Automated flow cytometry data analysis for diagnosis and discovery

Ryan Brinkman

Senior Scientist, BC Cancer Agency
Associate Professor, Department of Medical Genetics, UBC
Vancouver, British Columbia, Canada

MetroFlow / NYC
Oct 23, 2013

Computational Analysis Kicks Ass

Automated algorithms have reached a level of maturity that enables them to match and in many cases exceed the results produced by human experts.

Episode I: The Menace of R/BioConductor

Episode II: The Attack of FlowCAP

Episode III: The Revenge of flowType/RchyOptimyx for "Discovery"

Episode IV: The Hope of flowDensity for "Diagnosis"

Episode V: MIFlowCyt Strikes Back Against Data Annotation

Episode VI: The Return of flowRepository

I: Automated Flow Cytometry Data Analysis

Why should you care?

	1985	**2012**	Δ
Samples:	1	466	
Colours:	3	13	
Events:	50,000	400,000	
Data:			16,000x
CPU:	3 Mhz	600x 12 @3 GHz	
RAM:	2 MB	48 GB/node	
Power			7,000,000x
Fruit:			seeds, colour ($p<0.05$)*

Murphy *Cytometry* (1985) Aghaeepour *et al. Bioinformatics* (2012)

*Barone *BMJ* (2000)

Manual Analysis of "High-Dimensional" Data

Could possibly 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)

What Automated Analysis Needs to Deal With

- Large number of dimensions, events, samples
- Mutifactorial formats
- Need quick, robust, fully automated processing
- Need to maintain data & metadata relationships

- **No commercially available software solves these issues***

Bashashti *et al. Adv Bioinformatics* (2009)
PMID 20049163
*Robinson *et al. Expert Opinion Drug Discovery* (2012)
PMID 22708834
Le Meur *Curr Opin Biotechnol* (2013)
PMID 23062230

Solution: Free, Open Source Statistical Programming

- R is a free/libre open source, robust statistical programming environment for Windows, Mac & Linux that offers a wide range of statistical and visualization methods
- BioConductor provides R software modules for biological and clinical data analysis
- A **scripted** approach to high throughput data analysis
  - Non-interactive, self-documented, reproducible
  - Breaks problem into smaller pieces (packages)
  - Modules can plug-in & swap-out
- Integrates with other software tools via open data standards
  - Collaborative development

http://bioconductor.org

28 R packages for Flow Analysis

Data processing & visualization (12/28)

flowCore*	Read/write & process flow data
plateCore*	Analyze multiwell plates
flowUtils*	Import gates, transformation and compensation
flowQ*	Quality control of ungated data
flowStats*	Advanced statistical methods and functions
ncdfFlow	Advanced methods for large dataset processing
QUALIFIER	Quality control and assessment of gated data
flowViz	Visualization (*e.g.*, histograms, dot plots, density plots)
flowPlots*	Graphical displays with statistical tests
flowWorkspace*	Importing FlowJo workspaces
iFlow	GUI for exploratory analysis and visualization
flowTrans*	Estimates parameters for data transformation

(Co-)developed @ Brinkman lab

*Peer-reviewed manuscript available

17 R packages for Automated Gating

flowClust*	Clustering using t-mixture model with Box-Cox transformation
flowMerge*	flowClust + entropy-based merging
flowMeans*	k-means clustering and merging using the Mahalanobis distance
SamSpectral*	Efficient spectral clustering using density-based down-sampling
flowDensity	Supervised density-based gating for "diagnosis"
flowBin	Multi-tube binning for deep profiling
openCyto	Hierarchical Gating Pipeline for flow cytometry data
flowQB	Q&B analysis
flowPeaks*	Unsupervised clustering using k-means & mixture model
flowFP*	Fingerprint generation
flowPhyto*	Analysis of marine biology data
FLAME*	Multivariate finite mixtures of skew & tailed distributions
flowKoh	Self-organizing maps
NMF-curvHDR*	Density-based clustering and matrix factorization
flowCore/Stats*	Sequential gating and normalization w/ Beta-Binomial model
PRAMS*	2D Clustering and logistic regression
SPADE*	Density-based sampling, k-means & minimum spanning trees

(Co-)developed @ Brinkman lab

*Peer-reviewed manuscript available

2 Packages for Post-Gating "Significance" Assessment

flowType* Automated phenotyping using 1D gates extrapolated to multiple dimensions

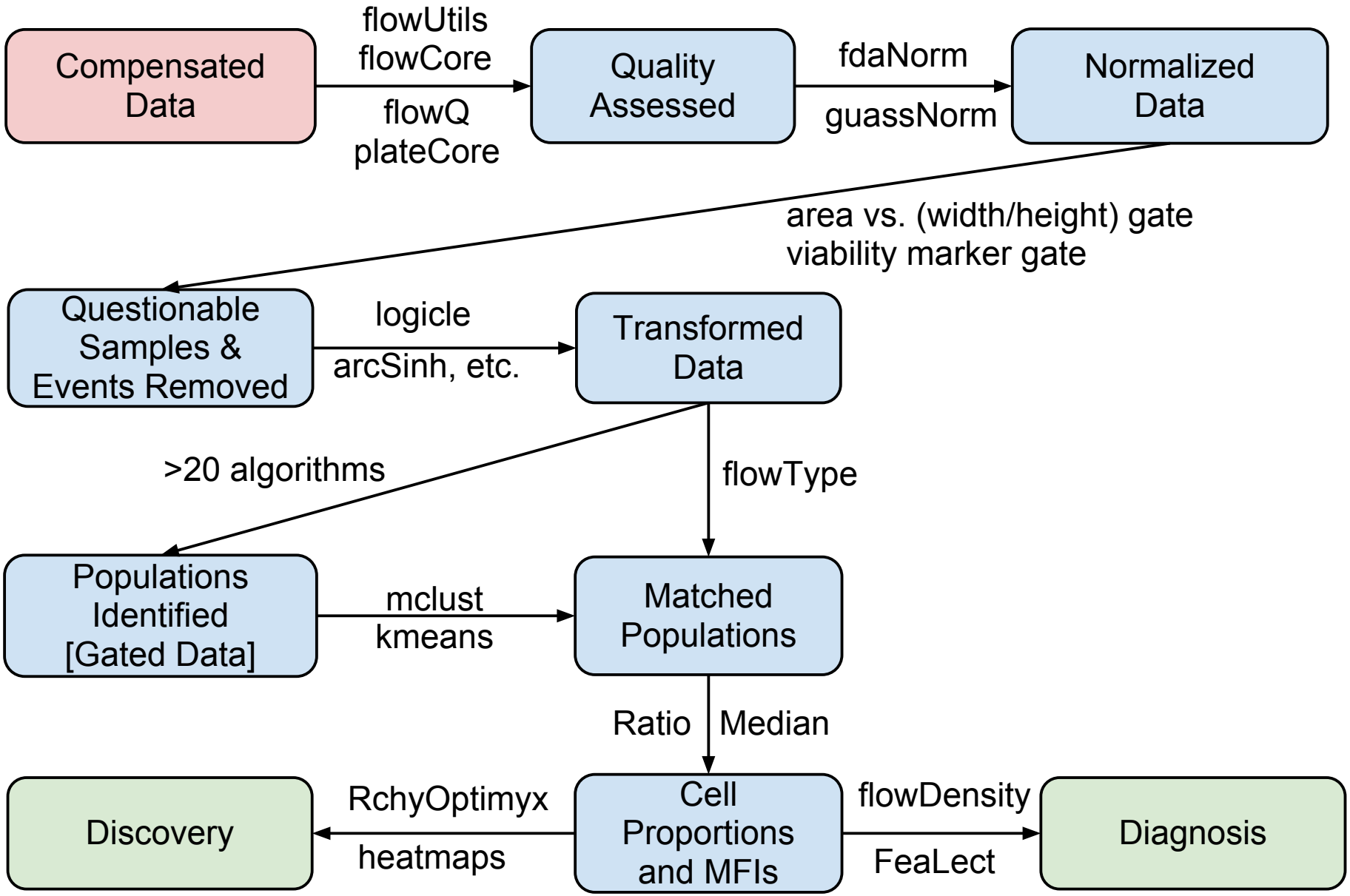
RchyOptimyx* Cellular hierarchies correlated with outcome of interest

Developed @ Brinkman lab

*Peer-reviewed manuscript available

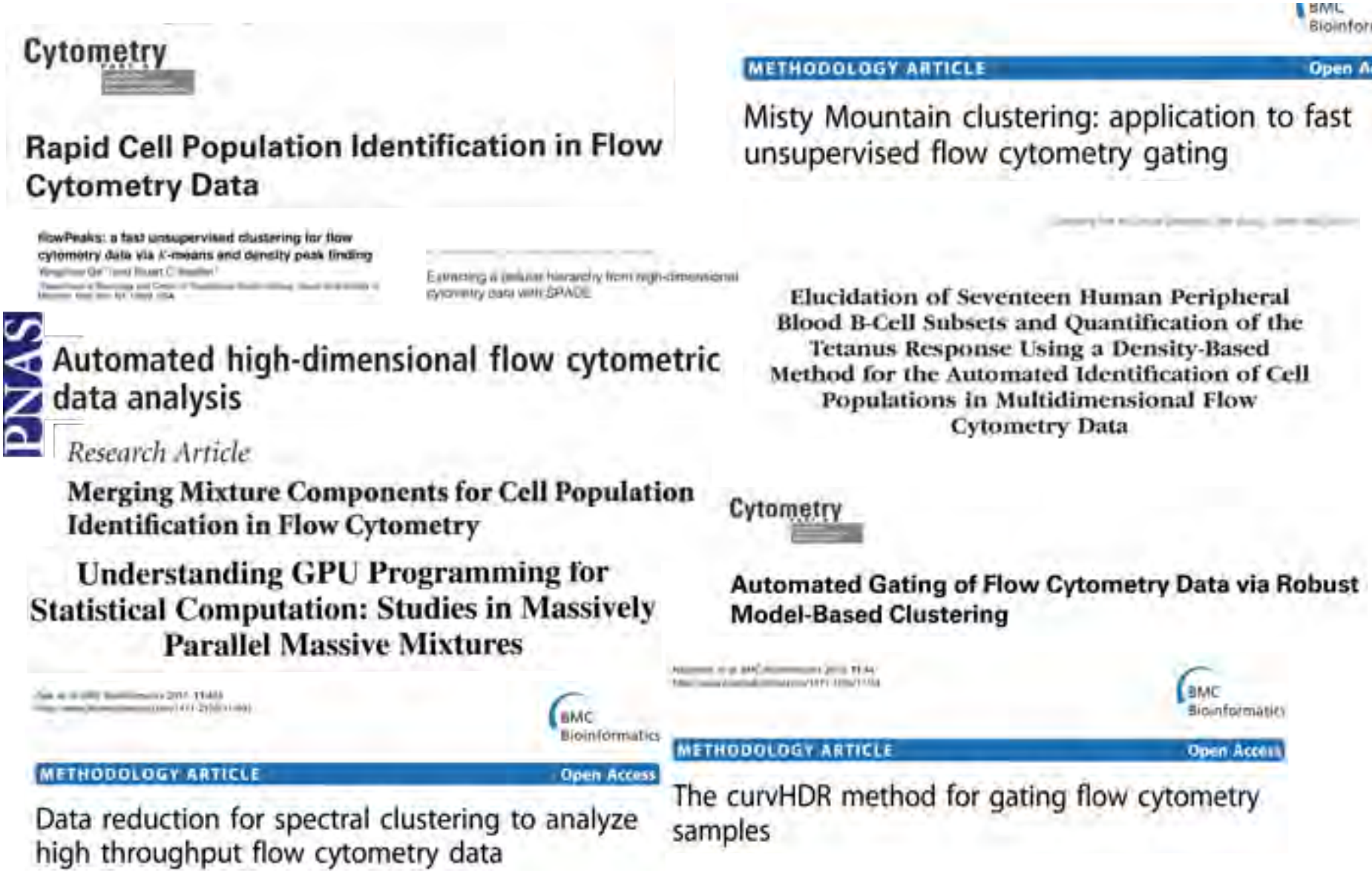
BioConductor's Open, Extensible Infrastructure

Packages are Interoperable & Interchangeable



Renaissance: >20 Automated Analysis Methods 2008

Which to use?



Episode II: The Attack of FlowCAP

Critical assessment of automated discovery and diagnosis

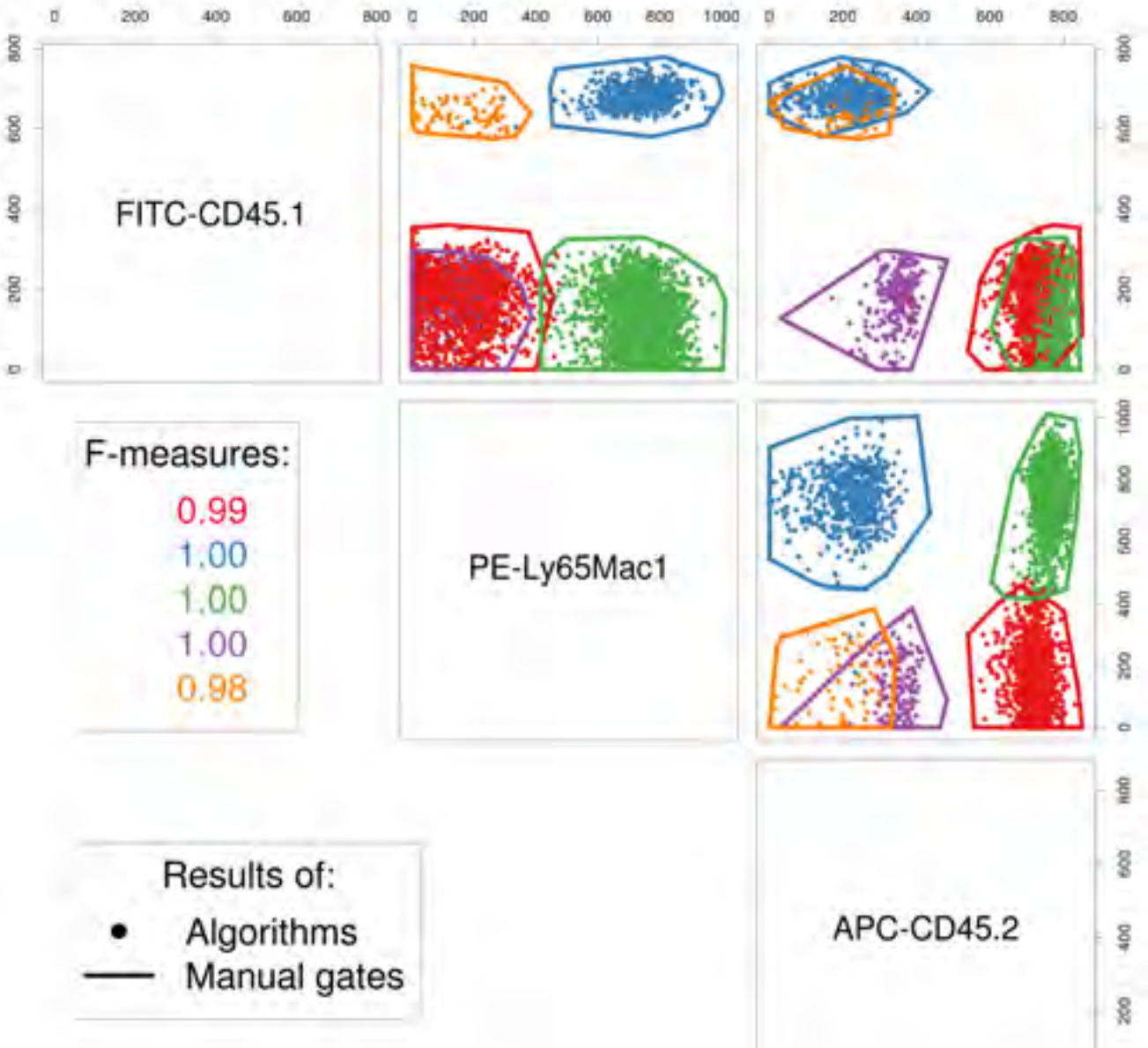
- Bring together everyone in the world developing automated tools to arrive at a consensus of the state of the art
- FlowCAP-1 & -2: All methods available as open source or pseudo-code along with data, analyses code
- FlowCAP-3: All data, analysis code freely available

*Aghaeepour *et al.*, *Nature Methods* (2013)

http://flowcap.flowsite.org

Consensus of FlowCAP1 analyses = humans' gating for discovery

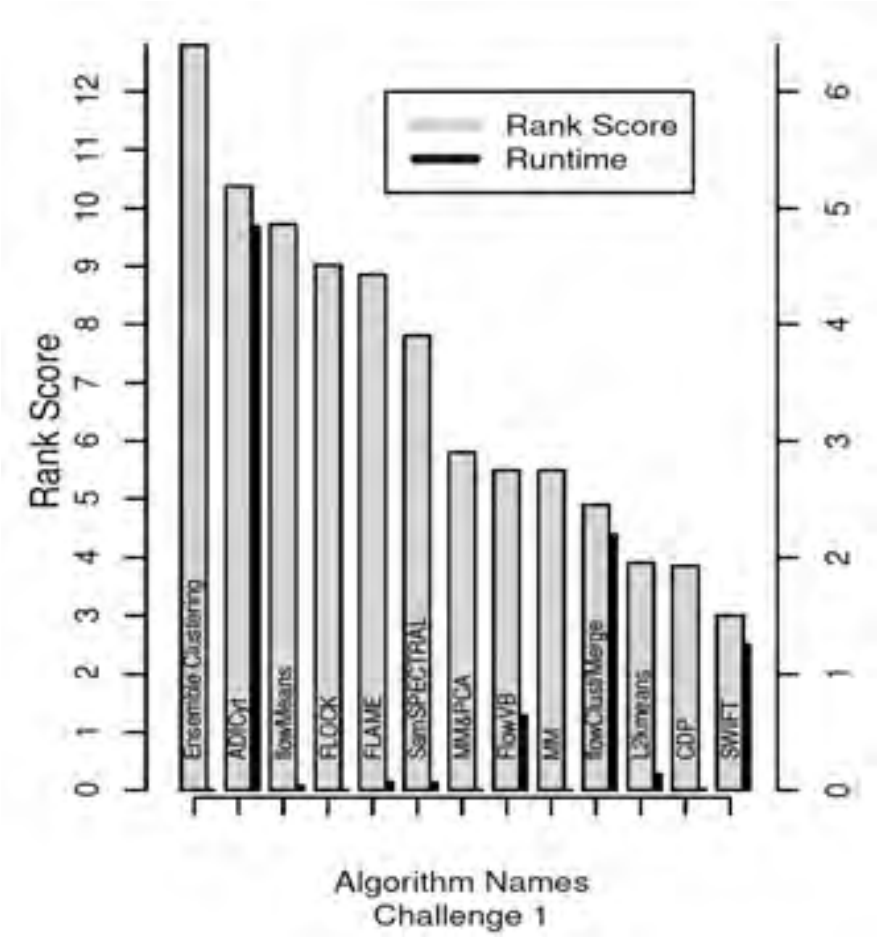
Individual performance can vary on specific cell populations



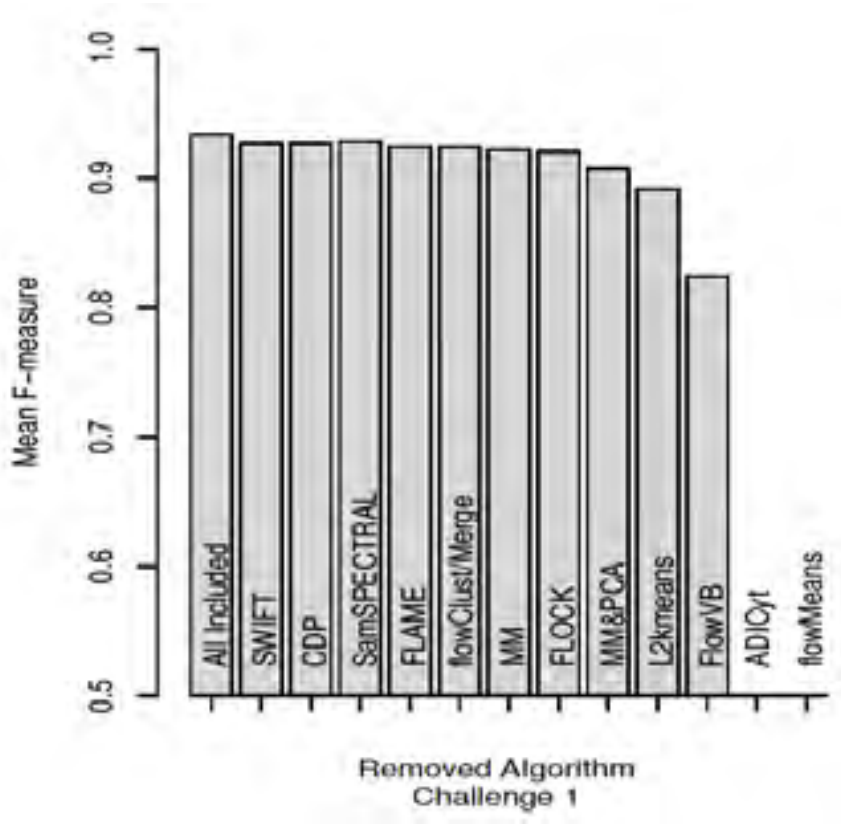
Best Unsupervised Results Obtained Using Multiple Approaches

Individual performance can vary on specific cell populations

Overall Ranking



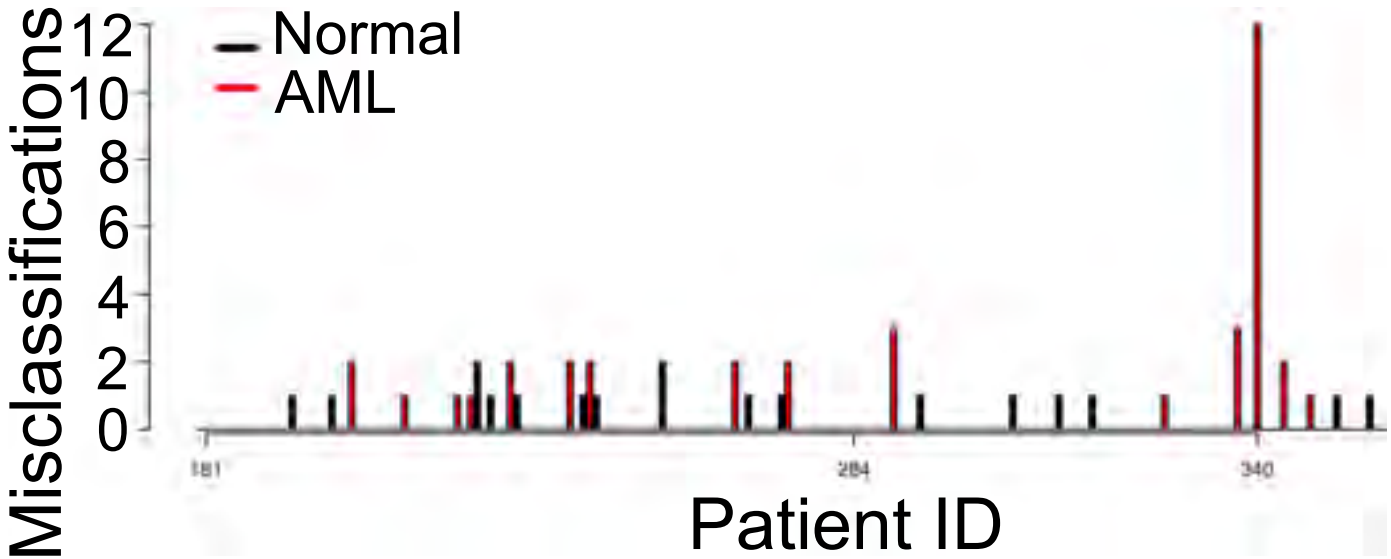
Removing Poor Performers



FlowCAP2: Tools for Diagnosis

- 8 tubes of 5 colour assays on 359 subjects;
  - 43 AML positive vs. 316 healthy donor
- 43 algorithms
  - Several performed perfectly

	Sensitivity	Specificity	Accuracy
flowType-FeaLect	1.00	1.00	1.00
flowPeakssvm	1.00	1.00	1.00
SPADE	1.00	1.00	1.00
... n=43	...	...	...

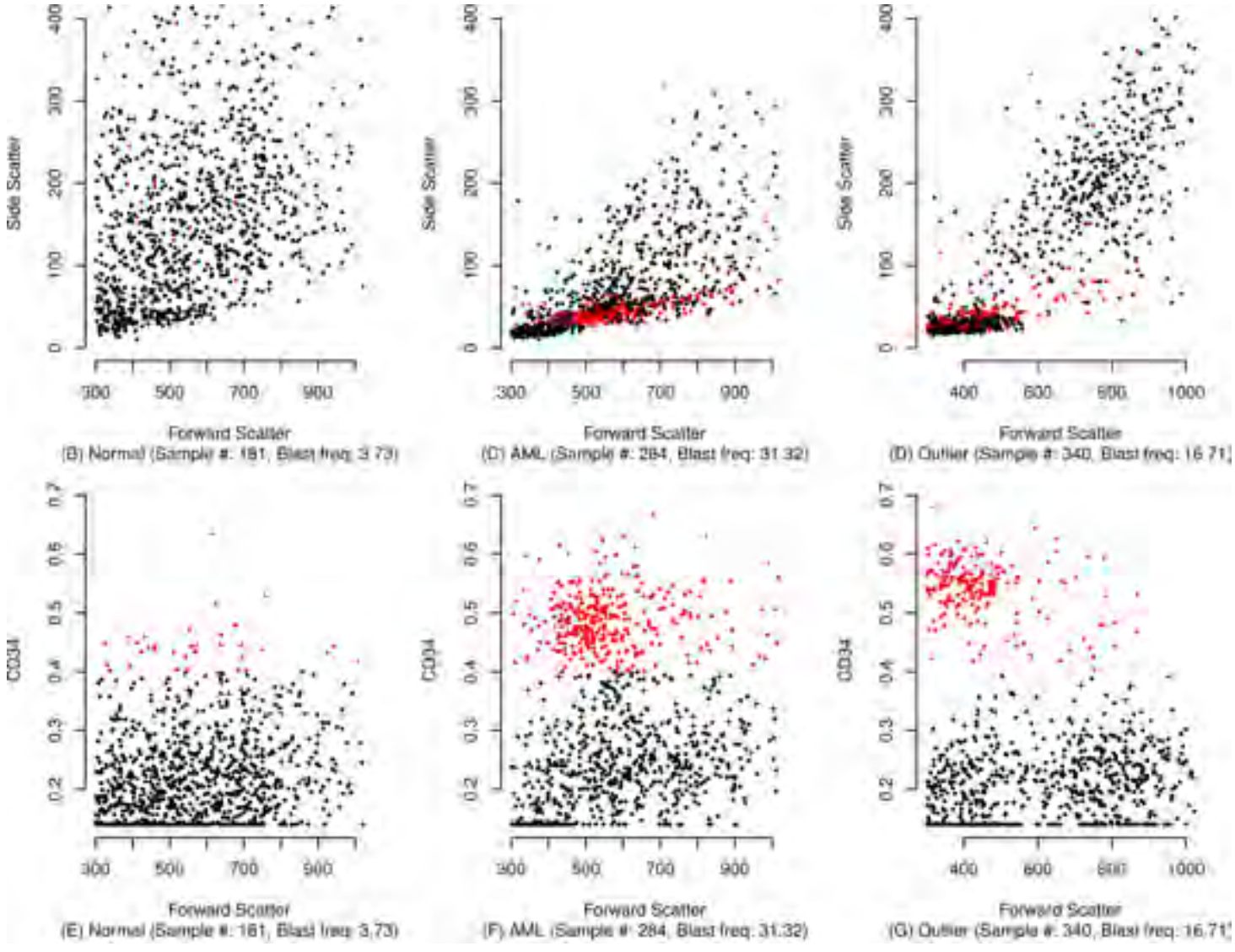


Developed @ Brinkman lab

FlowCAP2: AML Outlier

High grade myelodysplasia?

Flow underestimating blast frequency (hemodilution of specimen)?



Tools for Discovery in Practice

Immunologic and Virologic Events in Early HIV Infection Predict Subsequent Rate of Progression

Anuradha Ganesan,[1,a] Pratip K. Chattopadhyay,[2,a] Tess M. Brodie,[2] Jing Qin,[3] Wenjuan Gu,[4] John R. Mascola,[2] Nelson L. Michael,[5] Dean A. Follmann,[3] and Mario Roederer,[2] for the Infectious Disease Clinical Research Program HIV Working Group[b]

[1]National Naval Medical Center, Infectious Disease Clinical Research Program, Uniformed Services University, [2]Vaccine Research Center, National Institute of Allergy and Infectious Diseases, and [3]Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, Bethesda, [4]Biostatistics Research Branch, Scientific Application International Corporation–Frederick, Frederick, and [5]United States Military HIV Research Program, Walter Reed Army Institute of Research, Rockville, Maryland

United States Military HIV Natural History Study

- PBMCs of 466 HIV$^+$ personnel and beneficiaries from Army, Navy, Marines, and Air Force.
- 13 surface markers and KI-67 (cell proliferation).
- Clinical Data: Survival times including 135 events[a]

[a]An event is defined as progression to AIDS or initiation of HAART.

Manual Gating Results

- Frequency of **long-lived Memory Cells** ($CD127^+$) has a positive correlation.
- Frequency of **cells with high proliferation** ($KI\text{-}67^+$) has a negative correlation.
- Can we find what they have found? Can we find more?

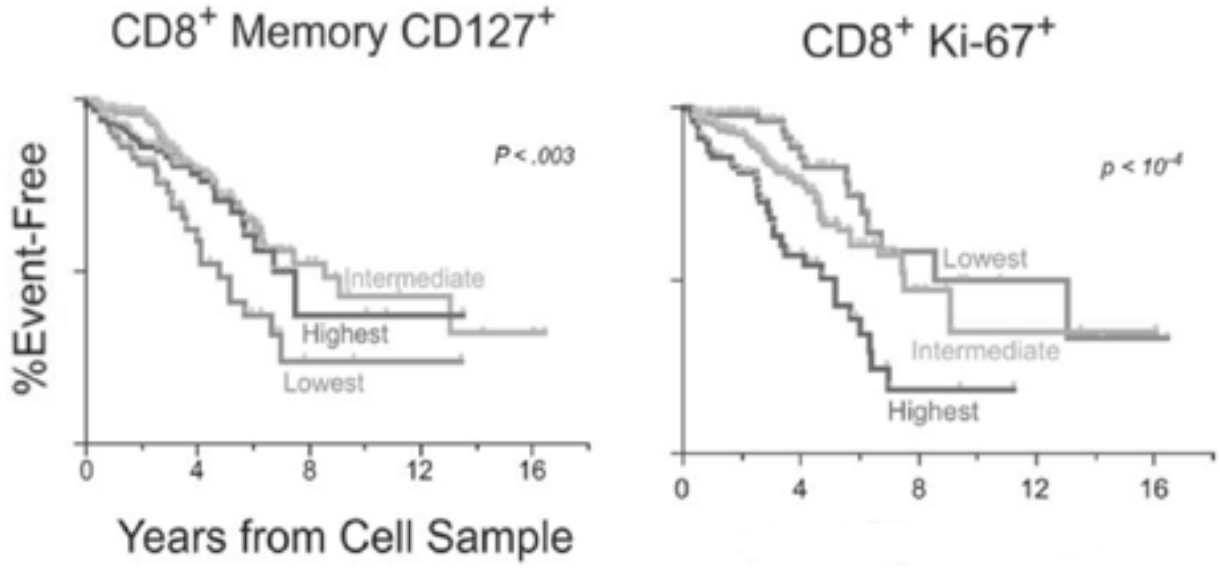
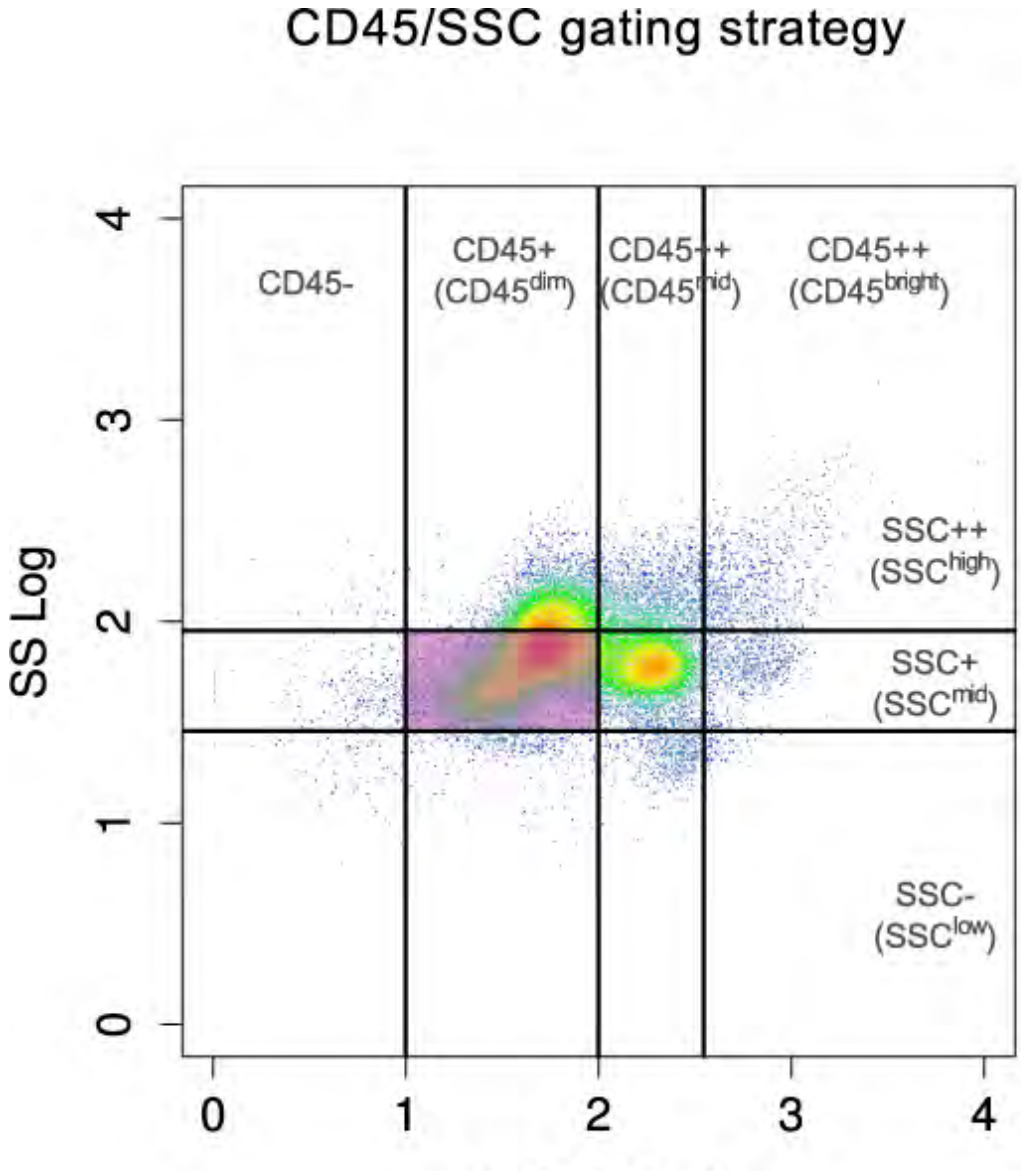


Figure from Ganesan *et. al.*, JID, 2011.

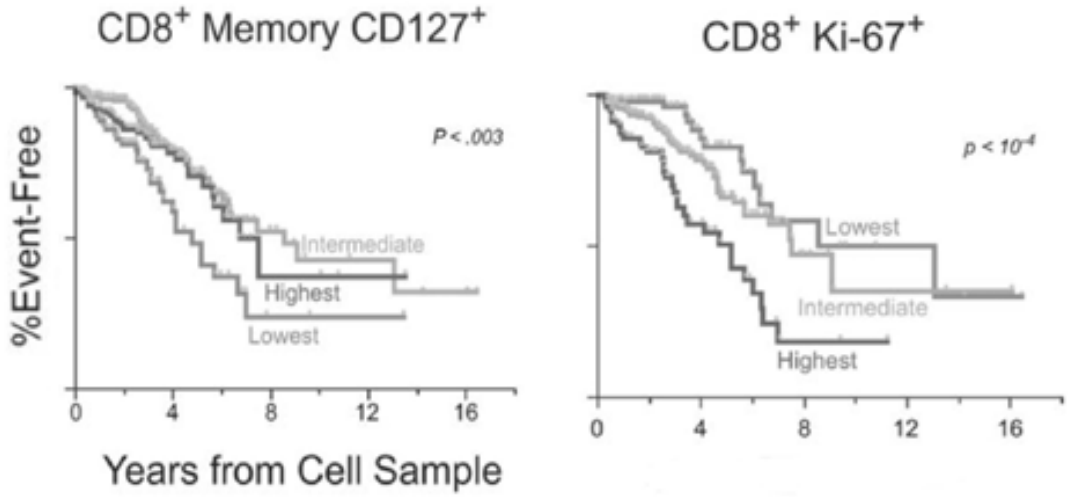
III: Tools for Discovery: flowType/RchyOptimyx

flowType2 Immunophenotype Extraction Concept

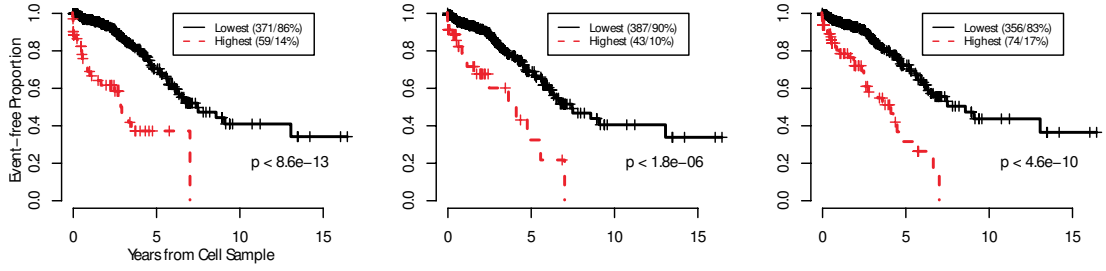
For ten markers: $3^{10} \approx 60,000$ possible cell populations



Manual analysis:

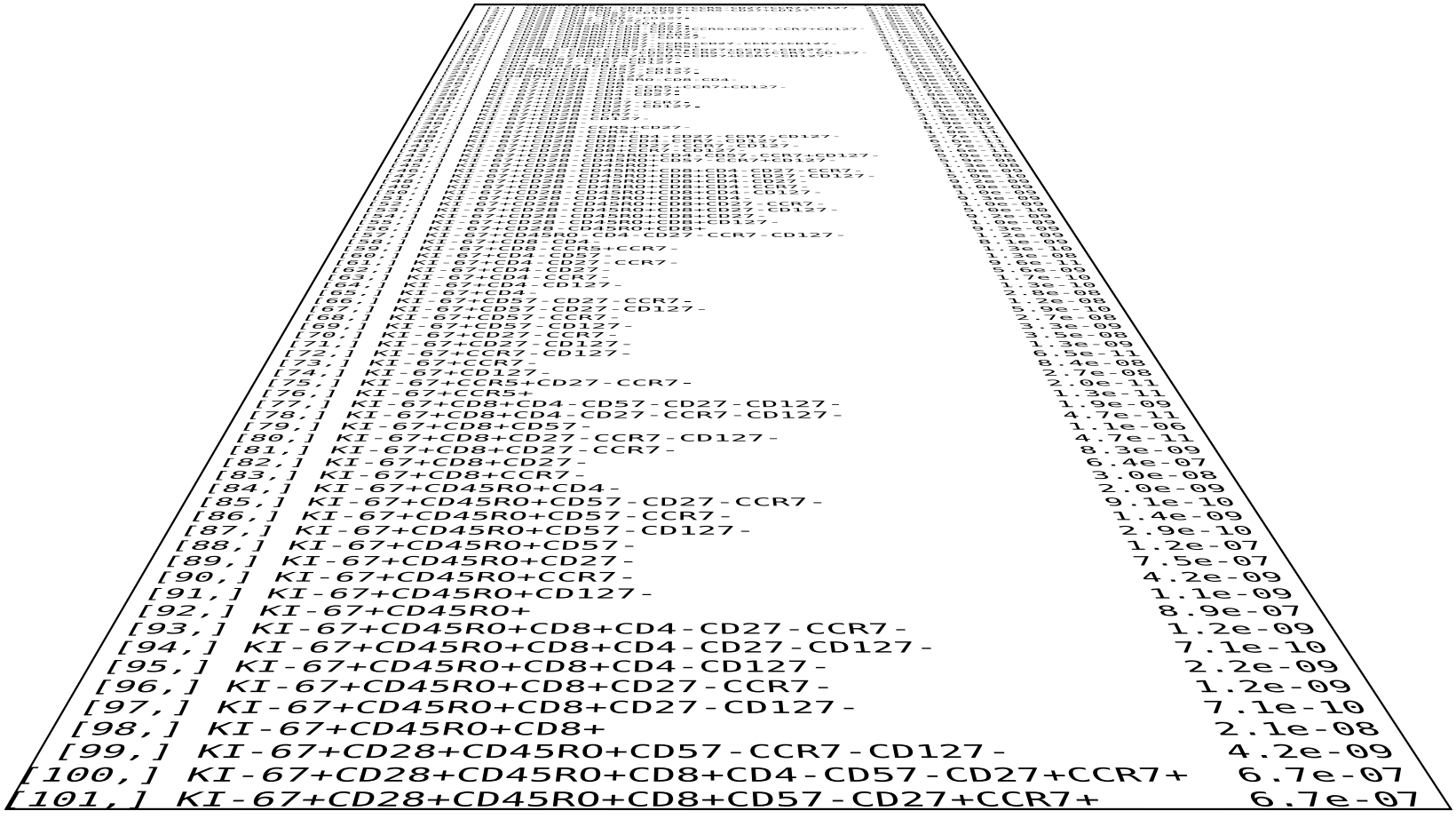


Computational analysis:



flowType: Aghaeepour *et. al.*, *Bioinformatics*, 2012

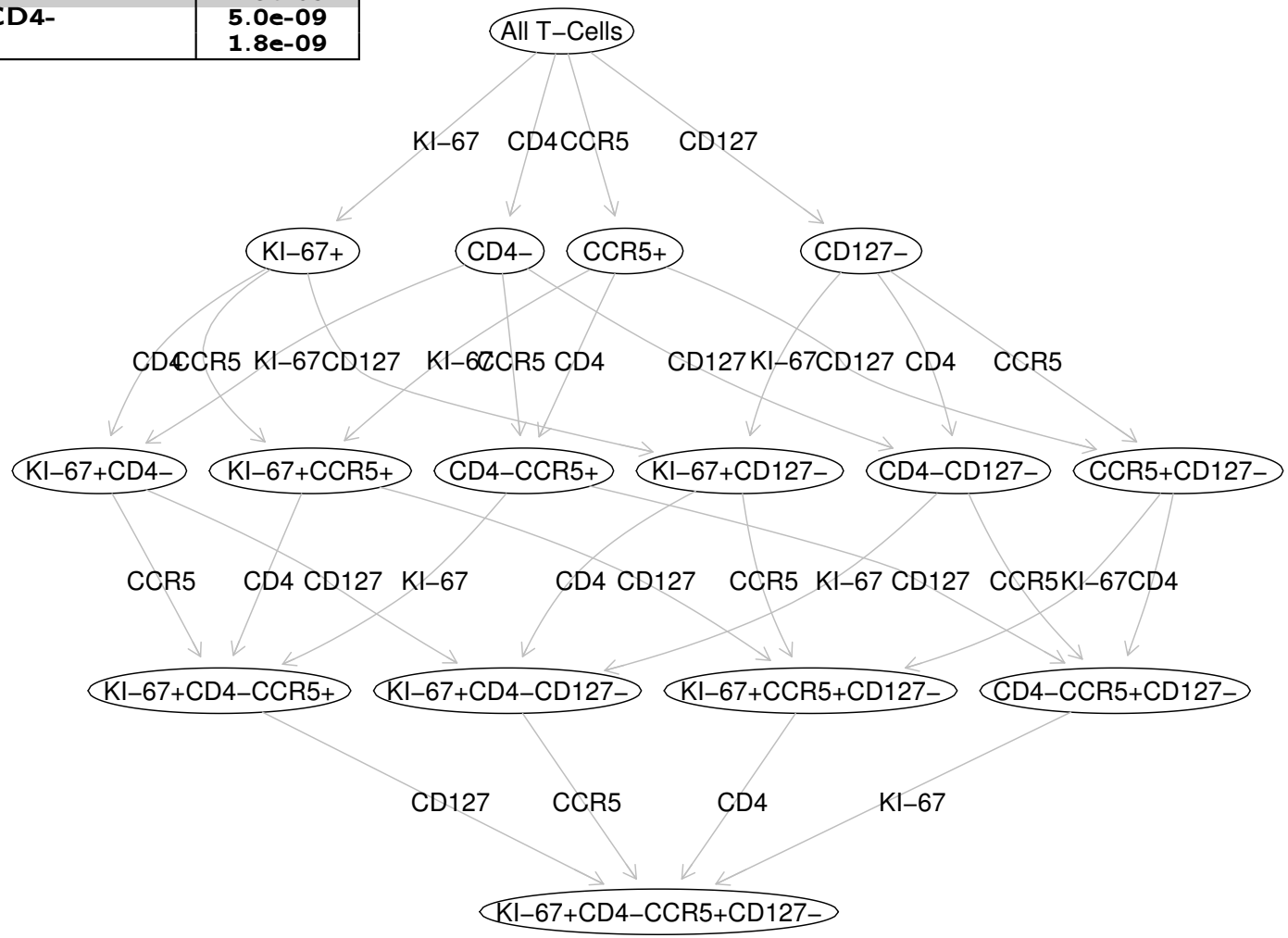
Problem with thorough analysis of large datasets: 60,000 total, 101 statistically significant immunophenotypes*



*Cox proportional hazards regression with bootstrapping & p-value adjustment

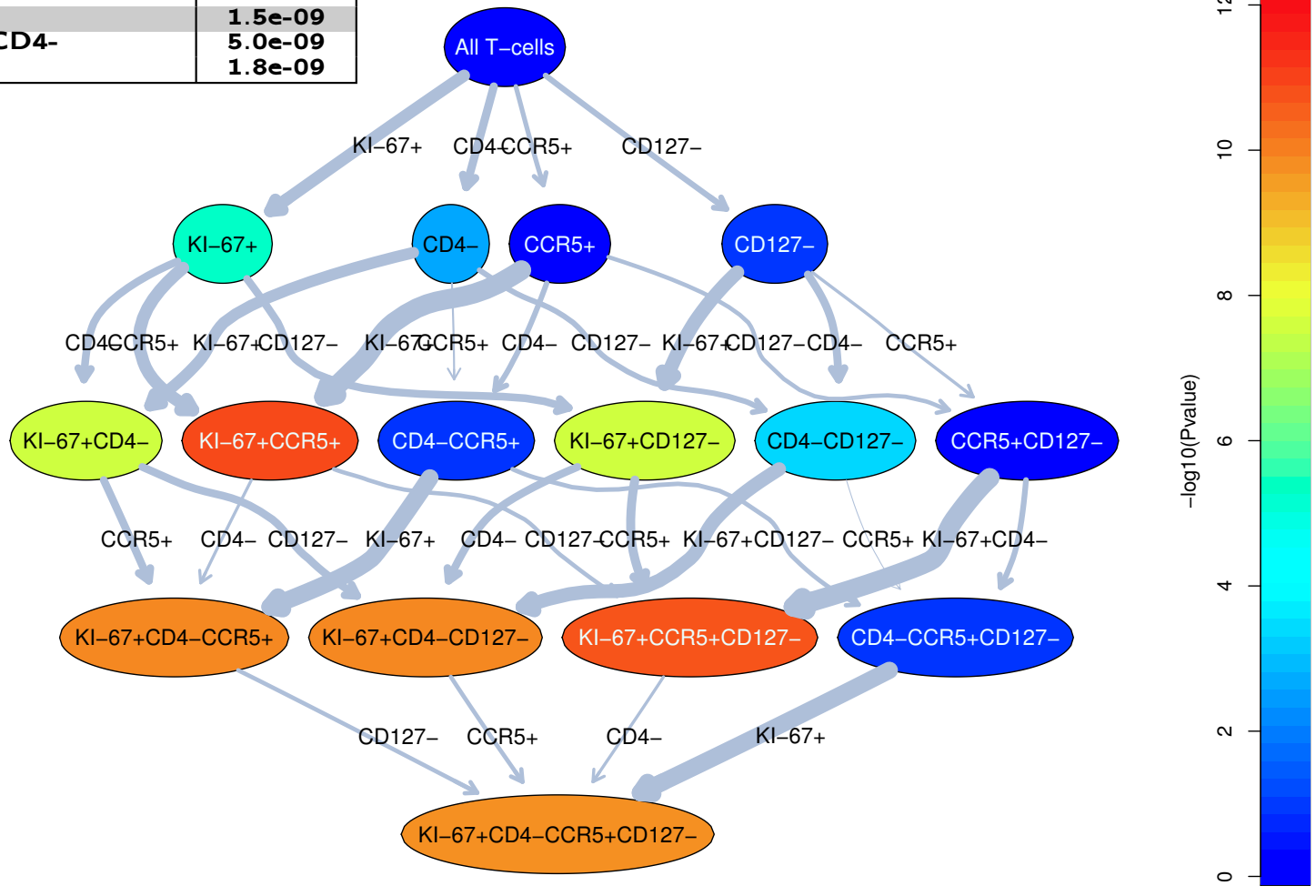
Cellular Hierarchy Concept

Immunophenotype	p
CD28-CD4-CD57-CD127-	3.3e-08
CD28-CD57-CCR7-CD127-	2.8e-07
CD28-CD8+CD57-CD127-	2.5e-07
CD28-CD45RO+CD4-CD57-	3.5e-07
KI-67+CD28-CD8-CCR5+CCR7+CD127-	2.6e-09
KI-67+CD28-CD4-CD57-	9.7e-09
KI-67+CD4-CCR5+CD127-	1.5e-09
KI-67+CD28-CD45RO-CD8-CD4-	5.0e-09
KI-67+CD28-CD4-CD27-	1.8e-09



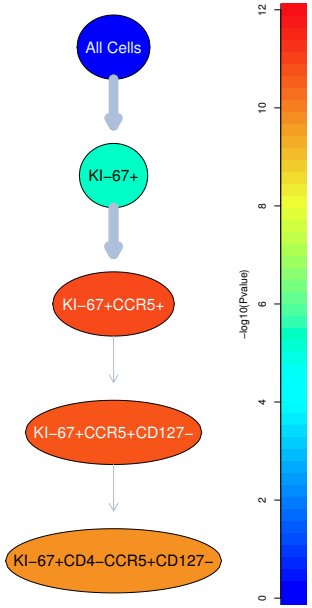
Cellular Hierarchy Concept

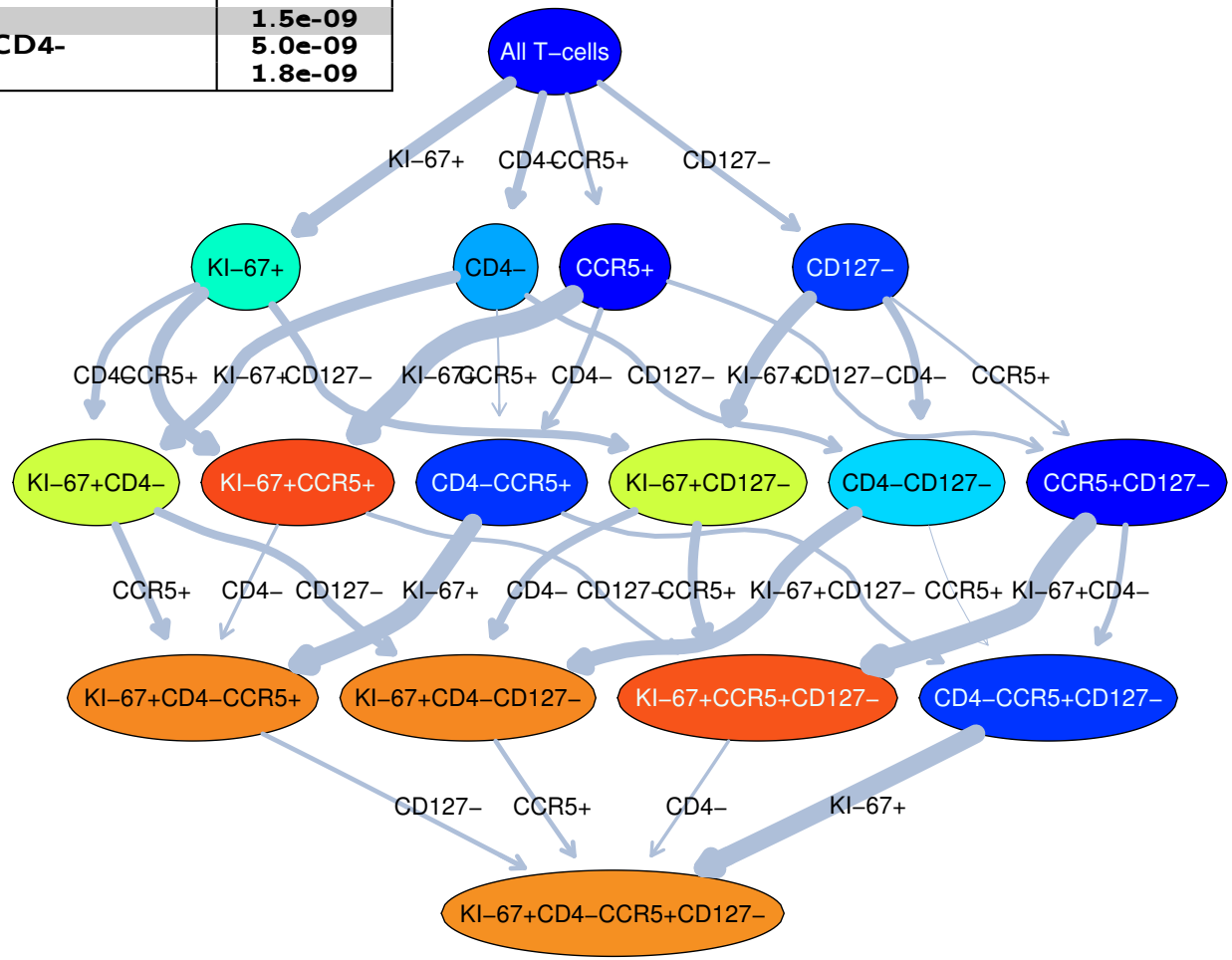
Immunophenotype	p
CD28-CD4-CD57-CD127-	3.3e-08
CD28-CD57-CCR7-CD127-	2.8e-07
CD28-CD8+CD57-CD127-	2.5e-07
CD28-CD45RO+CD4-CD57-	3.5e-07
KI-67+CD28-CD8-CCR5+CCR7+CD127-	2.6e-09
KI-67+CD28-CD4-CD57-	9.7e-09
KI-67+CD4-CCR5+CD127-	1.5e-09
KI-67+CD28-CD45RO-CD8-CD4-	5.0e-09
KI-67+CD28-CD4-CD27-	1.8e-09

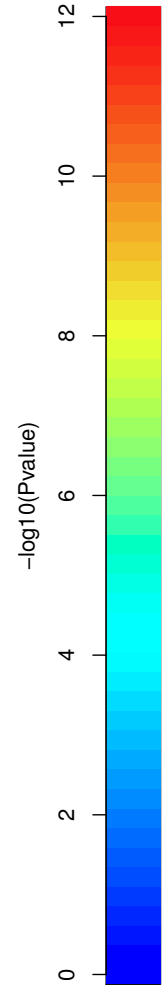


Cellular Hierarchy Concept

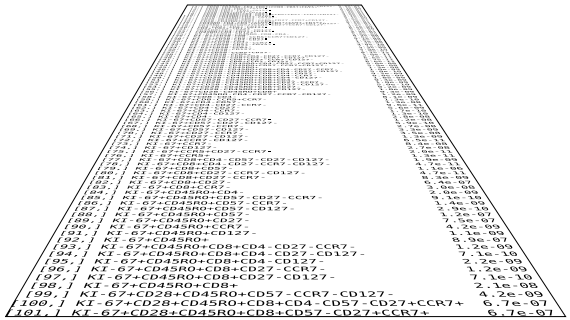
Immunophenotype	p
CD28-CD4-CD57-CD127-	3.3e-08
CD28-CD57-CCR7-CD127-	2.8e-07
CD28-CD8+CD57-CD127-	2.5e-07
CD28-CD45RO+CD4-CD57-	3.5e-07
KI-67+CD28-CD8-CCR5+CCR7+CD127-	2.6e-09
KI-67+CD28-CD4-CD57-	9.7e-09
KI-67+CD4-CCR5+CD127-	1.5e-09
KI-67+CD28-CD45RO-CD8-CD4-	5.0e-09
KI-67+CD28-CD4-CD27-	1.8e-09

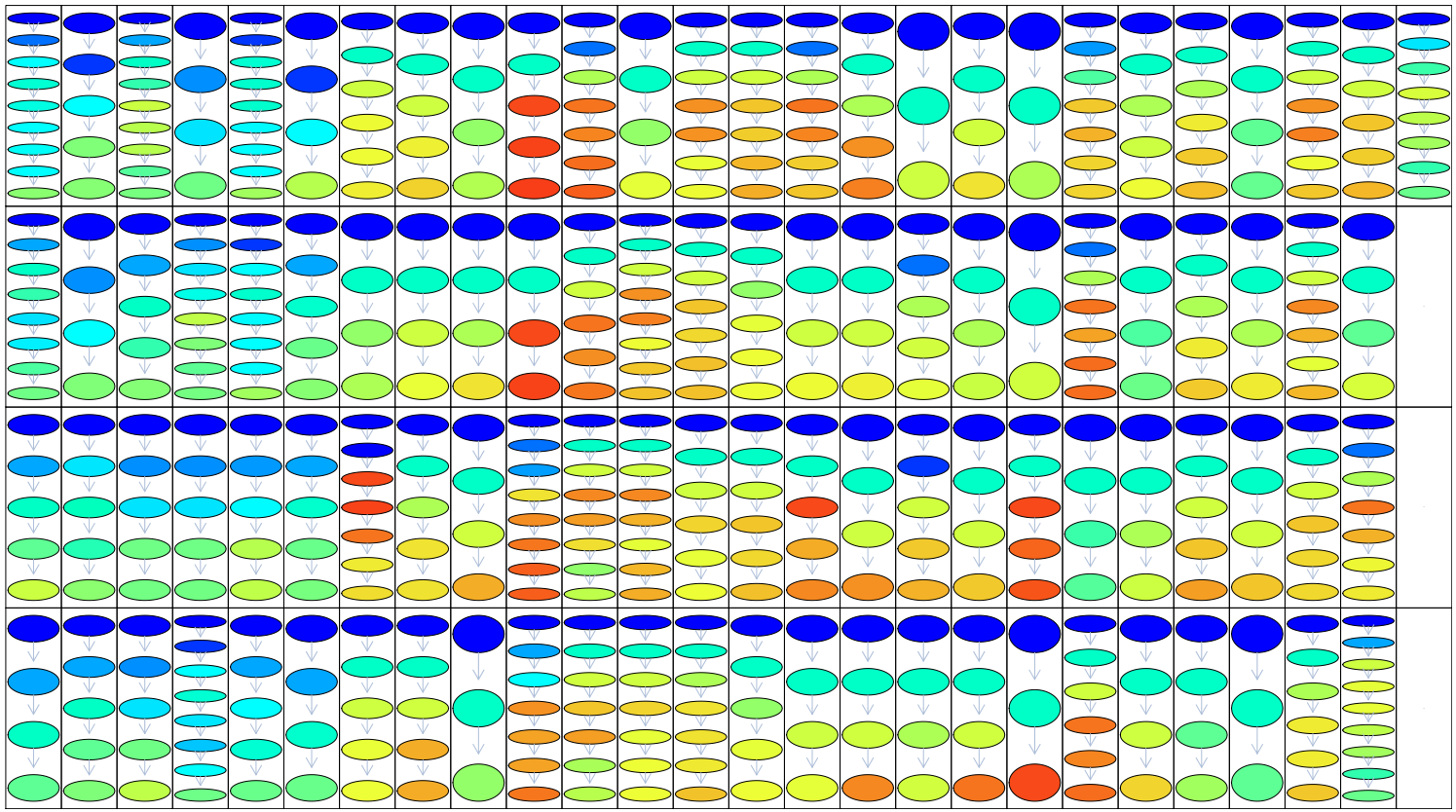




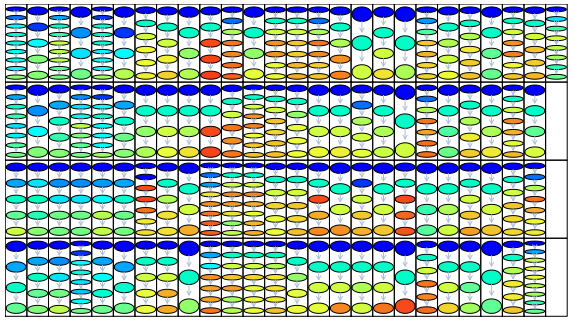


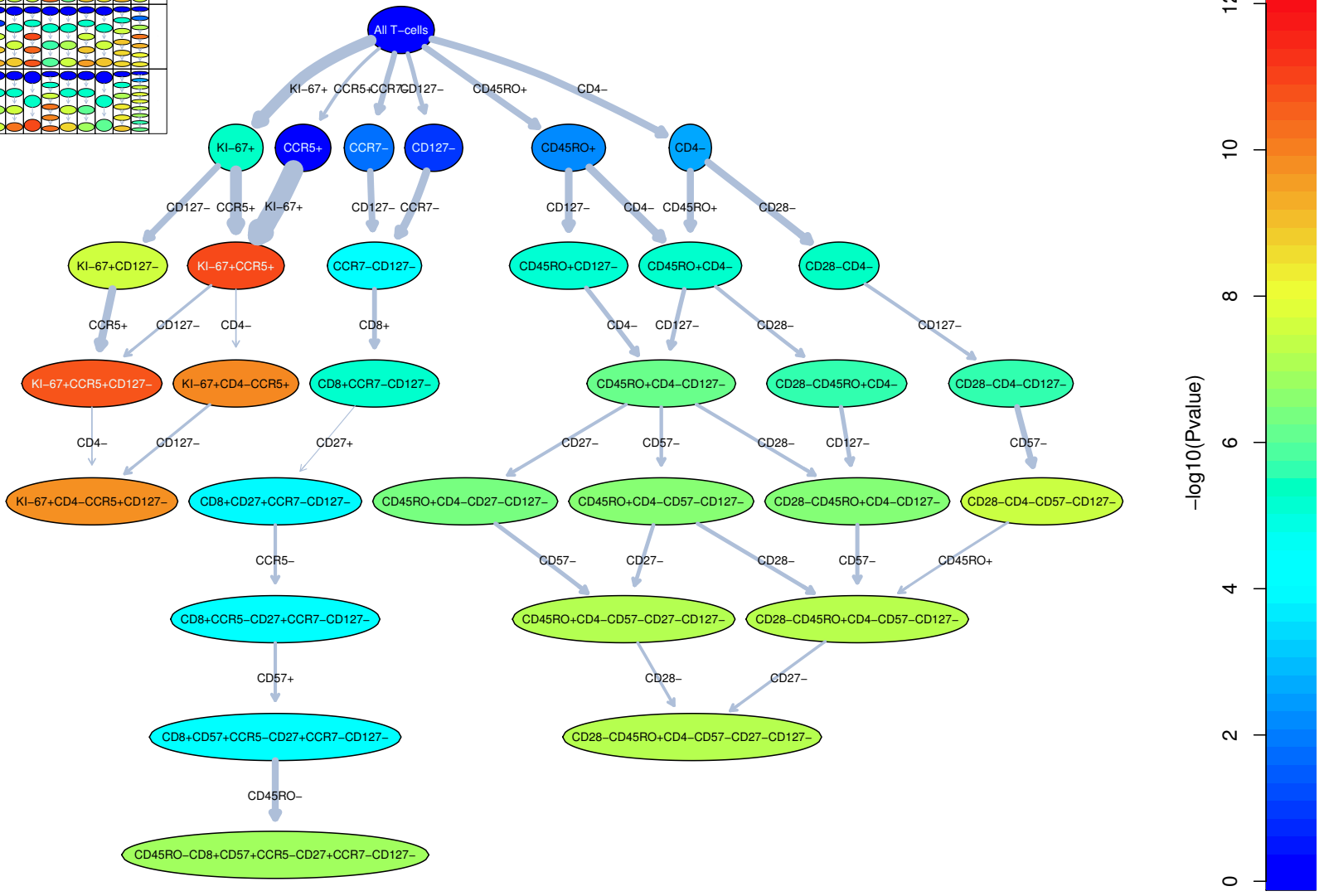
Best Hierarchy for Every Immunophenotype





Merge Hierarchies

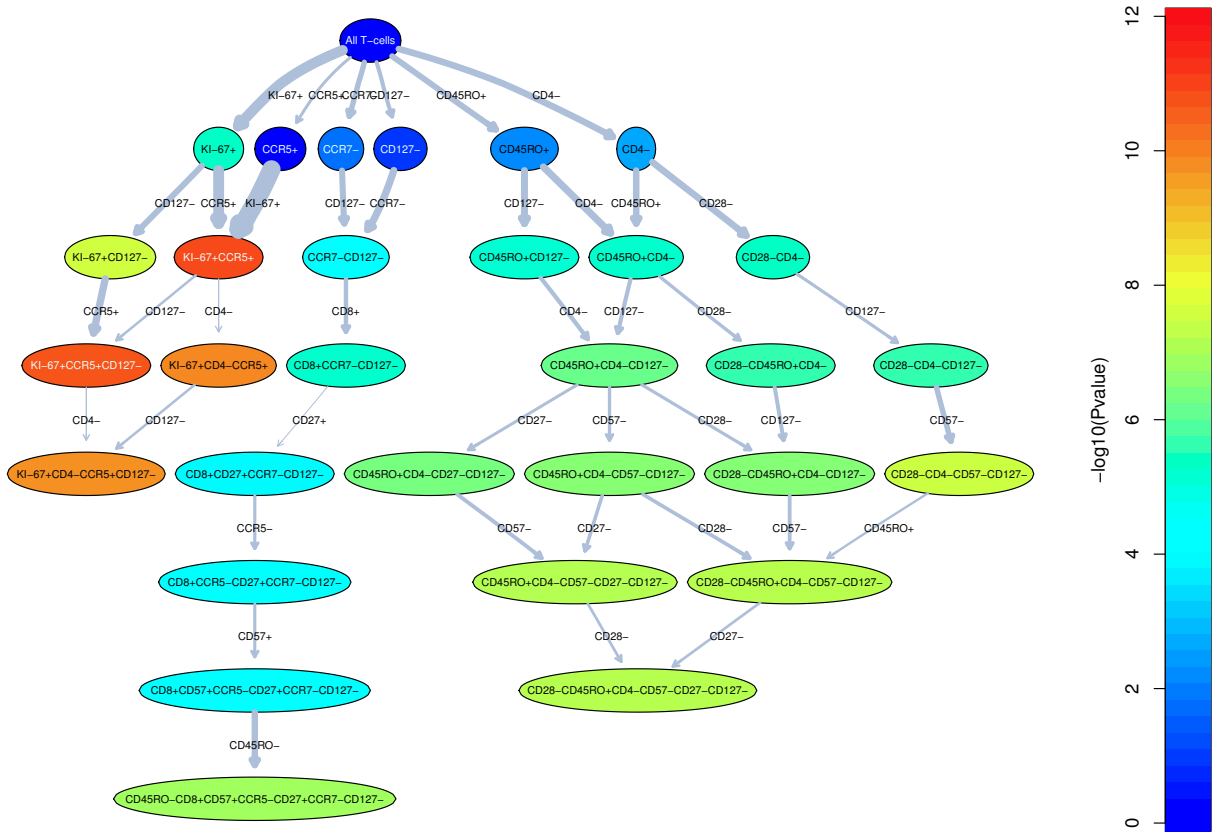




RchyOptimyx: Discover important immunophenotypes

Summarizing 466 patient/ 16 parameter dataset in 1 figure

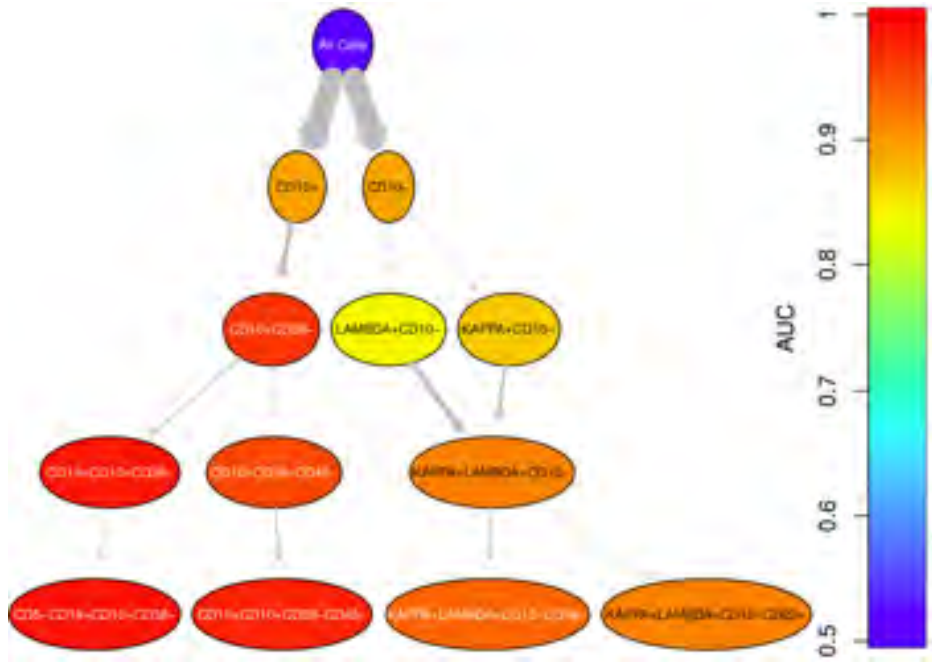
Annotate a large number of cell populations ID'd by other methods (*e.g.*, manual gating, SPADE, flowType) in terms of importance.



RchyOptimyx Example 2

52 Germinal center lymphoma vs. 48 Reactive lymphoid hyperplasia

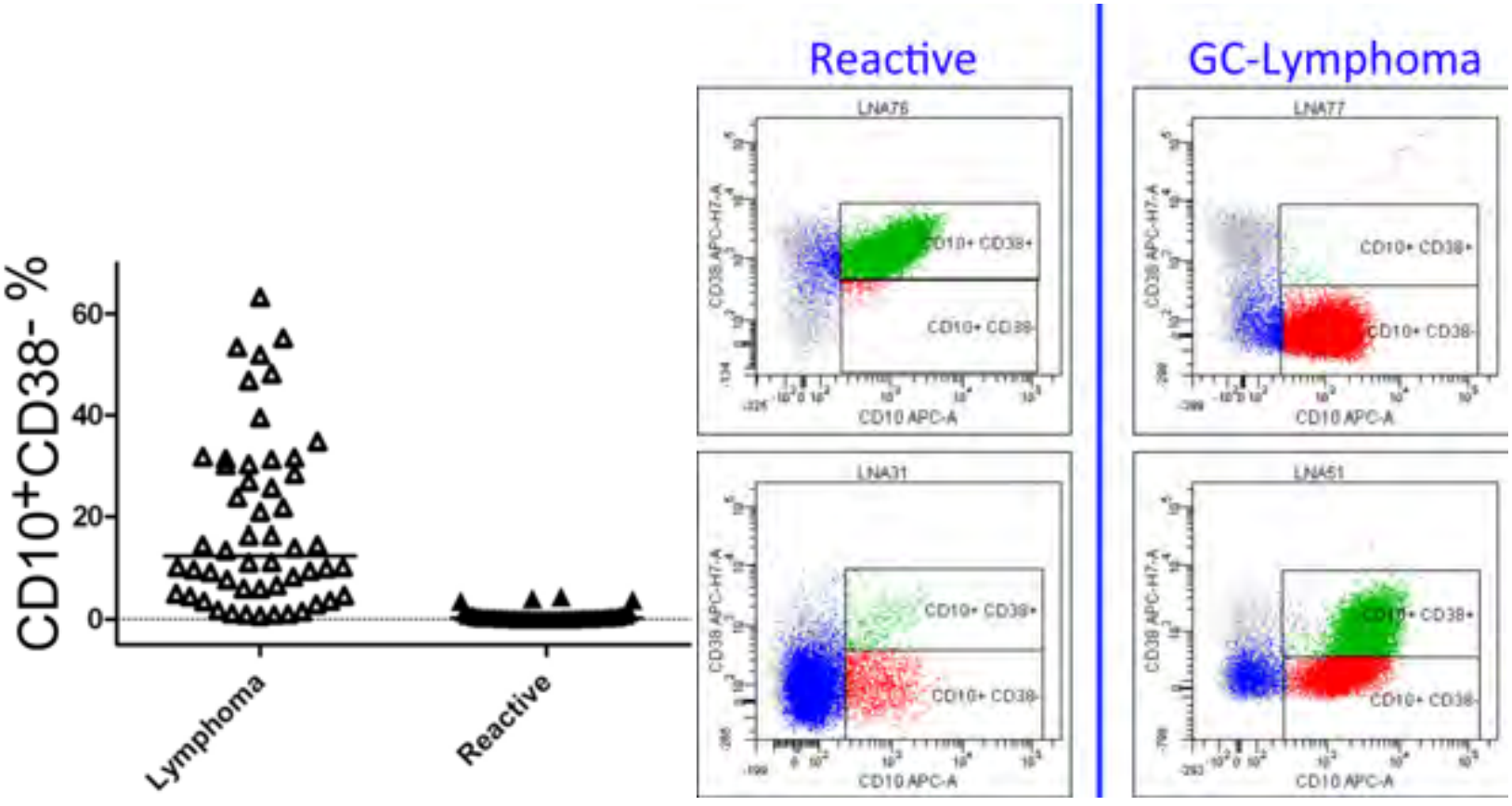
1. 5,660 phenotypes ID'd by flowType for 8 colour B-cell tube
2. ROC analysis to ID phenotypes with a strong predictive power
3. CD5-CD19+CD10+CD38- not ID'd by manual analysis



Craig et al., *Cytometry B*, 2013

Manual re-analysis: CD10+CD38-

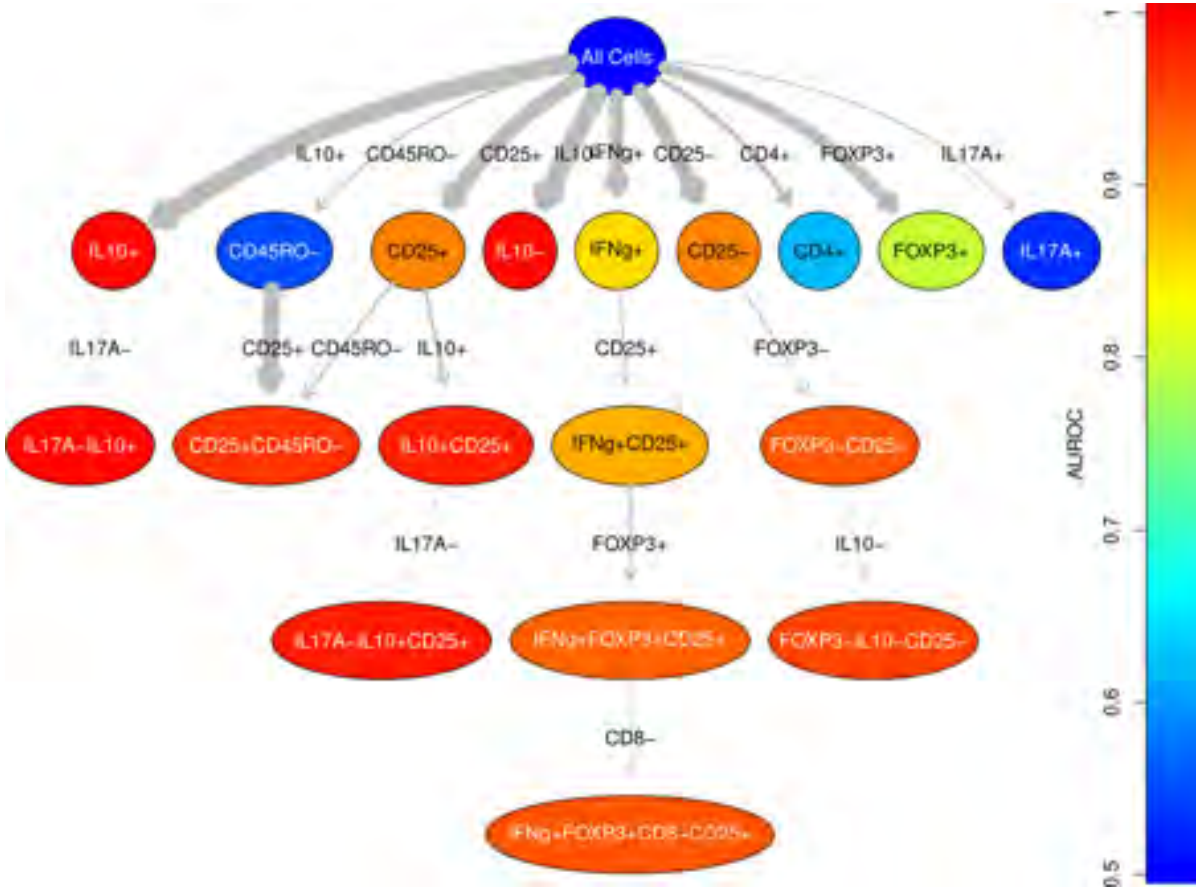
Specificity of 91.75%; Sensitivity of 65.4%



As seen first in Mantei and Wood, Flow Cytometric Evaluation of CD38 Expression Assists in Distinguishing Follicular Hyperplasia from Follicular Lymphoma, *Cytometry Part B*, 2009

RchyOptimyx Example 3

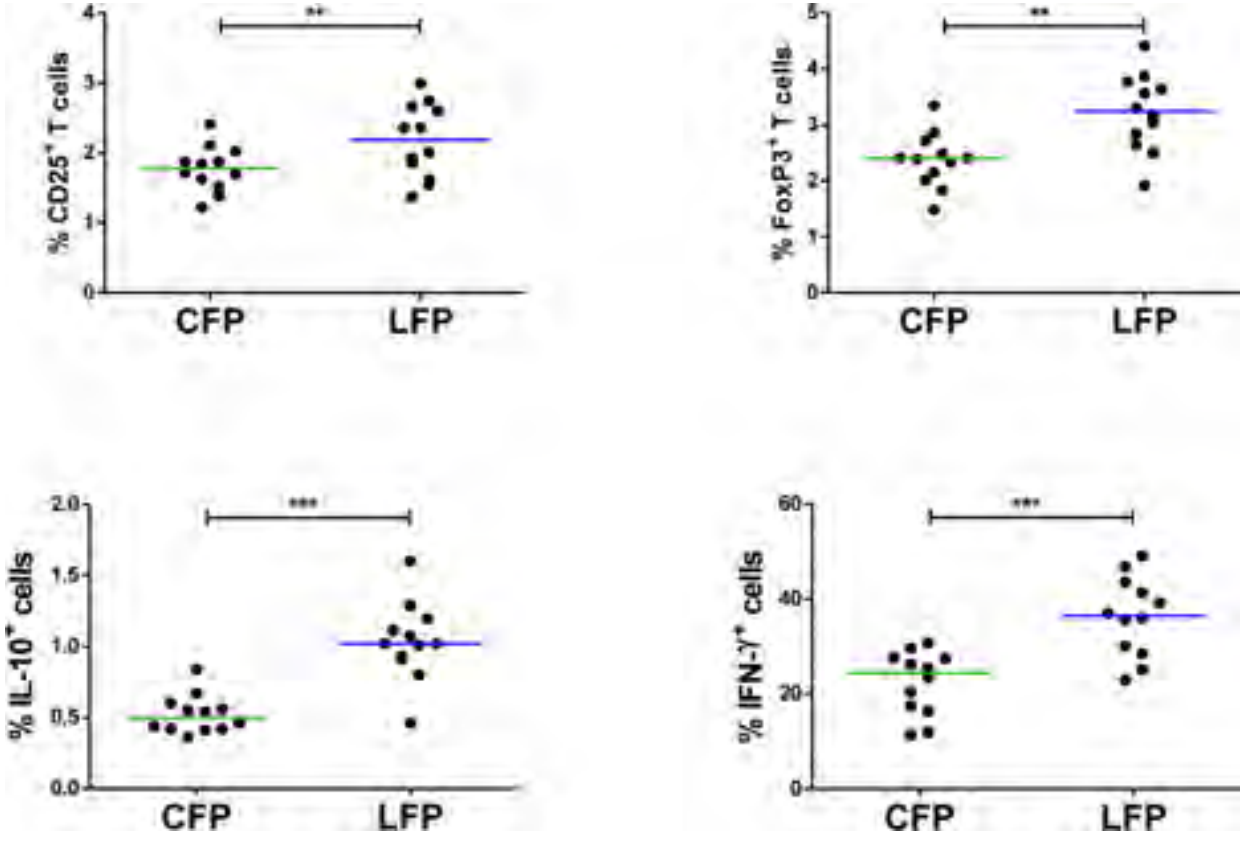
Most significant cell populations differentiating 2 treatment groups



Villanova et al. *PLoS ONE*, 2013

Liquid vs. Lyoplate Reagents

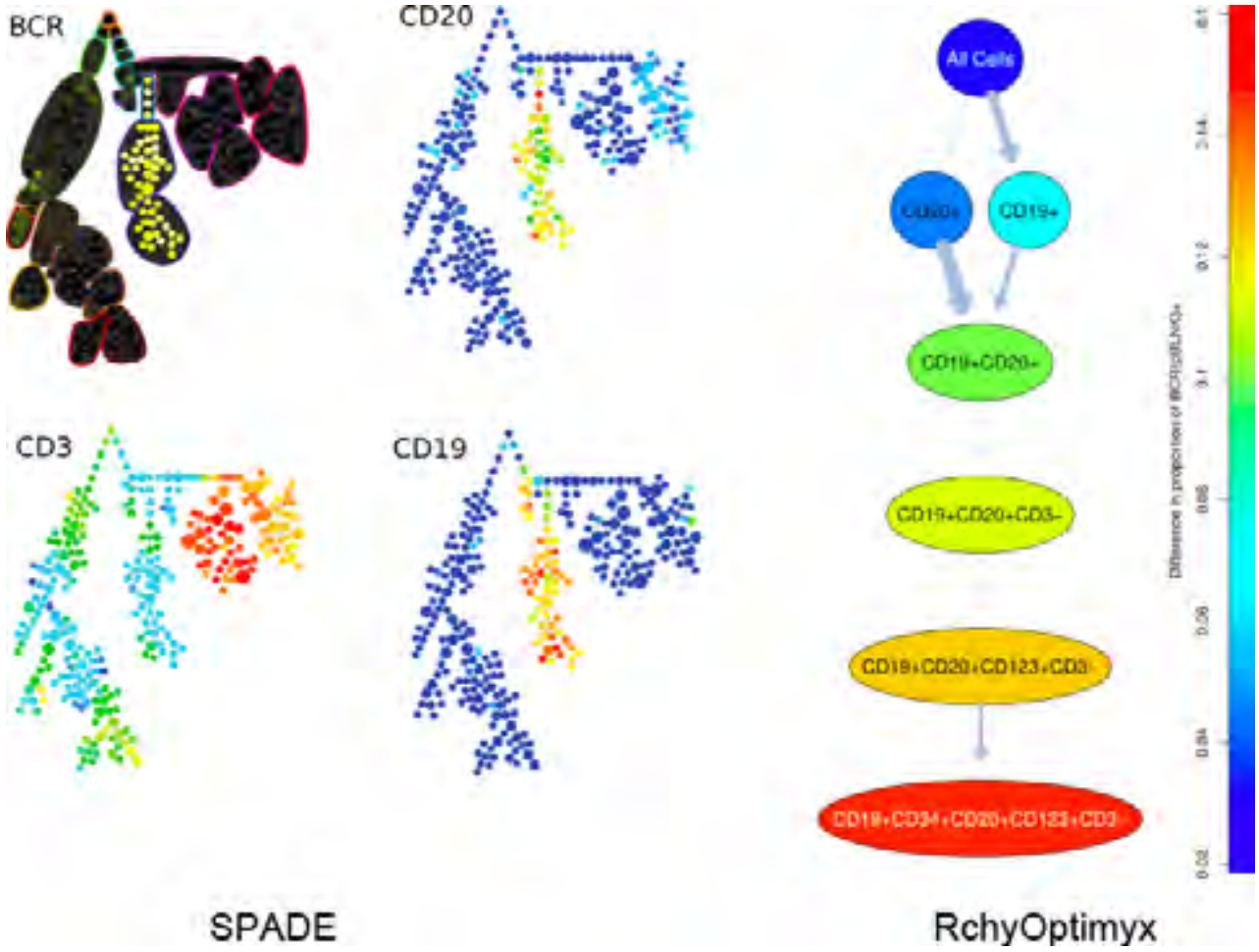
Manual validation of automated results



- Lyoplate: better detection of cytokines & activation markers
- Increased overall brightness

SPADE* vs. RcyhyOptimyx**

Bake off using same CyTOF dataset



*Qiu *et al. Nature Biotechnology*, 2011

**Aghaeepour *et al. Cytometry A*, 2012

SPADE* vs. RcyhyOptimyx** for Discovery

	SPADE	RchyOptimyx
CyTOF		
Relationships defined by distance		
Relationships defined by correlations		
Manual annotation required		
Good for gradual expression change		
Relationships involving several markers		
From the lab of the guy giving this talk		
Count	4	4

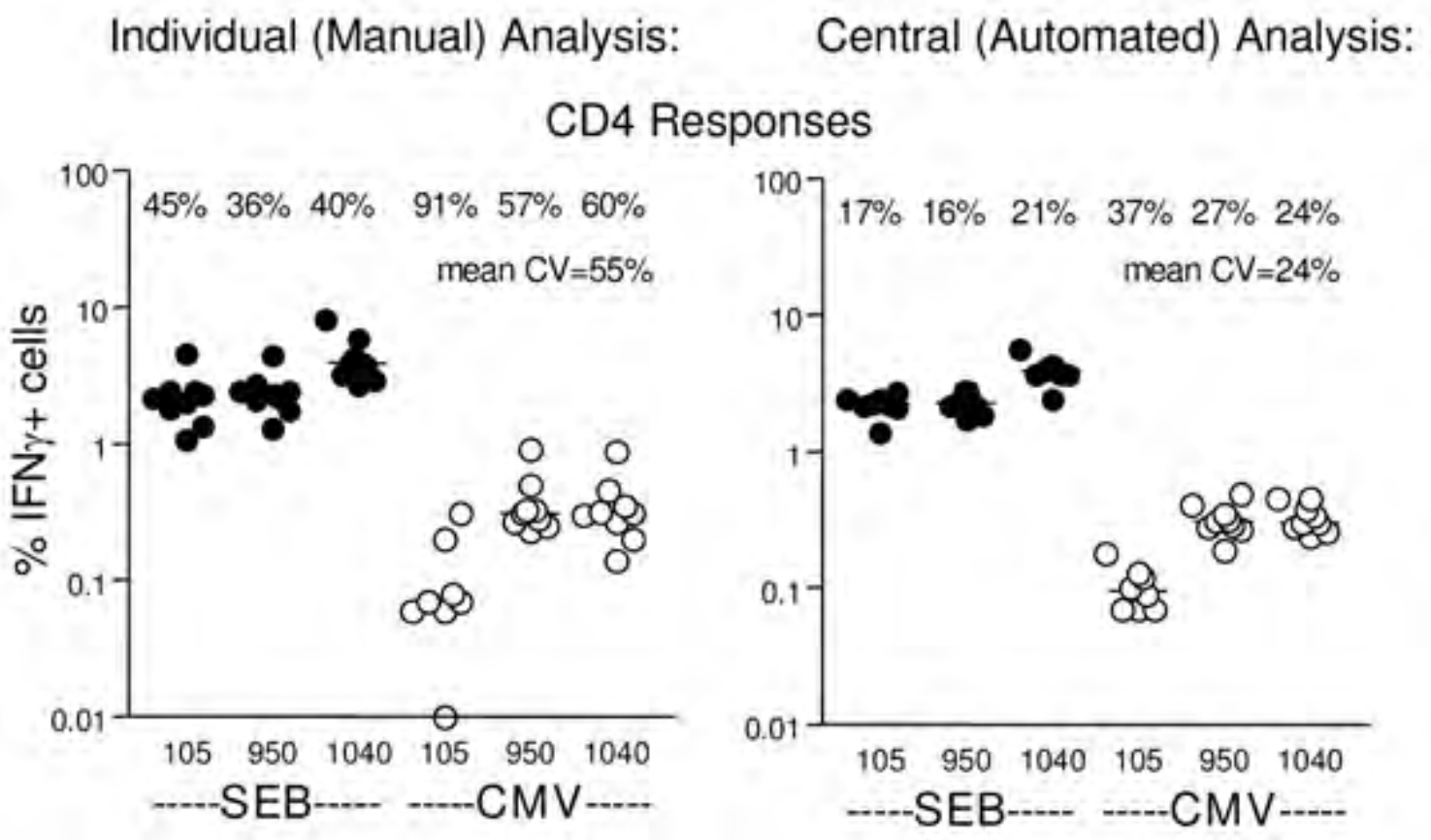
*Qiu *et al. Nature Biotechnology*, 2011

**Aghaeepour *et al. Cytometry A*, 2012

Episode IV: The Hope of flowDensity for "Diagnosis"

Reducing Variation in Repetitive Analyses

- Most variation in cross-center studies due to gating
  - Variation reduced by 30% using 1 manual gater

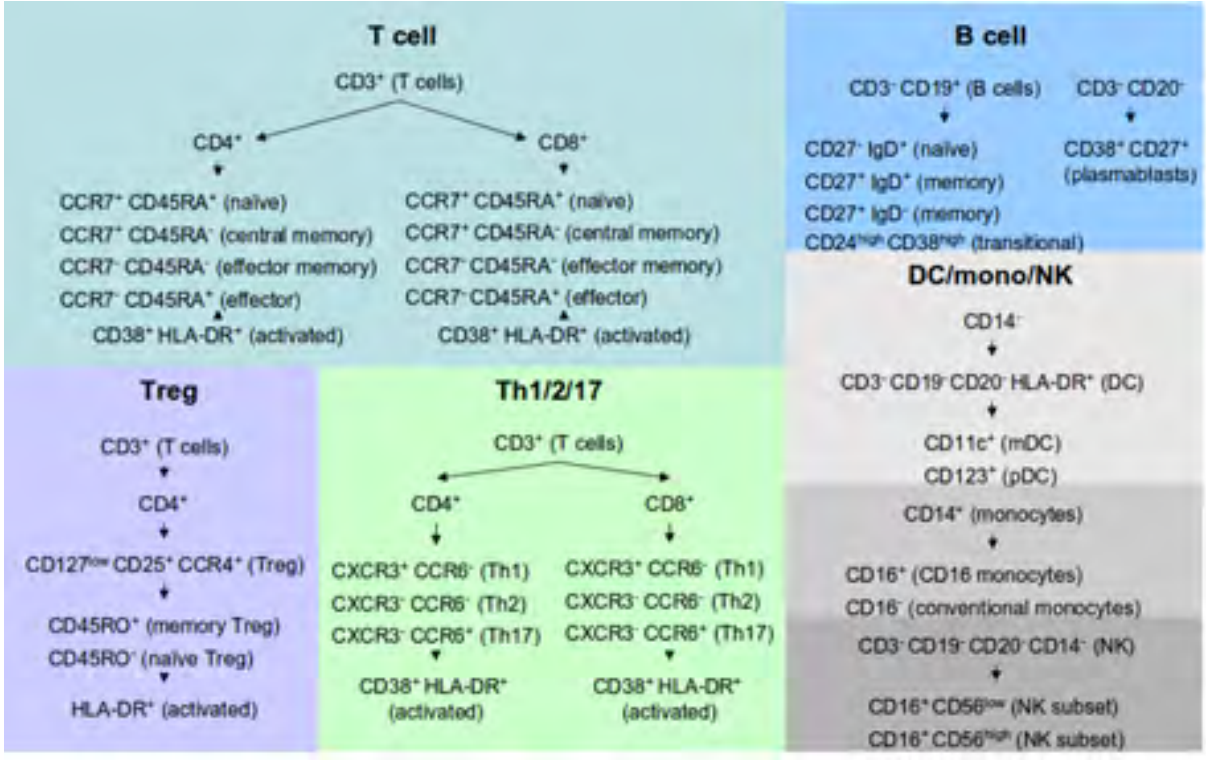


Maecker *et al. BMC Immunology*, 2005

Tools for Diagnosis in Practice

Reducing Variability of Analysis in Large Datasets

Hypothesis: Automated gating and lyophilized reagents would significantly reduce variability in large, multi-center studies

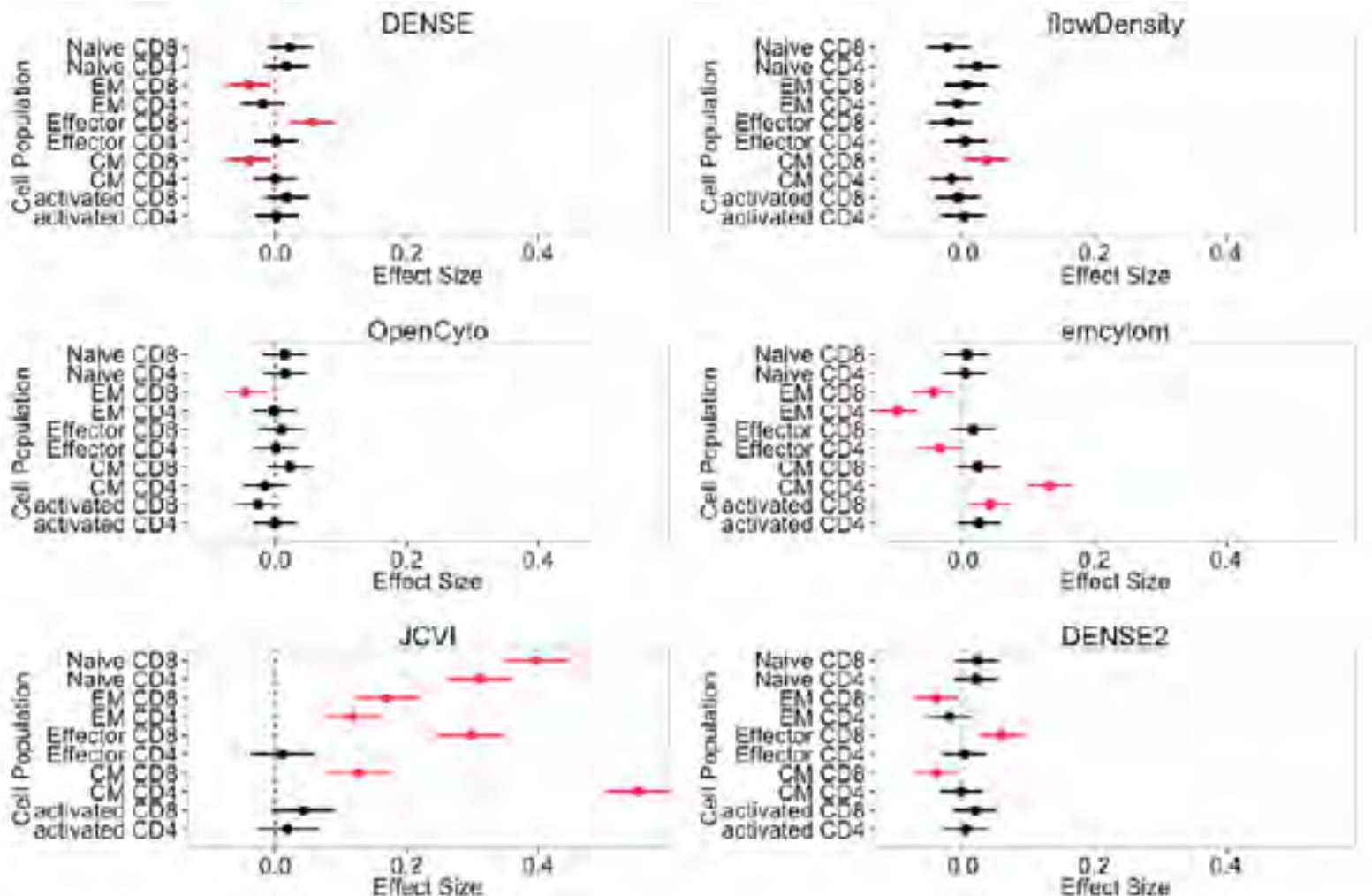


Maecker et al. *Nature Reviews Immunology*, 2012

FlowCAP-3 Directed Automated Analysis

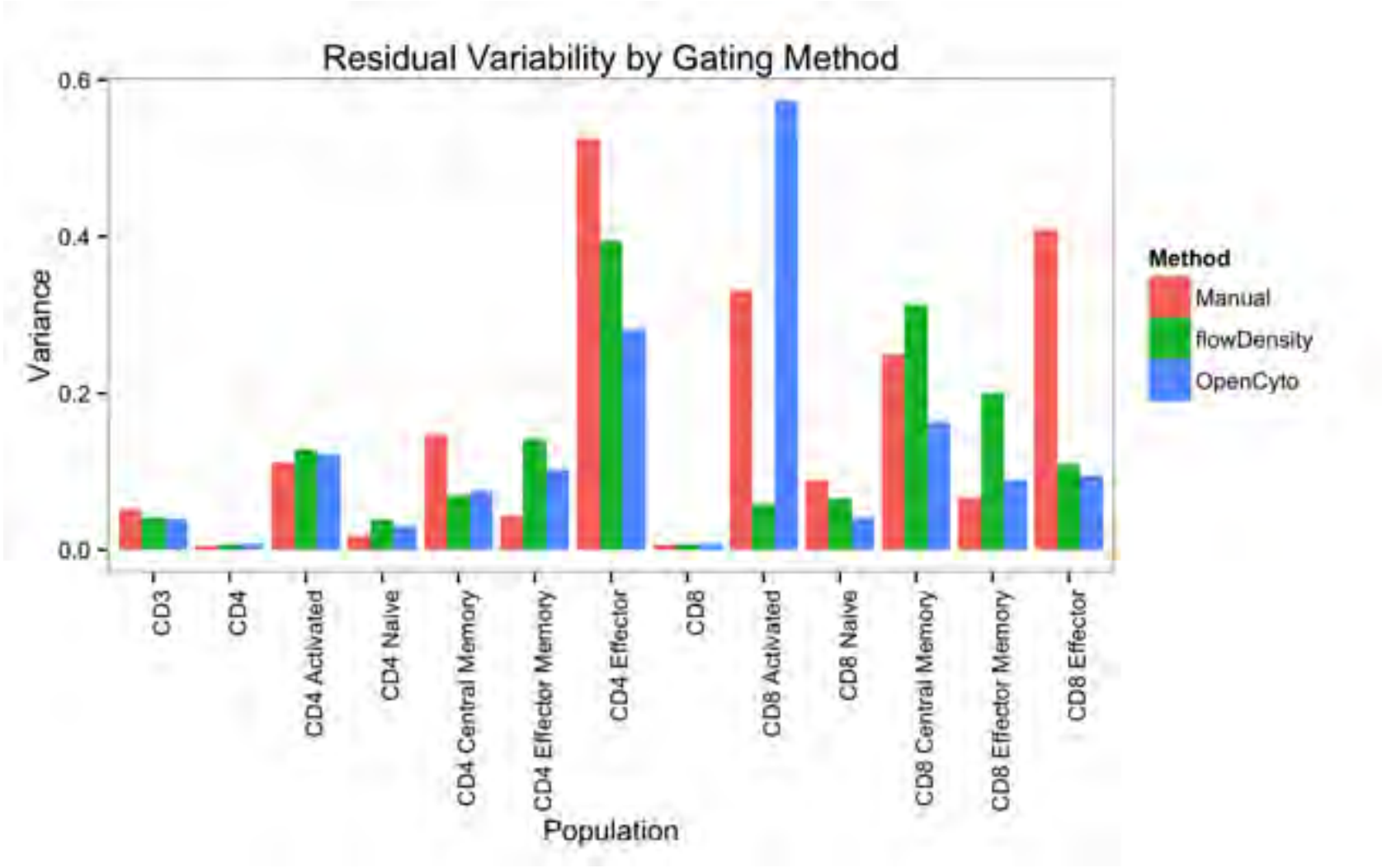
T cell panel (Corrected for donor and centre-level effects)

- 9 sites, 4 replicates of cryopreserved cells per site.
- Centralized gating of data based on a consensus best approach.
- Automated algorithms vs. the centralized gating.



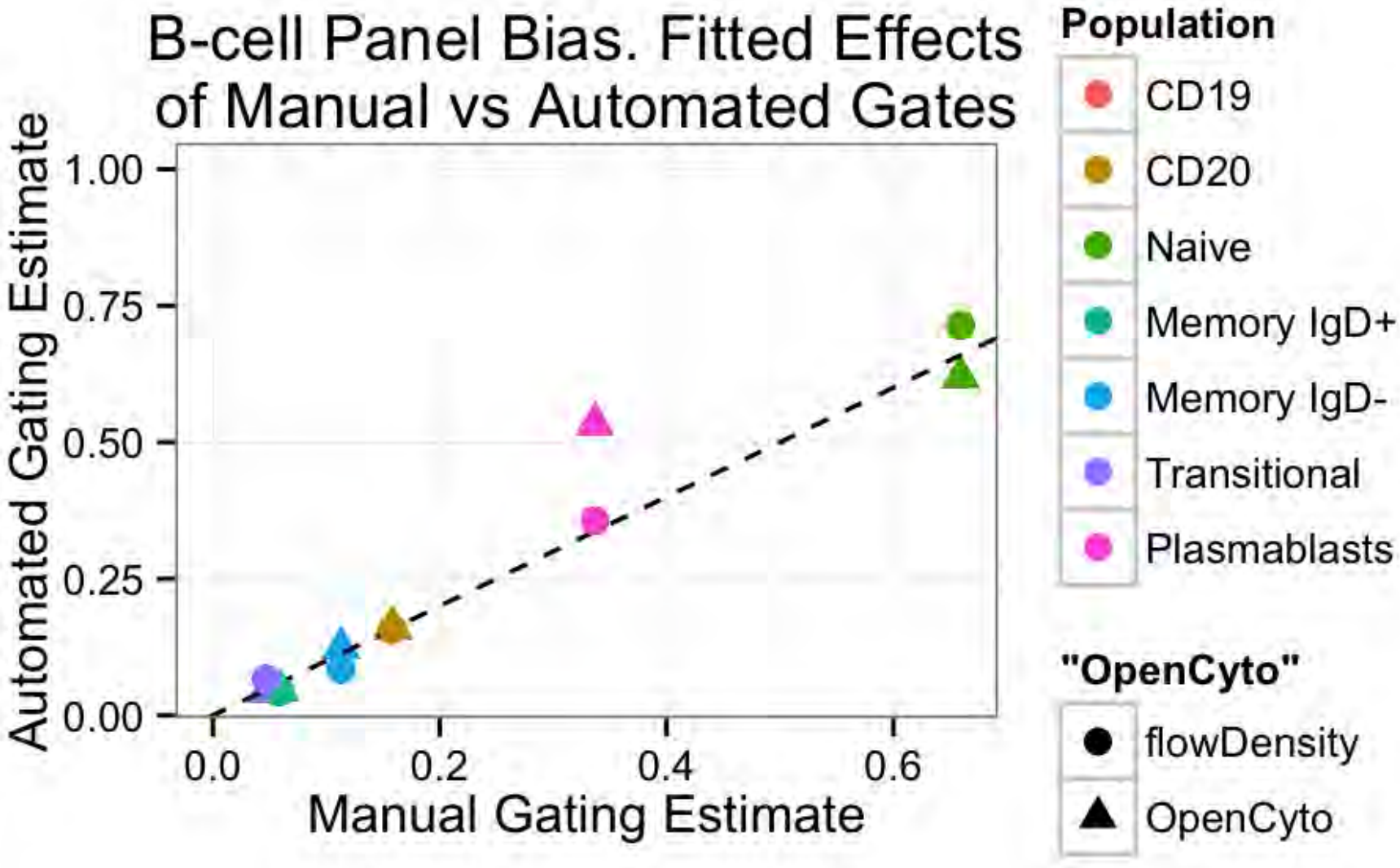
T-cell Panel: Variance vs. Centralized Gating

Centralized Gating vs. Automated Accounting for Center and Individual Effects

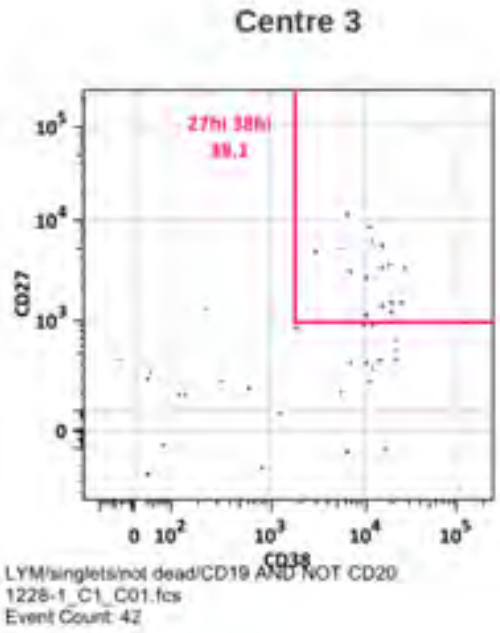


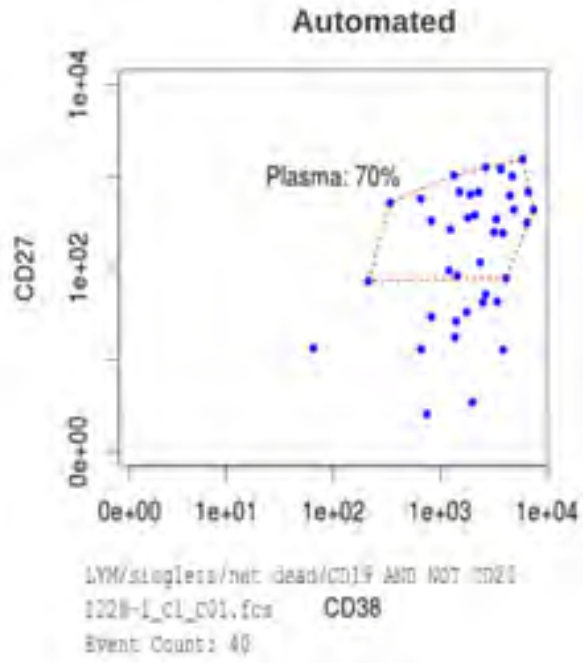
B-cell Panel

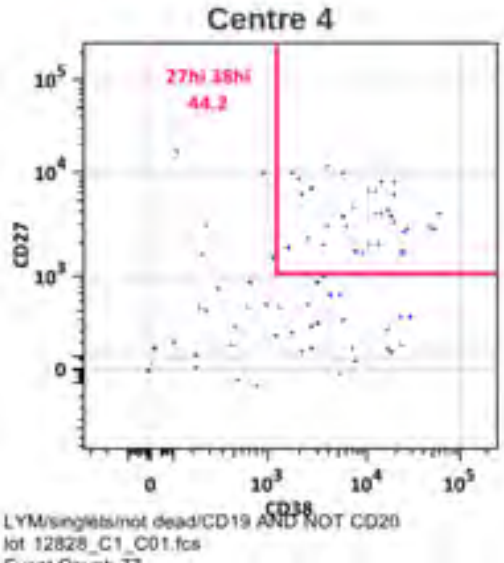
Centralized Gating vs. Automated

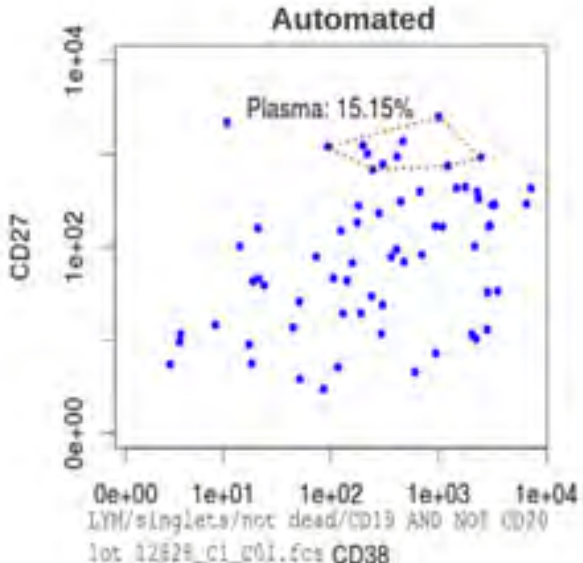


Several Gating Issues (*e.g.*, Plasmablasts-B Cells)



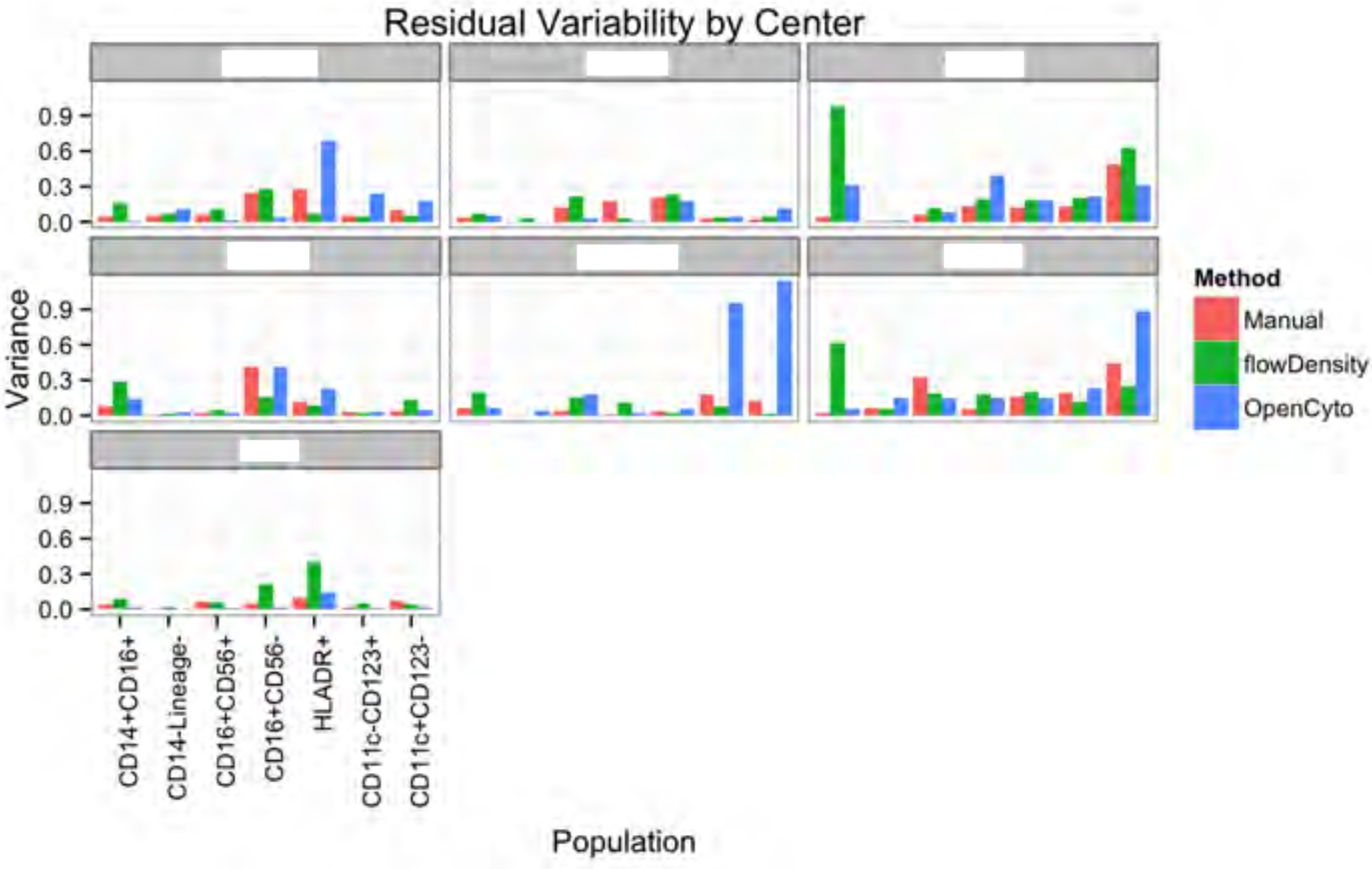






Center-level effects can be significant for some populations

DC/Mono/NK cells panel



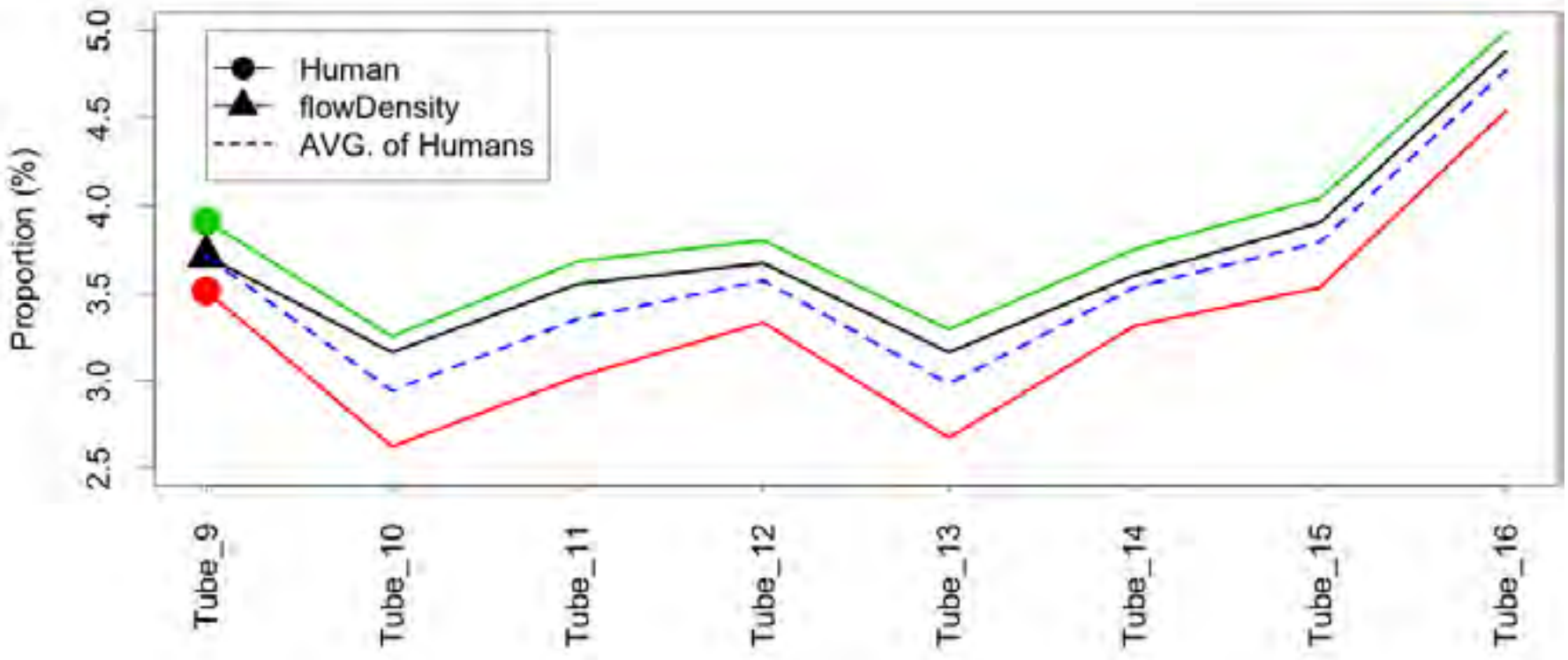
HIP-C conclusions
(for most cell populations)

- 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

flowDensity vs. 3 human experts for NK cells

Massive flow data generation

- 20,000 lines (2Fx1M) generated (1/gene) over next 5 years
- 2 x 10-12D FCS files for each of 60,000 mice
- 120,000 FCS files and 25 other phenotype measurements



flowType/RchyOptimyx and flowDensity

Automated analysis for discovery and diagnosis

- 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
  - Guaranteed* (lower CV, same range), or your money back
- 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

*When your gating strategy is supported by your data

Practical Considerations for Automated Analysis

- Don't waste your time on 12 clinical samples
  - Your study probably isn't sufficiently powered
- Don't waste your time on discovery using 6 colours
  - 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 is good
- Its OK to ask for help

Publicly Available High-Dimensional Datasets

	Microarray	**Flow Cytometry**	Δ
First publication :	1995	1965	30y
Public datasets:	879,722	58	15,000x
Fruit:			xkcd.com/388/

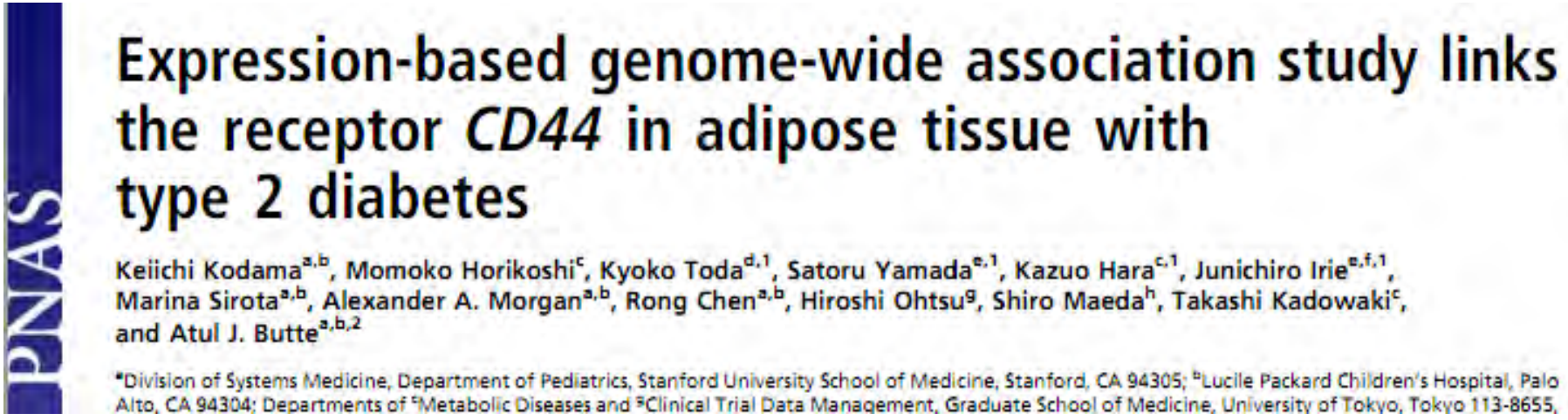
Schena *et. al. Science* (1995) Fulwyler *Science* (1965)

Why Share Your Data?

- Required by many funding agencies and scientific journals
- Promote open scientific inquiry and progress in the field
  - Re-exploration to test new hypotheses or algorithms
  - Independent validation and refutation of experimental findings
- Sharing detailed research data is associated with increased citation rate*
  - Publicly available data significantly ($p = 0.006$) associated with a 69% increase in citations, independently of journal impact factor, date of publication, and author country of origin using linear regression.
- Your mom told you so

*Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3): e308. doi:10.1371/journal.pone.0000308

Case in Point: Data Sharing Furthers Science



PNAS

Expression-based genome-wide association study links the receptor *CD44* in adipose tissue with type 2 diabetes

Keiichi Kodama[a,b], Momoko Horikoshi[c], Kyoko Toda[d,1], Satoru Yamada[e,1], Kazuo Hara[c,1], Junichiro Irie[e,f,1], Marina Sirota[a,b], Alexander A. Morgan[a,b], Rong Chen[a,b], Hiroshi Ohtsu[g], Shiro Maeda[h], Takashi Kadowaki[c], and Atul J. Butte[a,b,2]

[a]Division of Systems Medicine, Department of Pediatrics, Stanford University School of Medicine, Stanford, CA 94305; [b]Lucile Packard Children's Hospital, Palo Alto, CA 94304; Departments of [c]Metabolic Diseases and [g]Clinical Trial Data Management, Graduate School of Medicine, University of Tokyo, Tokyo 113-8655,

Kodamaa *et al.* Expression-based genome-wide association study links the receptor CD44 in adipose tissue with type 2 diabetes *PNAS* (2012)

Episode V: MIFlowCyt Strikes Back Against Data Annotation

What to Share?

- FCS files to facilitate re-analysis
- .. but a dump of FCS files is not enough
  - Data without context are not understandable to others
- Minimum Information about a Flow Cytometry Experiment
  - Outlines the minimum information required to report about flow cytometry experiments
  - Represents the community consensus
    - 33 coauthors from 19 institutions
    - ISAC Recommendation
  - Required/recommended by Cytometry A and Nature Publishing Group

Lee et al., MIFlowCyt: the Minimum Information about a Flow Cytometry Experiment. *Cytometry A*. 2008; 73(10): 926-930

MIFlowCyt Components

Experiment overview

- Purpose
- Keywords
- Experiment variables
- Date(s)
- Organization(s)
- Primary contact
- Quality control measures

Sample description

- Description
- Sample material
- Treatment
- Fluorescent reagents
- Source
- Biological samples: Organism with taxonomy, phenotype, genotype, age, gender, ...
- Location for environmental samples

Data analysis

- FCS data files
- Compensation and other transformations
- Gating details including gate description, statistics and boundaries or images or gate membership details

Instrumentation details

- Make
- Model
- User-adjustable components (e.g., detector voltages)
- Customized configurations

Episode VI: The Return of flowRepository

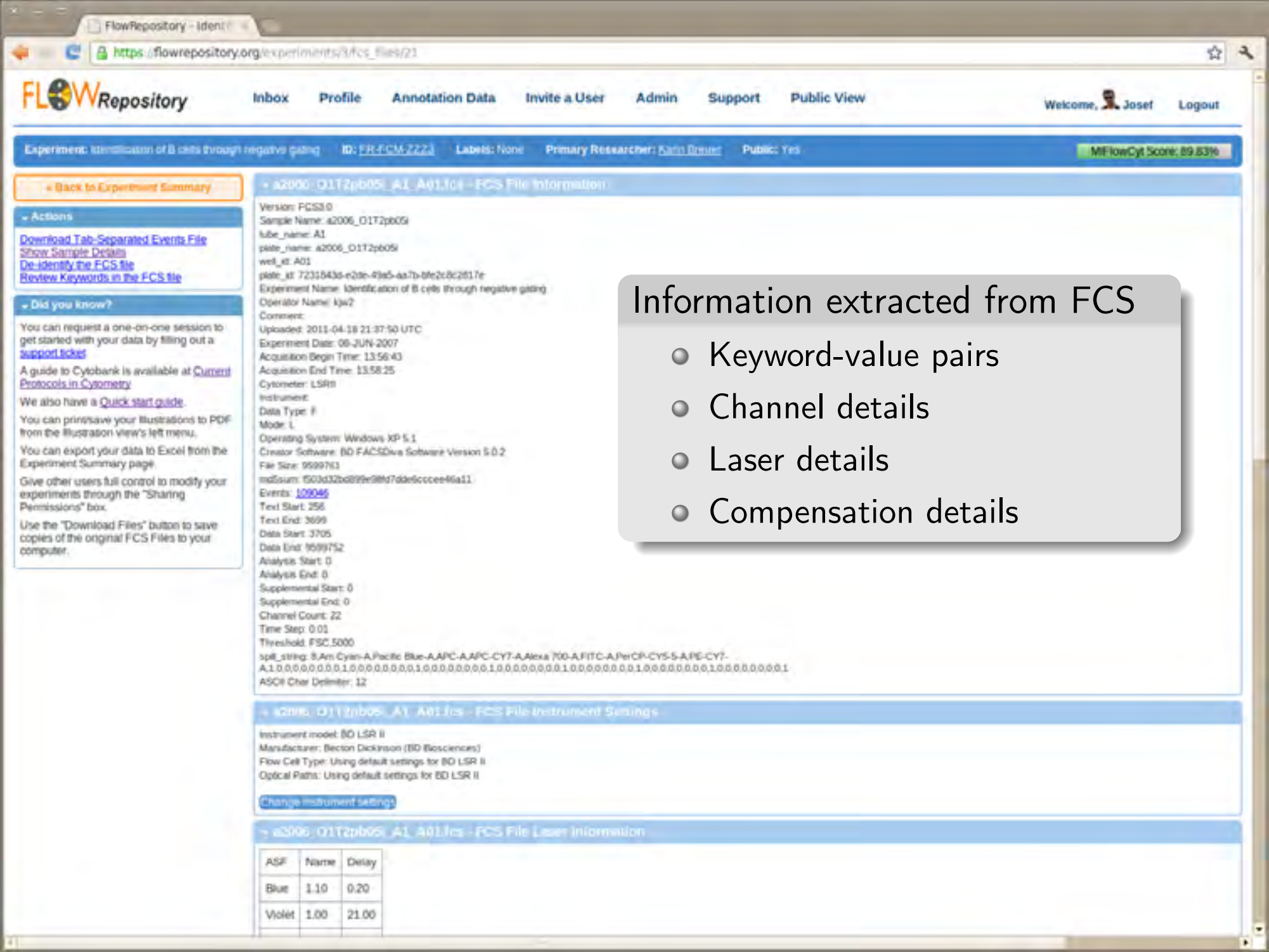
A free, public, online resource of MIFlowCyt-annotated FCS datasets

- Primarily data associated with peer-reviewed publications
- Web-based application created by extending Cytobank
  - Mainly to incorporate MIFlowCyt
- Complete code available as open source
  - Affero General Public License
- Supported by ISAC, ESCCA, ICCS, Wallace H. Coulter Foundation, British Columbia Cancer Agency
  - Ongoing maintenance in perpetuity through ISAC
  - Hosting through Carnegie Mellon University IT Department

Upload, Annotate and Share Your Own Dataset

Typical steps

1. Create a new "experiment"
2. Upload data (FCS files)
3. Prepare annotation templates
   - Or prepare spreadsheets with annotations
4. Annotate the experiment
   - Describe samples, sample sources and instrumentation
   - Provide experimental variables
   - Optionally also extract annotation from spreadsheets
5. Either analyze data online and create illustrations
   - Or upload third party analysis files
6. Review (and improve) your MIFlowCyt compliance
7. Share anonymously with reviewers (or make public)

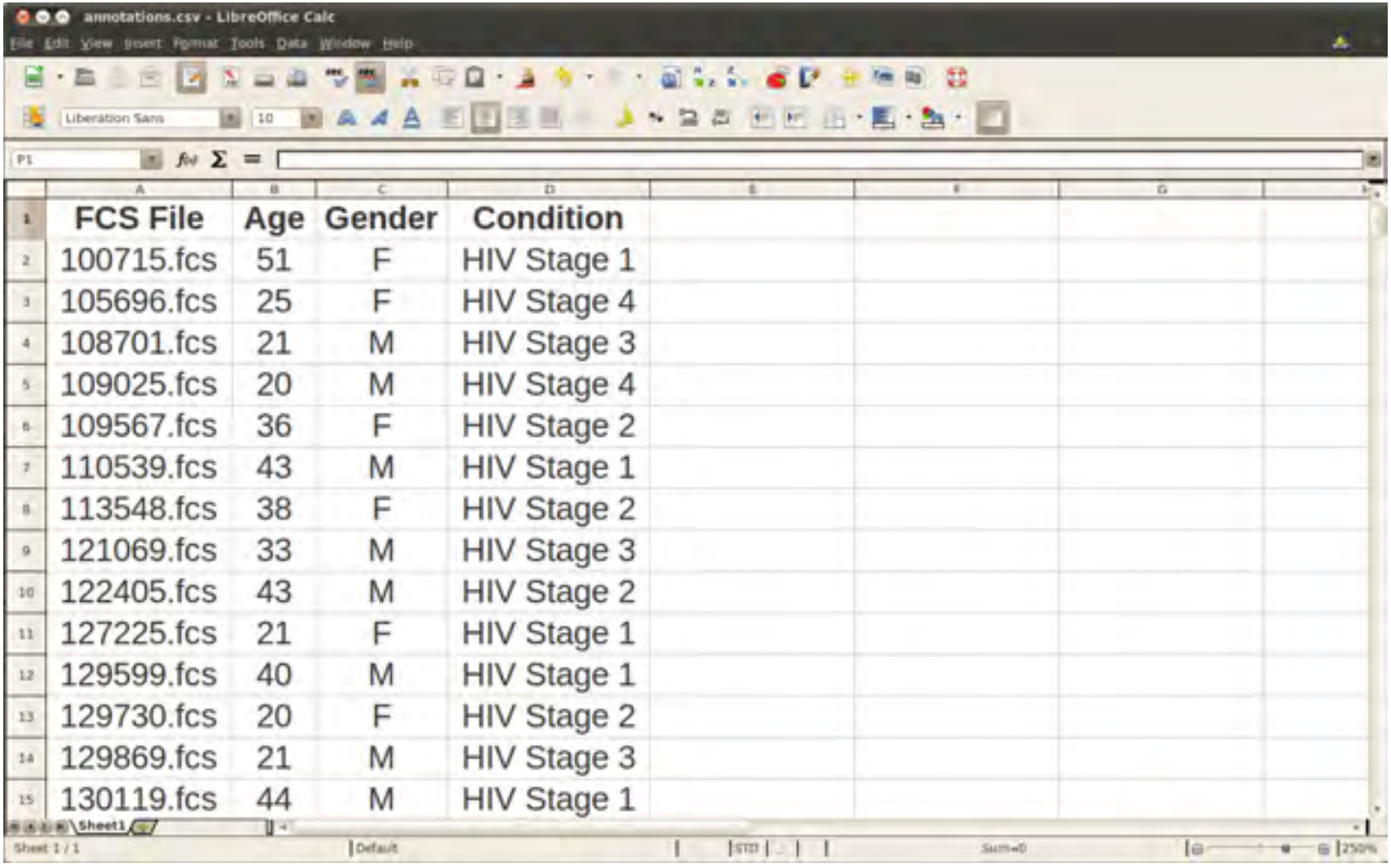

FlowRepository - Identif
https://flowrepository.org/experiments/3/fcs_files/21
FLOWRepository
Inbox
Profile
Annotation Data
Invite a User
Admin
Support
Public View
Welcome, Josef
Logout
Experiment: Identification of B cells through negative gating
ID: FR-FCM-ZZZ3
Labels: None
Primary Researcher: Karin Breuer
Public: Yes
MIFlowCyt Score: 89.83%
« Back to Experiment Summary
Actions
Download Tab-Separated Events File
Show Sample Details
De-identify the FCS file
Review Keywords in the FCS file
Did you know?
You can request a one-on-one session to get started with your data by filling out a support ticket
A guide to Cytobank is available at Current Protocols in Cytometry
We also have a Quick start guide.
You can print/save your illustrations to PDF from the Illustration view's left menu.
You can export your data to Excel from the Experiment Summary page.
Give other users full control to modify your experiments through the "Sharing Permissions" box.
Use the "Download Files" button to save copies of the original FCS Files to your computer.
a2006_O1T2pb05i_A1_A01.fcs - FCS File Information
Version: FCS3.0
Sample Name: a2006_O1T2pb05i
tube_name: A1
plate_name: a2006_O1T2pb05i
well_id: A01
plate_id: 7231843d-e2de-49a5-aa7b-bfe2c8c2817e
Experiment Name: Identification of B cells through negative gating
Operator Name: kjw2
Comment:
Uploaded: 2011-04-18 21:37:50 UTC
Experiment Date: 08-JUN-2007
Acquisition Begin Time: 13:56:43
Acquisition End Time: 13:58:25
Cytometer: LSRII
Instrument:
Data Type: I
Mode: L
Operating System: Windows XP 5.1
Creator Software: BD FACSDiva Software Version 5.0.2
File Size: 9599763
md5sum: f503d32bd899e08fd7dde6cccee46a11
Events: 109046
Text Start: 256
Text End: 3699
Data Start: 3705
Data End: 9599752
Analysis Start: 0
Analysis End: 0
Supplemental Start: 0
Supplemental End: 0
Channel Count: 22
Time Step: 0.01
Threshold: FSC,5000
spill_string: 8,Am Cyan-A,Pacific Blue-A,APC-A,APC-CY7-A,Alexa 700-A,FITC-A,PerCP-CY5-5-A,PE-CY7-A,1.0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1
ASCII Char Delimiter: 12
Information extracted from FCS
Keyword-value pairs
Channel details
Laser details
Compensation details
a2006_O1T2pb05i_A1_A01.fcs - FCS File Instrument Settings
Instrument model: BD LSR II
Manufacturer: Becton Dickinson (BD Biosciences)
Flow Cell Type: Using default settings for BD LSR II
Optical Paths: Using default settings for BD LSR II
Change instrument settings
a2006_O1T2pb05i_A1_A01.fcs - FCS File Laser Information
ASF
Name
Delay
Blue
1.10
0.20
Violet
1.00
21.00

Prepare Annotation Data – Create Sample Source Templates

- Different items required based on the sample source type
- Form changes accordingly
- Use ? for variable fields

Bulk Upload Your Annotations

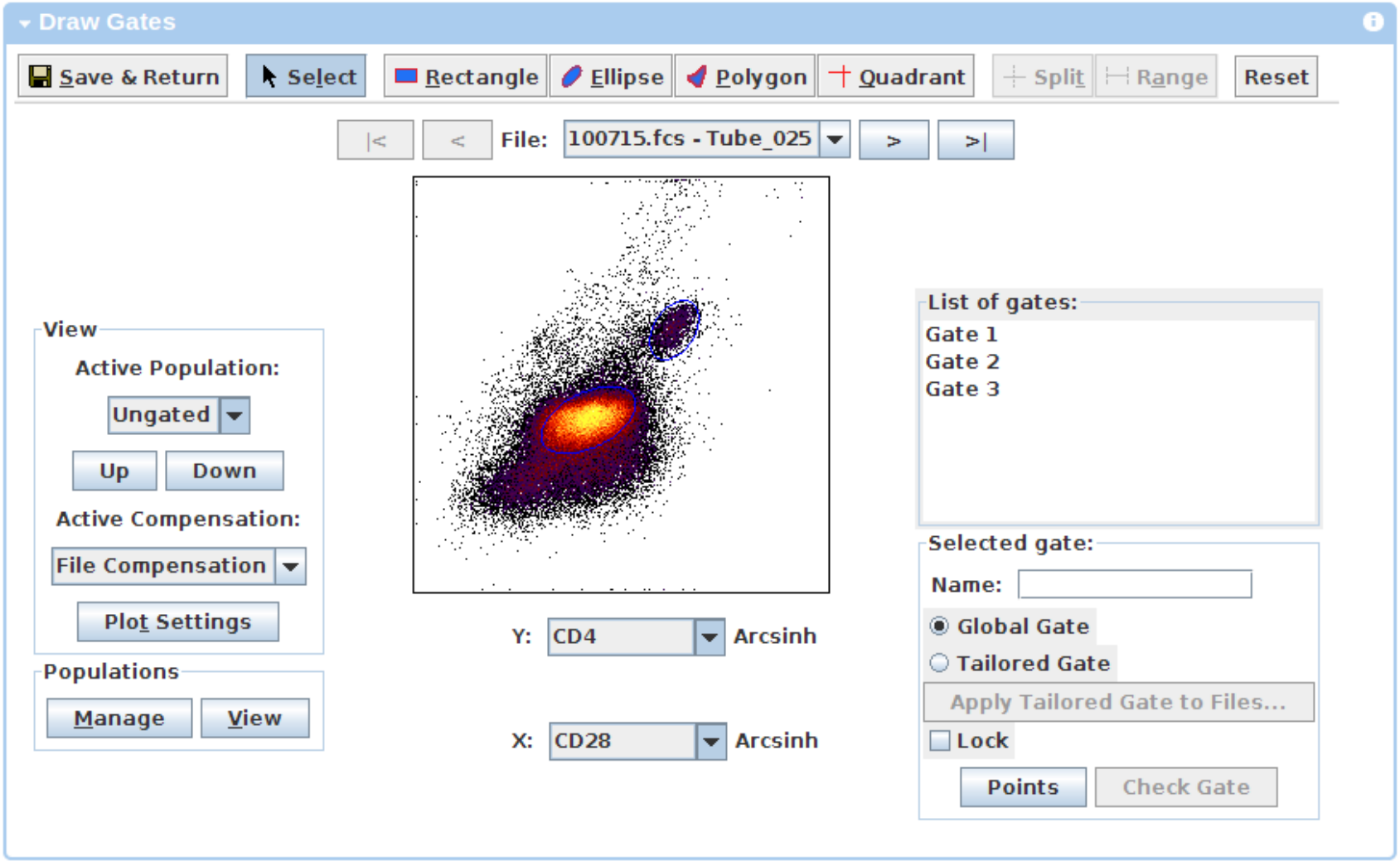
Use your favourite spreadsheet editor



	FCS File	Age	Gender	Condition
2	100715.fcs	51	F	HIV Stage 1
3	105696.fcs	25	F	HIV Stage 4
4	108701.fcs	21	M	HIV Stage 3
5	109025.fcs	20	M	HIV Stage 4
6	109567.fcs	36	F	HIV Stage 2
7	110539.fcs	43	M	HIV Stage 1
8	113548.fcs	38	F	HIV Stage 2
9	121069.fcs	33	M	HIV Stage 3
10	122405.fcs	43	M	HIV Stage 2
11	127225.fcs	21	F	HIV Stage 1
12	129599.fcs	40	M	HIV Stage 1
13	129730.fcs	20	F	HIV Stage 2
14	129869.fcs	21	M	HIV Stage 3
15	130119.fcs	44	M	HIV Stage 1

Analyze Data Online

- Draw your gates

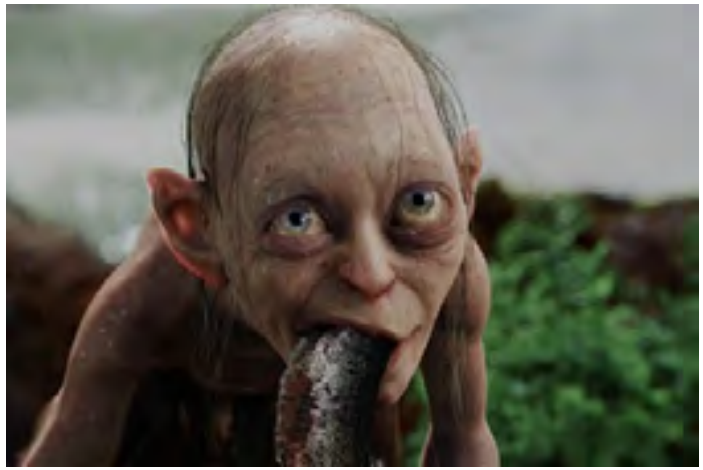


Acknowledgements

	R/BioConductor.org flow cytometry infrastructure
Genentech	Robert Gentleman, all BioConductor contributors
	FlowCAP
Coordinating Committee	Nima Aghaeepour (BCCA), Greg Finak (FHCRC), Raphael Gottardo (FHCRC), Tim Mosmann (U Rochester), Richard H. Scheuermann (UTSW)
Data providers and participants	flowcap.flowsite.org
	HIV/flowType/RchyOptimyx
BCCA	Nima Aghaeepour, Adrin Jalali, Kieran O'Neill, Habil Zare
HIV	Mario Roederer, Pratip K. Chattopadhyay
Lympoma	Fiona Craig
Lyoplates	Federica Villonova, Frank Nestle
	flowDensity
BCCA	Jafar Taghiyar, Mehrnoush Malekeshmaeili, Radina Droumeva
HIP-C	Holden Maecker/Phil McCoy and collaborating HIP-C Consortium centers
IMPC	Hervé Luche
	$
Funding	NIH (NIBIB, NIAID), HIP-C, TFRI & TFF, CCS, MSFHR

What next?

Tutorials: bioinformatics.ca
Web interface: GenePattern.org
Collaboration: rbrinkman@bccrc.ca



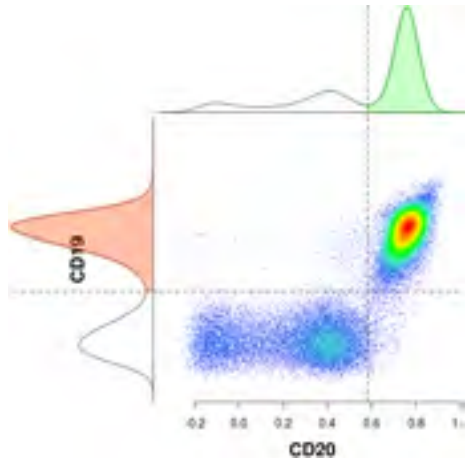
flowDensity Algorithm: 3 or more populations (peaks)

B-cells (CD3-CD19+)

1. Find all the peaks in the distribution

If $\geq$ 3 peaks:

2. Calculate metric as follows:
   - For each peak with index i, calculate the height of the peak, h_i, and its distance from the next adjacent peak, d_i
   - Calculate the metric $\frac{d_i}{h_i}$
3. Find the maximum peak
4. Pick the peak in the previous step and its next adjacent and find the minimum intersection point between the remaining two peaks on the density curve

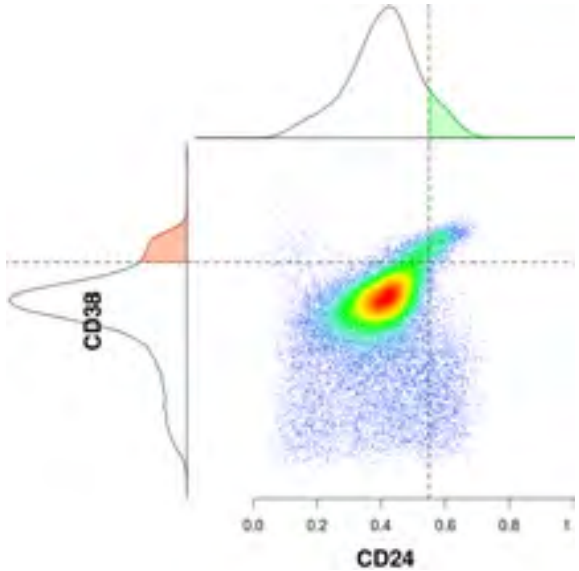


flowDensity Algorithm: 1 population (peaks)

Transitional cells ($CD24^{high}CD38^{high}$) using inflection point

If N peaks = 1:

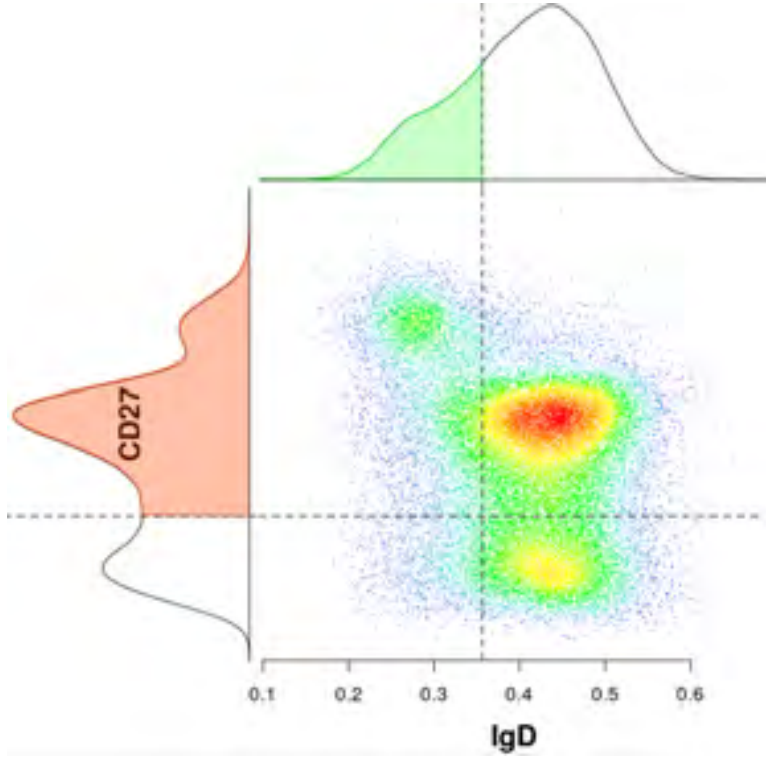
1. Determine the position of the gate by comparing the position of the peak p_x and the median m
2. If $m > p_x$ the gate will be after the peak; else before the peak
3. Try *trackSlope*, to find the inflection points and/or change in the slope of the density curve
4. Try *percentile*, if not set otherwise by user



flowDensity Algorithm: 1 population (peaks)

Memory B-cells (CD27+IgD-) using cut-off

5. If previous above options fail, best threshold is given by the max peak plus/minus one standard deviation



flowDensity Algorithm: 2 populations

If N peaks = 2:

1. Find the minimum intersection point between the two peaks on the density curve