

# Automated Bayesian Gating with OpenCyto

John A. Ramey, Ph.D.

Postdoc, Gottardo Lab

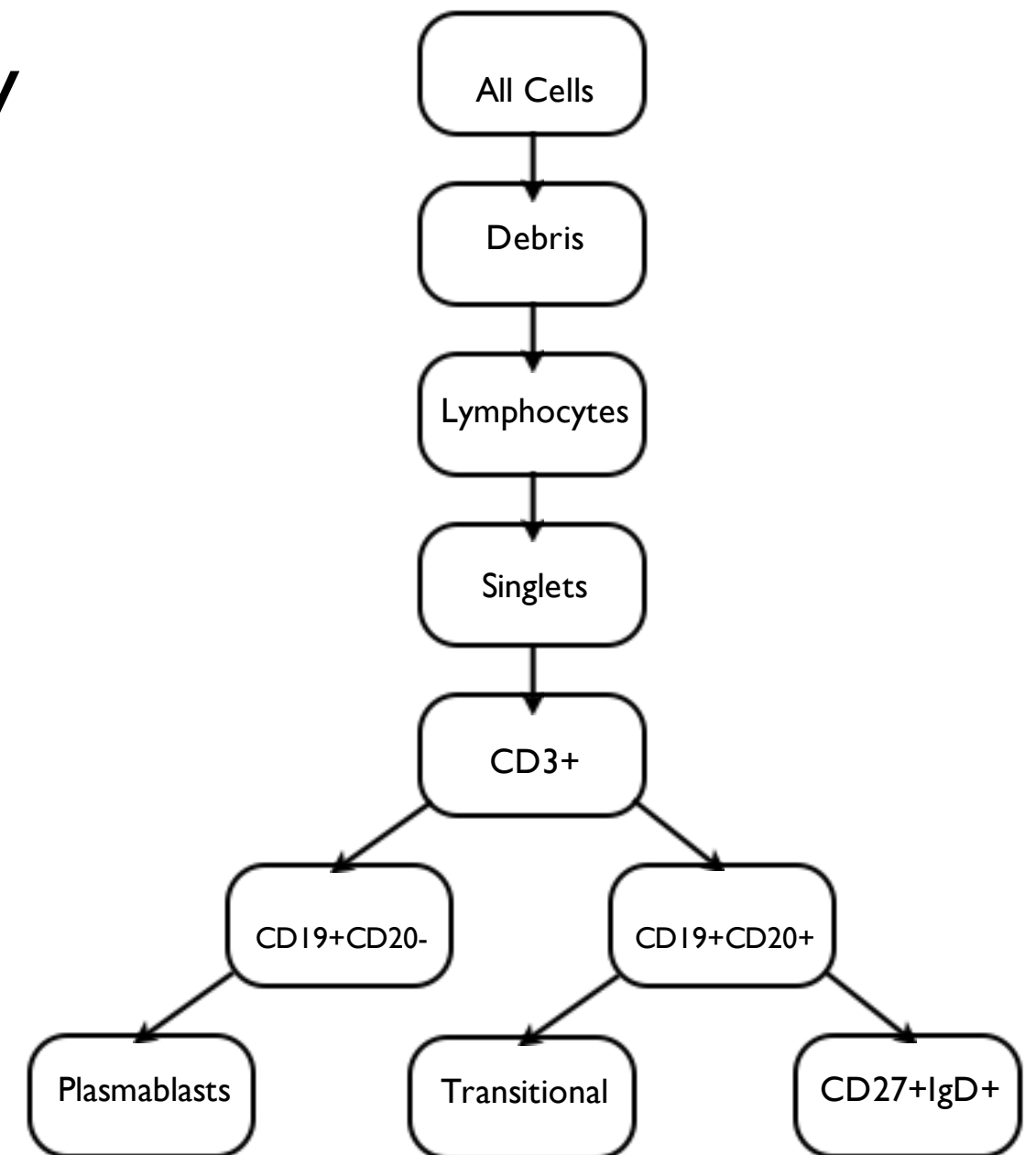
Fred Hutchinson Cancer Research Center

# OpenCyto Infrastructure

- Fast, robust automated gating
- Automated pipelines incorporating expert knowledge
- Fast processing of large data
- 1 GB max memory consumption
- C++ libraries and other technologies: netCDF, boost, serialization
- R Packages

# General Strategy

- Pipeline based on a specified gating hierarchy
- Data-derived gates for each sample using hierarchical gating
- Gate boundaries are data-derived
- Gating with Bayesian mixture models (flowClust 3.0)
- Priors are marker-specific, data-driven, and



# Challenge #3

- Pipeline followed the manual gating strategy
- Used flexible mixture models for negative peak fitting and quantiles for cytokine gates (rare populations)
- Extracted all Boolean subsets with associated proportions (features)
  - Example: (CD4) IL2+ and !IFNg+ and TNFa+
- LASSO-based classifier using the **glmnet** package, shrinkage parameter selected via cross-validation

# Challenge #3: Training Results

- Features selected: Antigen-specific T-cells

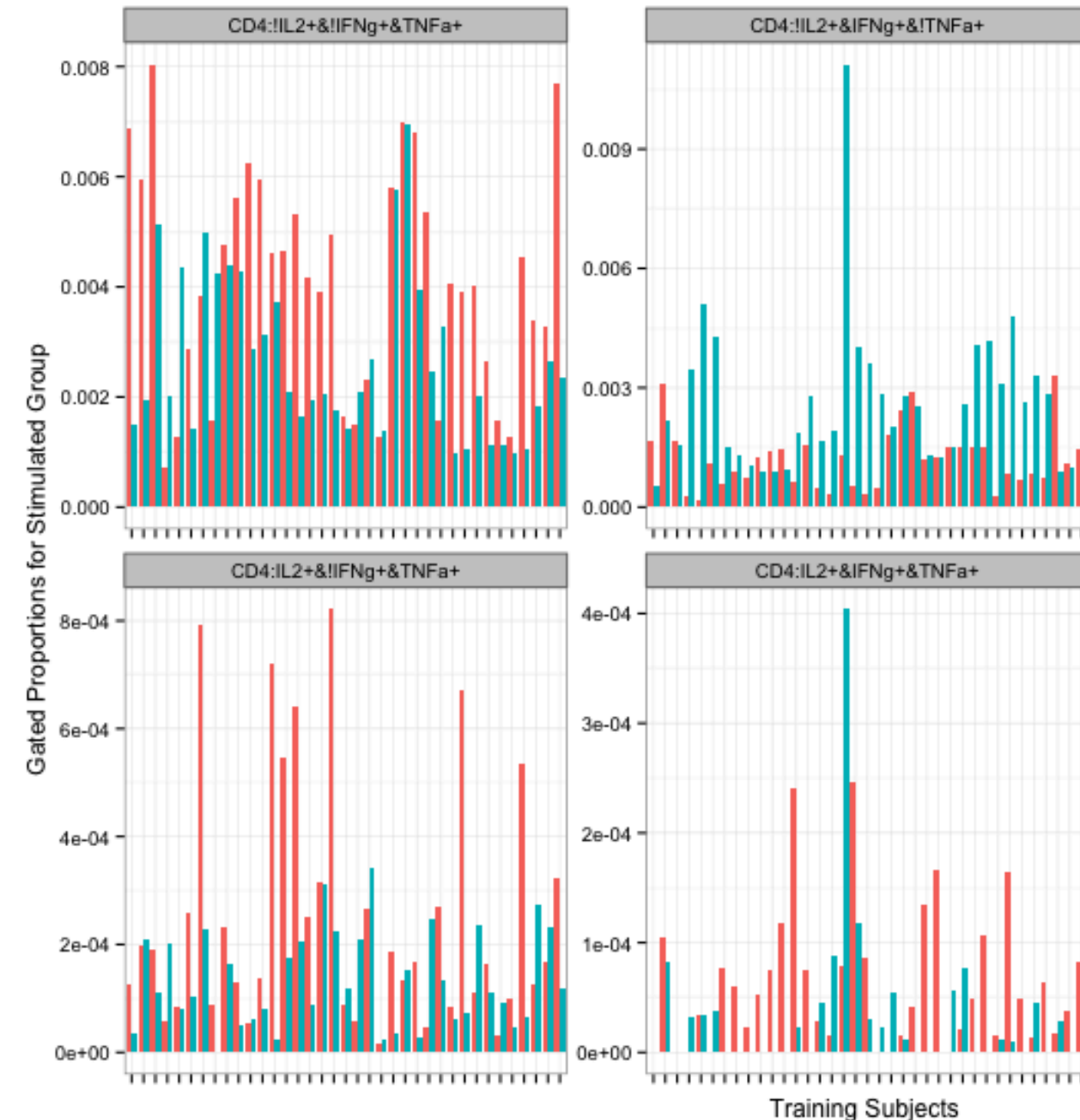
- IL2+ and !IFNg+ and TNFa+

- IL2+ and IFNg+ and TNFa+

- !IL2+ and !IFNg+ and TNFa+

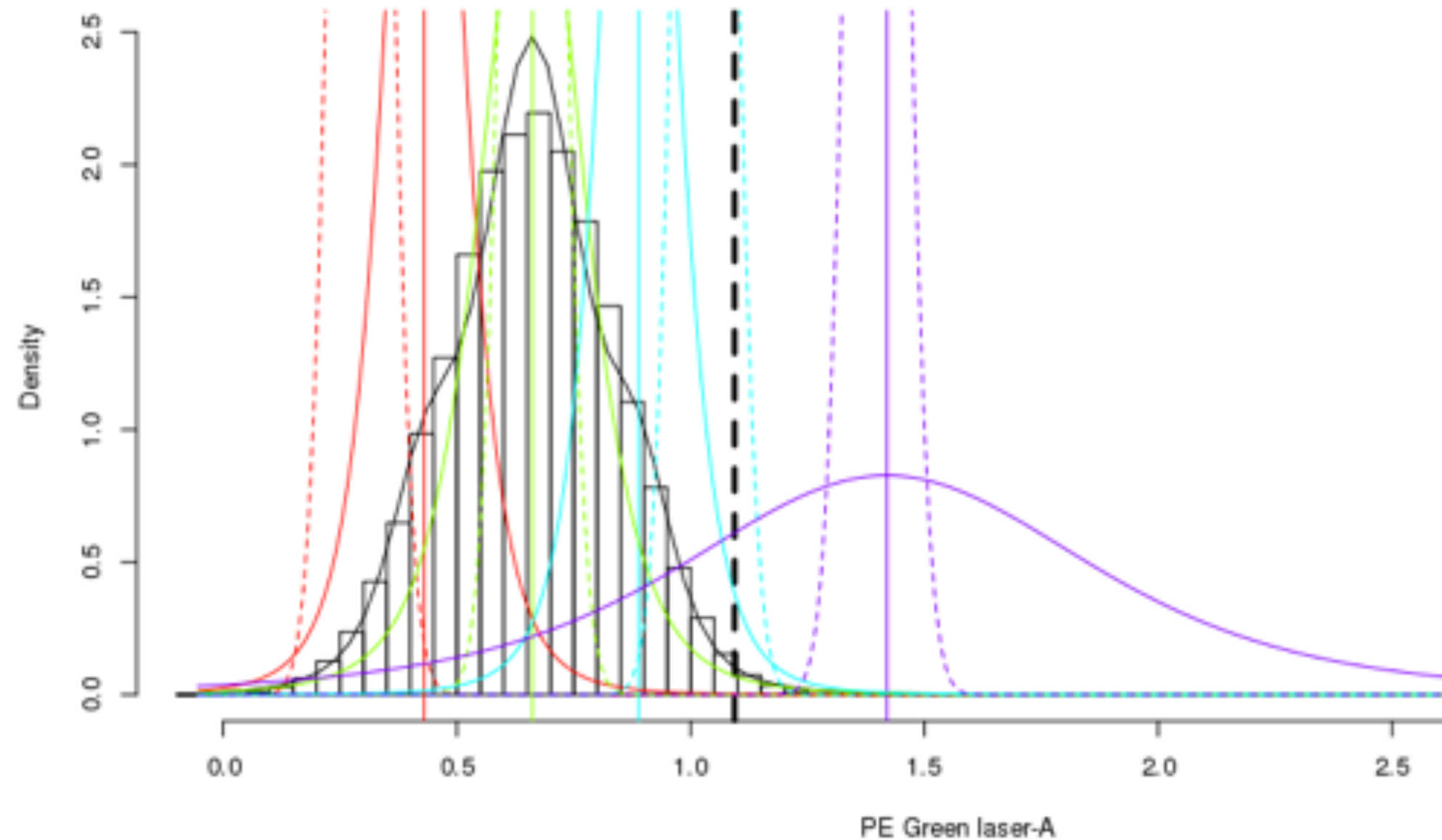
- !IL2+ and !IFNg+ and !TNFa+

- Classification separation from the



# Cytokine Gate - CD4/IL2+

- Negative population - 3 mixture components
- Positive population - 1 mixture component
- Prior means - dashed densities
- Posteriors - solid densities
- Gate - Black, vertical dashed line

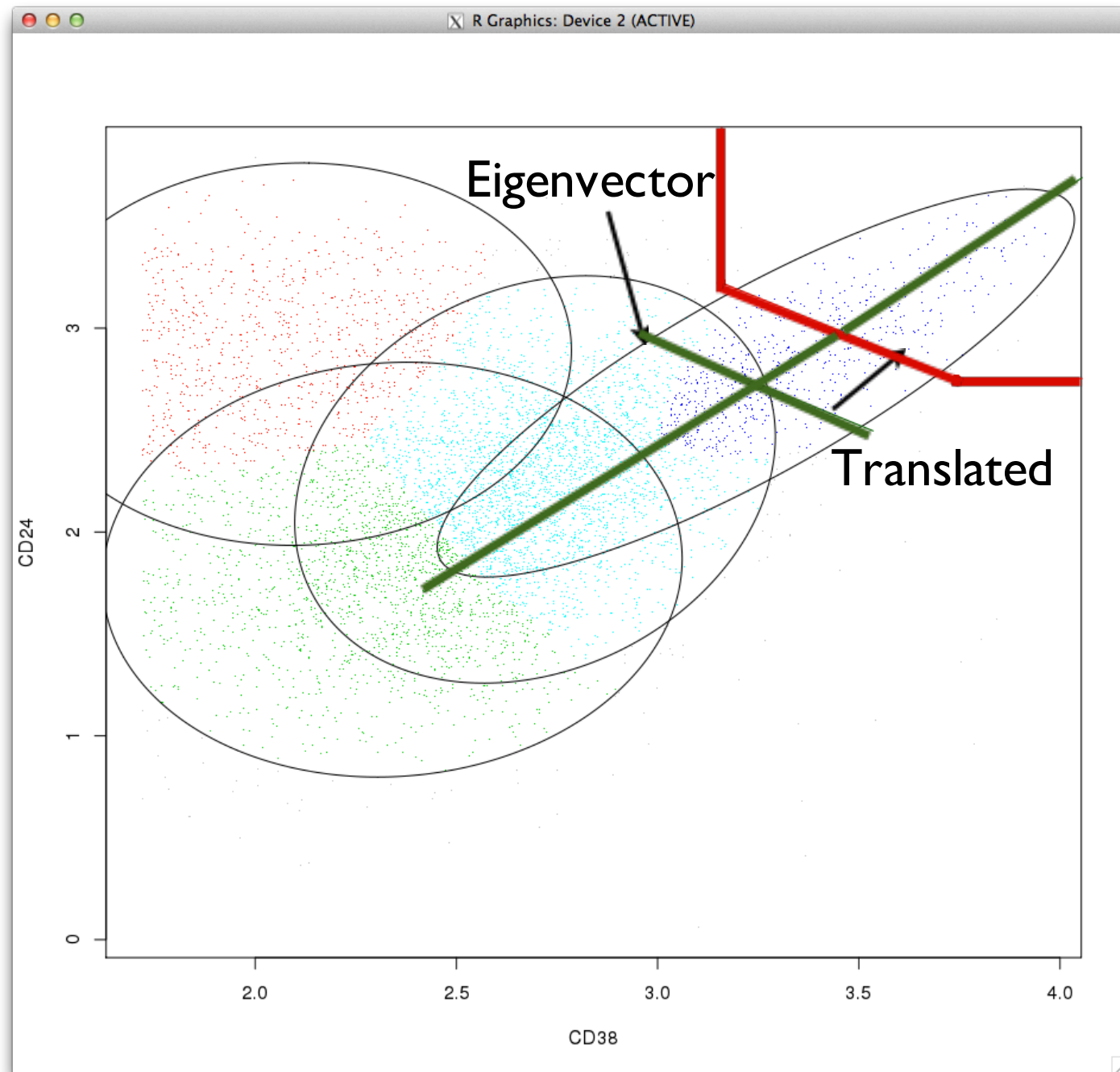


# Challenge #4

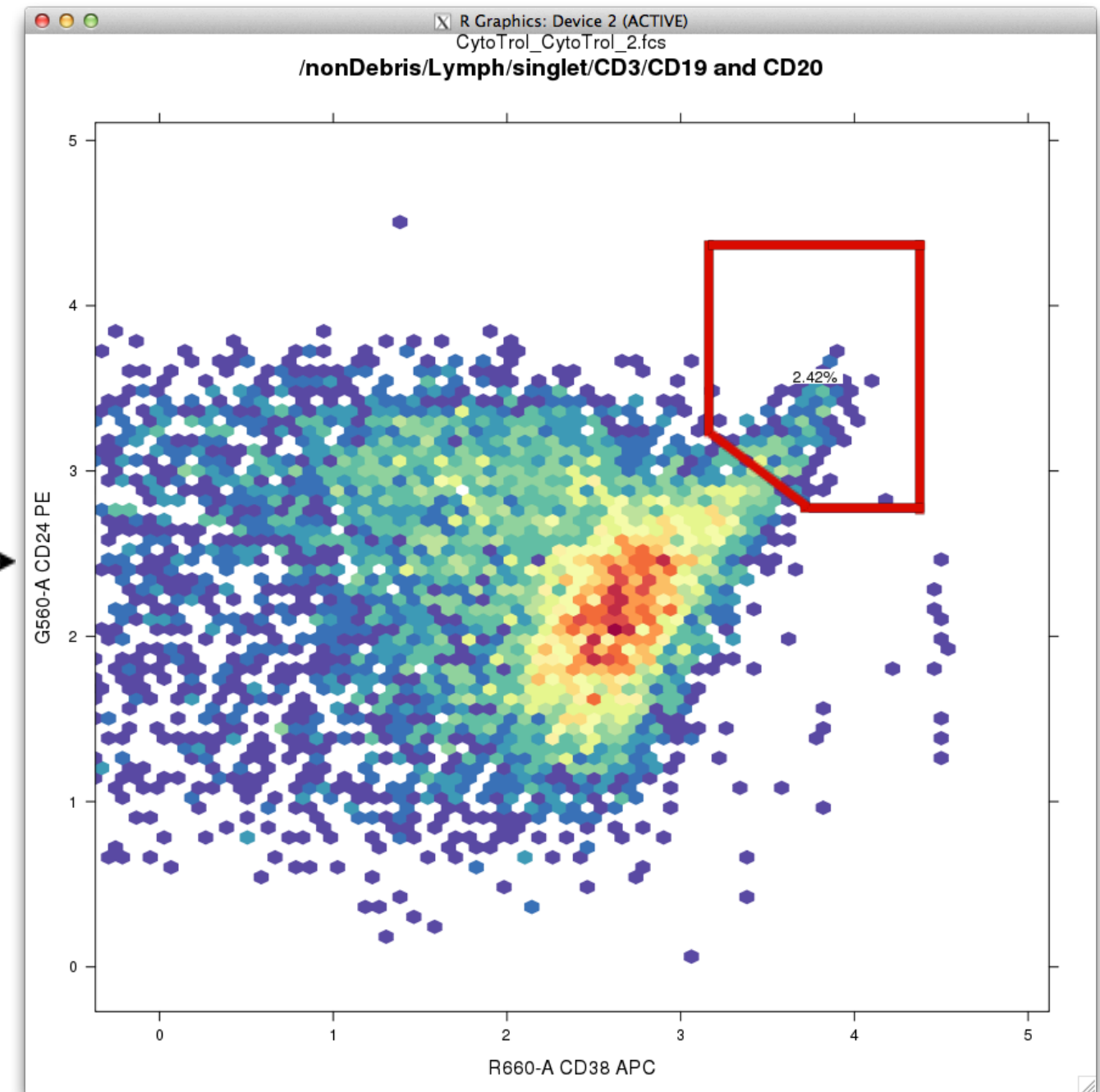
- Pipelines followed the manual gating strategy
- Marker-specific, data-driven priors
- Gate all centers <30 seconds
- B-Cell pipeline more difficult than T-Cell pipeline
- Difficult gates: Transitional, IgD+, Plasmablasts

# Difficult Gate: Transitional

Model Fit



Resulting Gate



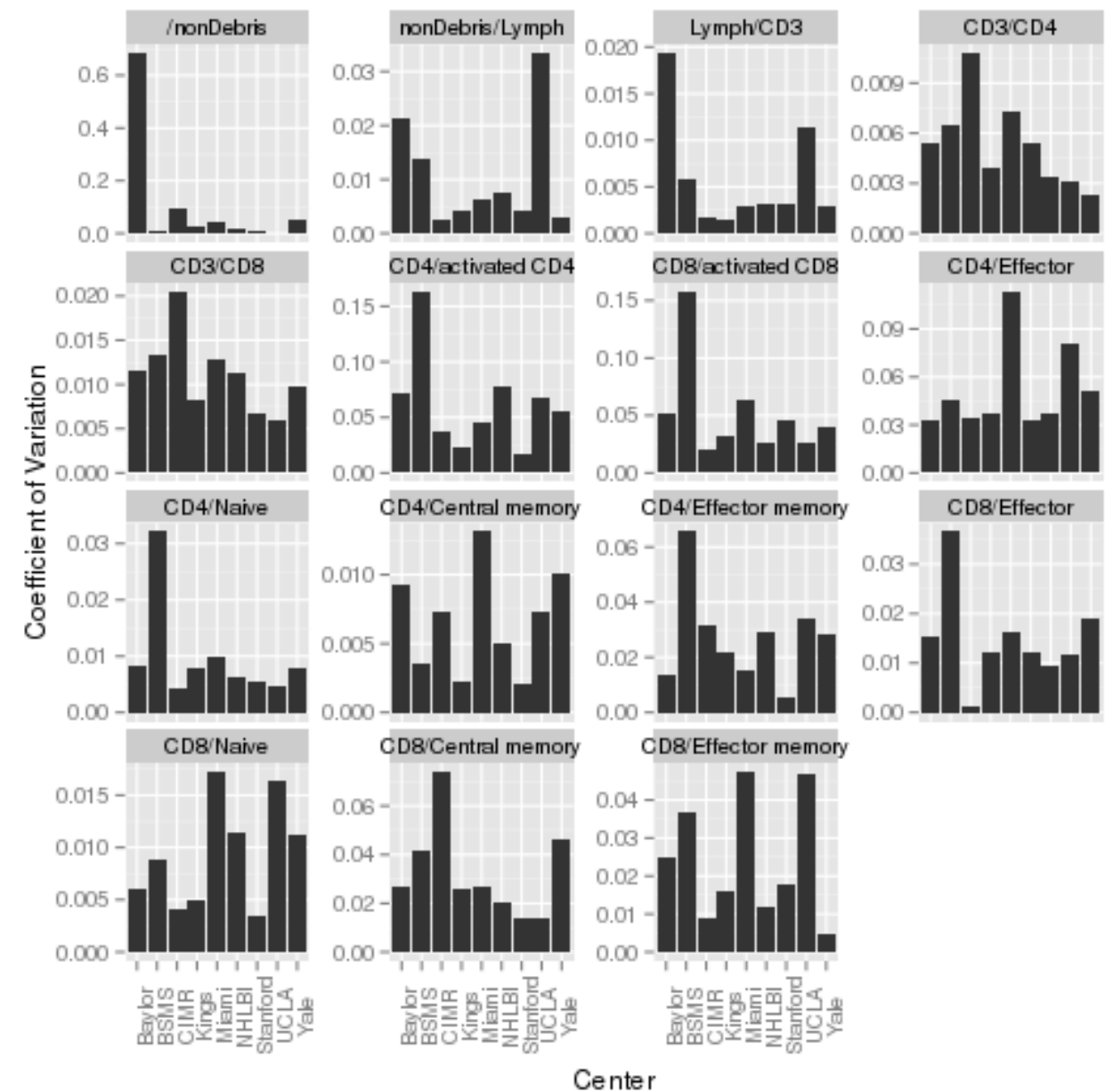
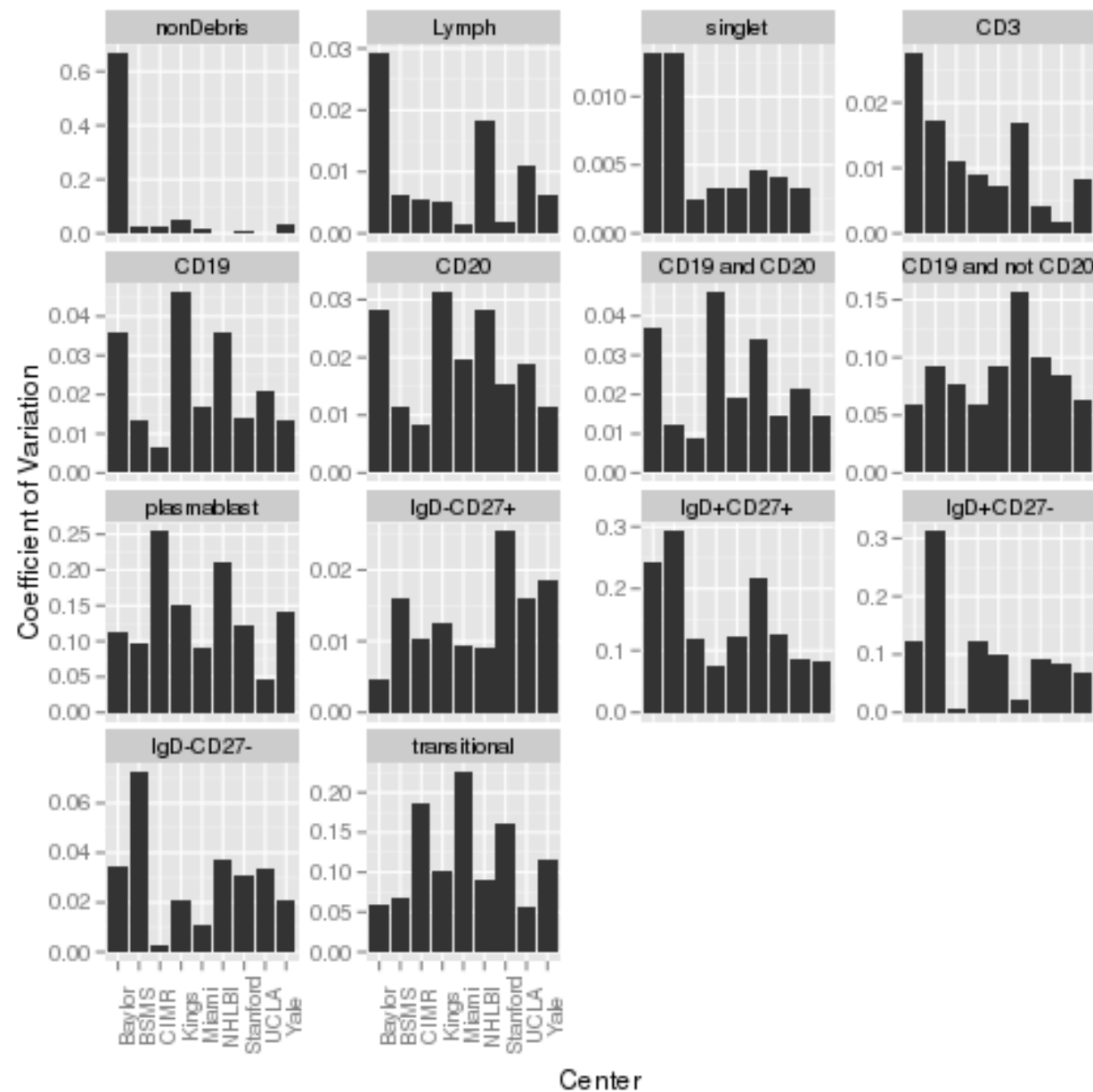


# Coefficients of Variation within Center

B-Cell

Most CV's < 0.05

T-Cell



# Conclusion

- OpenCyto:
  - Incorporates **expert** and **data-driven** prior knowledge
  - Yields **accurate** reproduction of manual gating schemes in an automated manner
  - Attains **robust**, accurate gating of rare cell populations
  - Is **flexible** - can be applied in fully automated gating scenarios. (i.e. learn priors from fully automated data)

# Acknowledgements

Funding

HIPC

NIH

NIAID

HVTN

R Package Development

Mike Jiang

Greg Finak