FlowCAP-I Debrief

Richard H. Scheuermann, Ph.D.

U.T. Southwestern Medical Center

FlowCAP-I is a success!

Participation in FlowCAP-I was better than expected

1st generation algorithms before better than expected

Interest in the stakeholder community was better than expected

Comparison against manual gating good choice for FlowCAP-I; but perhaps more selective for FlowCAP-II

Other positives

Easy to participate

Excellent responsiveness

Critical components

“standard” datasets

objective evaluation criteria

Challenge

Maintain the momentum

Learn from the experience

Get the word out

Ongoing support

Rapidly attain the goal of making computational algorithms an essential component of standard FCM data analysis

Room for improvement

Manual gating as “gold standard”

Handling of “outliers”

Evaluation metric

Dataset use case coverage

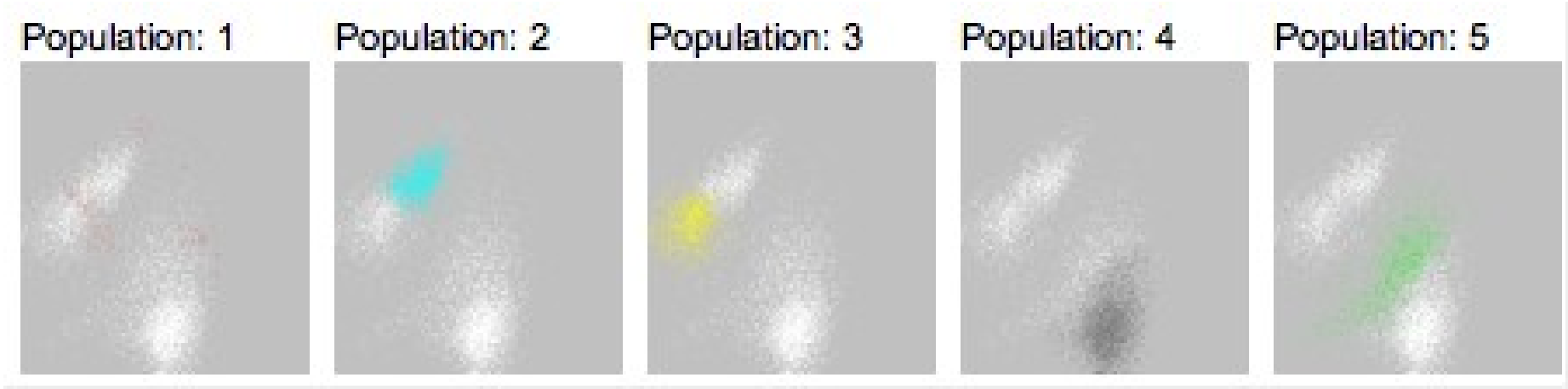
4 challenges

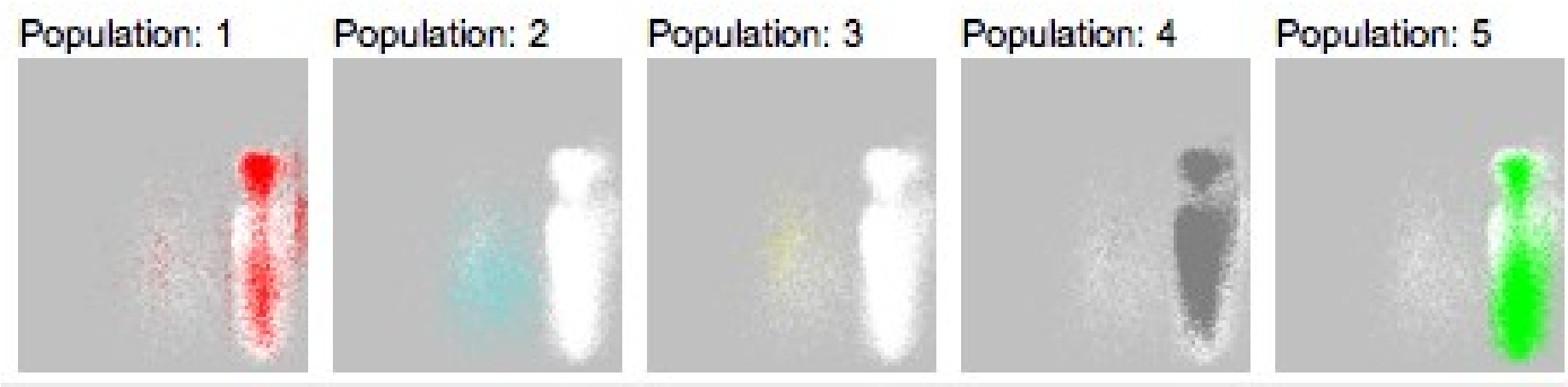
Sufficient time

Sufficient information

Others

The elephant in the room - Should manual gating be the gold standard?





Why weren't WNV and ND included in Challenge 3?

We knew that we didn't have a good estimate of "k"

Manual gating

Exhaustive gating vs. selective gating

Discovery (clustering) vs. classification

Need both

Manual gating needs to be done carefully by explicitly guiding the gaiter

Were outliers handled properly?

Every cell that was not included in the manual analysis by the human expert (due to noise or lack of biological interest) will be considered as an outlier for the purpose of this challenge. Algorithms will not be penalized for assigning an incorrect label to cells that are marked as outliers by these criteria. However, predicting biologically relevant (i.e., non-outlier) cells as outliers (with not assigning that cell to a cluster) will penalize the algorithm. Therefore, our advice is that the algorithms should analyze all

Outliers

Outliers/noise should not be excluded from the analysis if “k” is given, unless they are filtered from the dataset

One “k” for outliers may not be sufficient if included

Objective evaluation metrics

Is F-measure a sufficient metric?

Is time relevant?

Did we have sufficient datasets?

5 datasets

115 samples total

Maximum # of events = 100,000

Maximum # of markers = 10 + 2 (only 17,000 events)

Datasets did not represent the scope of the problem well

Relatively small number of events

Not enough high dimensional data

Cross sample comparison not included

Rare population use case not explicitly represented

Requested datasets

Scientific use cases

Detection of rare cell populations (e.g., minimal residual disease in cancer);

Enumeration of large numbers of distinct cell populations in high dimensional flow cytometry data (e.g., >10 colors);

Discovery of clinically relevant cell populations in patient cohorts associated with disease states (e.g. markers of autoimmune disease, survival indicators in lymphoma);

Did we need 4 different challenges?

Completely automated

Tuned algorithm

Population number

Trained algorithm (supervised classification)

Was there sufficient time for each challenge?

3 months for Challenge 1 and 2

3 weeks for Challenge 3

3 weeks for Challenge 4

Were the dataset descriptions sufficient for biological interpretation?

"Data sets should also be accompanied by metadata descriptions about the specimens and staining procedures used compliant with the MIFlowCyt data standard"

Difficult to arrive at any biological interpretation

Data formating issue

Need for better standardization of available file formats

Is the competition agreement reasonable?

Publishing the datasets provided by flowCAP is prohibited until the project publishes the results.

The datasets and results of the flowCAP project will be publicly available for any use after the summit.

Software submitted to flowCAP will remain confidential.

Participant won’t be identified (by name, group name, etc) in any materials without their approval