Basic data analysis and presentation

Tomáš Kalina

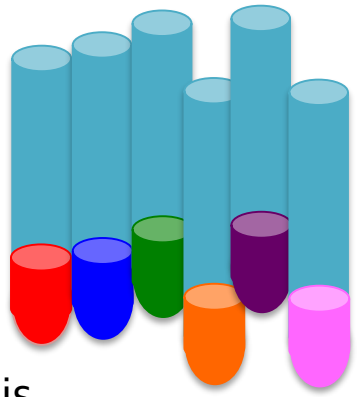
Charles University, 2nd Faculty of Medicine,
Prague, Czech Republic
Dpt. of Pediatric Hematology and Oncology

- Childhood Leukemia Investigation Prague

Competing features of cytometry

Many samples

- **Standardization**
- Ease of use
- Speed
- Automated analysis



Clinical 10 colors

Many parameters

Research 30+ colors
Spectral 20+ colors
Mass cytometry 40+ parameters

- **Computational analysis**
- Ease of use

Many cells, rare cells

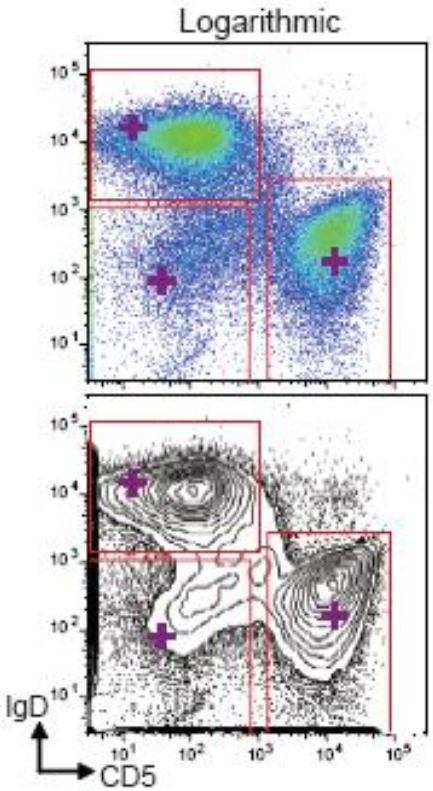
- **Speed**
- Automated analysis

0,650 Gb FCS file, 2mil events, 40 pars

Outline

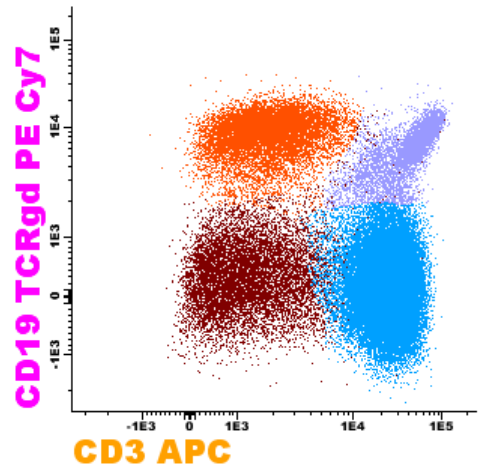
- Graphs and visualisation
- Gating controls
- FCS datafiles – keywords
- Complex data visualisation (data reduction: PCA, tSNE)

Data display - graphs



Pseudocolor plot

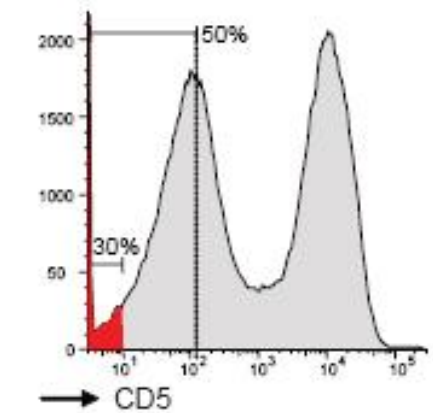
Contour plot



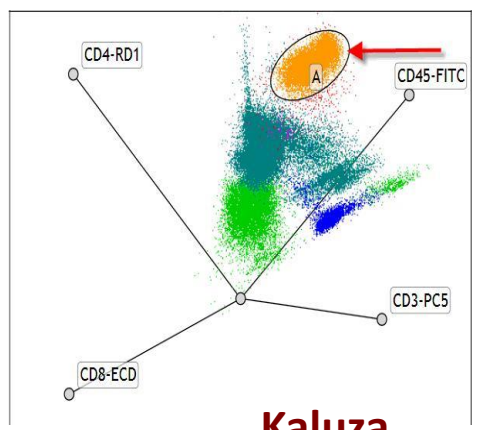
Dot plot

Infinicyt software

Histogram

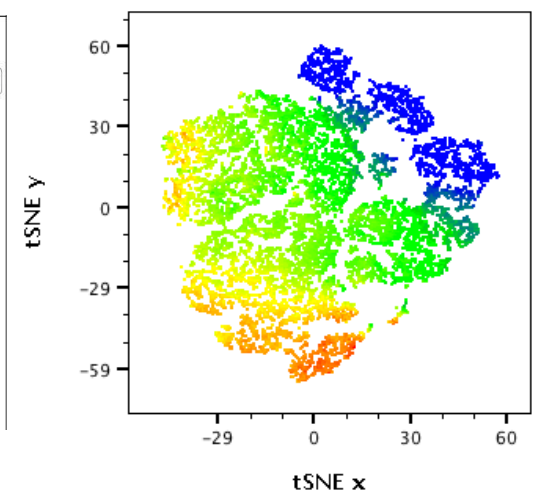


Radar Plot



Kaluza

t-SNE Plot with color map

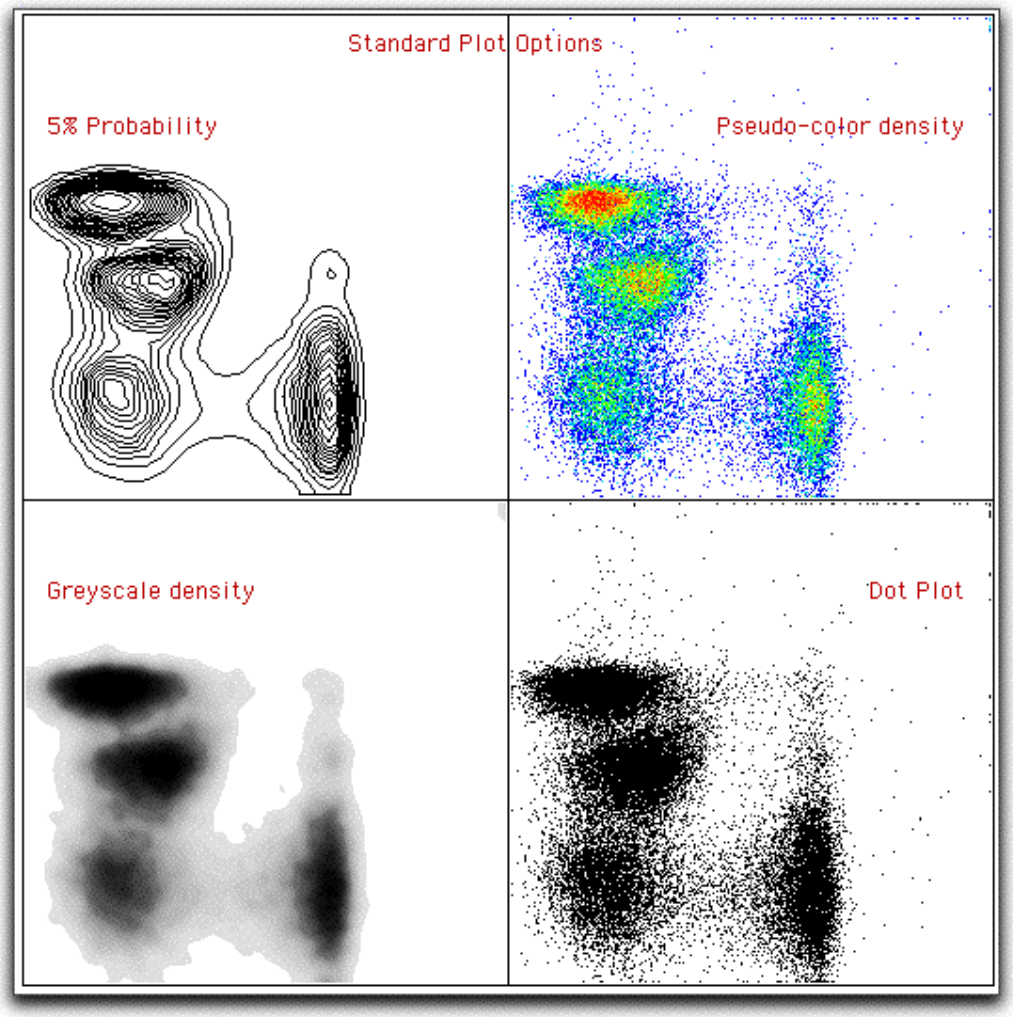


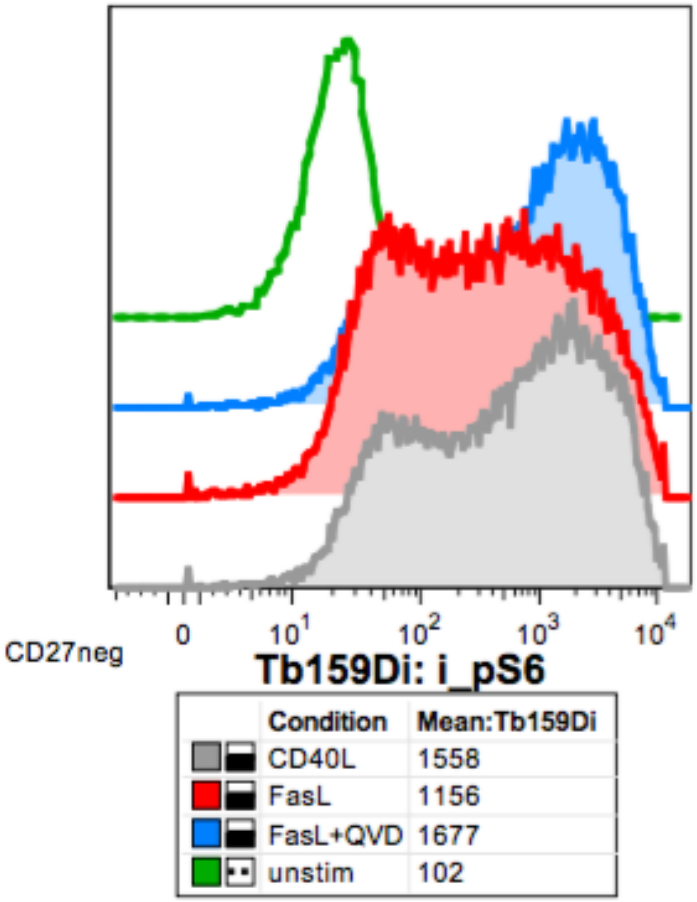
Color Map by CD45

Flowjo

Flowjo software (Mac)

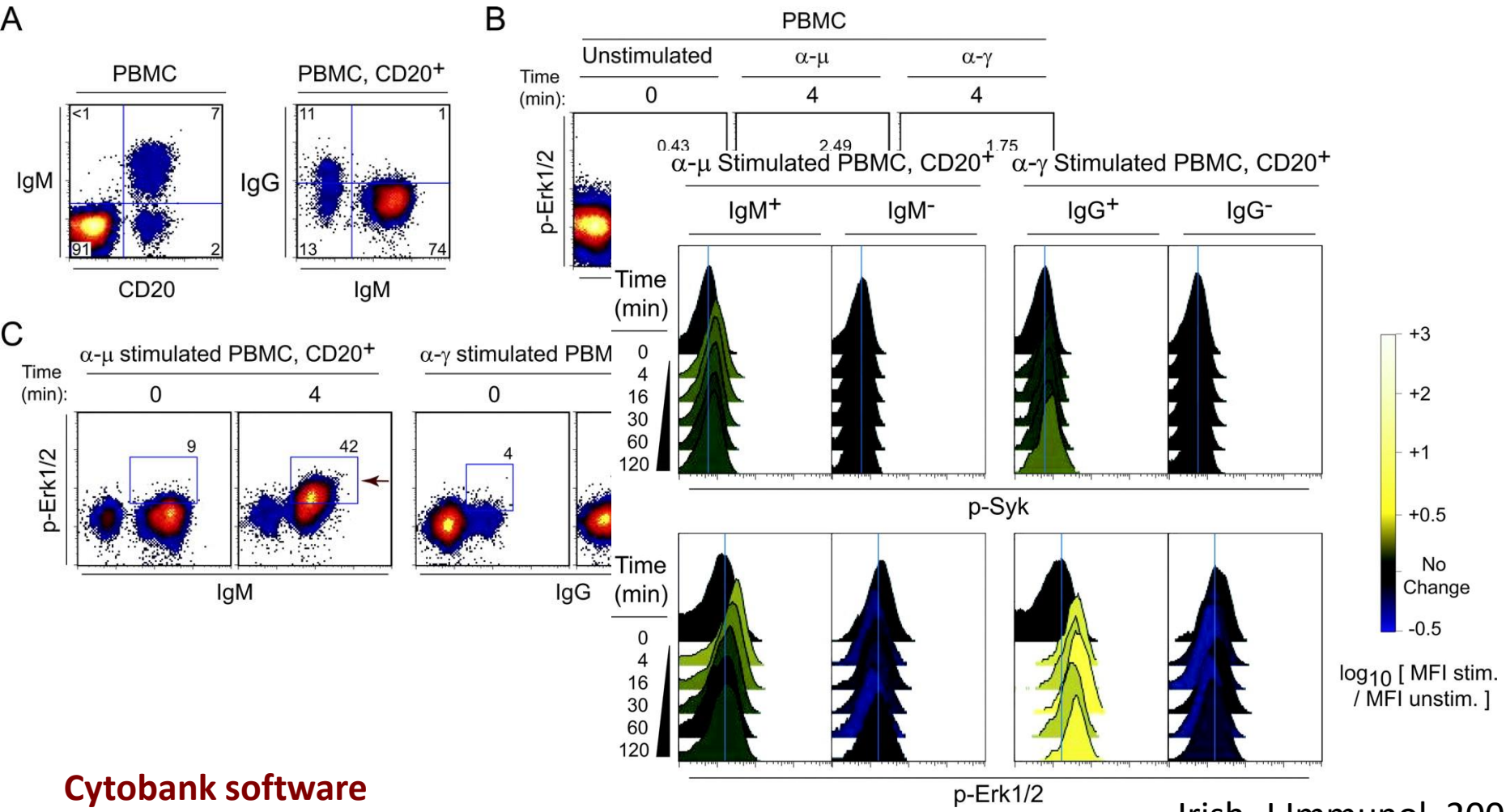
Basic graphs





FlowJo (Mac)

Overview gating and histogram comparison



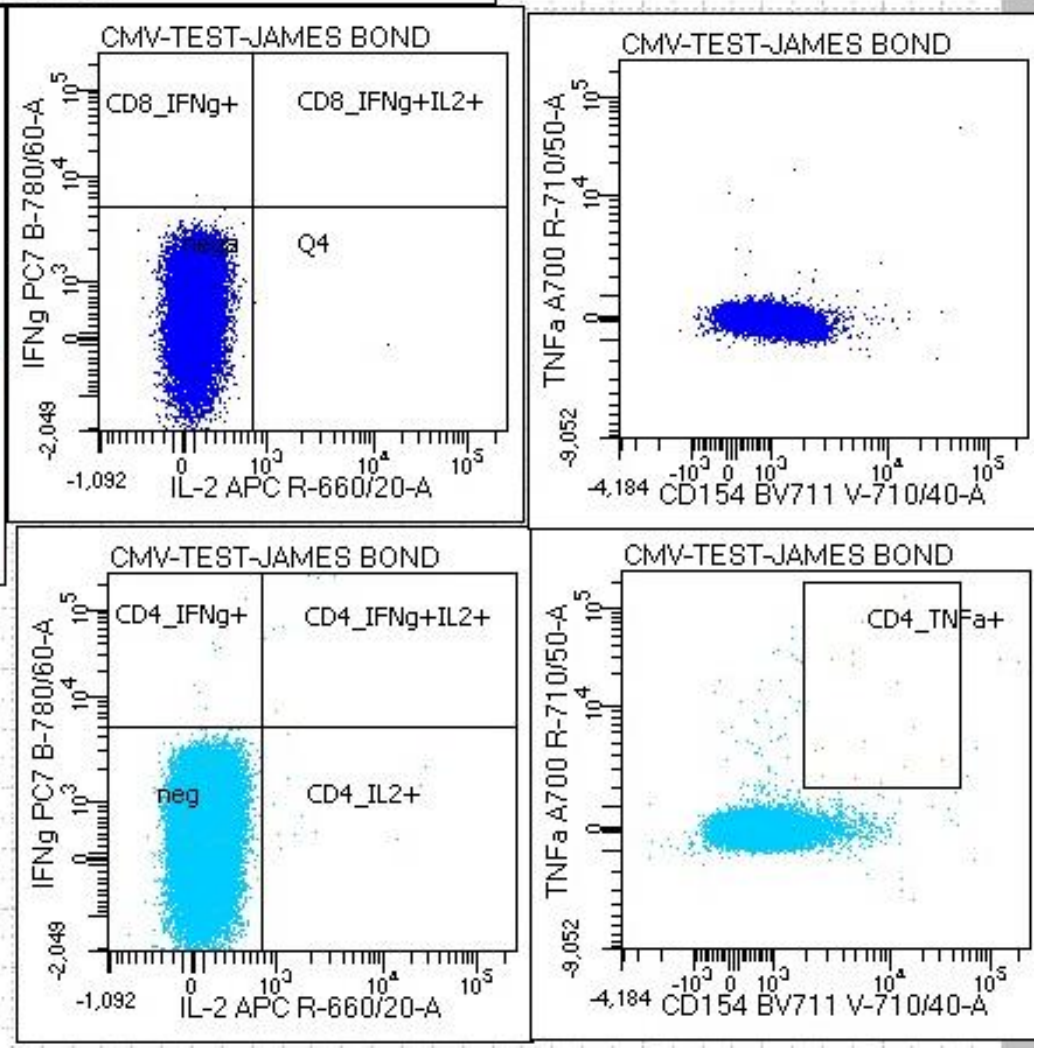
Cytobank software

Irish, J Immunol. 2006

Gating controls

Specimen Name: CMV-TEST
Tube Name: JAMES BOND
Record Date: Jan 17, 2018 4:1...
SAMPLE ID: CLIP44764cyto
PATIENT ID: Patient_007
STIMULACE: UNSTIM

Population	#Events	%Parent
CD8+	16,464	18.673
CD8_IFNg+	1	0.006
CD8_IFNg+IL2+	0	0.000
CD4+	69,683	79.033
CD4_IFNg+	7	0.010
CD4_TNFa+	18	0.026

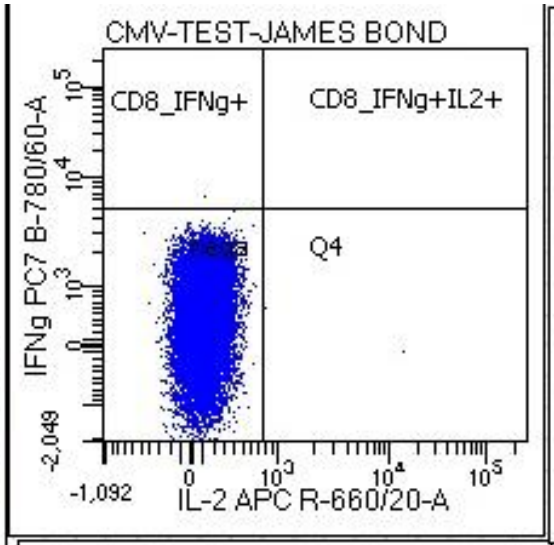


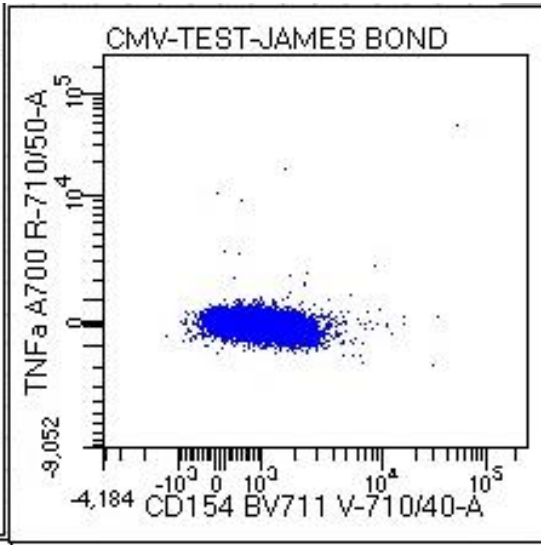
Diva software

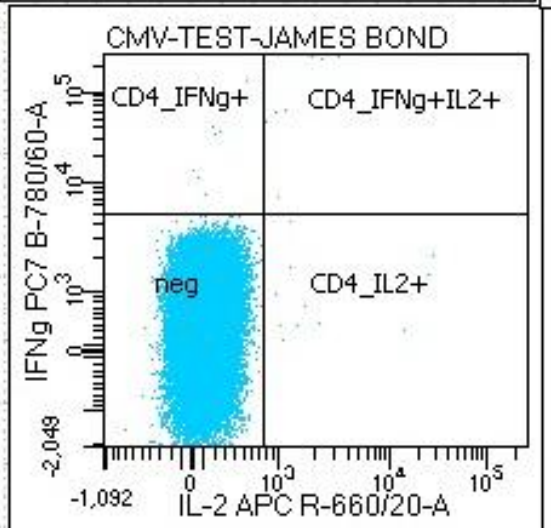
Gating controls

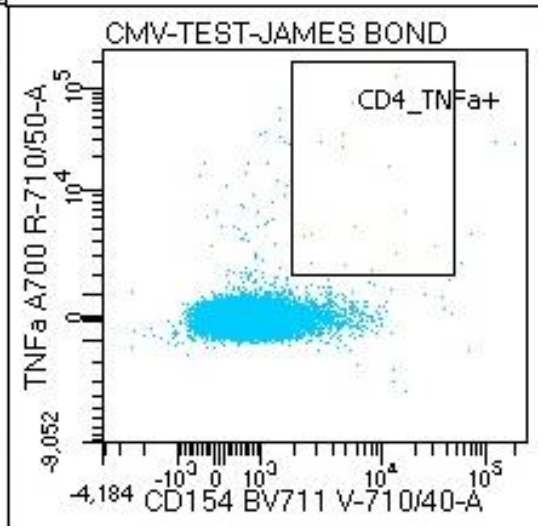
Specimen Name:	CMV-TEST
Tube Name:	JAMES BOND
Record Date:	Jan 17, 2018 4:1...
SAMPLE ID:	CLIP44764cyto
PATIENT ID:	Patient_007
dpTx:	75
STIMULACE:	UNSTIM

Population	#Events	%Parent
CD8+	16,464	18.673
CD8_IFNg+	1	0.006
CD8_IFNg+IL2+	0	0.000
CD4+	69,683	79.033
CD4_IFNg+	7	0.010
CD4_TNFa+	18	0.026





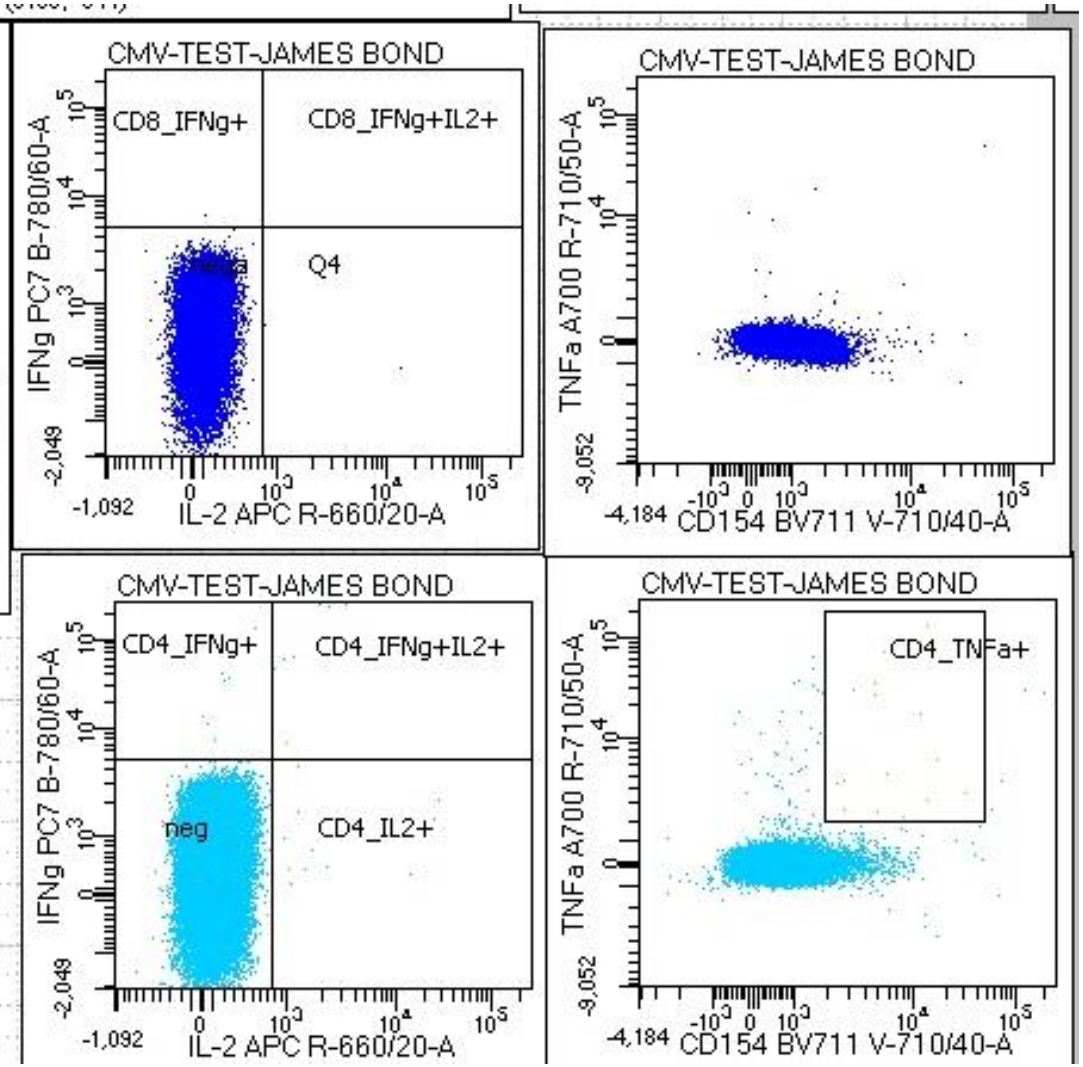




Gating controls

Specimen Name:	CMV-TEST
Tube Name:	JAMES BOND
Record Date:	Jan 17, 2018 4:1...
SAMPLE ID:	CLIP44764cyto
PATIENT ID:	Patient_007
STIMULACE:	UNSTIM

Population	#Events	%Parent
CD8+	16,464	18.673
CD8_IFNg+	1	0.006
CD8_IFNg+IL2+	0	0.000
CD4+	69,683	79.033
CD4_IFNg+	7	0.010
CD4_TNFa+	18	0.026



Basic concepts

- FCS datafiles
  - What are they, how to use Keywords

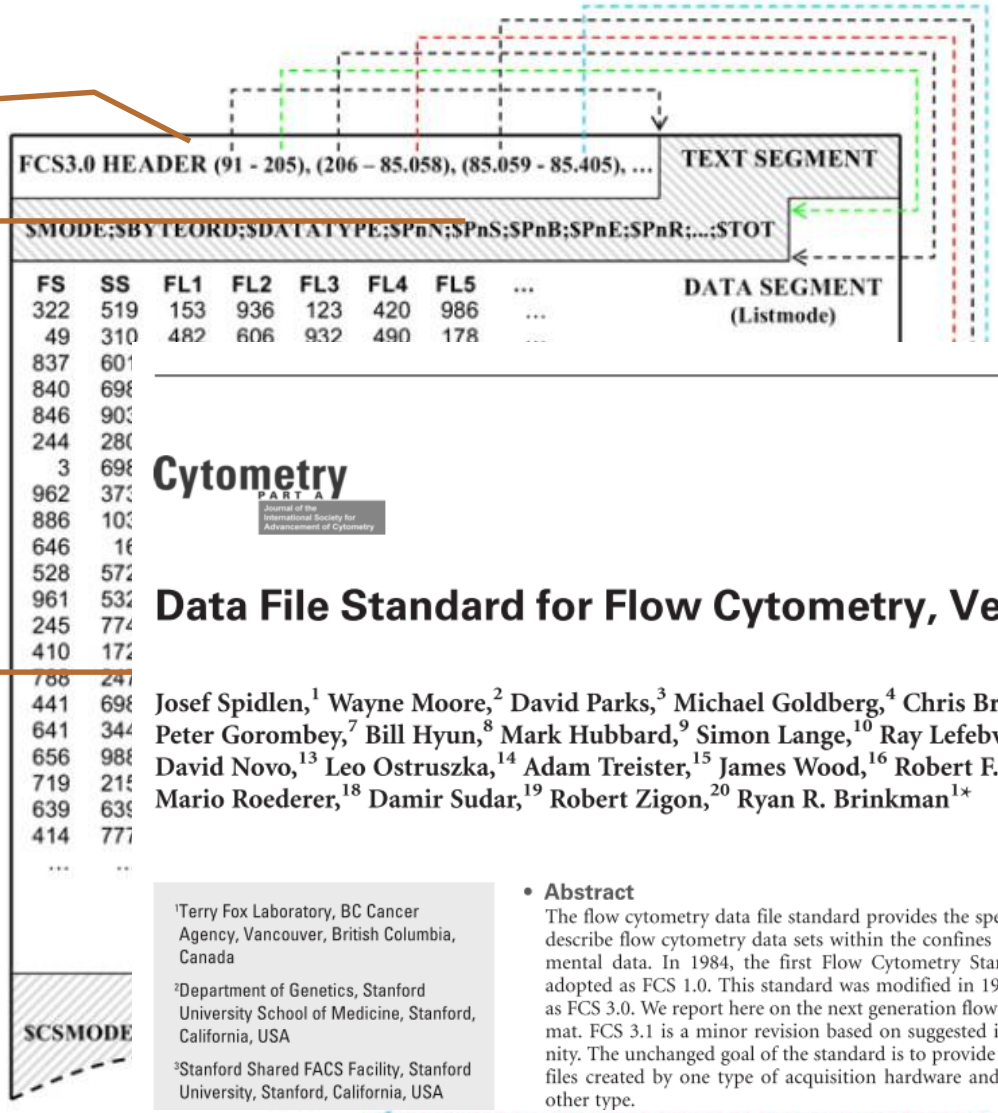
Keywords are essential for proper data annotation:
Data cohorts

What is a FSC datafile ?

HEADER (FSC 3.0) etc
(text)

KEYWORDS
($DATE, $TOT,
$SAMPLEID.....
$P1S, $P1N, $P1V,)

DATA
row = cell
Column = parameter
(compressed)



ORIGINAL ARTICLE

Cytometry PART A
Journal of the International Society for Advancement of Cytometry

Data File Standard for Flow Cytometry, Version FCS 3.1

Josef Spidlen,[1] Wayne Moore,[2] David Parks,[3] Michael Goldberg,[4] Chris Bray,[5] Pierre Bierre,[6] Peter Gorombey,[7] Bill Hyun,[8] Mark Hubbard,[9] Simon Lange,[10] Ray Lefebvre,[11] Robert Leif,[12] David Novo,[13] Leo Ostruszka,[14] Adam Treister,[15] James Wood,[16] Robert F. Murphy,[17] Mario Roederer,[18] Damir Sudar,[19] Robert Zigon,[20] Ryan R. Brinkman[1*]

[1]Terry Fox Laboratory, BC Cancer Agency, Vancouver, British Columbia, Canada

[2]Department of Genetics, Stanford University School of Medicine, Stanford, California, USA

[3]Stanford Shared FACS Facility, Stanford University, Stanford, California, USA

Abstract

The flow cytometry data file standard provides the specifications needed to completely describe flow cytometry data sets within the confines of the file containing the experimental data. In 1984, the first Flow Cytometry Standard format for data files was adopted as FCS 1.0. This standard was modified in 1990 as FCS 2.0 and again in 1997 as FCS 3.0. We report here on the next generation flow cytometry standard data file format. FCS 3.1 is a minor revision based on suggested improvements from the community. The unchanged goal of the standard is to provide a uniform file format that allows files created by one type of acquisition hardware and software to be analyzed by any other type.

KEYWORDS in FCS

Parameters in FCS (“labels”):

$P1N	… par name….	FL1	FITC	**B-530-30**
$P1S	… par stain ….	CD3 FITC	CD3	**CD3 FITC**
$P1V	. par voltage ..	550	550	**550**

Labels (Diva)

BD FACSDiva Software - Administrator (BV-6c wo Yellow561-2012)

File Edit View Experiment Populations Worksheet Cytometer HTS Help

Browser - DEPLECE_TCRab_CD19

Name	Date
6-KOMBI-NEW	1/13/16 4:13:11 PM
DEPLECE_TCRab_CD19	1/20/16 1:13:50 PM
Cytometer Settings	
Global Worksheets	
L160120_BC	
TCRab_CD19	1/20/16 11:44:22 AM
CD34	1/20/16 1
TCRab_CD19_n1	1/20/16 1
Compensation Controls	
OLGA	
MS	
VERONIKA	
DANIELA	
TOMAS GROH	
ELISKA	
MISA	
IVANA	
ASRAF	
KATKA	
Honza	
DAN	
TITRACE	
FELICITA	
IMUNOLOGIE	

pector - TCRab_CD19

Tube | Labels | Acq. | Cytometer Settings | Keywords

B-530/30	TCRgd FITC
B-575/26	TCRab PE
B-695/40	
B-780/60	CD45 PC7
R-660/20	CD3 APC
V-450/50	DAPI
B-610/20	CD20 PE-E610
780/60	

Cytometer - LSRII (1)

Status | Parameters | Threshold | Laser | Compensation | Ratio

Enable Compensation | Clear

Fluorochrome	- % Fluorochrome	Spectral Overlap
B-575/26	B-530/30	14.40
B-695/40	B-530/30	3.57
B-780/60	B-530/30	0.28
R-660/20	B-530/30	0.12
V-450/50	B-530/30	0.00

Experiment Layout

Labels | Keywords | Acquisition

Quick Entry

Label

Name	Label	Label	Label	Label	Label
DEPLECE_TCR_CD19					
L160120_BC					
TCRab_CD19	B-530/30 TCRgd FITC	B-575/26 TCRab PE	B-695/40	B-780/60 CD45 PC7	R-660/20 CD3 APC
CD34	B-530/30 Syto16	B-575/26 CD3 PE	B-695/40 CD45 PerCP	B-780/60	R-660/20 CD34 APC
TCRab_CD19_n1	B-530/30 TCRgd FITC	B-575/26 TCRab PE	B-695/40	B-780/60 CD45 PC7	R-660/20 CD3 APC

Labels

Name

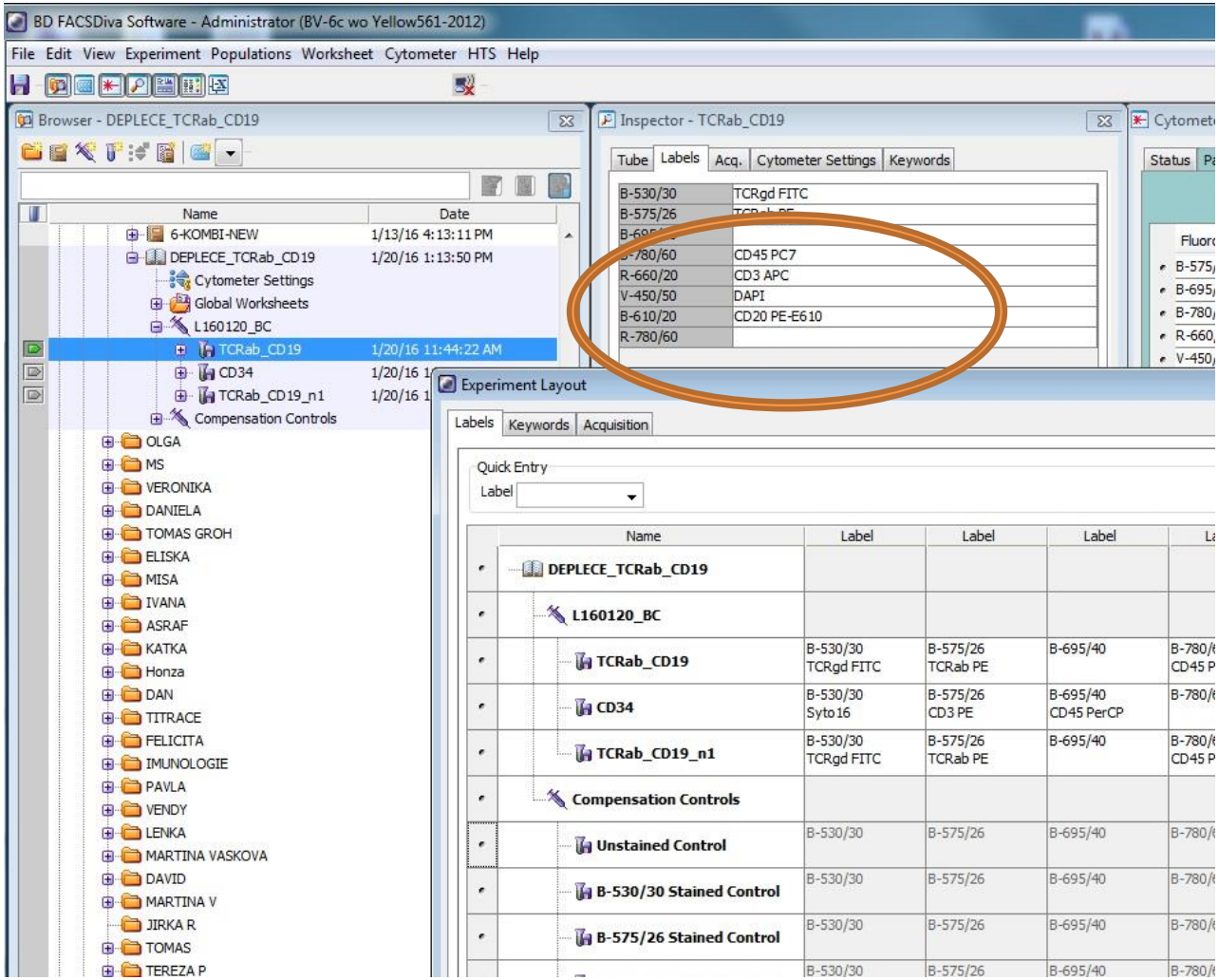
List by user

Administrator

BD Defined

Export data as:

FCS 3.0



Other formats: Navios/LMD

“LMD”
-> double FCS file

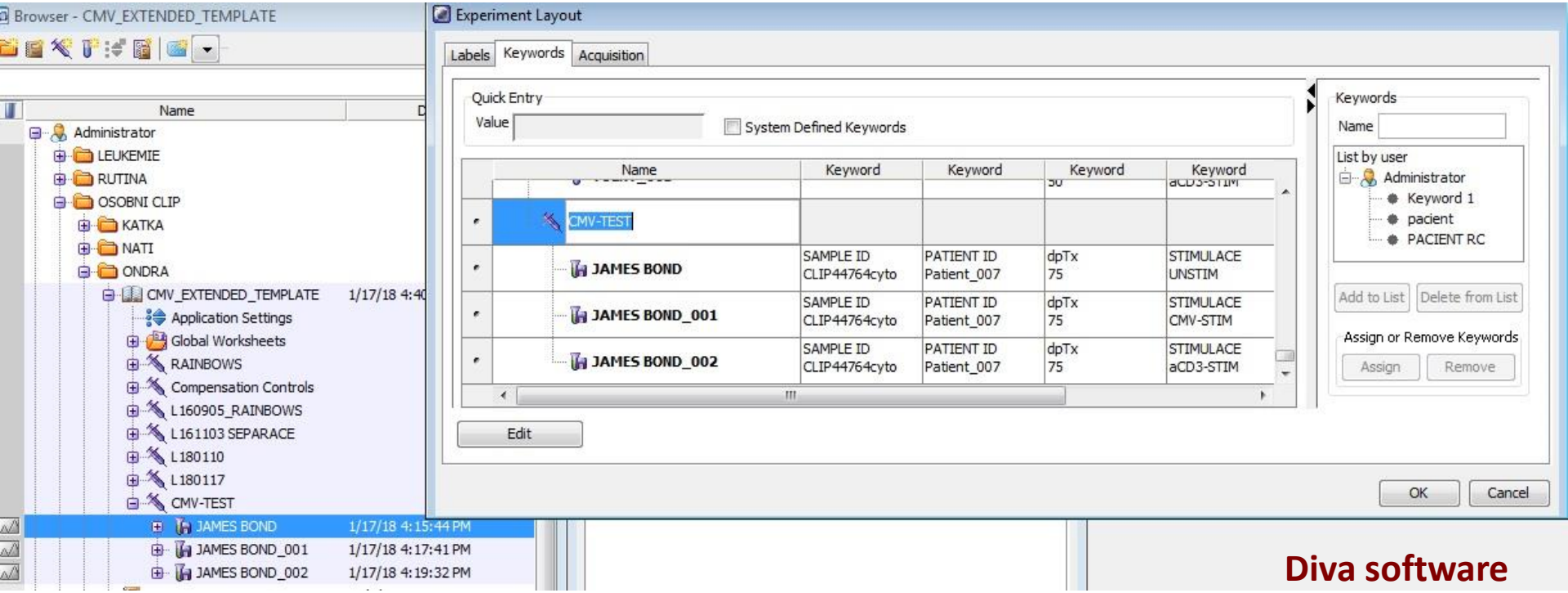
First piece:
FCS 2.0 file – simple, less resolution, fixed comp

Second piece:
FCS 3.0 file – full resolution data, without “keywords”

Annotation of FCS

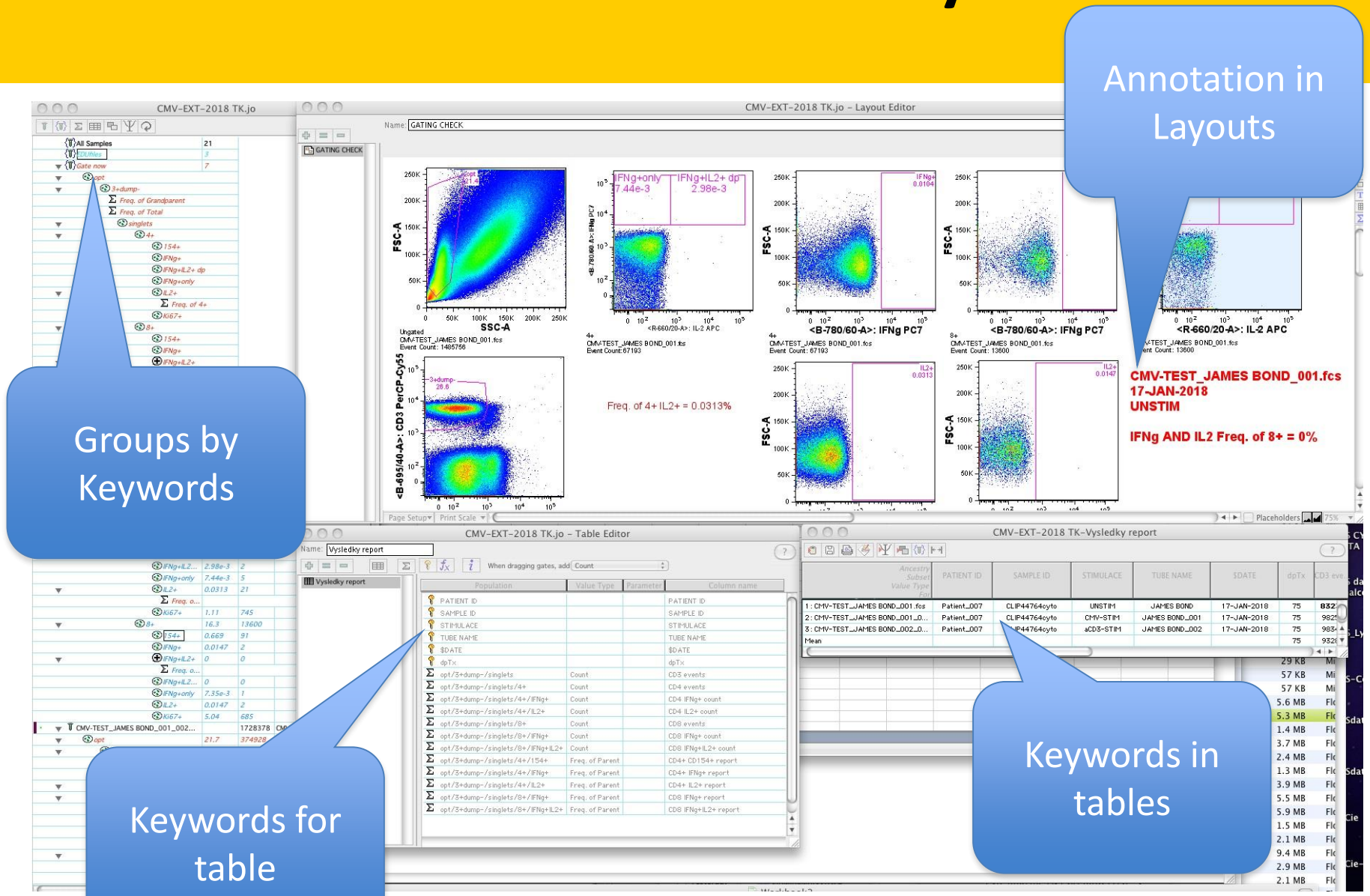
Keywords in datafiles:

Sample ID
Patient ID
Condition
Day post TX



Diva software

Annotations in Analysis

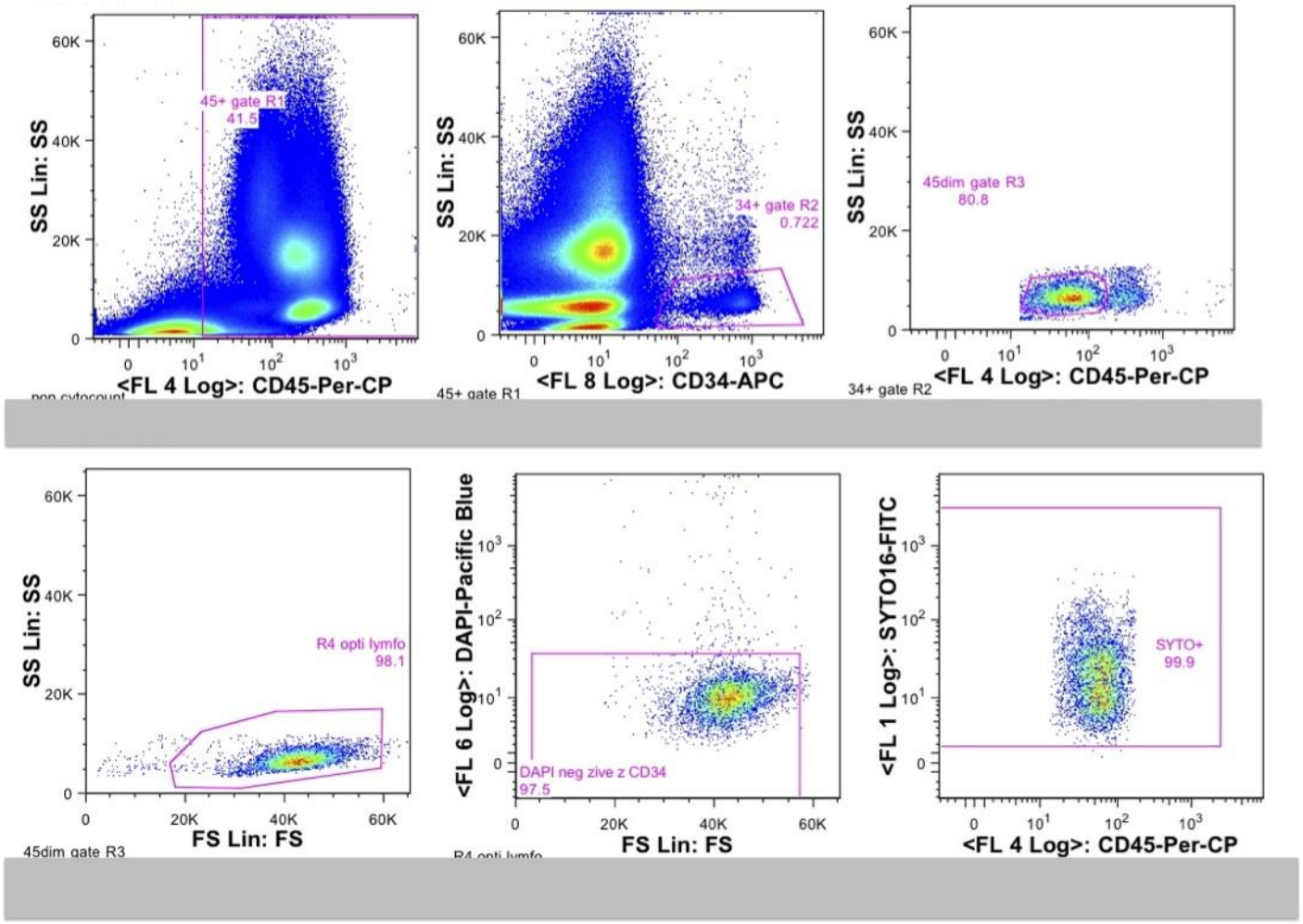


Flowjo software (Mac)

Gating controls summary

- Isotype controls (sticky cells)
- FMO controls (multicolor, compensated data)
- Sample comparisons
  (stimulated, unstimulated)

ISHAGE protocol
Layout in FlowJo

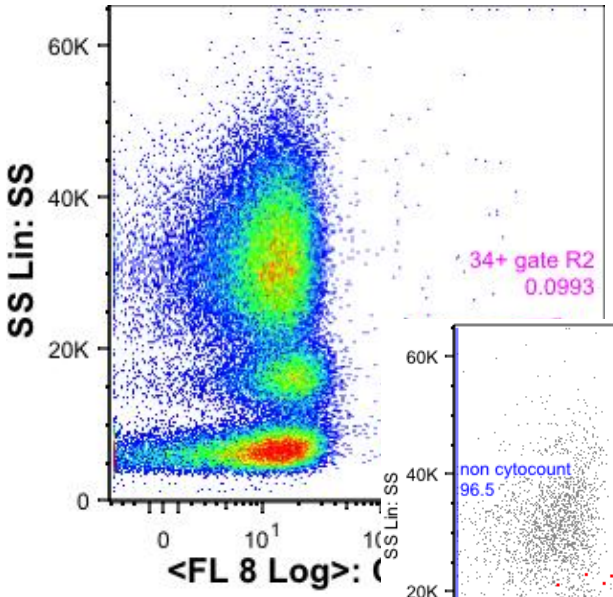


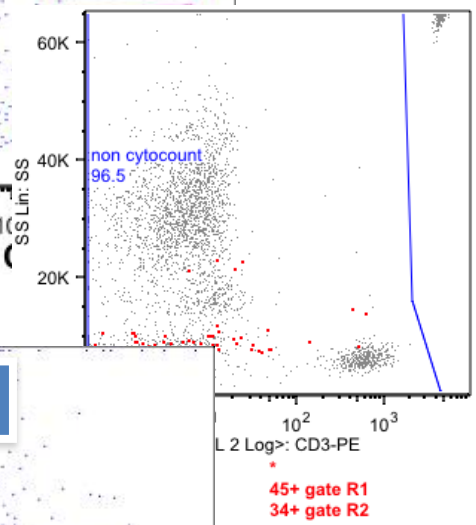
CD34+DAPIneg of CD45+ gate R1 = 0.558%

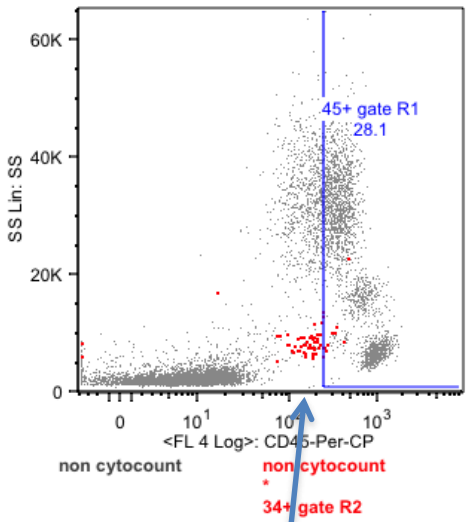
DAPI NEG Freq. of Parent = 97.5%

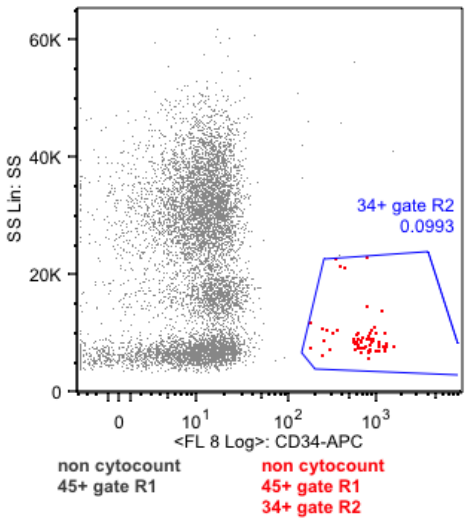
Backgating

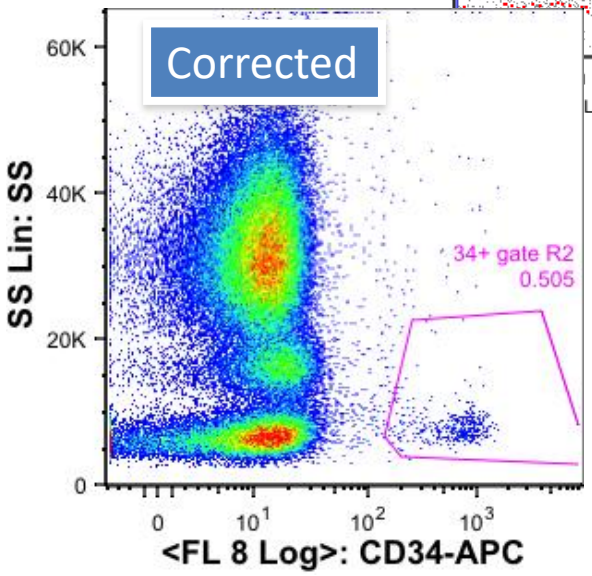
Check parental gates by backgating











Wrong position of CD45+ gate!!! by backgating

Create formulas

Create Table

Drag & Drop

Calculate table

ISHAGE TK 2016.jo

All Samples 10
CYAN TEMPLATE
LSR TEMPLATE

Name	Statistic	#Cells
		3000000
Accucount		
Koule		
non cytocount		
45+ gate R1	41.5	1243362
3+	23.5	292671
Count		
Freq. of Parent		
34+ gate R2	0.722	8982
45dim gate R3	80.8	7256
R4 opti lymfo	98.1	7121
Count		
Freq. of 45+ gate R1		
DAPI neg zive z CD34	97.5	6942
Freq. of Parent		
Freq. of 45+ gate R1		
SYTO+	99.9	6936
wbl-live	98.4	1223954
_SYTO16_3_45_34_DAPI_2.fcs		3000000
_SYTO16_3_45_34_DAPI_3.fcs		3000000
_SYTO16_IgG1_45_IgG1_DAPI_4_2...		3000000
_3_8_45_4_DAPI_STEP_5.fcs		3000000
45+ gate R1		
3+		
dapi-		
4+		
8+		
8+		

Name: ISHAGE v3 Cyan Accucount

When dragging gates, add Frequency of Parent

Population	Value Type	Parameter	Column name
$DATE			$DATE
$BTIM			$BTIM
PATIENT ID			PATIENT ID
non cytocount/45+ gate R1/34+ gate R2/45dim gate R3/R4 opti lymfo/DAPI neg zive z CD34	Freq. of	45+ gate R1	zive CD34+ z 45+
non cytocount/45+ gate R1/3+	Freq. of Parent		CD3+ z CD45+
non cytocount/45+ gate R1/34+ gate R2/45dim gate R3/R4 opti lymfo/DAPI neg zive z CD34	Freq. of Parent		viabilita CD34
non cytocount	Count		Celkem eventu
Formula	(<Column "CD3 eventu">/<Colum...		Conc CD3+ bb/ml
Formula	(<Column "zive CD34 eventu">/<...		Conc. CD34+ bb/ml

Create Formula for Table

Column Name: Conc. CD34+ bb/ml

(<Column "zive CD34 eventu">/<Column "Accucount">)*(51011/50)*1000

Insert Function: Choose... elp on functions

Insert Reference to Colum

To use a column's values in the formula, you must give the column a unique name. Only uniquely-named columns are shown in this menu.

Missing input values result in a blank output result

Cancel Create Formula

ISHAGE TK 2016-ISHAGE v3 Cyan Accucount

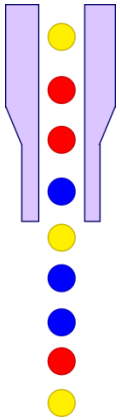
Ancestry / Subset / Value Type / For	$DATE	$BTIM	PATIENT...	zive CD...	CD3+ z ...	viabilita...	Celkem ...	Conc... bb ml	Conc... bb ml	WBC bb mL	zive CD...	CD34 ev...	CD3 eve...	CD45 ev...	Accucount	WBC-live	CD34+ ...	CD34+
	26 J...16	09:...:45		0.559	23.5	97.5	3.e6	1.06e8	2.5e6	4.5e8	6942	7121	2.93e5	1.24e6	2816	98.5	0.573	•
	26 J...16	09:...:18		0.586	23.1	97.9	3.e6	1.06e8	2.7e6	4.6e8	7239	7392	2.85e5	1.23e6	2739	•	0.599	•
	26 J...16	09:...:29		0.557	23.5	97.4	3.e6	1.02e8	2.4e6	4.35e8	7162	7354	3.02e5	1.29e6	3015	•	0.571	•
Mean				0.567	23.4	97.6	3.e6	1.05e8	2.5e6	4.48e8	7114	7289	2.93e5	1.25e6	2857	98.5	0.581	0
StdDev				0.0162	0.231	0.265	0	2.31e6	1.5e5	1.26e7	154	147	8505	32146	142	0	0.0156	0

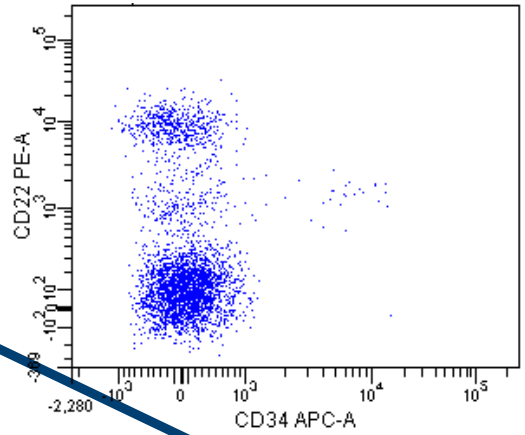
Curse of dimensionality

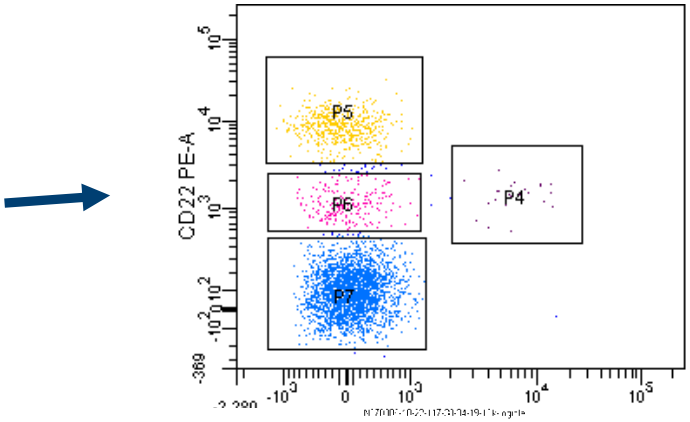
More parameters => more information

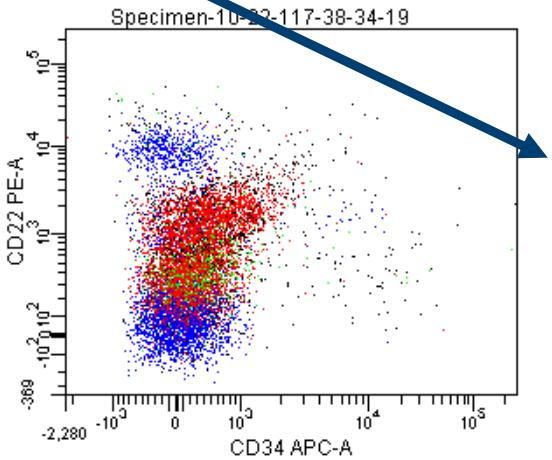
➔ More difficult to understand it

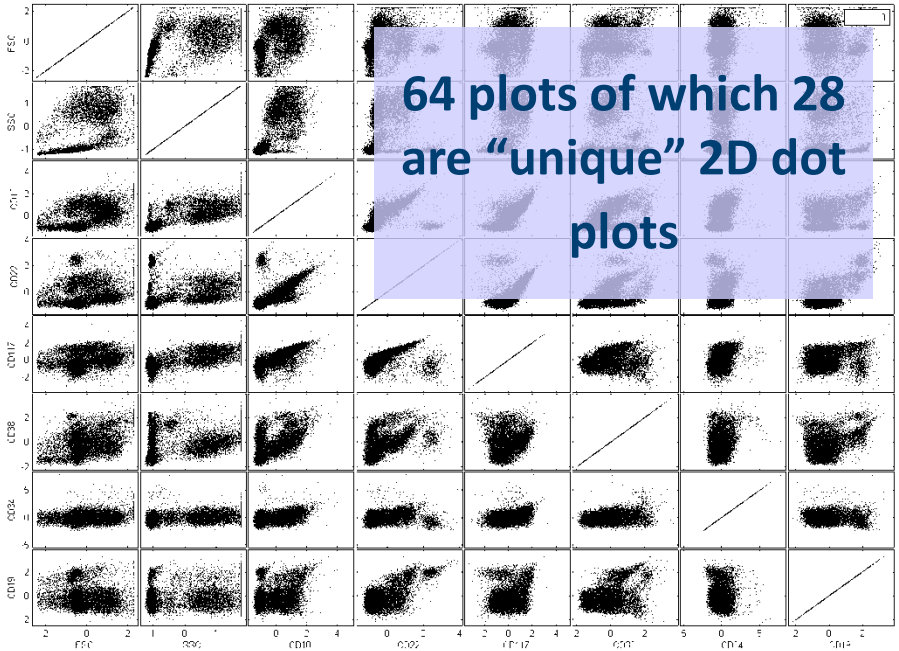
8-color data – perplexed!











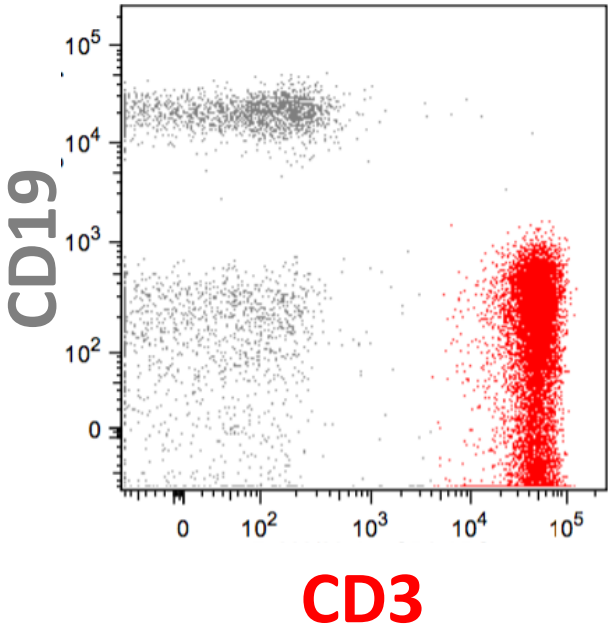
64 plots of which 28 are “unique” 2D dot plots

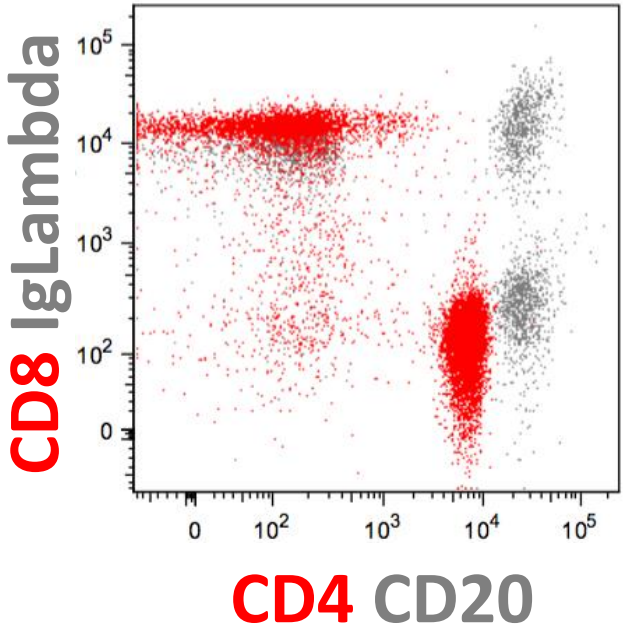
Slide Karel Fišer

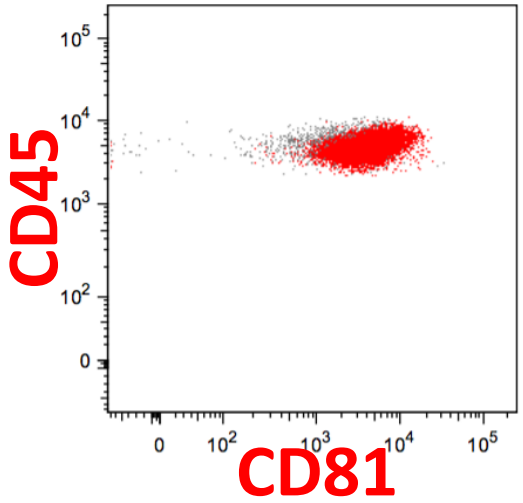
Projection of n-dimensions to 2D space

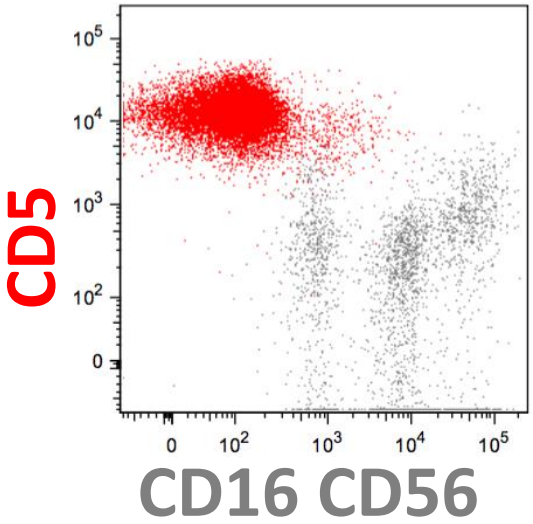
Lymphocytes

8 color data



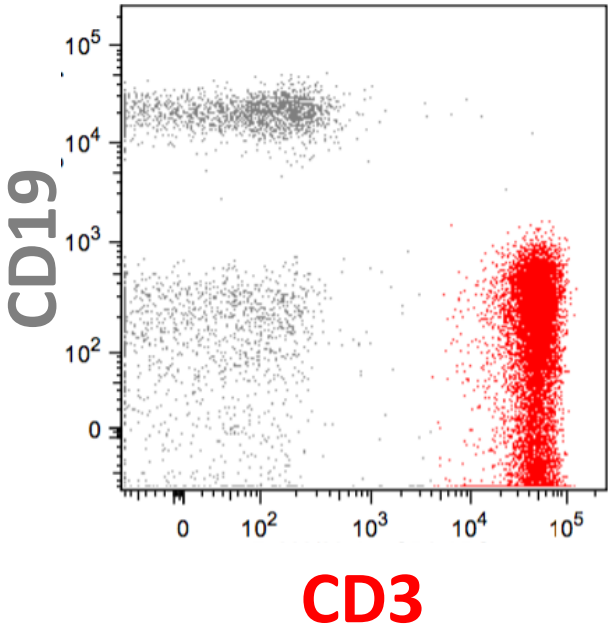




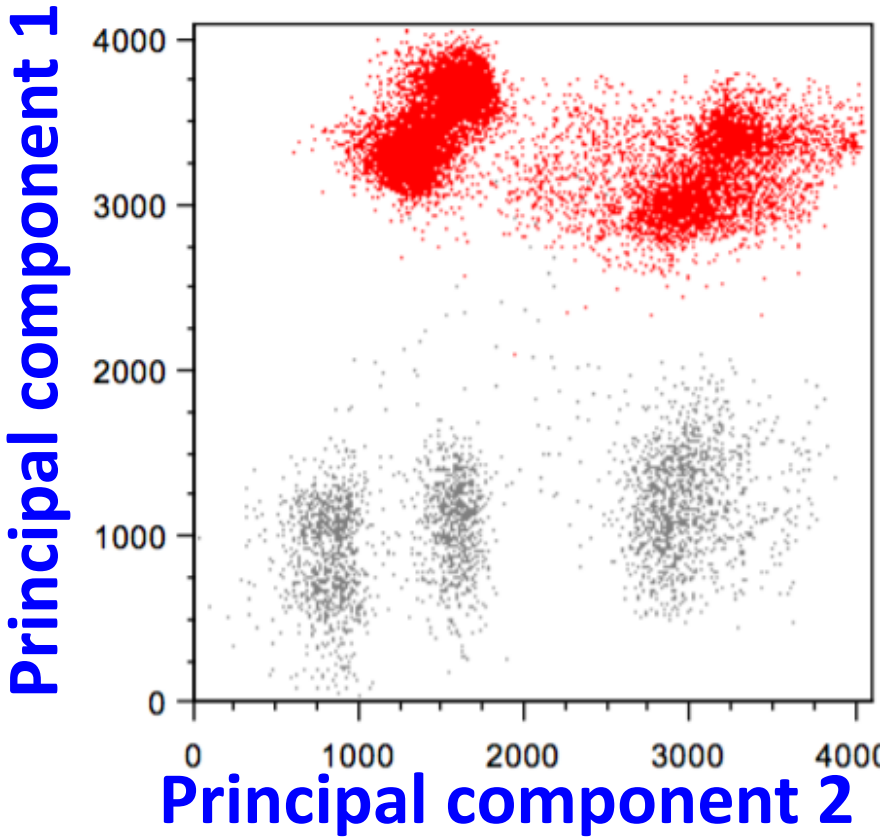


Principal component analysis

Lymphocytes



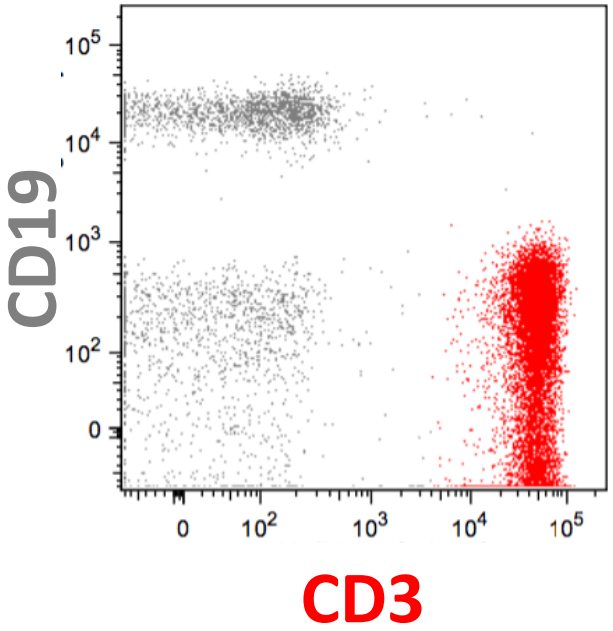
8 color data

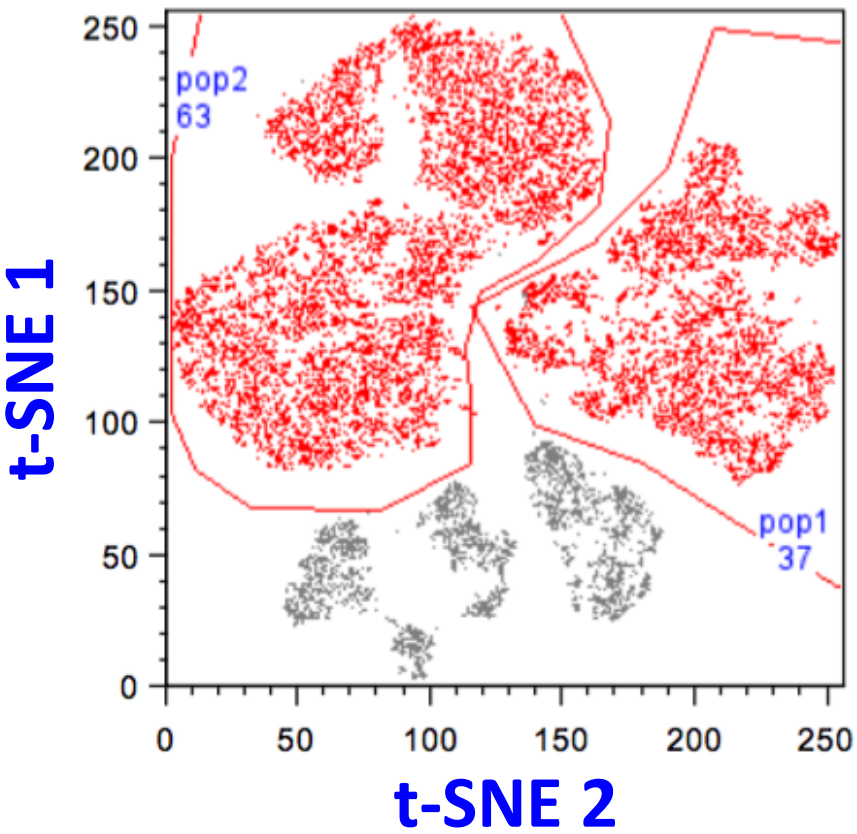


t-Distributed Stochastic Neighbor Embedding (t-SNE)

Lymphocytes

8 color data



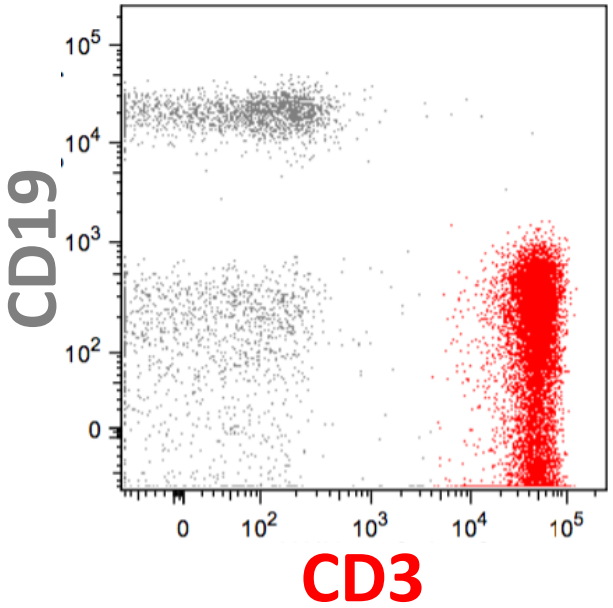


Flowjo software (Mac)

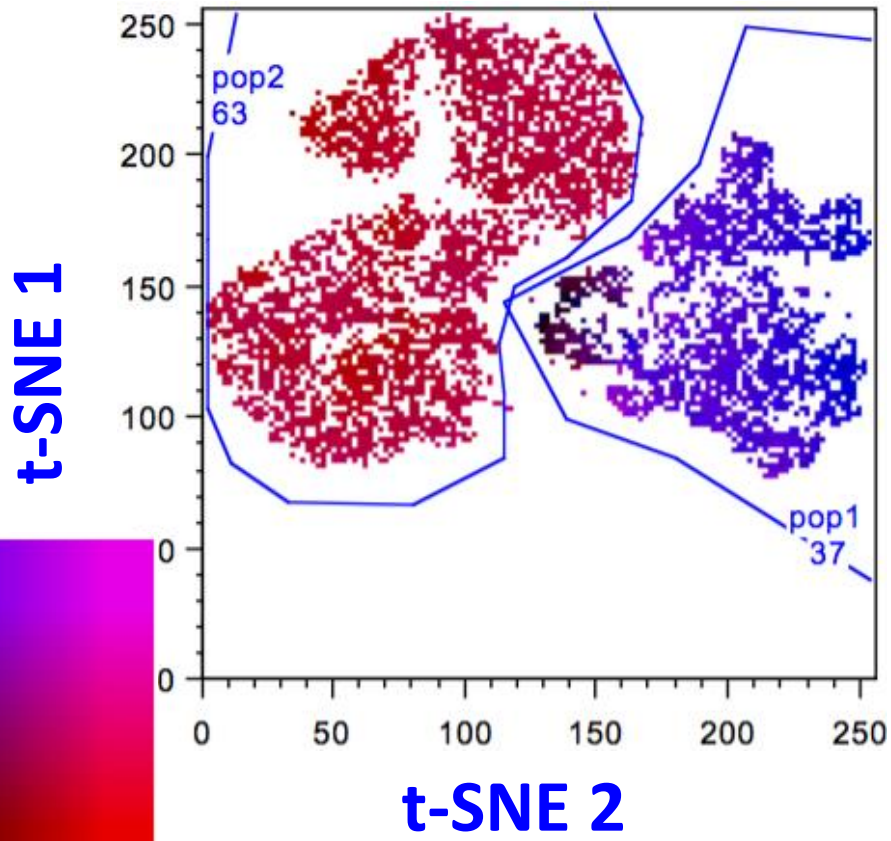
t-Distributed Stochastic Neighbor Embedding (t-SNE)

Lymphocytes

8 color data



Parameter as color:



Flowjo software (Mac)

LEUKEMIA (t-SNE)

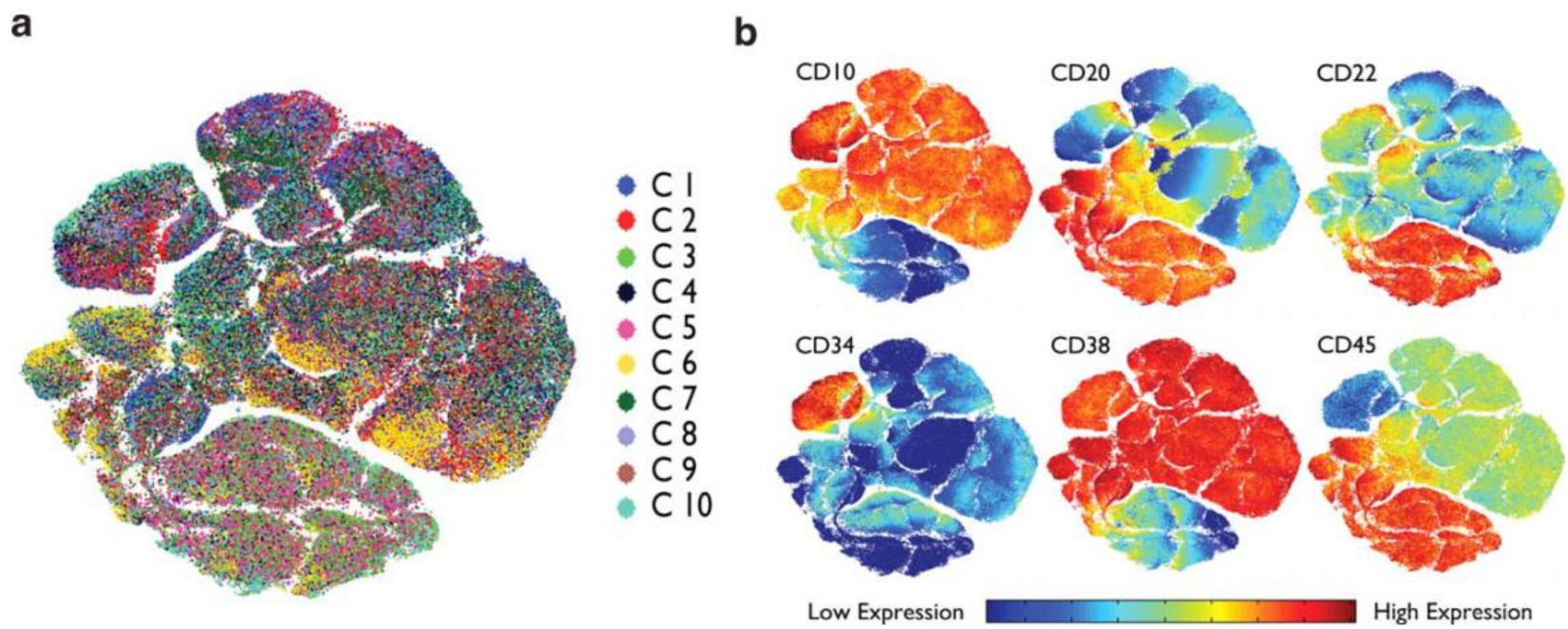
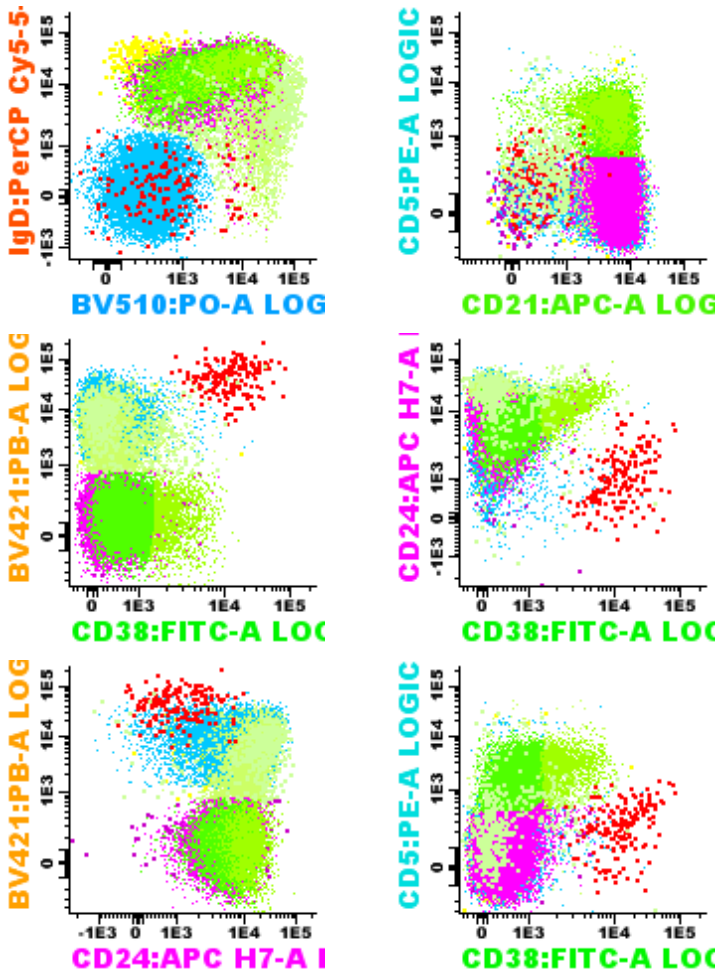


FIG. 1. viSNE map of normal bone marrow B cells labeled with a single 8-color antibody combination (tube A). Panel **A**: Each of the 10 control bone marrow samples used to construct the map (C1 through C10) is identified individually with a unique color. Each point in the viSNE map represents an individual cell colored by sample identity. Panel **B**: viSNE map shown in Panel (A), in which each cell is colored to reflect intensity of antigen expression for six different markers in tube A. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

DiGiuseppe, J. a, Tadmor, M. D. & Pe'er, D. Detection of minimal residual disease in B lymphoblastic leukemia using viSNE. Cytom. Part B Clin. Cytom. 88, 294–304 (2015).

B cells' subsets (t-SNE)



Infinicyt software

IMMATURE

NAIVE CELLS

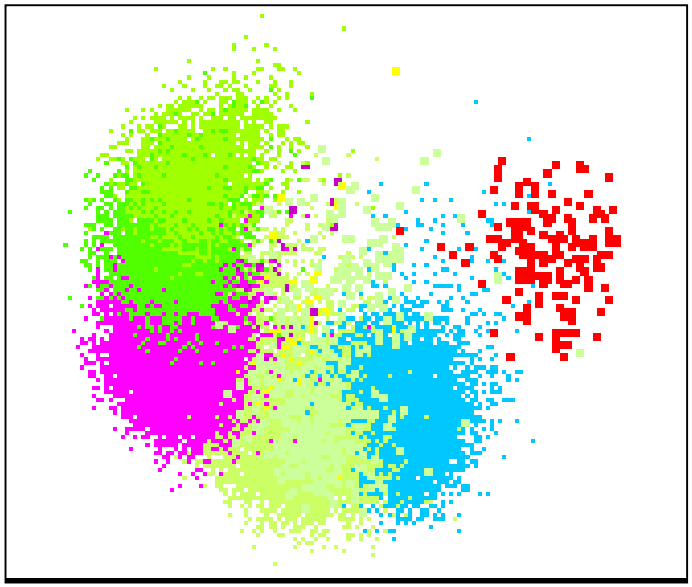
NAIVE CELLS

IGM+IGD+
MEMORY CELLS

IGM-IGD-
MEMORY CELLS

PLASMABLAST CELLS

Principal component analysis



Fast, linear

Hiding small

Projection of n-dimensions to 2D space

Principal component analysis

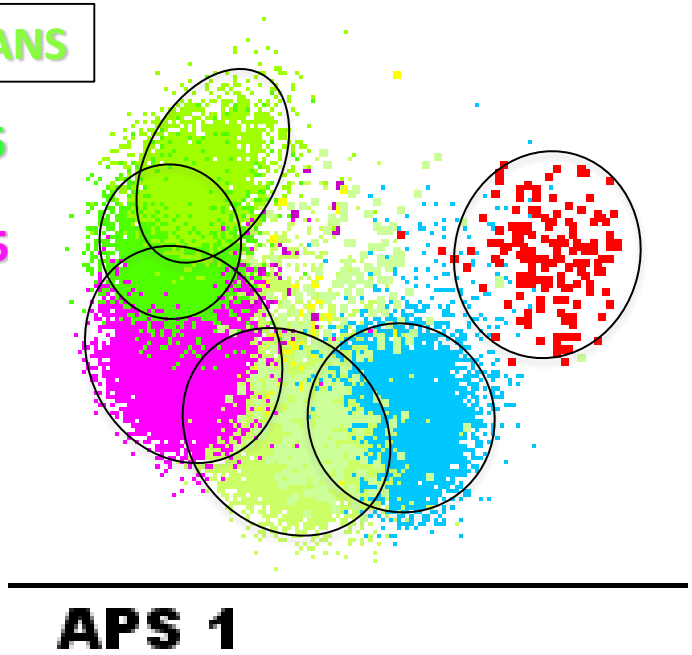
IMMATURE/TRANS

NAIVE CELLS CD5

NAIVE CELLS CD5

IGM+IGD+ MEMORY CELLS

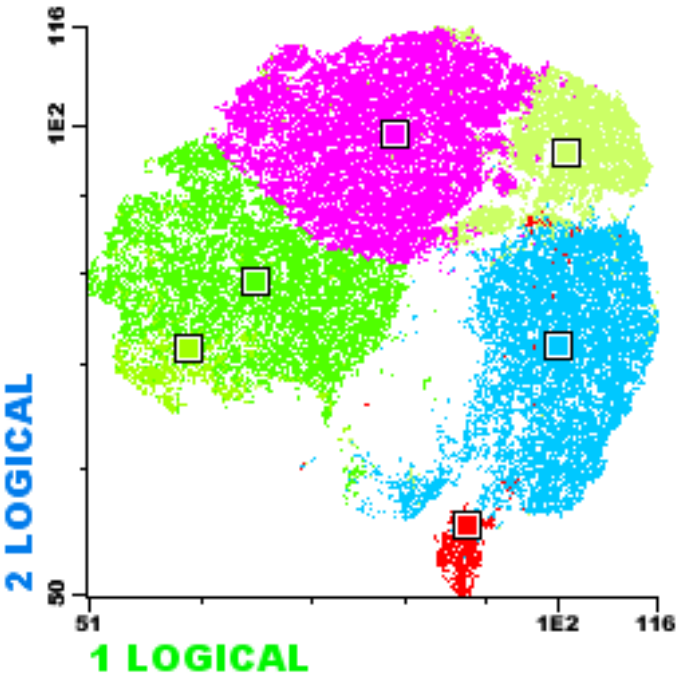
IGM-IGD- MEMORY CELLS



PLASMABLAST CELLS

Infinicyt software / R

t-SNE 2Dmap



Representing all events

Slow, non-linear

Software - overview

- FlowJo (Mac & Win) – batch gating, statistics/tables, overlays, stacked histograms, t-SNE (PCA)
- Diva – layouts/batch
- Kaluza – Radar plot, Tree plots, layouts ?
- Infinicyt – PCA, databases - automated analysis
- Cytobank – ViSNE, SPADE, histograms
- Summit
- R-project

- FCS Express
- Flowing software
- Cyflogic
- Cytospec and PlateAnalyzer

Hardware

Dataserver (few Tb)

PC

- SSD disk (prevent overfilling)
- RAM (16GB or more)
- Fast graphics
- Good monitor

Tip: Software and hardware is expensive, add it to instrument purchase

Sharing the knowledge

- Quality description of experiments

ORIGINAL ARTICLE

Cytometry PART A
Journal of the International Society for Advancement of Cytometry

MIFlowCyt: The Minimum Information About a Flow Cytometry Experiment

Jamie A. Lee,[1†] Josef Spidlen,[2†] Keith Boyce,[3] Jennifer Cai,[1] Nicholas Crosbie,[4] Mark Dalphin,[5] Jeff Furlong,[6] Maura Gasparetto,[2] Michael Goldberg,[7] Elizabeth M. Goralczyk,[8] Bill Hyun,[9] Kirstin Jansen,[6] Tobias Kollmann,[10] Megan Kong,[1] Robert Leif,[11] Shannon McWeeney,[12,13,14] Thomas D. Moloshok,[8] Wayne Moore,[15] Garry Nolan,[16] John Nolan,[17] Janko Nikolich-Zugich,[18] David Parrish,[3] Barclay Purcell,[19] Yu Qian,[1] Biruntha Selvaraj,[19] Clayton Smith,[2] Olga Tchuvatkina,[7] Anne Wertheimer,[20] Peter Wilkinson,[21] Christopher Wilson,[6] James Wood,[22] Robert Zigon,[23] The International Society for Advancement of Cytometry Data Standards Task Force, Richard H. Scheuermann,[1,24] Ryan R. Brinkman[2*]

- FCS data availability

https://flowrepository.org/