FlowCAP - History

Richard H. Scheuermann, Ph.D.

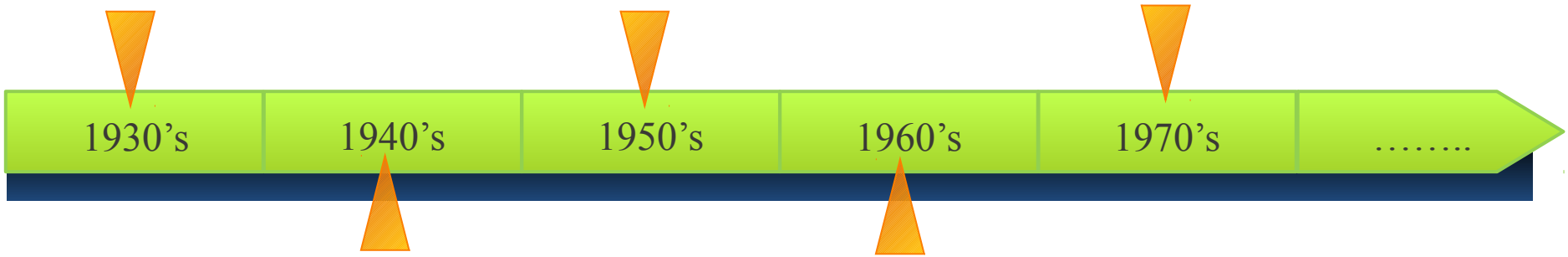
U.T. Southwestern Medical Center

Brief History of Cytometry

Microspectrophotometry to measure DNA & RNA content in cancer cells
Torbjorn Caspersson, Karolinska Inst.

Term *Analytical Cytometry* coined
Francis Schmitt, MIT
Coulter Counter
Wallace Coulter

Differential counter – HemalogD
Ornstein & Kamentsky
Fluorescence-Activated Cell Sorting
Herzenberg & Becton-Dickinson
Society for Analytical Cytology founded



Laminar flow and light scatter measurement
Gucker et al, Northwestern

Hematology counter w/fluorescence
Hallermann et al., Leitz
Interface w/minicomputers
George Wied, U. Chicago
Gunter Bahr, AFIP
Peter Bartels, U. Arizona

Selected from "The Evolution of Cytometers"
Shapiro, HM (2004) Cytometry Part A 58A:13-20.

FCM instrumentation & reagents

ELSEVIER

Journal of Immunological Methods 243 (2000) 77–97

JIM Journal of Immunological Methods

www.elsevier.nl/locate/jim

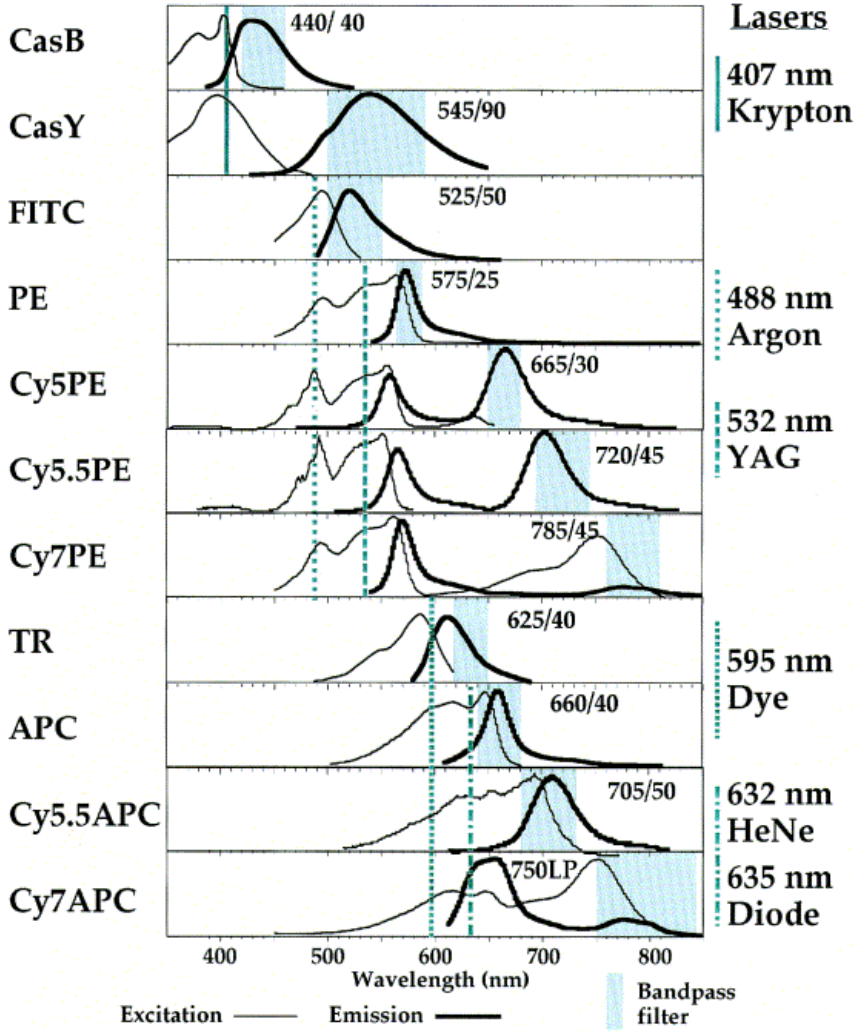
A practical approach to multicolor flow cytometry for immunophenotyping

Nicole Baumgarth[a,1], Mario Roederer[b,*]

[a]Department of Genetics, Stanford University Medical School, Beckman Center B-007, Stanford, CA 94305-5318, USA

[b]Department of Stomatology, UCSF, San Francisco, CA, USA

FCM can measure many parameters simultaneously, e.g., BD LSR-II can produce data for up to 19 parameters for every cell in a given sample



Flow Cytometry (FCM)

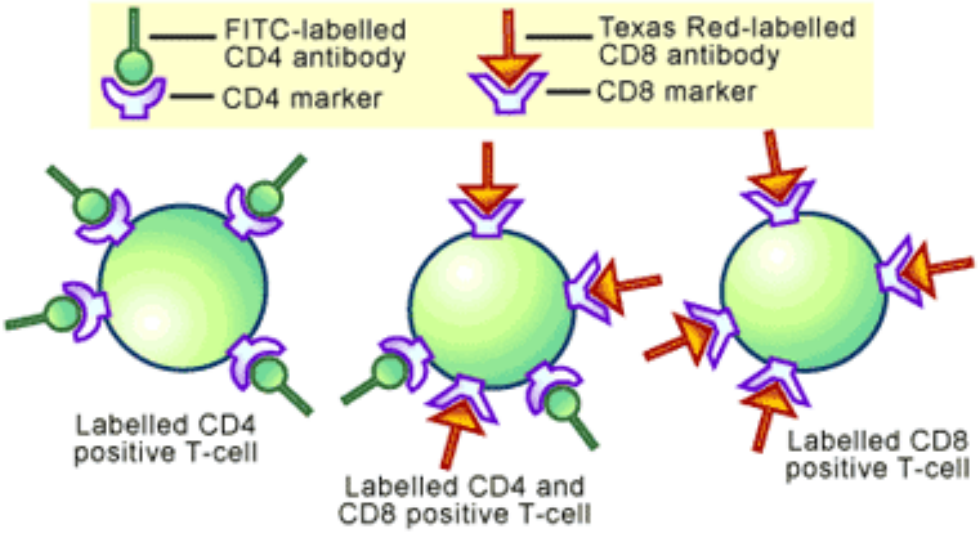
a.k.a. Fluorescence Activated Cell Sorting (FACSTM)

Method:

Stain cell population with fluorescent reagents that bind to specific molecules, e.g. fluorescein-conjugated anti-CD40 antibodies

Measure fluorescence properties of each cell using flow cytometer

Direct and indirect measurement of individual cell characteristics, e.g. cell size, membrane protein expression, secreted protein expression, cell cycle state, DNA pl



Uses of Flow Cytometry (FCM)

Differences in cell populations between specimens

Study of normal cell activation, differentiation and function

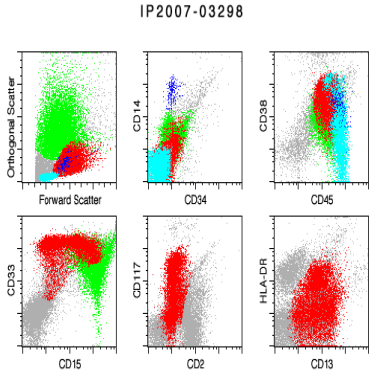
Study of abnormal cell activation, differentiation and function

Isolate cells from mixture based on their molecular characteristics

Diagnostics - leukemia, lymphoma, myeloproliferative disorders

Novel biomarkers

normal leukemia

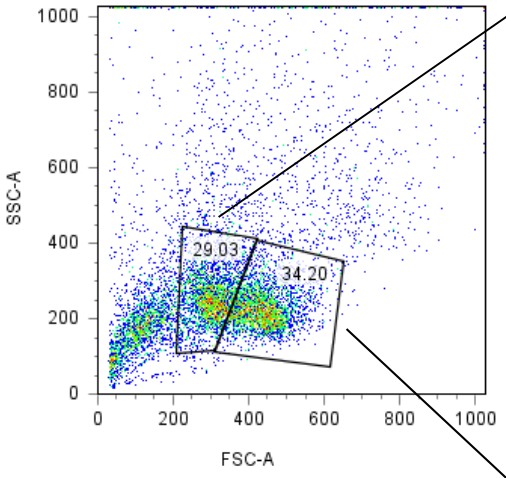


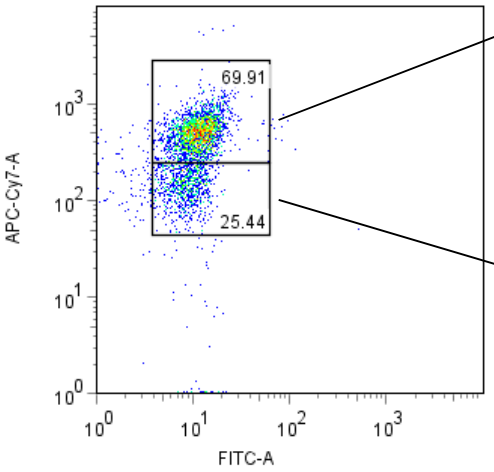
Red - Myeloblasts
Green - Granulocytes
L. Blue - Monocytes

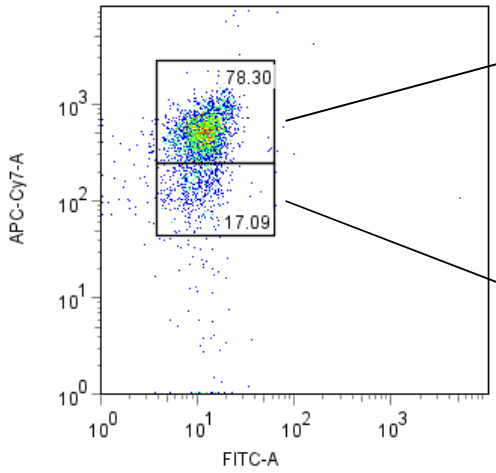
Traditional Flow Cytometry Analysis

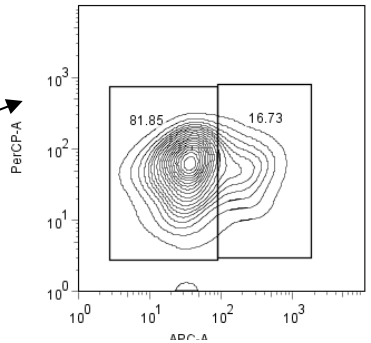
Goal - group together cells with similar characteristics

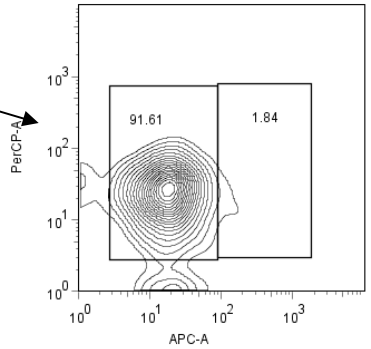
Traditional approach - manual gating 2D at a time

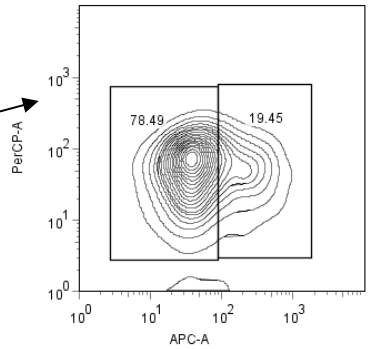


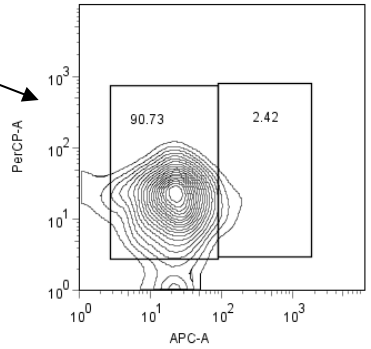




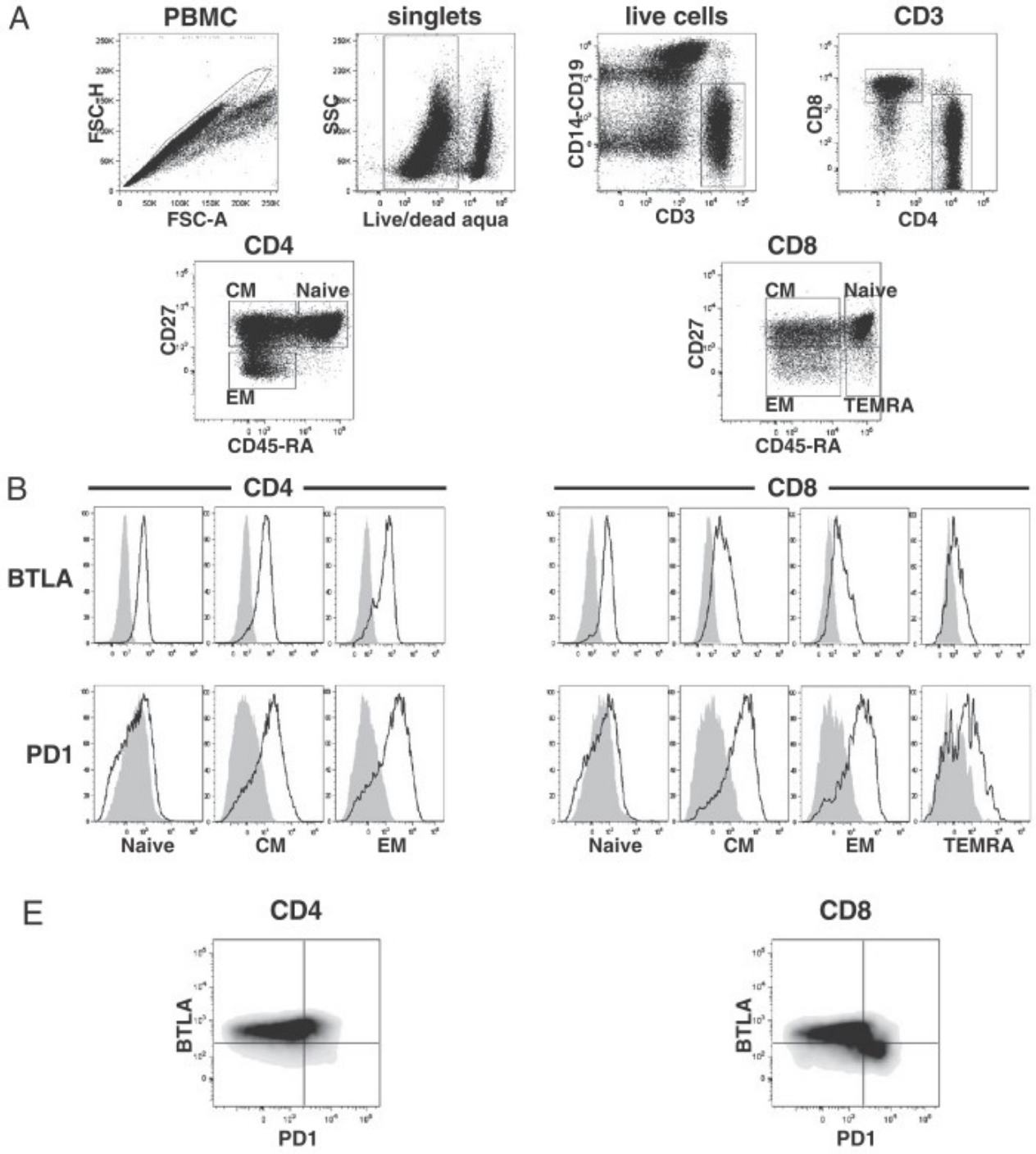








- Subjective
- Time-consuming
- Doesn't handle overlapping distributions well
- Sensitive to slight difference in fluorescence intensity distributions between samples
- Requires at least one 2D plot that clearly segregates populations in question


A
PBMC
FSC-H
FSC-A
singlets
SSC
Live/dead aqua
live cells
CD14-CD19
CD3
CD3
CD8
CD4
CD4
CM
Naive
EM
CD27
CD45-RA
CD8
CM
Naive
EM
TEMRA
CD27
CD45-RA
B
CD4
CD8
BTLA
PD1
Naive
CM
EM
Naive
CM
EM
TEMRA
E
CD4
BTLA
PD1
CD8
BTLA
PD1

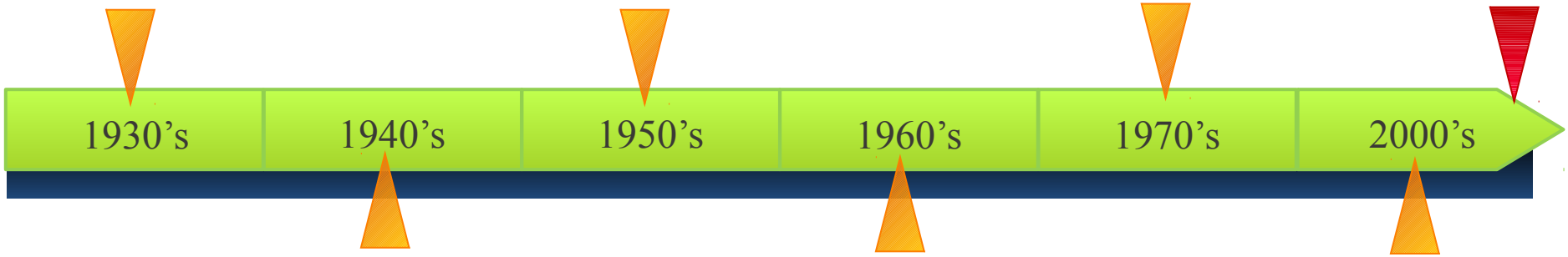
Brief History of Cytometry

Microspectrophotometry to measure DNA & RNA content in cancer cells
Torbjorn Caspersson, Karolinska Inst.

Term *Analytical Cytometry* coined
Francis Schmitt, MIT
Coulter Counter
Wallace Coulter

Differential counter – HemalogD
Ornstein & Kamentsky
Fluorescence-Activated Cell Sorting
Herzenberg & Becton-Dickinson
Society for Analytical Cytology founded

FlowCAP-I



Laminar flow and light scatter measurement
Gucker et al, Northwestern

Hematology counter w/fluorescence
Hallermann et al., Leitz
Interface w/minicomputers
George Wied, U. Chicago
Gunter Bahr, AFIP
Peter Bartels, U. Arizona

K means clustering applied to FCM data

Selected from "The Evolution of Cytometers"
Shapiro, HM (2004) Cytometry Part A 58A:13-20.

Improved Approaches

Identifying cell populations automatically, objectively, and quickly in multi-dimensional flow cytometry data (eliminate manual gating)

Quantitatively compare the identified populations across different samples and across different experiments

Characteristics of FCM Data

Data sets are:

Large (and various) size

From hundreds to millions of events

Multidimensional

19 parameter instrument already available

Noise and Outlier

Dead cells and dirt

Populations are different in:

shapes

Elongated, ellipsoid, spherical, banana shapes…

densities

Some cell populations are relatively sparse even on 2D space

compositions

Events that pile up on axis can change data distribution

positions

Some are very close while others are far away

sizes

From several events to hundreds of thousands events

Hindawi Publishing Corporation
Advances in Bioinformatics
Volume 2009, Article ID 247646, 12 pages
doi:10.1155/2009/247646

Research Article

Merging Mixture Components for Cell Population Identification in Flow Cytometry

Greg Finak,[1] Ali Bashashati,[2] Ryan Brinkman,[2] and Raphaël Gottardo[1,3]

[1] Computational Biology Unit, Clinical Research Institute of Montreal, 110 Pine Avenue West, Montreal, QC, Ca
[2] Terry Fox Laboratory, BC Cancer Research Center, Vancouver, BC, Canada V5Z 1L3
[3] Département de Biochimie, Université de Montreal, Montreal, QC, Canada

Cytometry Part B (Clinical Cytometry) 78B (Suppl. 1):S69–S82 (2010)

Elucidation of Seventeen Human Peripheral Blood B-Cell Subsets and Quantification of the Tetanus Response Using a Density-Based Method for the Automated Identification of Cell Populations in Multidimensional Flow Cytometry Data

Yu Qian,[1,2] Chungwen Wei,[3] F. Eun-Hyung Lee,[3] John Campbell,[4] Jessica Halliley,[3] Jamie A. Lee,[1] Jennifer Cai,[1] Y. Megan Kong,[1] Eva Sadat,[1] Elizabeth Thomson,[1] Patrick Dunn,[4] Adam C. Seegmiller,[1] Nitin J. Karandikar,[1] Christopher M. Tipton,[3] Tim Mosmann,[3] Iñaki Sanz,[3] and Richard H. Scheuermann[1,2]*

[1]Department of Pathology, University of Texas Southwestern Medical Center, Dallas, Texas 75390
[2]Division of Biomedical Informatics, University of Texas Southwestern Medical Center, Dallas, Texas 75390
[3]Department of Medicine, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642
[4]Health Information Systems, Northrop Grumman, Rockville, Maryland 20850

Cytometry

Statistical Mixture Modeling for Cell Subtype Identification in Flow Cytometry

Cliburn Chan,[1]* Feng Feng,[1] Janet Ottinger,[2] David Foster,[3] Mike West,[4] Thomas B. Kepler[1,4,5]

Cytometry Part A • 73A: 693–701, 2008

BMC Bioinformatics

BioMed Central

flowClust: a Bioconductor package for automated gating of flow cytometry data

BMC Bioinformatics 2009, **10**:145 doi:10.1186/1471-2105-10-145

Kenneth Lo (c.lo@stat.ubc.ca)
Florian Hahne (fhahne@fhcrc.org)
Ryan R Brinkman (rbrinkman@bccrc.ca)
Raphael Gottardo (raphael.gottardo@ircm.qc.ca)

Zare *et al. BMC Bioinformatics* 2010, **11**:403
http://www.biomedcentral.com/1471-2105/11/403

BMC Bioinformatics

METHODOLOGY ARTICLE Open Access

Data reduction for spectral clustering to analyze high throughput flow cytometry data

Habil Zare[1,2], Parisa Shooshtari[1,2], Arvind Gupta[3], Ryan R Brinkman[2,4]*

Automated high-dimensional flow cytometric data analysis

Saumyadipta Pyne[a], Xinli Hu[a,1], Kui Wang[b,1], Elizabeth Rossin[a,1], Tsung-I Lin[c], Lisa M. Maier[a,d], Clare Baecher-Allan[d], Geoffrey J. McLachlan[b,e], Pablo Tamayo[a], David A. Hafler[a,d,2], Philip L. De Jager[a,d,f,2], and Jill P. Mesirov[a,2,3]

[a]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge MA 02142; [b]Department of Mathematics and [e]Institute for Molecular Bioscience, University of Queensland, St. Lucia, Queensland, 4072, Australia; [c]Department of Applied Mathematics, National Chung Hsing University, Taichung 402, Taiwan; [d]Division of Molecular Immunology, Center for Neurologic Diseases, Brigham and Women's Hospital and Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115; and [f]Partners Center for Personalized Genetic Medicine, Boston, MA 02115

www.pnas.org/cgi/doi/10.1073/pnas.0903028106 PNAS | May 26, 2009 | vol. 106 | no. 21 | 8519–8524

Cytometry

Automated Gating of Flow Cytometry Data via Robust Model-Based Clustering

Kenneth Lo,[1]* Ryan Remy Brinkman,[2] Raphael Gottardo[1]

Cytometry Part A • 73A: 321–332, 2008

Cytometry

A Statistical Pattern Recognition Approach for Determining Cellular Viability and Lineage Phenotype in Cultured Cells and Murine Bone Marrow

John Quinn,[1,2] Paul W. Fisher,[3] Renold J. Capocasale,[3] Ram Achuthanandam,[3] Moshe Kam,[4] Peter J. Bugelski,[1,3] Leonid Hrebien[4]*

Cytometry Part A • 71A: 612–624, 2007

Naumann *et al. BMC Bioinformatics* 2010, **11**:44
http://www.biomedcentral.com/1471-2105/11/44

BMC Bioinformatics

METHODOLOGY ARTICLE Open Access

The curvHDR method for gating flow cytometry samples

Ulrike Naumann[1], George Luta[2], Matthew P Wand[3]*

December 2008

FlowCAP

Flow Cytometry: Critical Assessment of Population Identification Methods (FlowCAP)

The goal of FlowCAP is ***to advance the development of computational methods*** for the identification of cell populations of interest in flow cytometry data. FlowCAP will ***provide the means to objectively test these methods***, first by comparison to manual analysis by experts using common datasets, and second by comparison to synthetic data sets having known properties.

FlowCAP will consist of three parts:

1) The collection of de-identified data sets for prediction from the experimental community that will be shared among the algorithm development community as a common reference for analysis;

2) The collection of population subset predictions (gates) from the computational biology community derived from these common reference data sets using existing and novel algorithmic approaches; and

3) The assessment and discussion of the results in comparison with the manual gating gold standard.

FlowCAP-I Time Line

Release of materials for challenge 1 and 2: 01 MAR 2010

Submission deadline for challenge 1 and 2: 30 JUN 2010

Release of materials for challenge 3: 30 JUN 2010

Submission deadline for challenge 3: 21 JUL 2010

Release of materials for challenge 4: 21 JUL 2010

Submission deadline for challenge 4: 15 AUG 2010

Public release of the results: 15 SEP 2010

FlowCAP summit: 21-22 SEP 2010

Datasets

Diffuse Large B-cell Lymphoma (DLBCL) – lymph node biopsies from patients treated at the British Columbia Cancer Agency between 2003 and 2008. These patients were histologically confirmed to have diffuse large B-cell lymphoma (DLBCL). This dataset is provided by BCCRC.

Symptomatic West Nile Virus (WNV) – peripheral blood mononuclear cells from patients with symptomatic West Nile virus infection stimulated in-vitro with peptide pools of the WNV polyprotein. This dataset is provided by the

Dataset	#Samples	#Events	#Colors	Analyte-Reporter	Provider
GvHD	12	14,000	4	CD4-FITC CD8b-PE CD3PerCP CD8-APC	BCCRC & TreeStar
DLBCL	30	5,000	3	CD3-Cy5 CD5-FITC CD19-PE	BCCRC
HSCT	30	10,000	4	CD45.1-FITC Ly65/Mac1-PE Dead cells-PI CD45.2-APC	BCCRC
WNV	13	100,000	6	IFNg-PEA CD3-PECy5 CD4-PECy7 IL17-APC CD8-AF700 Free amine-CFSE	McMaster
ND	30	17,000	10	CD56-Q605 CD8-AF700 CD45-PECy5 CD3/CD14-PECy7 Proprietary-FITC, PerCPCy5, PacificBlue, PacificOrange, APC, PE	Amgen

Four Competitions

Challenge 1: Automated Algorithms

Compare results from automated gating algorithms for exploratory analysis on a wide range of FCM samples against the manual gating benchmark. Software used in this challenge should not have any free parameters (if you have a free parameters it must be set to a single value for all of the datasets). For this challenge, participants will use software that, given only a FCS file and no other information, produces a population membership label (or set of labels with likelihoods) for each event.

Challenge 2: Tuned Algorithms

Compare results from automated gating algorithms for exploratory analysis on a wide range of FCM samples against the manual gating benchmark. Software used in this challenge may have free parameters that can be manually adjusted before running (i.e., you can submit an algorithm with some free parameters for each dataset).

Challenge 3: Assignment of Cells to Populations with Pre-defined Number of Populations

Compare the ability of the algorithms to assign correct labels to cells when the number of expected populations is known, against the manual gating benchmark.

Challenge 4: Supervised Clustering Trained using Manual Gates

In this challenge a few files with manual gates (i.e., membership labels) will be provided to the participants for tuning their algorithms for each dataset. The tuned software can then be run on the remaining data files; the results will be compared against the manual gating benchmark.

Let N be the number of data points, C the set of classes, K the set of clusters and n_{ij} be the number of members of class $c_i \in C$ that are elements of cluster $k_j \in K$.

$$F(C,K) = \sum_{c_i \in C} \frac{|c_i|}{N} \max_{k_j \in K} \{F(c_i,k_j)\} \quad (3)$$

$$F(c_i,k_j) = \frac{2 * R(c_i,k_j) * P(c_i,k_j)}{R(c_i,k_j) + P(c_i,k_j)}$$

$$R(c_i,k_j) = \frac{n_{ij}}{|c_i|}$$

$$P(c_i,k_j) = \frac{n_{ij}}{|k_j|}$$

Figure 1: Calculation of clustering F-measure

V-Measure: A conditional entropy-based external cluster evaluation measure

Andrew Rosenberg and Julia Hirschberg
Department of Computer Science
Columbia University
New York, NY 10027
{amaxwell,julia}@cs.columbia.edu

Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 410–420, Prague, June 2007.

FlowCAP Summit 2010 Agenda

Day 1

Competition participant presentations

Day 2

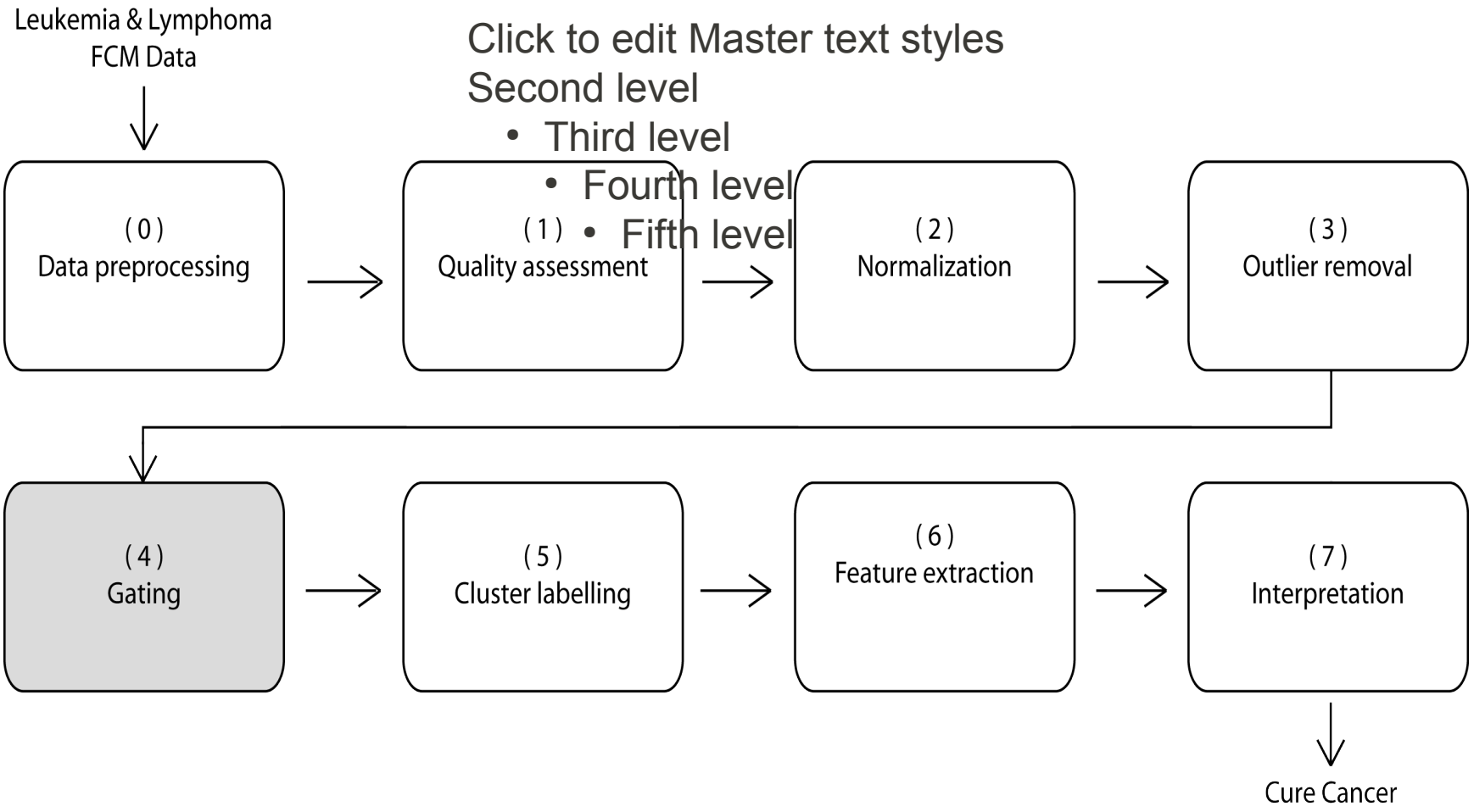
Keynote Presentation – Mario Roederer

FlowCAP-I results

FlowCAP-I debrief

FlowCAP-II planning

FCM Analysis Workflow



Acknowledgments

National Institute of Allergy and Infectious Diseases

FlowCAP Organizing Committee

Nima Aghaeepour for competition results analysis

FlowCAP-I participants

Mark Smith

FlowCAP Summit 2010 attendees