Automated
Flow Cytometry Analysis

R. BURKE SQUIRES

License and Sources

LICENSE

This material is licensed under the **Creative Commons Attribution-ShareAlike 4.0 License**.

SOURCES

Ryan Brinkman, PhD

Bioinformatics.ca
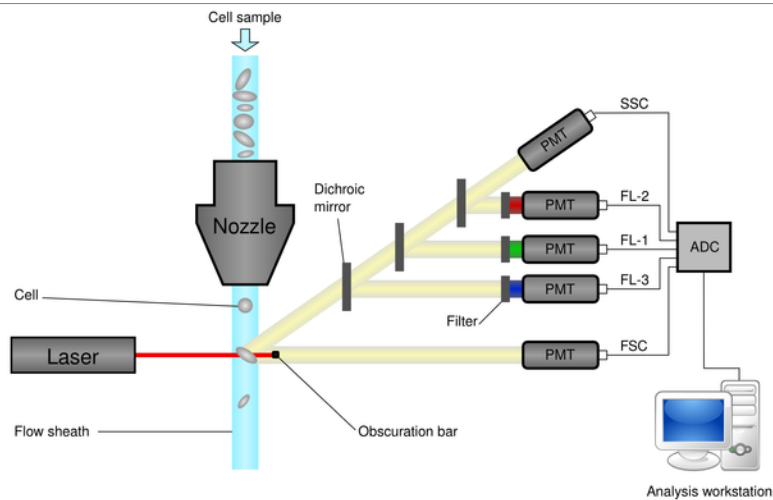
Outline

Flow Cytometry Review

The State of Automated Flow Cytometry Analysis
◦ Flow Cytometry Standards and Analysis (FlowCAP)

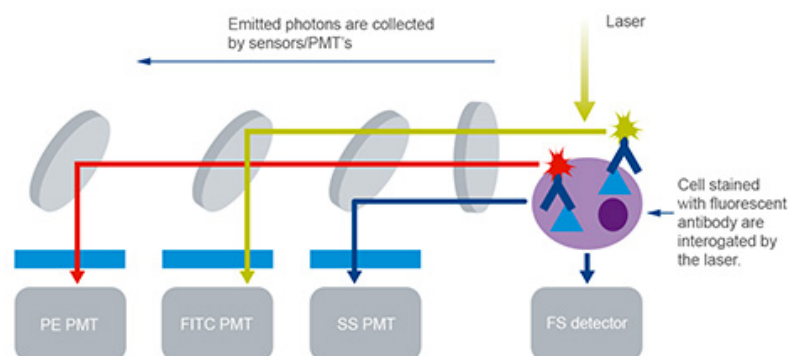Automated Flow Cytometry Analysis

Flow Cytometry Review

Schematic Diagram Of A Flow Cytometer

Fluorescent Light Is Filtered So That Each PMT Detects A Specific Wavelength

Fluorochromes

Fluorochrome

Abcam
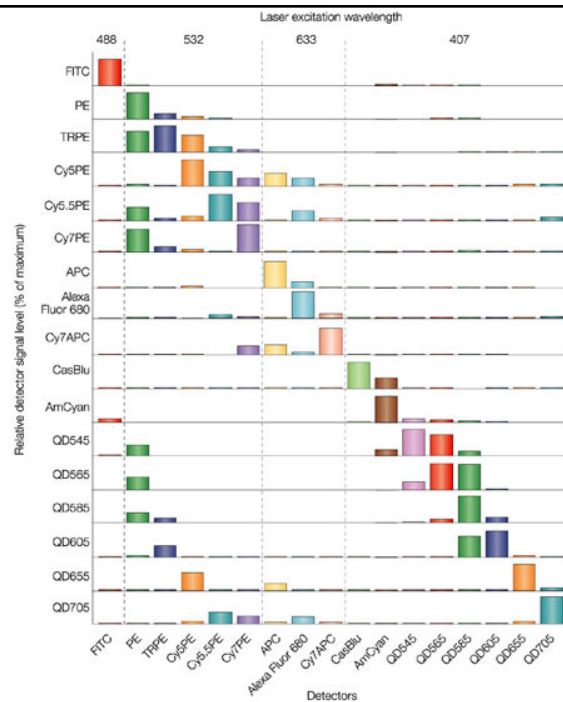
Manual Gating (polygons) Vs. Automated Gating (colored dots)

F-measures:
0.86
1.00
0.99

Results of:
- Algorithms
—Manual gates

FITC-CD45.1

PE-Ly65Mac1

APC-CD45.2
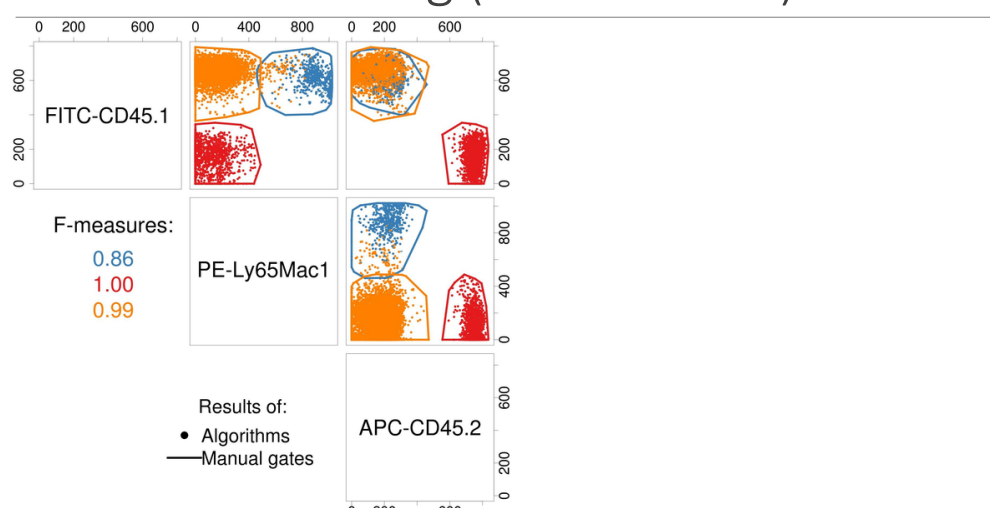
O'Neill K, Aghaeepour N, Špidlen J, Brinkman R (2013) Flow Cytometry Bioinformatics. PLOS Computational Biology 9(12): e1003365. doi:10.1371/journal.pcbi.1003365
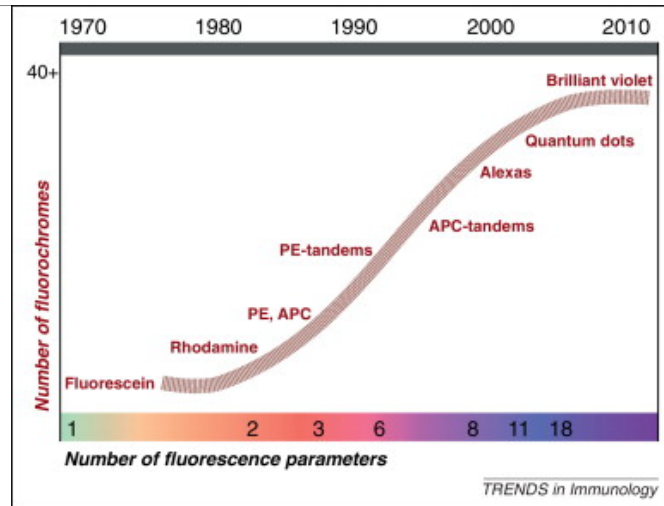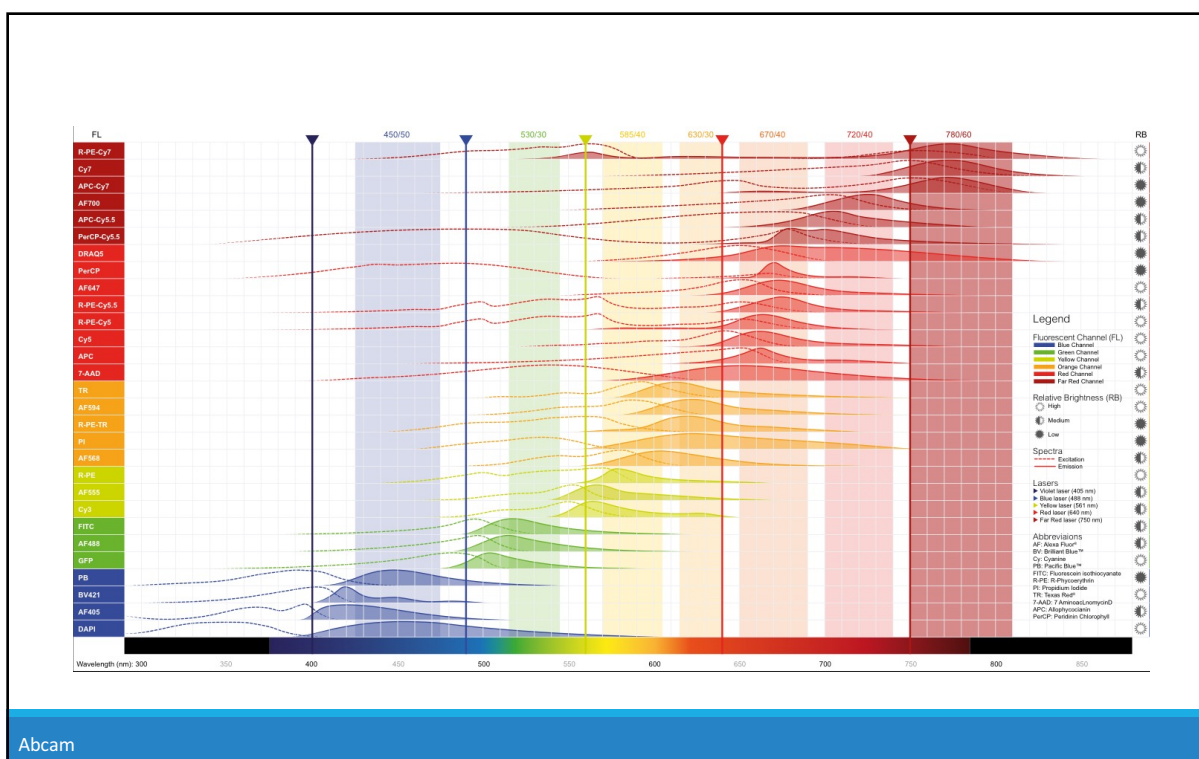http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003365

PLOS | COMPUTATIONAL BIOLOGY

Automated
Flow Cytometry Analysis

Automated Flow Cytometry Analysis

"Automated algorithms for flow cytometry data analysis have reached a level of maturity that enables them to match and in many cases exceed the results produced by human experts."
◦ Aghaeepour et al., Nature Methods, 2013

Supervised gating (diagnosis): OpenCyto, flowDensity
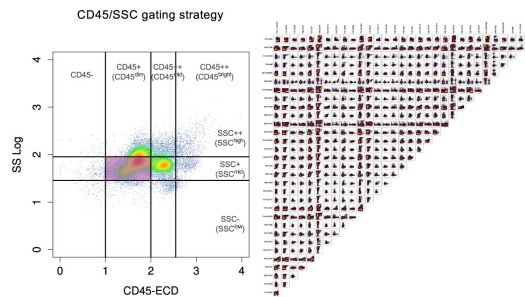◦ Malek et al., Bioinformatics, 2014

Unsupervised biomarker discovery: FlowReMi, flowType/RchyOptimyx
◦ O'Neill et al., Bioinformatics, 2014

Ryan Brinkman

Big Flow Cytometry Data

AUTOMATED ANALYSIS OF FLOW CYTOMETRY
DATA IN R/BIOCONDUCTOR

CD45/SSC gating strategy

	1985[1]	2012[2]	2016+[3]
Samples	1	466	77,000
Dimensions	5	13	50
Cells	50,000	400,000	1,000,000
Datasets	2.5^5	2.5^9	4^11

[1] Murphy *Cytometry* (1985)
[2] Aghaeepour *et al. Bioinformatics* (2012)
[3] International Mouse Phenotyping Consortium (2015)

Ryan Brinkman

Can Manual Analysis of
"High-Dimensional" Data Be Improved Upon?

Time consuming, especially for "discovery"

Analysis guided by history with limited, intuitive exploration

Rarely (ever?) examine entire multidimensional dataset

Significant cross-individual variability (>10%)

No appropriate statistical basis to assess relative significance

Not fun (?)

"Unfortunately, the use of three or more independent fluorescent parameters complicates the analysis of the resulting data significantly." Murphy Cytometry (1985)

"Despite the technological advances in acquiring [30] parameters per single cell, methods for analyzing multidimensional single-cell data remain inadequate." Qiu et al. Nature Biotechnology (2011)

Ryan Brinkman

>50 Peer-reviewed, Free, Open Source Software Tools

45 R/BioConductor for data analysis

- A scripted approach to high throughput analysis
- Non-interactive, self-documented, repr
- Breaks problem into smaller pieces (pa
- Modules can plug-in & swap-out
- Collaborative, cross-platform developm environment

9 additional software tools

- Java, Python, Matlab, C++
- Stand alone (single problem/solution)

Immunity
Letter

CelPress

Thinking Outside the Gate:
Single-Cell Assessments in Multiple Dimensions

Pia Kvistborg,[1,19] Cécile Gouttefangeas,[2,19] Nima Aghaeepour,[3] Angelica Cazaly,[4] Pratip K. Chattopadhyay,[5] Cliburn Chan,[6] Judith Eckl,[7] Greg Finak,[8] Sine Reker Hadrup,[9] Holden T. Maecker,[10] Dominik Maurer,[11] Tim Mosmann,[12] Peng Qiu,[13] Richard H. Scheuermann,[14,15] Marij J.P. Welters,[16] Guido Ferrari,[17] Ryan R. Brinkman,[8,20,*] and Cedrik M. Britten[18,20,21,*]

Ryan Brinkman

BioConductor
- Open, Extensible Infrastructure
- Packages are Interoperable & Interchangeable

Ryan Brinkman
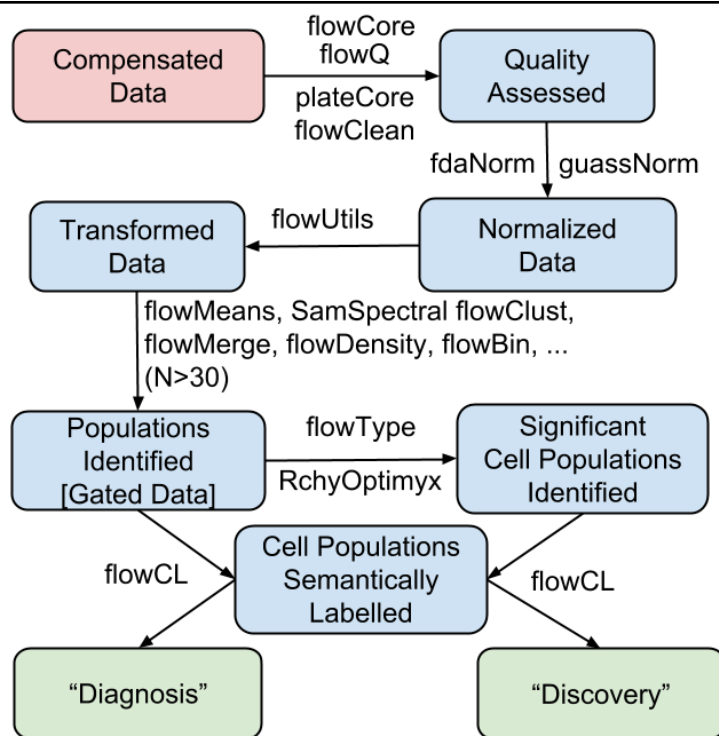
Which Automated Analysis Methods to Use?

Ryan Brinkman

Flow Cytometry Standards and Analysis (FlowCAP)

FlowCAP (Critical Assessment of Automated Analysis Methods)

Community-based evaluation of flow bioinformatics tools

FlowCAP-I (2010): Matching manual gating (1st gen tools)

FlowCAP-II (2011): (Too simple) sample classification

FlowCAP-III (2012): 2nd generation automated gating tools
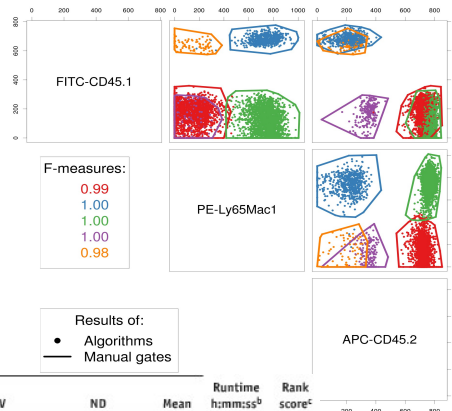
FlowCAP-IV (2014): (Hard) Biomarker discovery

http://flowcap.flowsite.org

Ryan Brinkman

FlowCAP-I: Unsupervised gating (discovery) = humans'

Individual performance can vary on specific cell populations (might not matter)
- *Aghaeepour et al., Nature Methods (2013)
- http://flowcap.flowsite.org

15 different tools on 5 datasets

FITC-CD45.1

F-measures:
0.99
1.00
1.00
1.00
0.98

PE-Ly65Mac1

APC-CD45.2

Results of:
• Algorithms
— Manual gates

	GvHD	DLBCL	HSCT	WNV	ND	Mean	Runtime h:mm:ss[b]	Rank score[c]
Challenge 1: completely automated								
ADICyt	0.81 (0.72, 0.88)	0.93 (0.91, 0.95)	0.93 (0.90, 0.96)	0.86 (0.84, 0.87)	0.92 (0.92, 0.93)	0.89	4:50:37	52
flowMeans	0.88 (0.82, 0.93)	0.92 (0.89, 0.95)	0.92 (0.90, 0.94)	0.88 (0.86, 0.90)	0.85 (0.76, 0.92)	0.89	0:02:18	49

Ryan Brinkman

There is No Single Best Gating Solution (2010)

Ryan Brinkman
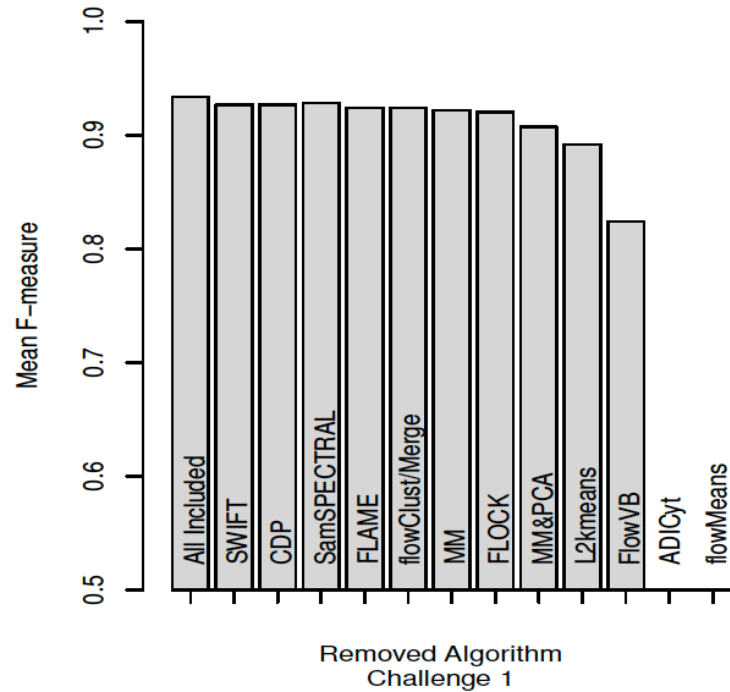
FlowCAP-II:
Tools for Clinical Classification
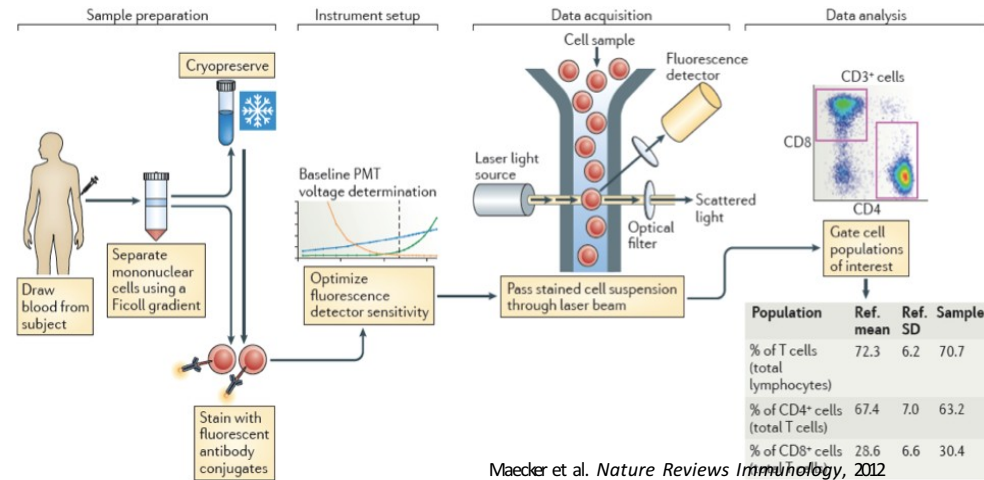
Several algorithms performed perfectly

There were misclassifications

	Sensitivity	Specificity	Accuracy
flowType-FeaLect	1.00	1.00	1.00
flowPeaks	1.00	1.00	1.00
SPADE	1.00	1.00	1.00
...	...	...	...

Ryan Brinkman

FlowCAP-III: Reducing Variability in Translational Immunology

Sample prep (BD lyoplates), Instrumentation, Acquisition, Automated Analysis

Maecker et al. *Nature Reviews Immunology*, 2012

Ryan Brinkman

FlowCAP-III:
Supervised Analysis for "Diagnostics"

2 automated tools can match human gating

9 clinical sites, 4 replicates of cryopreserved cells per site.

Centralized gating of data based on a consensus best approach.

Automated algorithms vs. centralized gating.
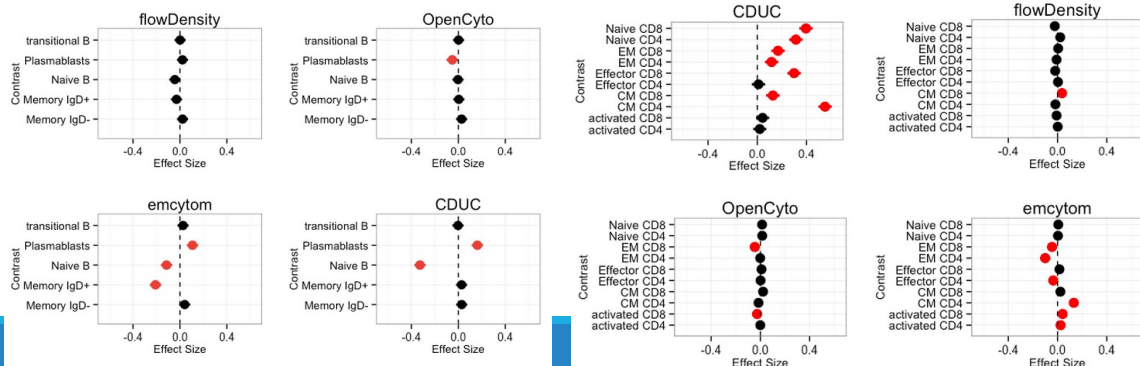
Ryan Brinkman

Gating Transitionals From Lymphocytes

MANUAL: 9.38   AUTO: 10.56
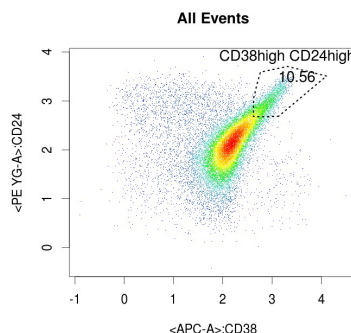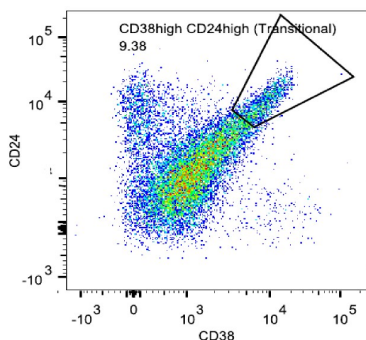
Ryan Brinkman

FlowCAP-III Conclusion:
Stop Manually Gating

Supervised gating can match manual analysis when it is data-driven

Automated gating is unbiased relative to manual gating

Variability is as low or lower than manual gating

Even when biased, the bias is associated with populations that have low cell counts and CV is lower than manual gating

Not following SOPs can result in large variability

*Finak *et al., Nature Scientific Reports* (2016)

Ryan Brinkman

FlowCAP-IV:
Unsupervised Clustering and Classification

Biomarker discovery

388 patients (split training and test)

14 parameter data

Predict survival time (onset of HIV) & identify biomarkers

Thorough manual analysis (NIH/VRC) had failed to identify any biomarker

*Aghaeepour *et al., Cytometry A* (2016)

Ryan Brinkman

FlowCAP-IV:
Unsupervised Clustering and Classification

Two similar methods had significant results on test data

- flowDensity / flowType
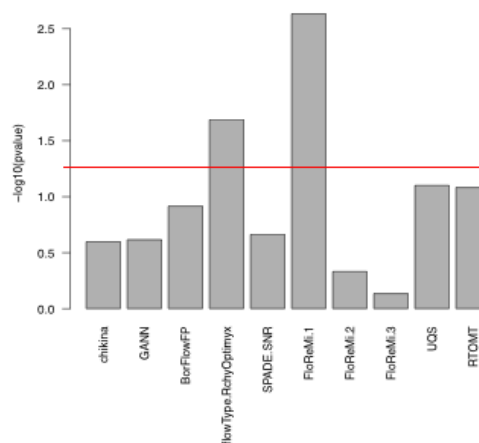
*p-value not adjusted for testing multiple cell populations, but not algorithms

Ryan Brinkman

Automated Analysis For Discovery & Diagnosis In Big Flow Cytometry Data

FLOWDENSITY:
PIPELINE FOR DIAGNOSIS

◦ Finds what you want to find, how you want find it
◦ Based on density estimation techniques
◦ Seconds per FCS file
◦ Identical to the manual practice of 2D gating

FLOWTYPE / RCHYOPTIMYX:
PIPELINE FOR DISCOVERY

◦ You split FCS files into groups
◦ Pipeline finds best cell populations that correlate with that split
◦ One graph summary of very large datasets
◦ Can be used as input to large multi-group studies

Ryan Brinkman

What You Can Do To Get Started?

What You Can Do To Get Started

Practical Considerations for Automated Analysis
◦ Don't waste your time on 12 clinical samples
◦ Your study probably isn't sufficiently powered for unsupervised analysis
◦ Don't waste your time on automated discovery using 6 colors
  ◦ Automated analysis will find everything you found by hand
◦ Good bioinformatics can't save bad data
◦ Discovery analysis is hypothesis generating
  ◦ Finding cell populations that don't "make sense" will happen
◦ Its OK to ask for help (and to attend training! ☺)

Ryan Brinkman

Learning R: BioConductor.org

Ryan Brinkman

RStudio

R in FlowJo

GenePattern

Ryan Brinkman

GenePattern - Example

Ryan Brinkman

ImmPort - FLOCK
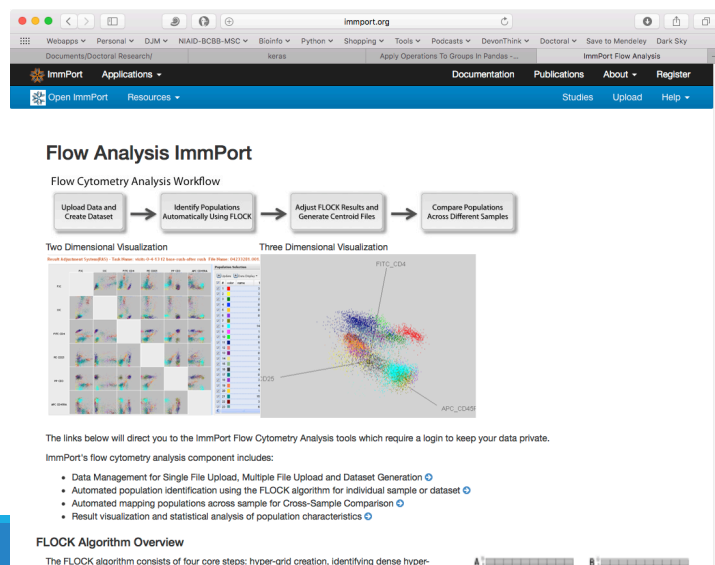
Resources

Intro to Flow Cytometry
- http://www.abcam.com/protocols/introduction-to-flow-cytometry
- https://www.bio-rad-antibodies.com/introduction-to-flow-cytometry.html

Flow Cytometry GitHub Repository / Videos
- https://github.com/bioinformatics-ca/other_workshops/tree/master/flow_cytometry_2013
- http://bioinformatics-ca.github.io/flow_cytometry_2013/

R Programming
- http://www.cyclismo.org/tutorial/R/
- https://cran.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro.pdf

Repository
- https://flowrepository.org/

Unix
- http://www.ee.surrey.ac.uk/Teaching/Unix/
- http://www.rain.org/%7Ekummel/unix.html

References

Aghaeepour, N., & Jalali, A. (2016). RchyOptimyx: Gating Hierarchy Optimization for Flow Cytometry.

Aghaeepour, N., Nikolic, R., Hoos, H. H., & Brinkman, R. R. (2011). Rapid cell population identification in flow cytometry data. Cytometry Part A, 79A(1), 6–13. http://doi.org/10.1002/cyto.a.21007

Brinkman, R. R., Aghaeepour, N., Finak, G., Gottardo, R., Mosmann, T., & Scheuermann, R. H. (2016, January). Automated analysis of flow cytometry data comes of age. Cytometry Part A, 89(1), 13–15. http://doi.org/10.1002/cyto.a.22810

De Rosa, S. C., Brenchley, J. M., & Roederer, M. (2003). Beyond six colors: A new era in flow cytometry. Nature Medicine, 9(1), 112–117. http://doi.org/10.1038/nm0103-112

Hahne, F., LeMeur, N., Brinkman, R. R., Ellis, B., Haaland, P., Sarkar, D., … Gentleman, R. (2009). flowCore: a Bioconductor package for high throughput flow cytometry. BMC Bioinformatics, 10, 106. http://doi.org/10.1186/1471-2105-10-106

Janes, M. R., & Rommel, C. (2011). Next-generation flow cytometry. Nature Biotechnology, 29(7), 602–604. http://doi.org/10.1038/nbt.1919

Malek, M., Taghiyar, M. J., Chong, L., Finak, G., Gottardo, R., & Brinkman, R. R. (2015). flowDensity: reproducing manual gating of flow cytometry data by automated density-based cell population identification. Bioinformatics, 31(4), 606–607. http://doi.org/10.1093/bioinformatics/btu677

O'Neill, K., Aghaeepour, N., Špidlen, J., Brinkman, R., & Nolan, G. (2013). Flow Cytometry Bioinformatics. PLoS Computational Biology, 9(12), e1003365. http://doi.org/10.1371/journal.pcbi.1003365

Pyne, S., Hu, X., Wang, K., Rossin, E., Lin, T.-I., Maier, L. M., … Mesirov, J. P. (2009). Automated high-dimensional flow cytometric data analysis. Proceedings of the National Academy of Sciences of the United States of America, 106(21), 8519–24. http://doi.org/10.1073/pnas.0903028106

Raveche, E., Abbasi, F., Yuan, Y., Salerno, E., Kasar, S., & Marti, G. E. (2010). Introduction to Flow Cytometry. In Flow Cytometry in Drug Discovery and Development (pp. 3–21). http://doi.org/10.1002/9780470910085.ch1

Verschoor, C. P., Lelic, A., Bramson, J. L., & Bowdish, D. M. E. (2015). An introduction to automated flow cytometry gating tools and their implementation. Frontiers in Immunology, 6(JUL), 380. http://doi.org/10.3389/fimmu.2015.00380