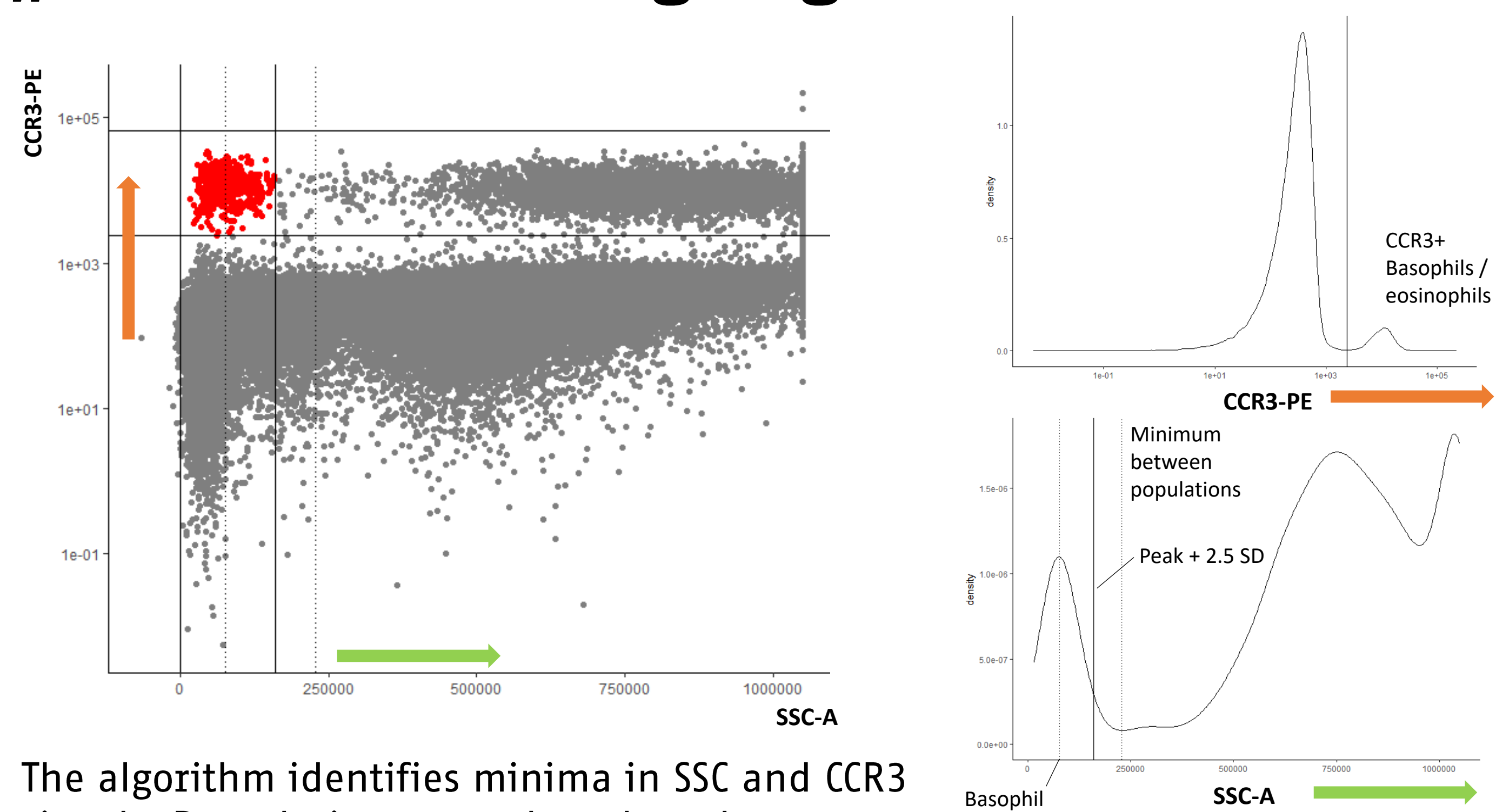
Basophil identification error rates



Basophil activation results (fraction of CD63+)



„MinMax“ Gating Algorithm



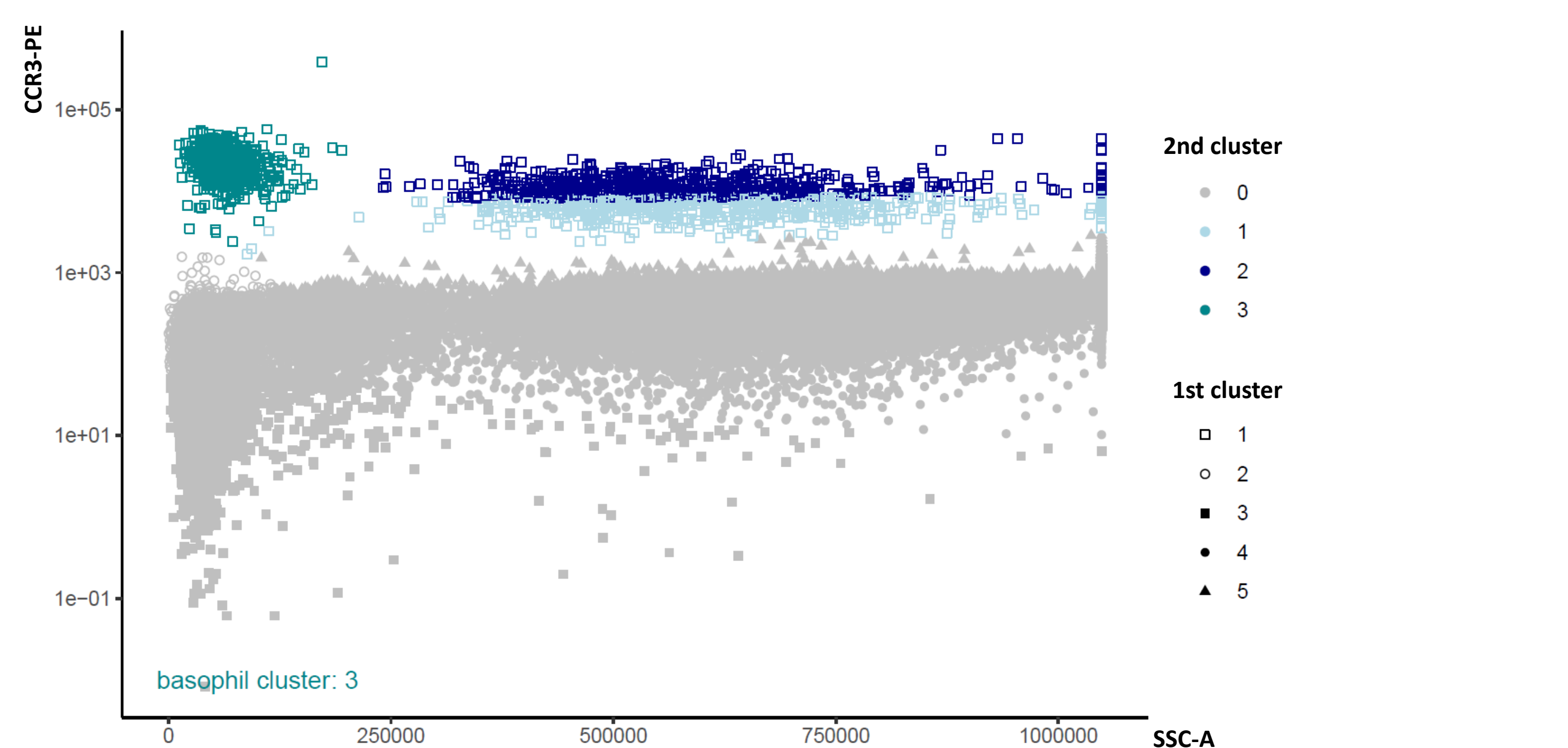
The algorithm identifies minima in SSC and CCR3 signals. Boundaries are set based on the basophil peak maximum + 2.5xSD.

K-Means Clustering Algorithm

2 step process:

- 1st clustering to identify CCR3^{high} subpopulation
- 2nd clustering to identify SSC^{low} subpopulation

Clustering is performed on log10-transformed, centered, and scaled data.



Random Forest Tree Classification

Machine Learning Procedure:

- Training set built on 40 randomly selected BAT measurements with manual annotation.
- Non-basophils downsampled to 2000 per BAT for training Random Forest
- Random Forest function is applied to 884 BAT datasets and compared to manual gating

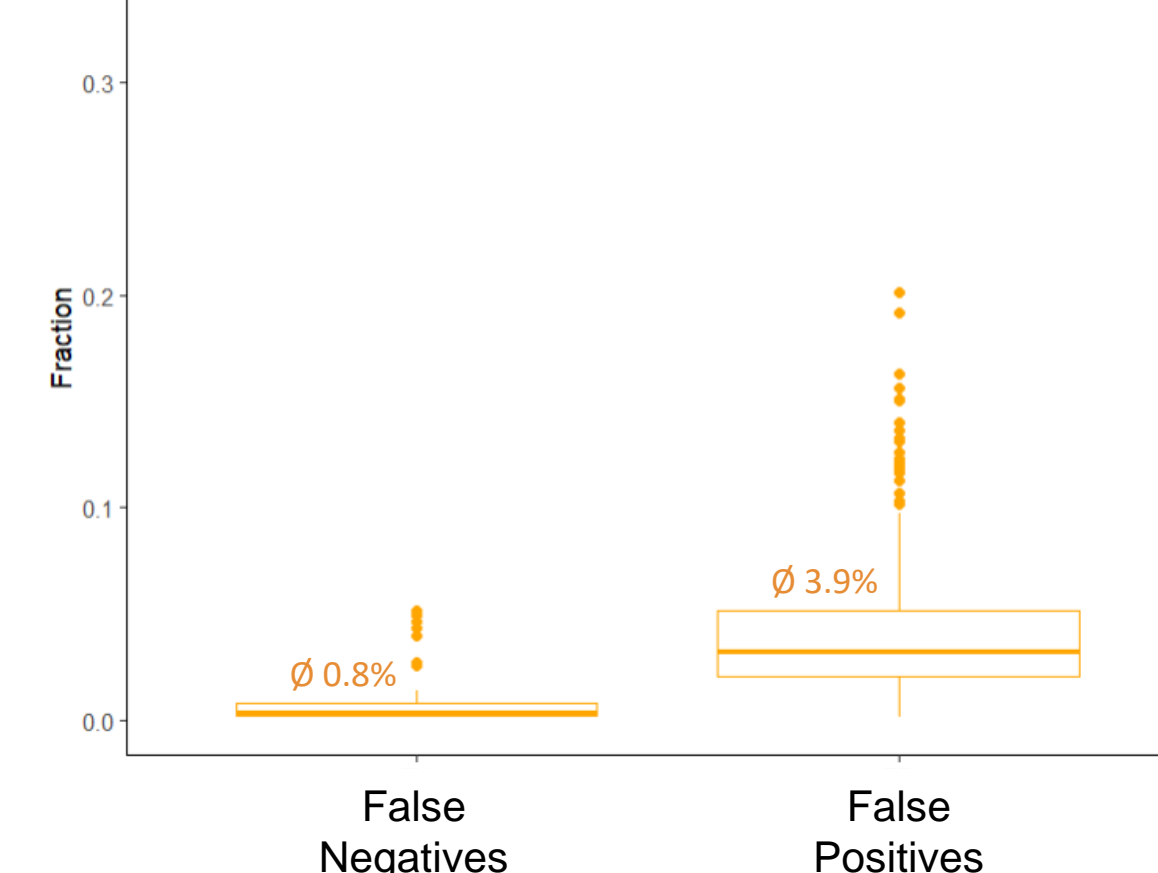
Model Descriptors

OOB estimate of error rate: 0.03%

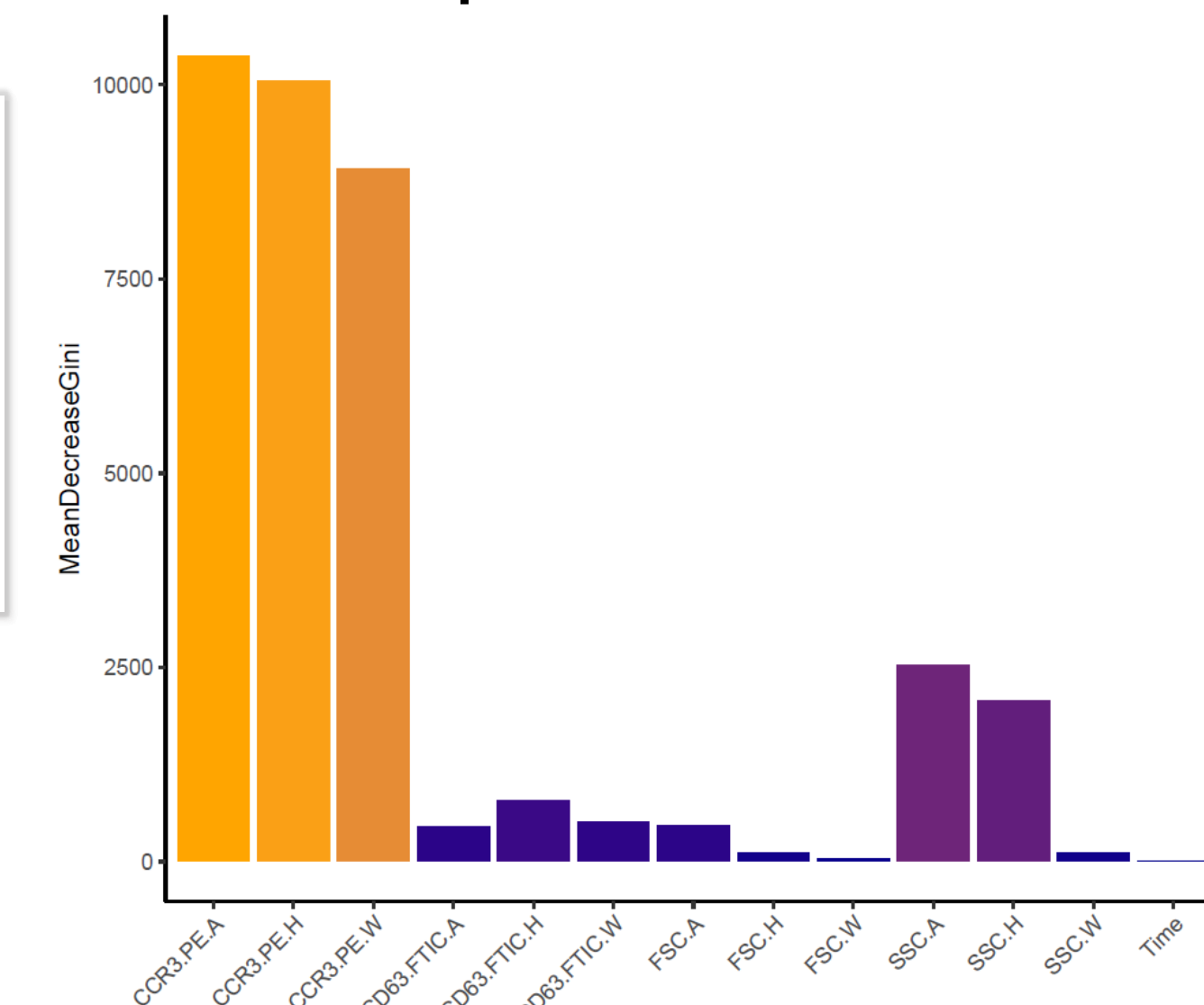
Confusion matrix:

	FALSE	TRUE	class.error
FALSE	79975	25	0.0003125000
TRUE	10	23597	0.0004236032

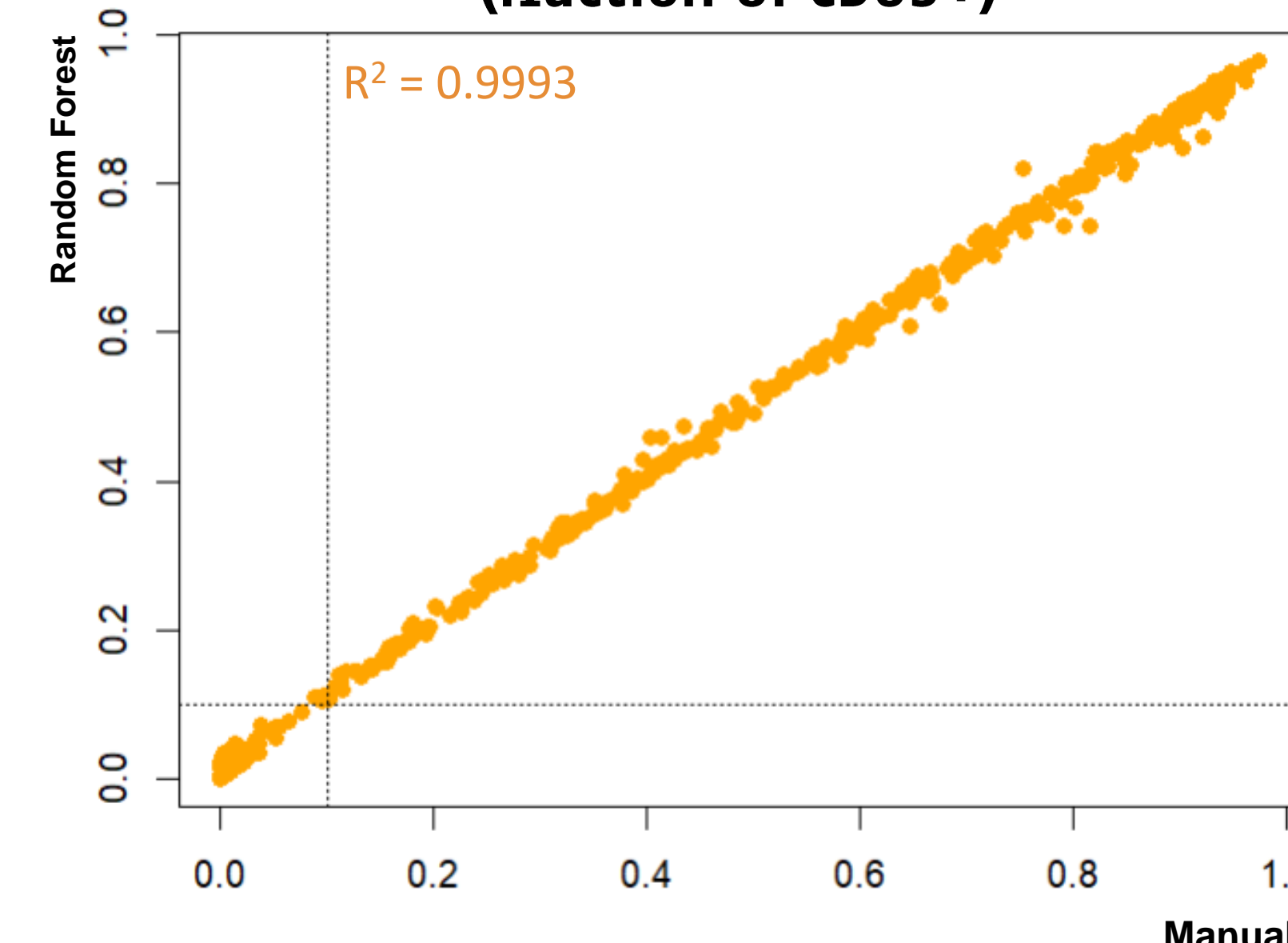
Misclassification Rates (compared to manual gating)



Feature Importance



Basophil activation results (fraction of CD63+)



CONCLUSIONS:

- Automated gating of basophils is highly comparable to manual gating and has very little effect on the result of the BAT.
- The „MinMax“ gating method slightly outperforms k-means clustering and random forest tree classification.