

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

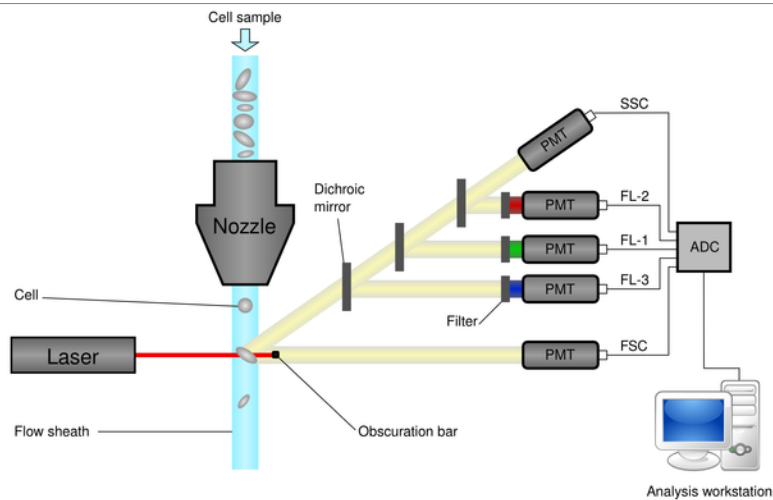
The State of Automated Flow Cytometry Analysis

- Flow Cytometry Standards and Analysis (FlowCAP)

Automated Flow Cytometry Analysis

Flow Cytometry Review

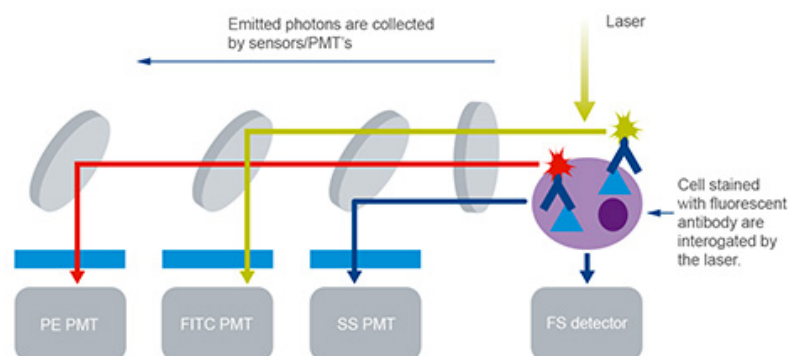
Schematic Diagram Of A Flow Cytometer



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>

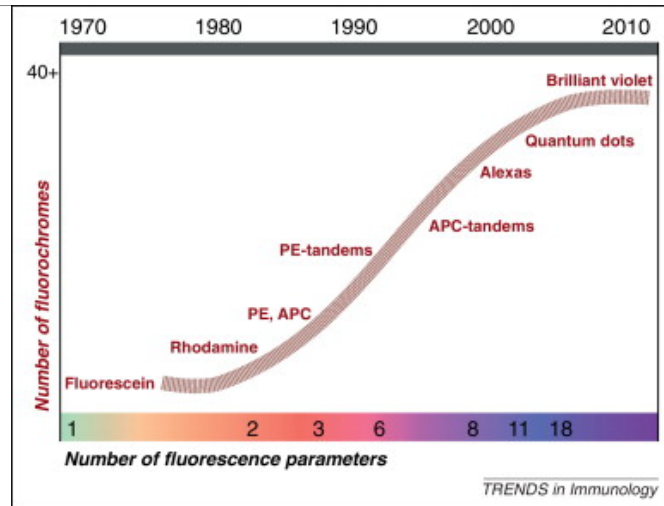


Fluorescent Light Is Filtered So That Each PMT Detects A Specific Wavelength



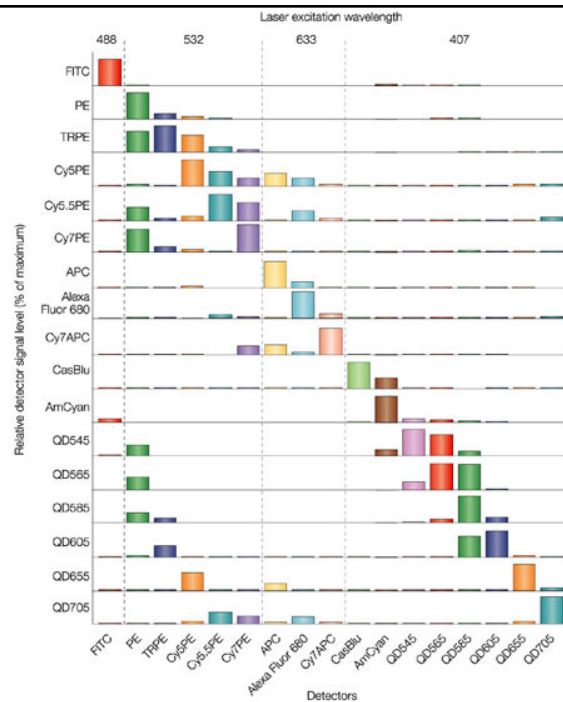
<http://www.abcam.com/protocols/introduction-to-flow-cytometry>

Fluorochromes

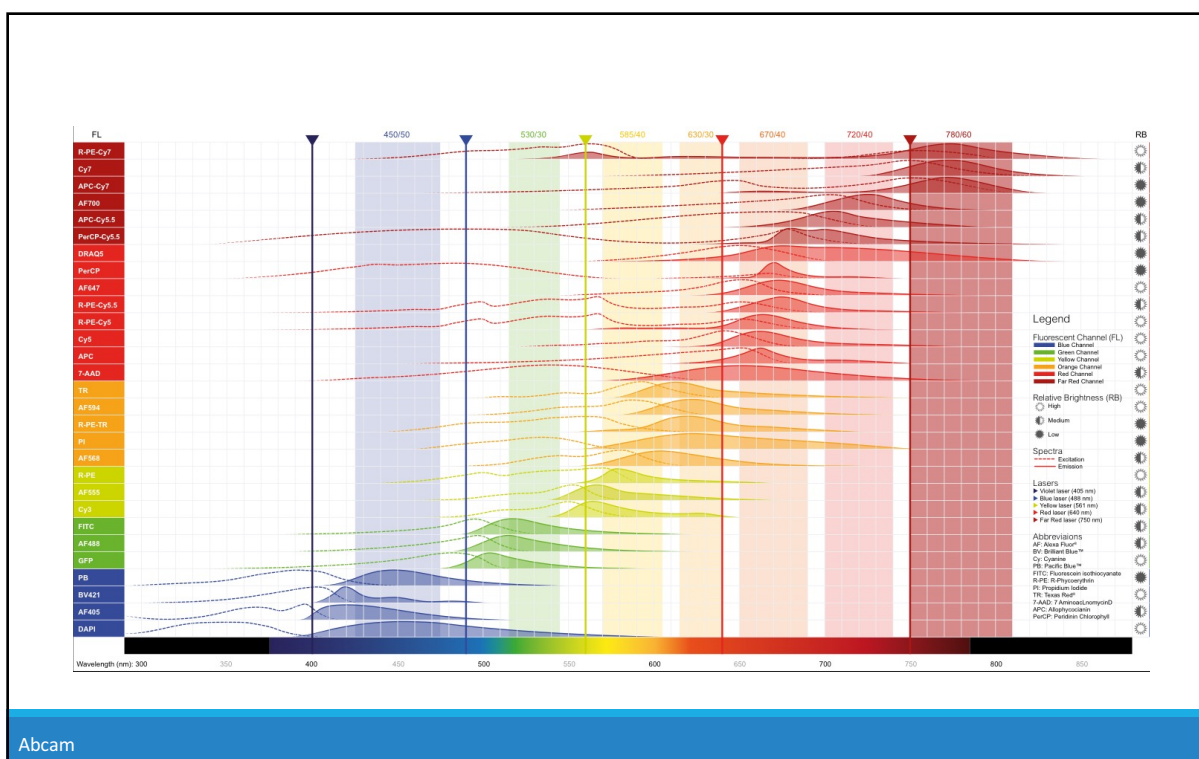


Bendall, S. C., Nolan, G. P., Roederer, M. & Chattopadhyay, P. K. A Deep Profiler's Guide to Cytometry. *Trends in Immunology* **33**, 323–332 (2012).

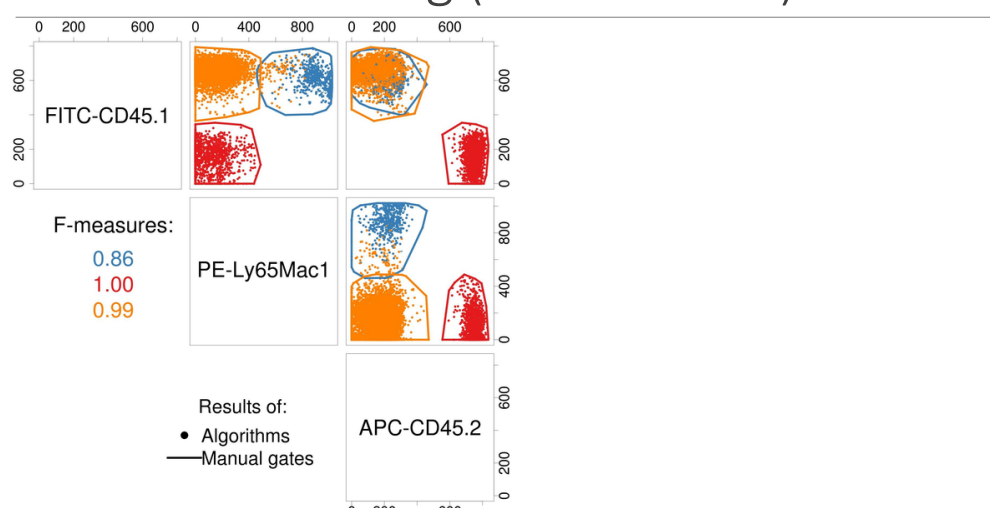
Fluorochrome



Nature Reviews | Immunology



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



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>

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

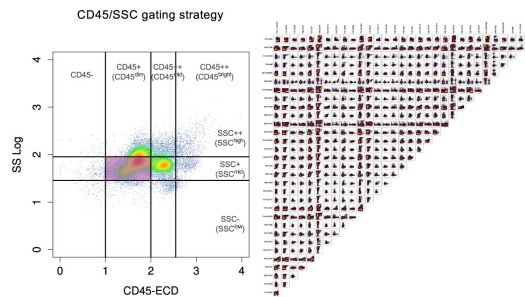
Unsupervised biomarker discovery: FlowReMi, flowType/RchyOptimyx

- O’Neill et al., Bioinformatics, 2014

Big Flow Cytometry Data

AUTOMATED ANALYSIS OF FLOW CYTOMETRY DATA IN R/BIOCONDUCTOR

	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 ⁴	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, reproducible
- Breaks problem into smaller pieces (packages)
- Modules can plug-in & swap-out
- Collaborative, cross-platform development environment

Immunity
Letter

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

Pia Kvistborg,^{1,2} Cécile Goutellange,^{3,10} Nima Aghaeepour,³ Angelica Caraly,⁴ Pratip K. Chattopadhyay,⁵ Cilbum Chan,⁶ Judith Eckl,⁷ Greg Finak,⁸ Sine Raker Hadrup,⁹ Holden T. Maack,¹⁰ Dominik Maurer,¹¹ Tim Moermann,¹² Peng Qu,¹³ Richard H. Scheuermann,^{14,15} Marj J.P. Wellers,¹⁶ Guido Ferrarini,¹⁷ Ryan R. Brinkman,^{1,2,18} and Cedrik M. Britten^{19,20,21}

CellPress

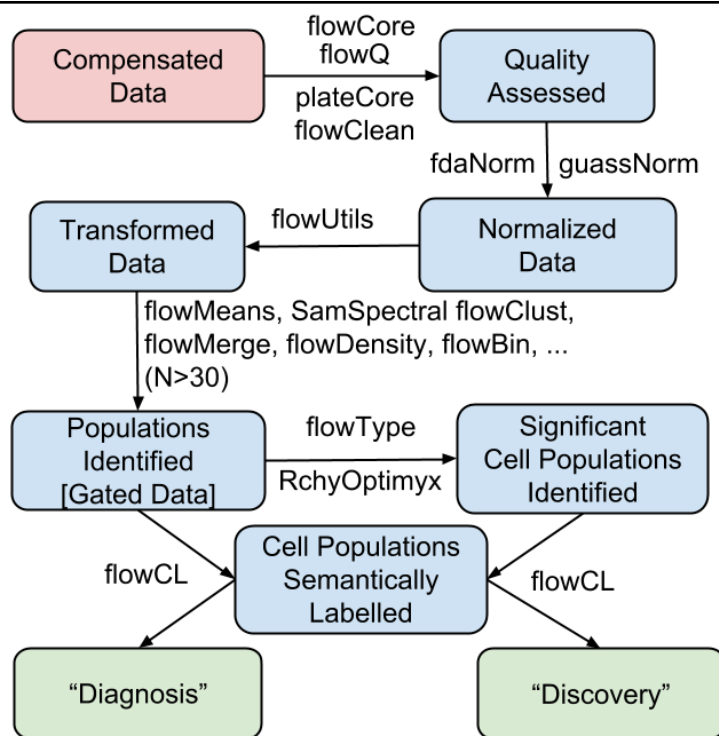
9 additional software tools

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

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>

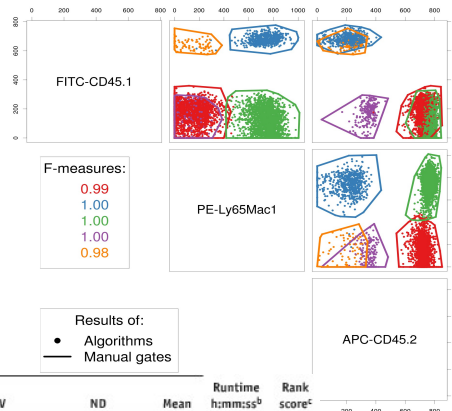
Ryan Brinkman

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

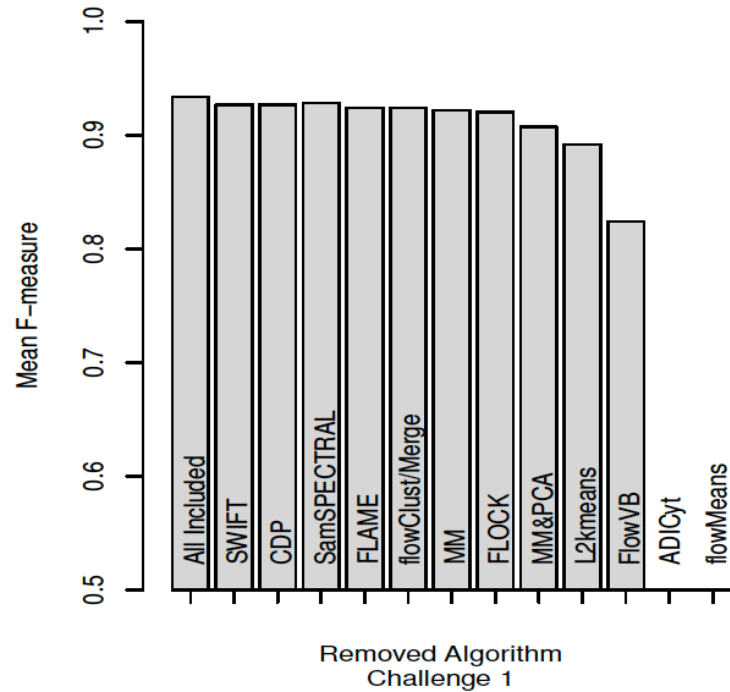


	GvHD	DLBCL	H5CT	WNV	ND	Mean	Runtime h:mm:ss ^a	Rank score ^b
Challenge 1: completely automated								
ADICyt	0.81 (0.72, 0.88)	0.93 (0.91, 0.95)	0.93 (0.90, 0.96)	0.86 (0.84, 0.87)	0.92 (0.92, 0.93)	0.89	4:50:37	52
FlowMeans	0.88 (0.82, 0.93)	0.92 (0.89, 0.95)	0.92 (0.90, 0.94)	0.88 (0.86, 0.90)	0.85 (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

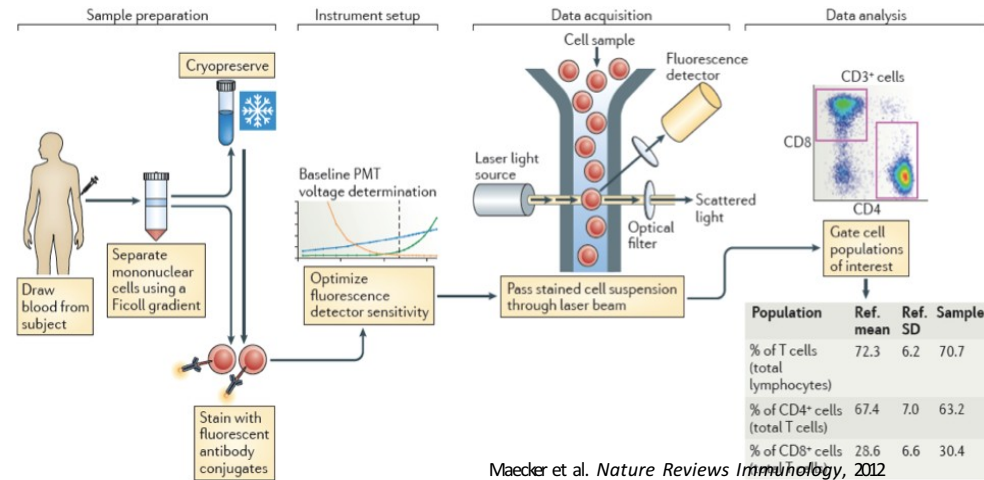
There were misclassifications

	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 Iyoplates), Instrumentation, Acquisition, Automated Analysis



Maecker et al. *Nature Reviews Immunology*, 2012

Ryan Brinkman

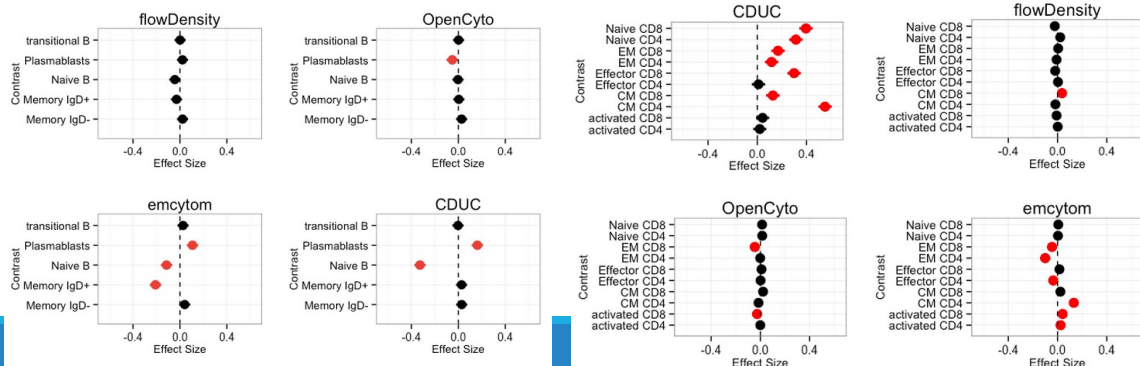
FlowCAP-III: Supervised Analysis for “Diagnostics”

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.

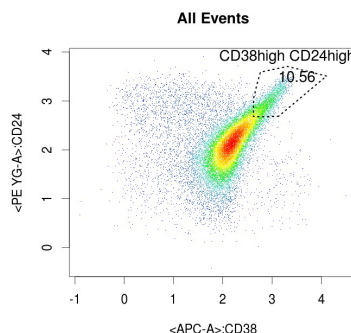
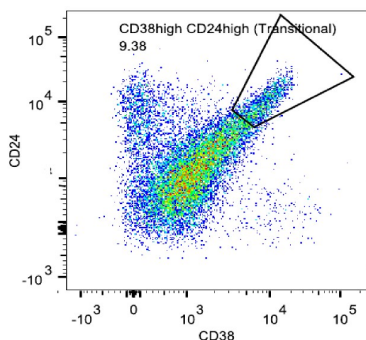


Ryan Brinkman

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

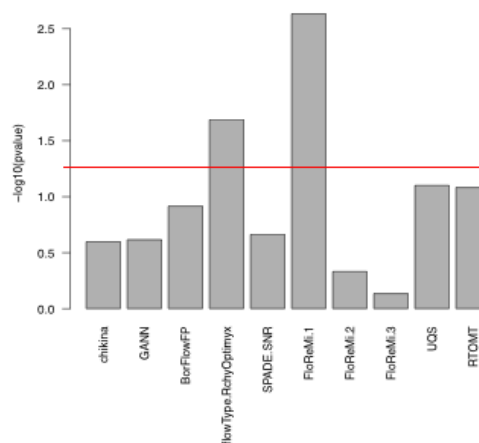
*Aghaeepour *et al.*, *Cytometry A* (2016)

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

The screenshot shows the Bioconductor.org website. The header includes the Bioconductor logo and navigation links: Home, Install, Help, Developers, and About. A search bar is also present. The main content area is divided into two columns. The left column is titled 'About Bioconductor' and describes the software as an open-source tool for genomic data analysis, mentioning its R-based nature and active community. The right column is titled 'Use Bioconductor for...' and lists several categories of tools: Microarrays, Sequence Data, Variants, Annotation, and High Throughput Assays, each with a brief description of their capabilities.

About Bioconductor

Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data. Bioconductor uses the R statistical programming language, and is open source and open development. It has two releases each year, 554 software packages, and an active user community. Bioconductor is also available as an [Amazon Machine Image \(AMI\)](#).

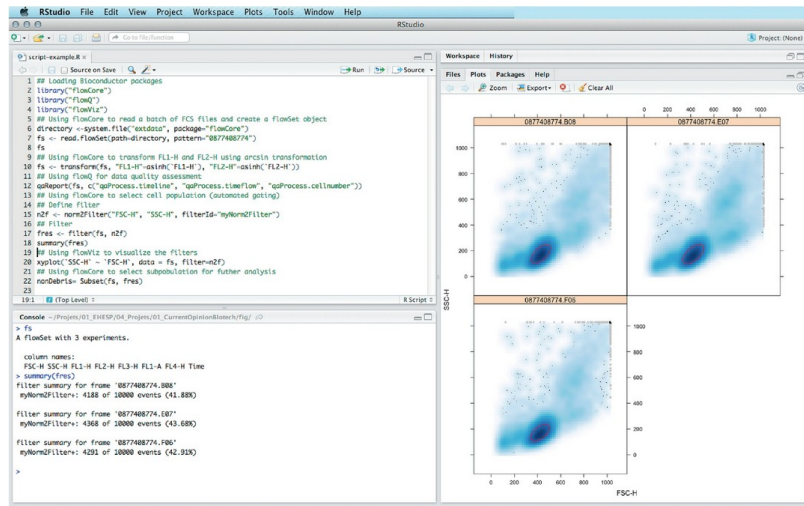
Join us at [BioC 2012](#), July 24-25.

Use Bioconductor for...

- Microarrays**
Import Affymetrix, Illumina, Nimblegen, Agilent, and other platforms. Perform quality assessment, normalization, differential expression, clustering, classification, gene set enrichment, genetical genomics and other workflows for expression, exon, copy number, SNP, methylation and other assays. Access GEO, ArrayExpress, Biomart, UCSC, and other community resources.
- Variants**
Read and write VCF files. Identify structural location of variants and compute amino acid coding changes for non-synonymous variants. Use SIFT and PolyPhen database packages to predict consequence of amino acid coding changes.
- Sequence Data**
Import fastq, fastq, ELAND, MAQ, BWA Bowtie, BAM, gff, bed, wig, and other sequence formats. Trim, transform, align and manipulate sequences. Perform quality assessment, ChIP-seq, differential expression, RNA-seq, and other workflows. Access the Sequence Read Archive.
- Annotation**
Use microarray probe, gene, pathway, gene ontology, homology and other annotations. Access GO, KEGG, NCBI, Biomart, UCSC, vendor, and other sources.
- High Throughput Assays**
Import, transform, edit, analyze and visualize flow cytometric, mass spec, HTqPCR, cell-based, and other assays.

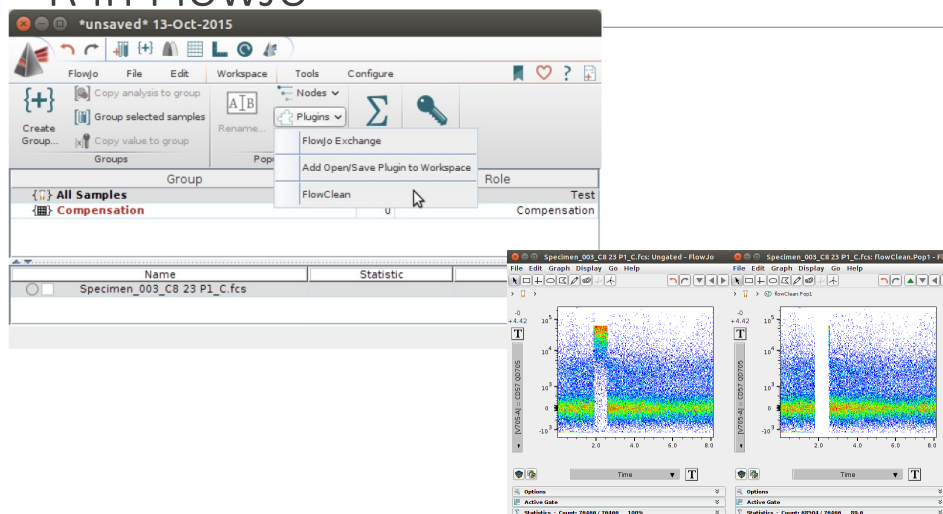
Ryan Brinkman

RStudio



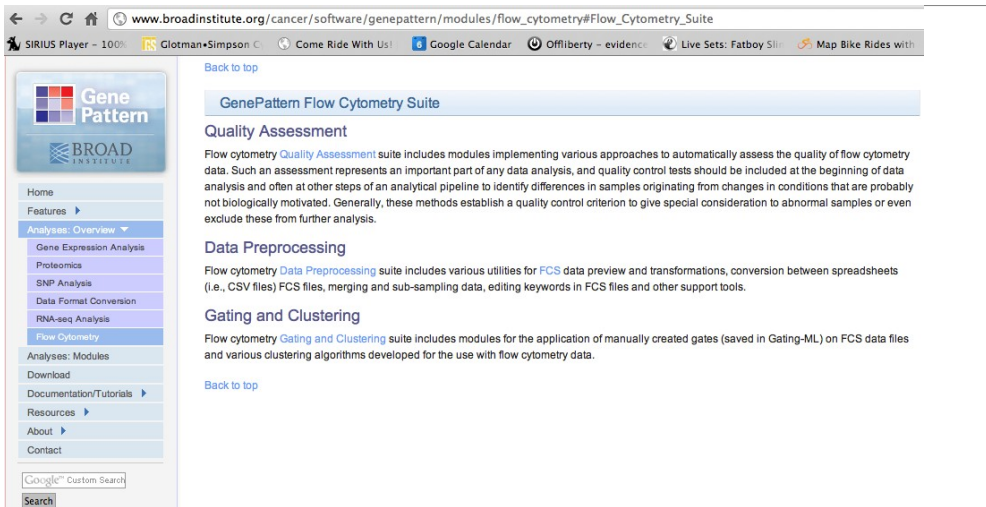
Ryan Brinkman

R in FlowJo



Ryan Brinkman

GenePattern



The screenshot shows a web browser window with the address bar displaying www.broadinstitute.org/cancer/software/genepattern/modules/flow_cytometry#Flow_Cytometry_Suite. The browser's address bar also shows several open tabs: SIRIUS Player - 100%, Glotman+Simpson, Come Ride With Us!, Google Calendar, Offberty - evidence, Live Sets: Fatboy Slim, and Map Bike Rides with. The GenePattern logo is visible in the top left corner of the page. The main content area is titled "GenePattern Flow Cytometry Suite" and contains two sections: "Quality Assessment" and "Data Preprocessing". The "Quality Assessment" section describes a suite of modules for automatically assessing the quality of flow cytometry data. The "Data Preprocessing" section describes a suite of modules for FCS data preview and transformations, conversion between spreadsheets (i.e., CSV files) FCS files, merging and sub-sampling data, editing keywords in FCS files and other support tools. The "Gating and Clustering" section describes a suite of modules for the application of manually created gates (saved in Gating-ML) on FCS data files and various clustering algorithms developed for the use with flow cytometry data. A sidebar on the left contains a navigation menu with links to Home, Features, Analyses: Overview, Gene Expression Analysis, Proteomics, SNP Analysis, Data Format Conversion, RNA-seq Analysis, Flow Cytometry, Analyses: Modules, Download, Documentation/Tutorials, Resources, About, and Contact. A Google Custom Search box is located at the bottom of the sidebar. The footer of the page displays "Broad Institute Cancer Program" and "©2006-2012 Broad Institute".

GenePattern

Flow Cytometry Suite

Quality Assessment

Flow cytometry [Quality Assessment](#) suite includes modules implementing various approaches to automatically assess the quality of flow cytometry data. Such an assessment represents an important part of any data analysis, and quality control tests should be included at the beginning of data analysis and often at other steps of an analytical pipeline to identify differences in samples originating from changes in conditions that are probably not biologically motivated. Generally, these methods establish a quality control criterion to give special consideration to abnormal samples or even exclude these from further analysis.

Data Preprocessing

Flow cytometry [Data Preprocessing](#) suite includes various utilities for [FCS](#) data preview and transformations, conversion between spreadsheets (i.e., CSV files) FCS files, merging and sub-sampling data, editing keywords in FCS files and other support tools.

Gating and Clustering

Flow cytometry [Gating and Clustering](#) suite includes modules for the application of manually created gates (saved in Gating-ML) on FCS data files and various clustering algorithms developed for the use with flow cytometry data.

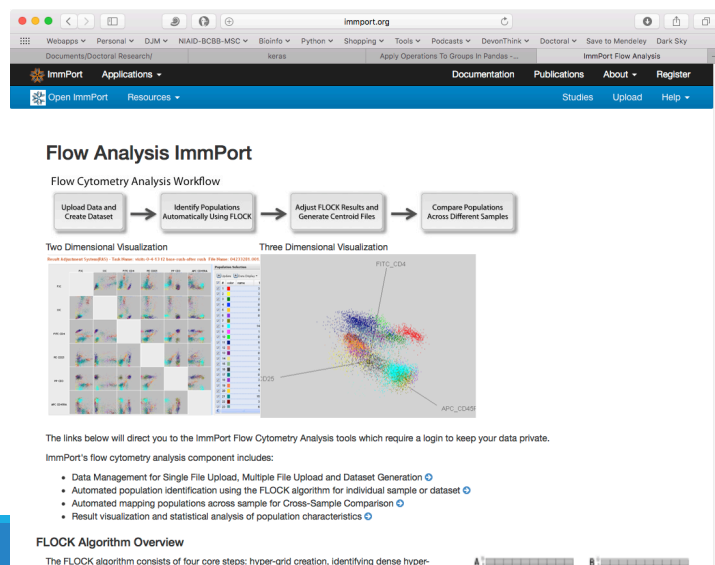
Broad Institute
Cancer Program

©2006-2012 Broad Institute

Ryan Brinkman

GenePattern - Example

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/%7Emkummel/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>