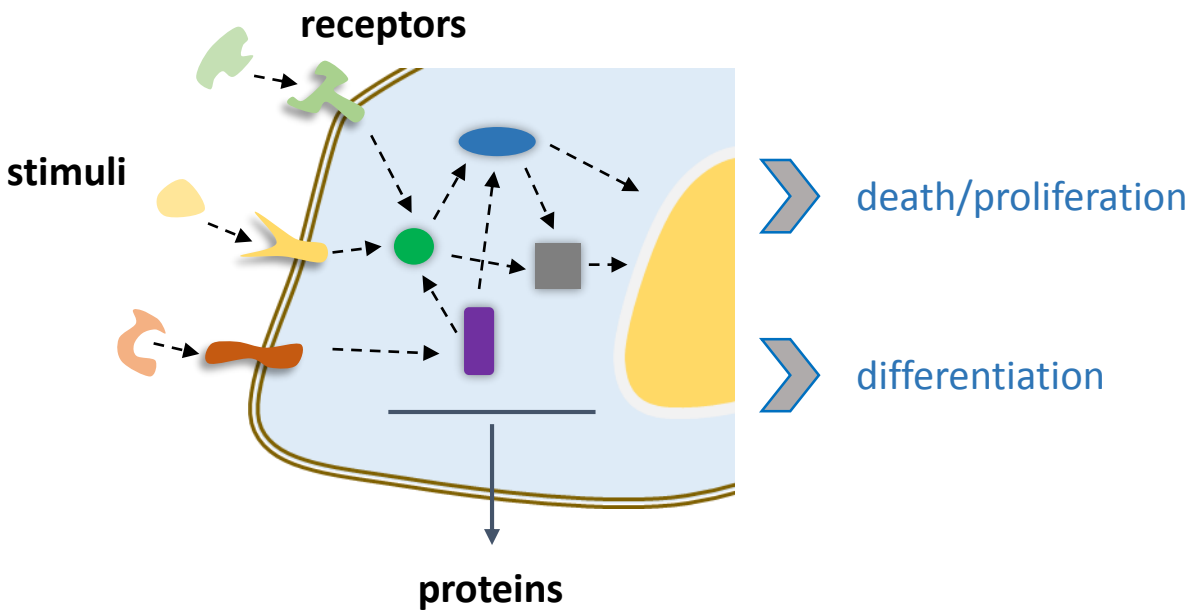
Single cell (mass) cytometry

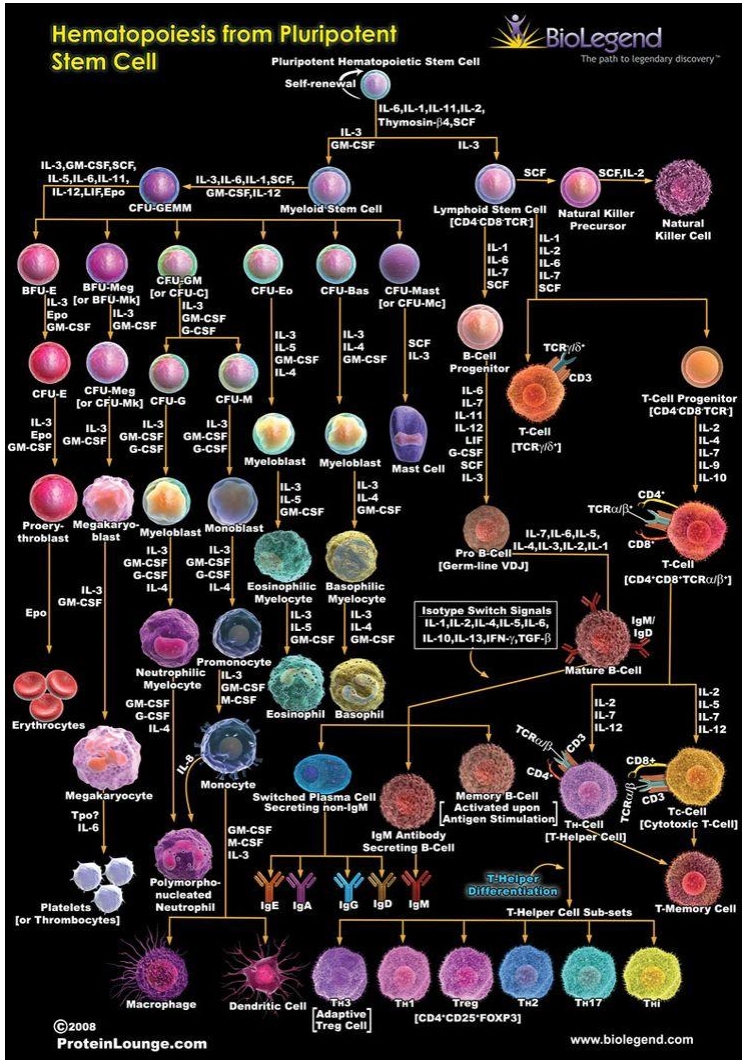
Introduction to R for Bioinformatics

Presented by: Giorgos Papoutsoglou, PhD

University of Crete
Computer Science Department

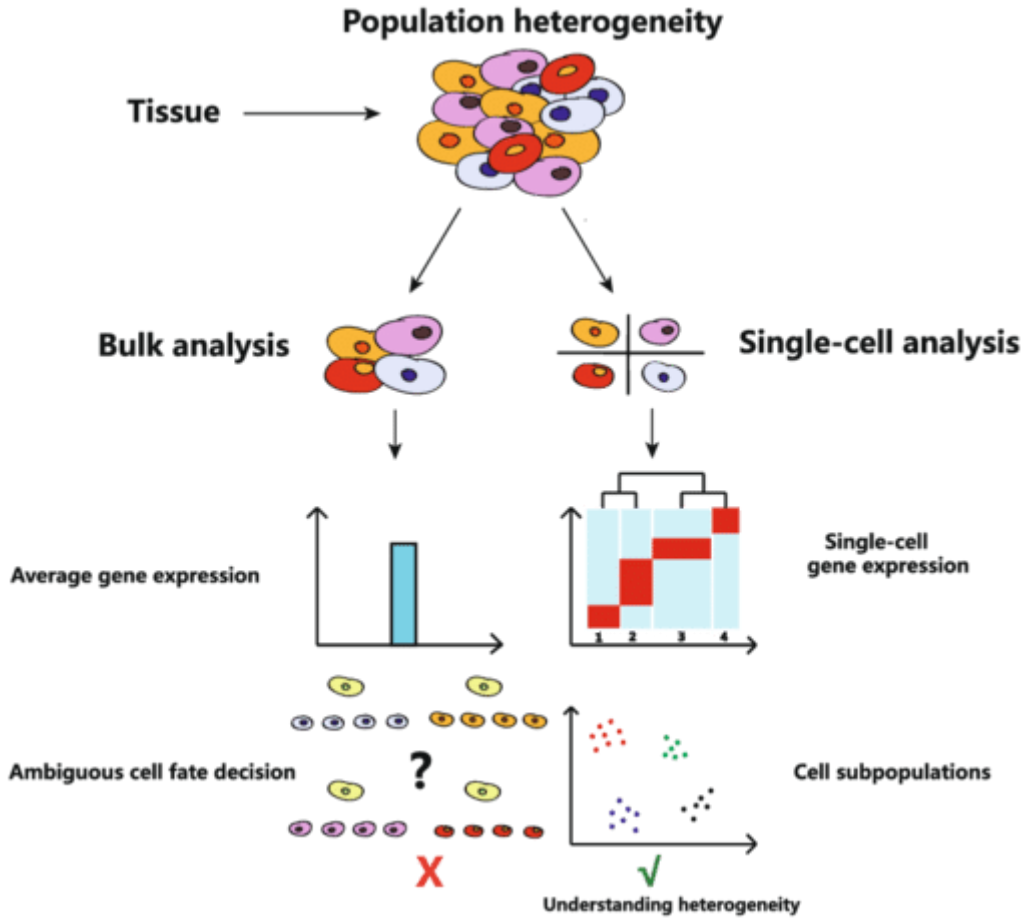
Background: Cell development





Why single-cell data ?

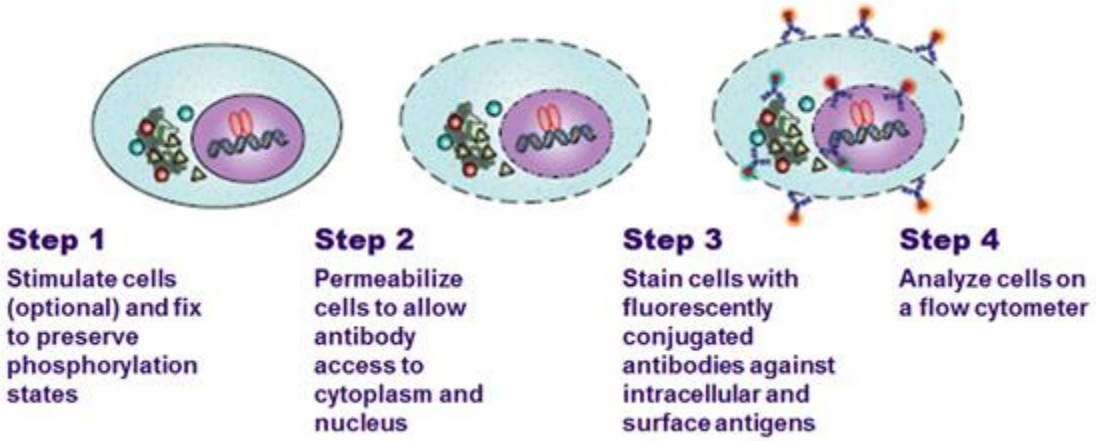
- Find new cell populations / functions
- Decipher system heterogeneity and general system monitoring
- Disease measurement, prediction, understanding
- Network / inference / co-variance analysis
- Single cell systems ordering (revealing unanticipated order within a mixutre)



Ye, *J. Hematol. Oncol.*, 2017

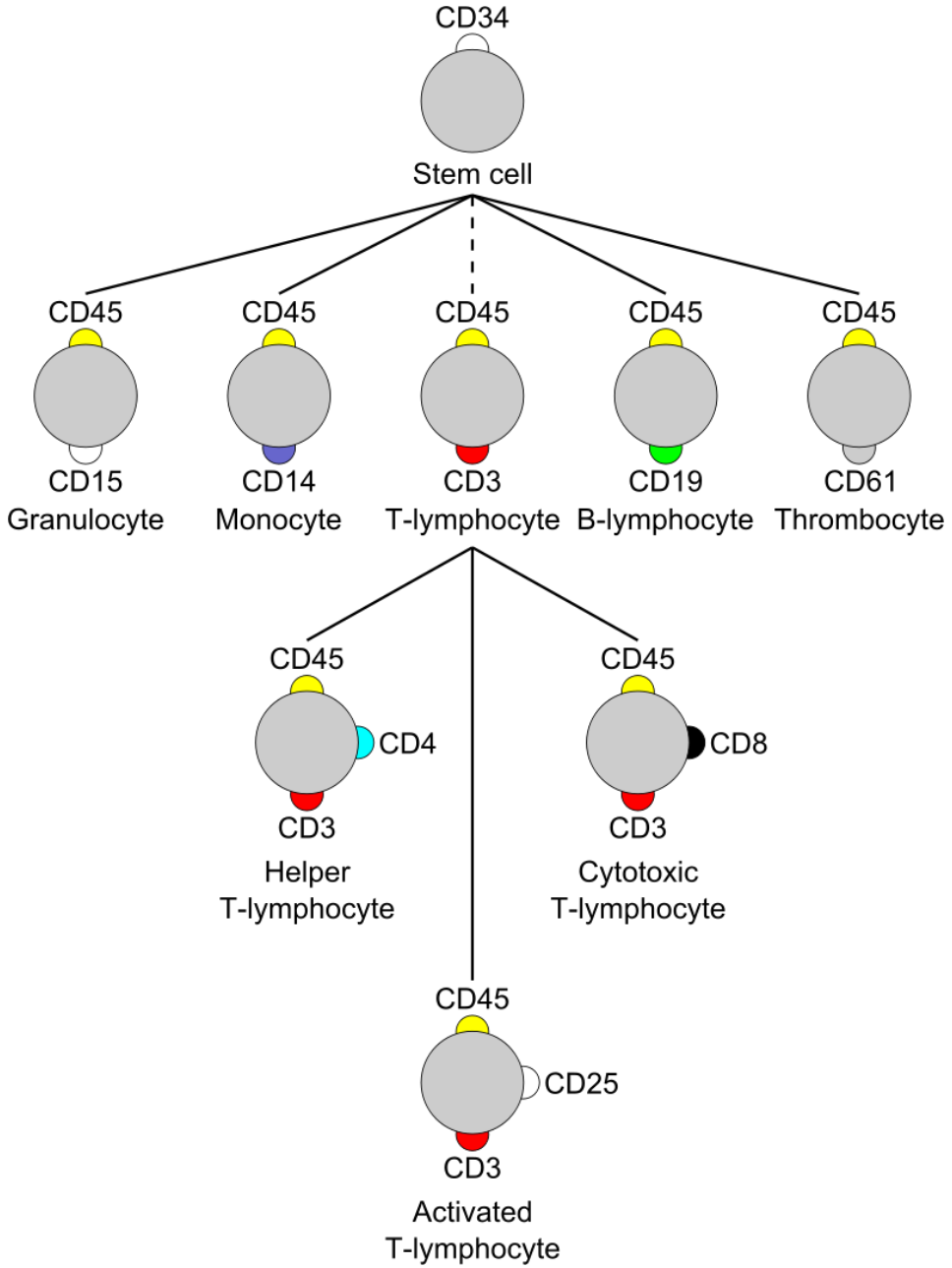
Study cell populations

Use of fluorescent tags coupled to antibodies, able to bind to specific cell targets



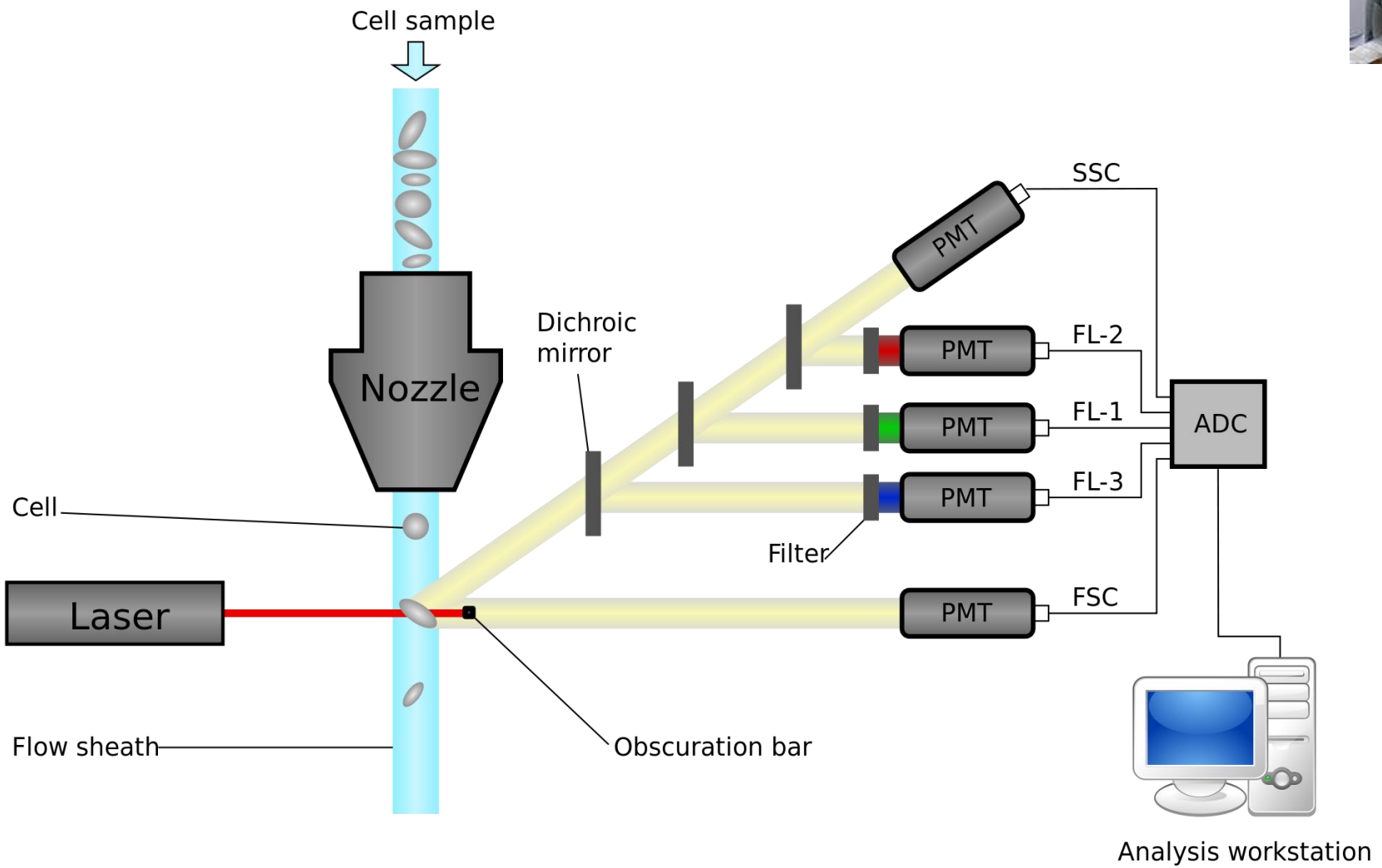
www.bdbiosciences.com

For example, in *immunophenotyping* the targets are specific surface proteins



en.wikipedia.org

Fluorescent cytometry



https://commons.wikimedia.org

Data files

- FCS 2.0 and FCS 3.0 conventions
- Contain all of the measurements (FSC, SSC, FL1…) for each individual cell processed in a given sample

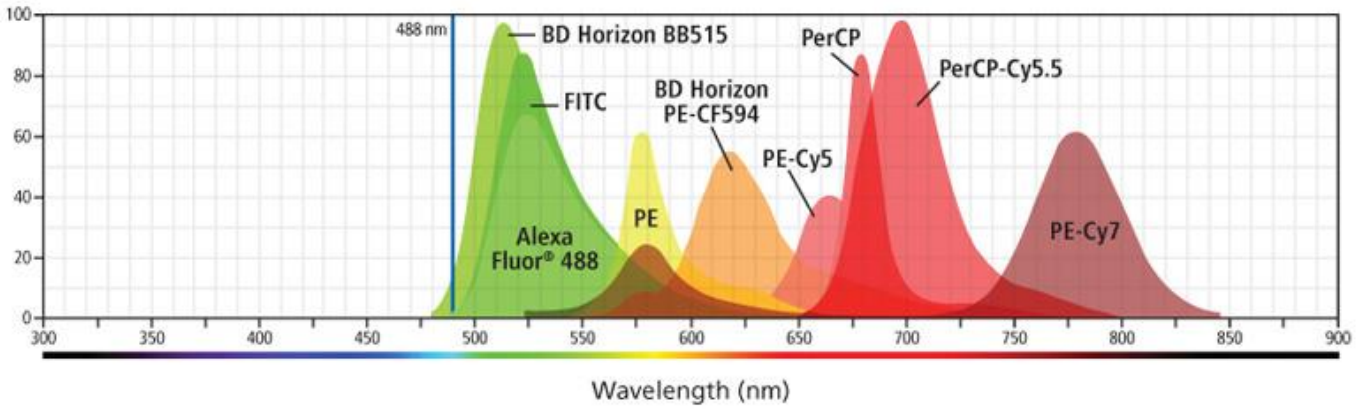
	FSC	SSC	FL1	…
Cell 1	##	##	##	
…				
cell *N*	##	##	##	

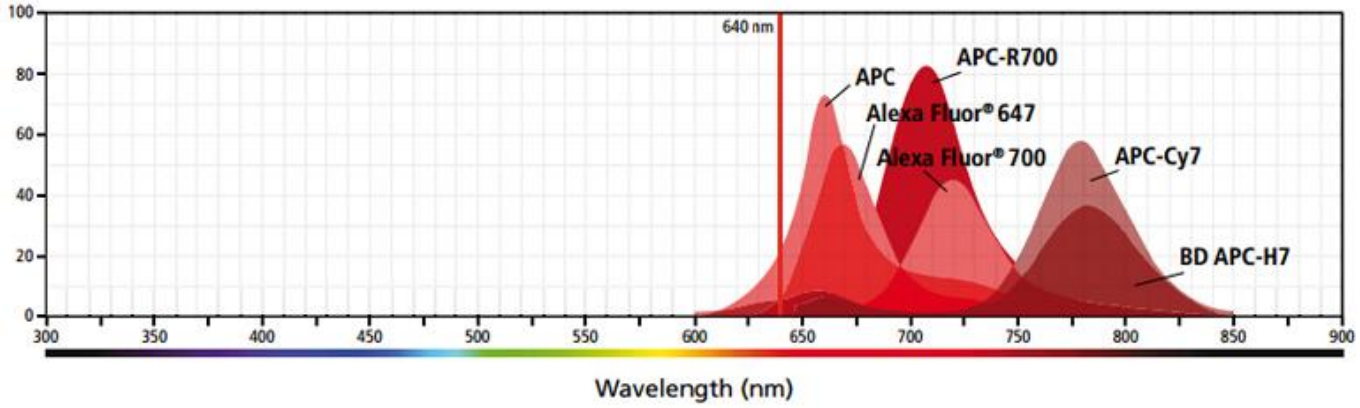
Bottlenecks

Channel Overlap (*spillover*)

- Up to 12 colors can be “routine”
- 17 colors have been reported
- High background

Variable dynamic range

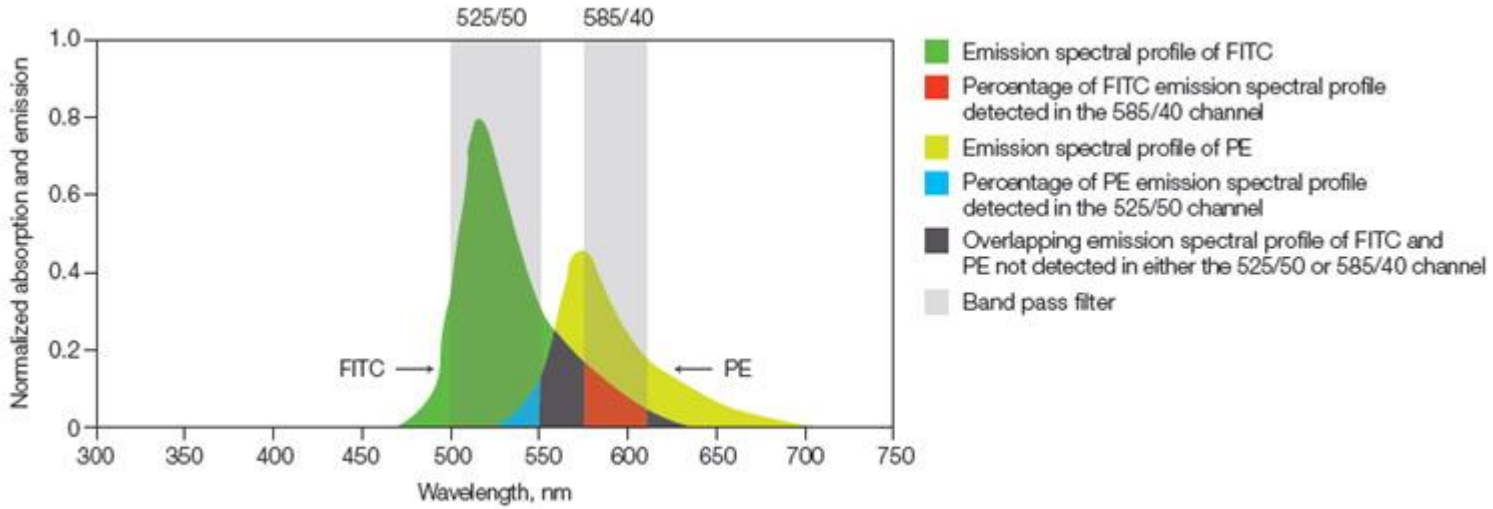




www.bdbiosciences.com

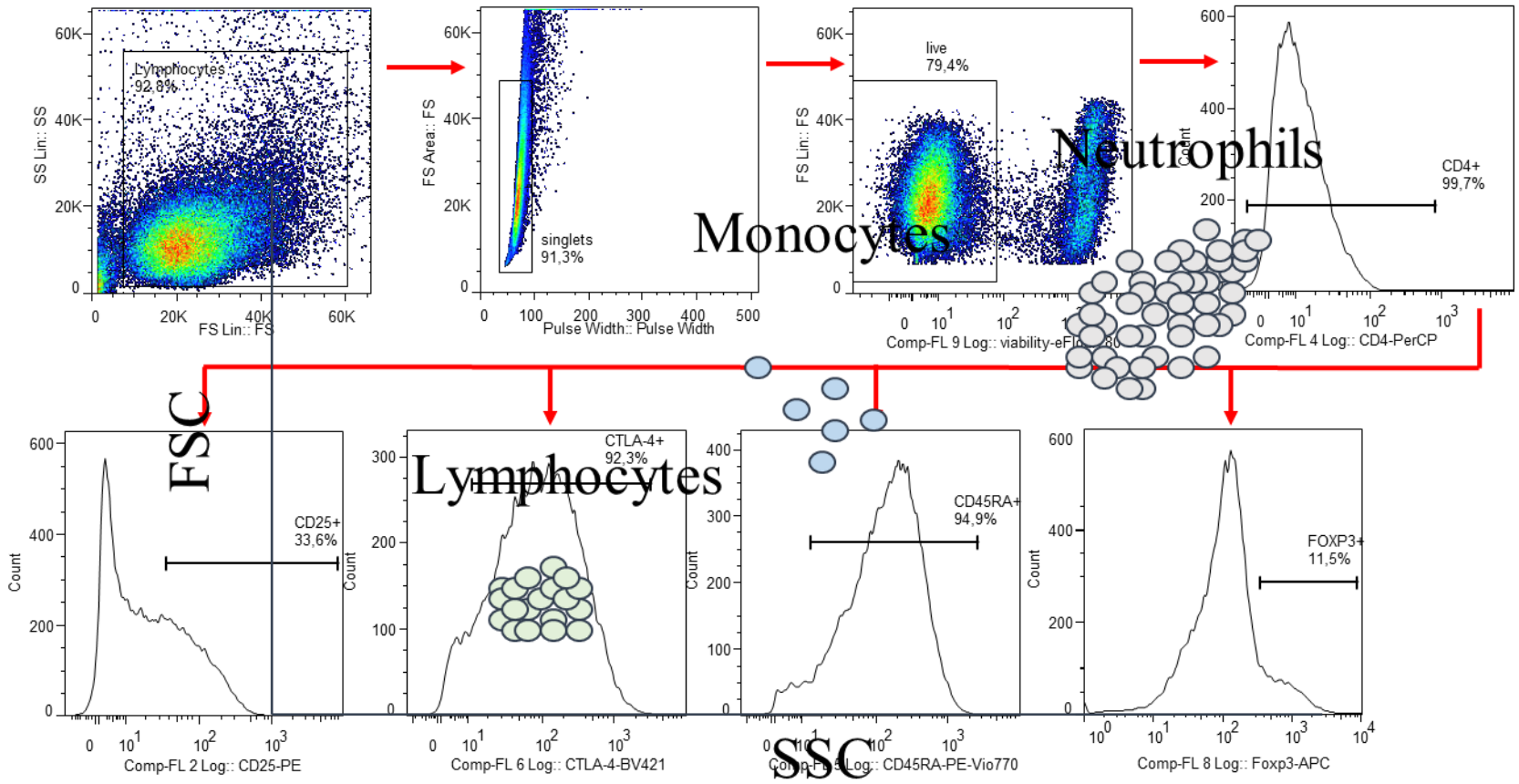
Compensation

- Mathematical procedure to remove the spillover i.e. to account for spectral overlap and to measure the photons deriving from one fluorophore into multiple detectors
- Single stained controls are required for all fluorophores used to reveal the level of spectral overlap for each detector



www.bio-rad-antibodies.com

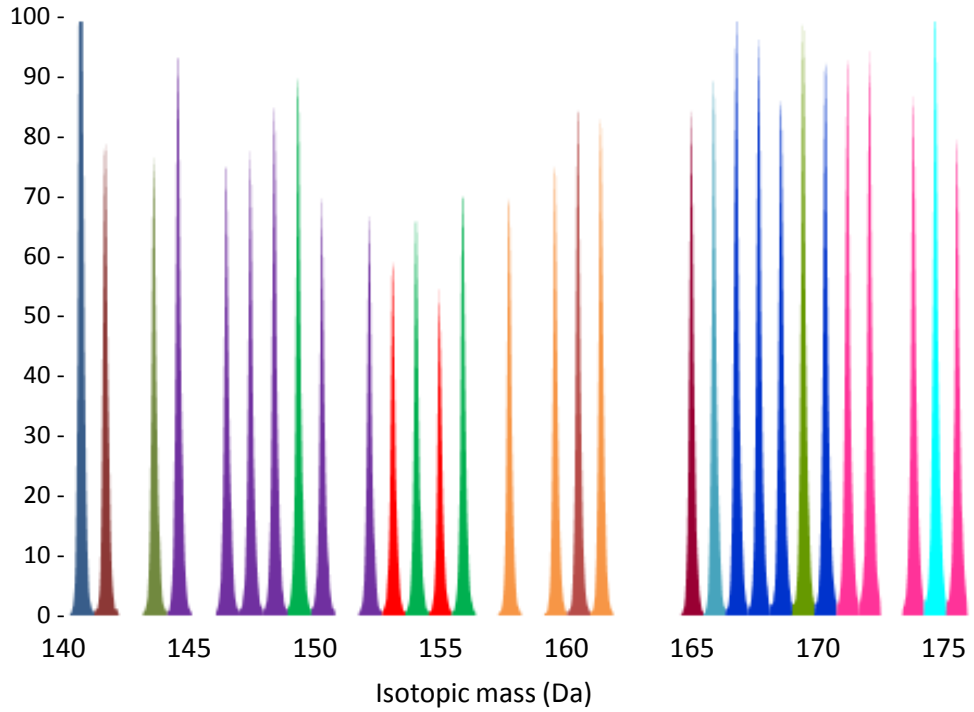
Data analysis: Gating



Papoutsoglou, et. al., *NAR*, 2017

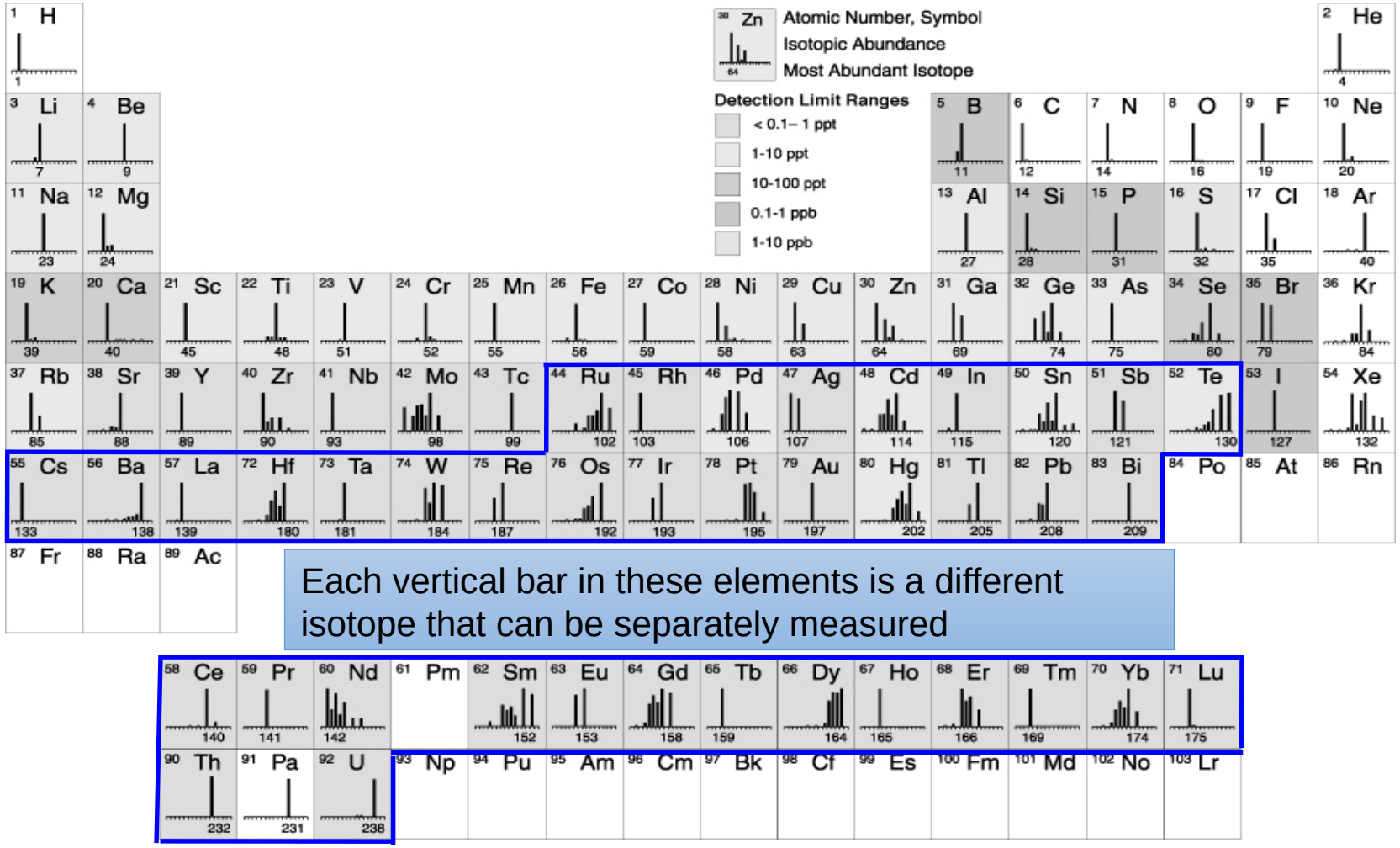
Mass cytometry

- Up to 100 non-biological elemental mass channels
- No compensation required
- Zero background



Adapted from S. Bendall, http://web.stanford.edu/class/immunol206a/

How do you get > 50 parameters?



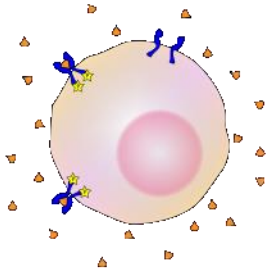
Adapted from S. Bendall, http://web.stanford.edu/class/immunol206a/

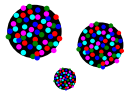
Data preparation

- Tag antibodies with different metals (chelation)
- Crosslink the proteins (freeze the inner of the cell)
- Permeabilize the membrane (make holes)
- Label cell proteins with chelated antibodies
- Spike beads in the samples for internal standard correction (normalization)
- Label samples with different barcodes to use them together

Workflow

Perturbations



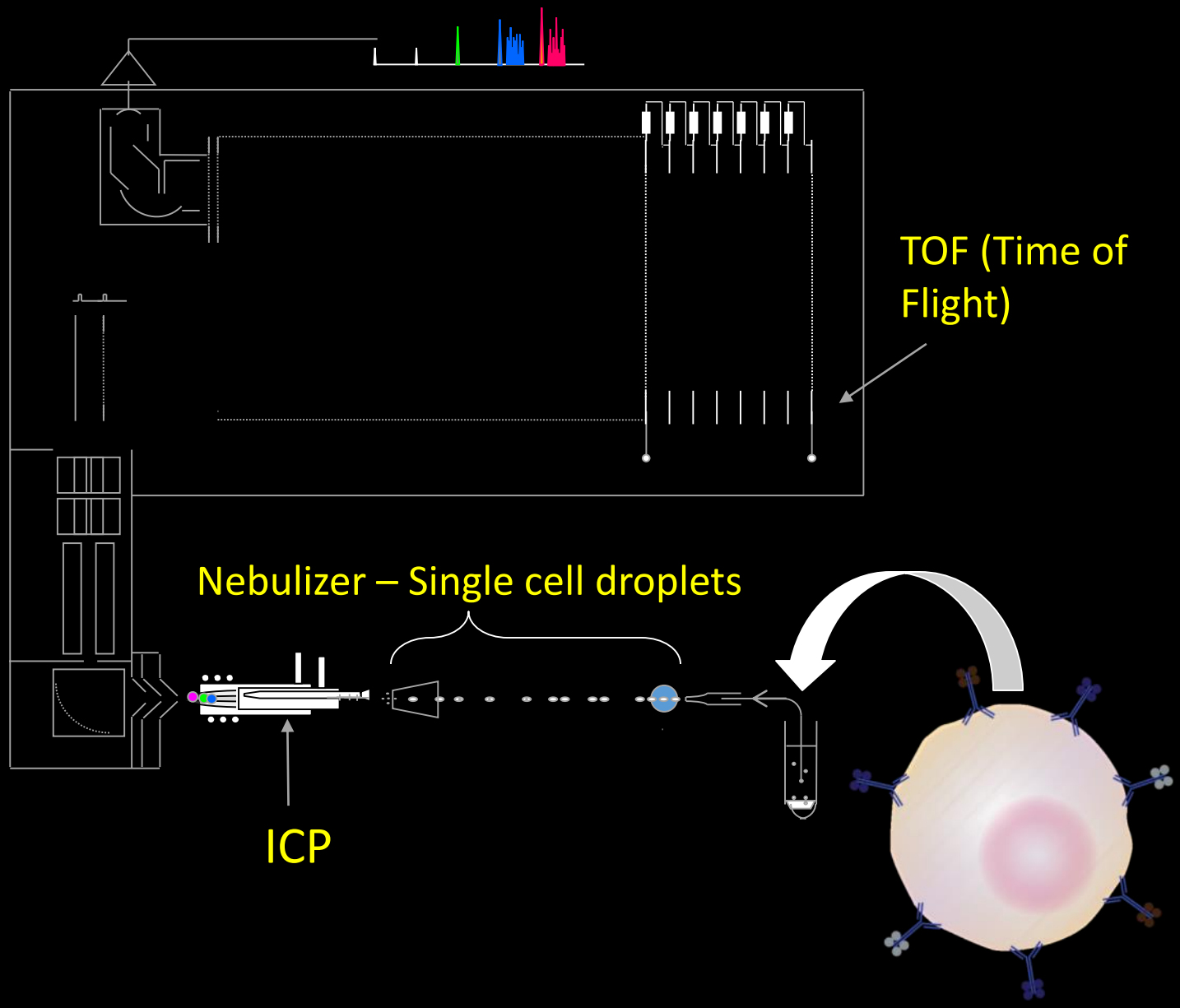


Beads

Adapted from R. Finck, http://web.stanford.edu/class/immunol206a/

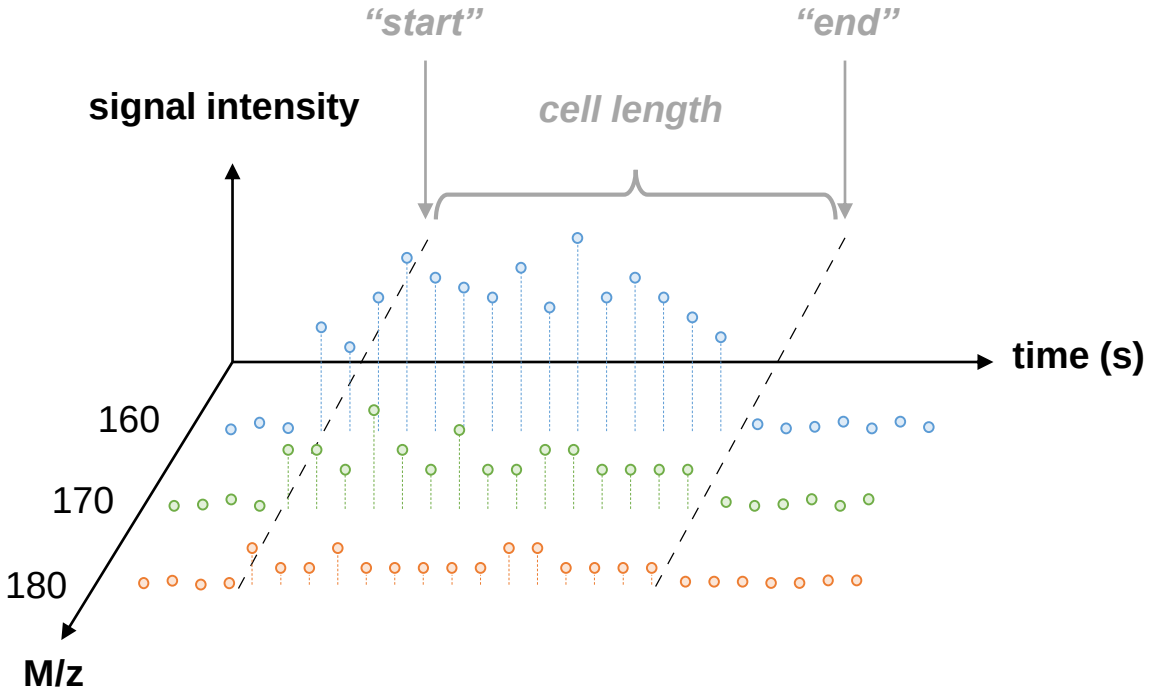
CyTOF: A prototype schematic

Bandura D, *et al*. Anal Chem. 2009



Adapted from S. Bendall, http://web.stanford.edu/class/immunol206a/

Data collection



.IMD file

time (s)	ch. 160	ch. 170	ch. 180
10	25	11	2
11	20	18	4
12	28	16	3
...	...	...	...
59	2	3	3
…	…	…	…

−

Detection threshold
3

=

time (s)	ch. 160	ch. 170	ch. 180
10	22	8	-1
11	17	15	1
12	25	13	0
...	...	...	...
59	-1	0	0
…	…	…	…

if <0, set to 0

find event "*start*", "*end*"

cell length = "end"-"start"

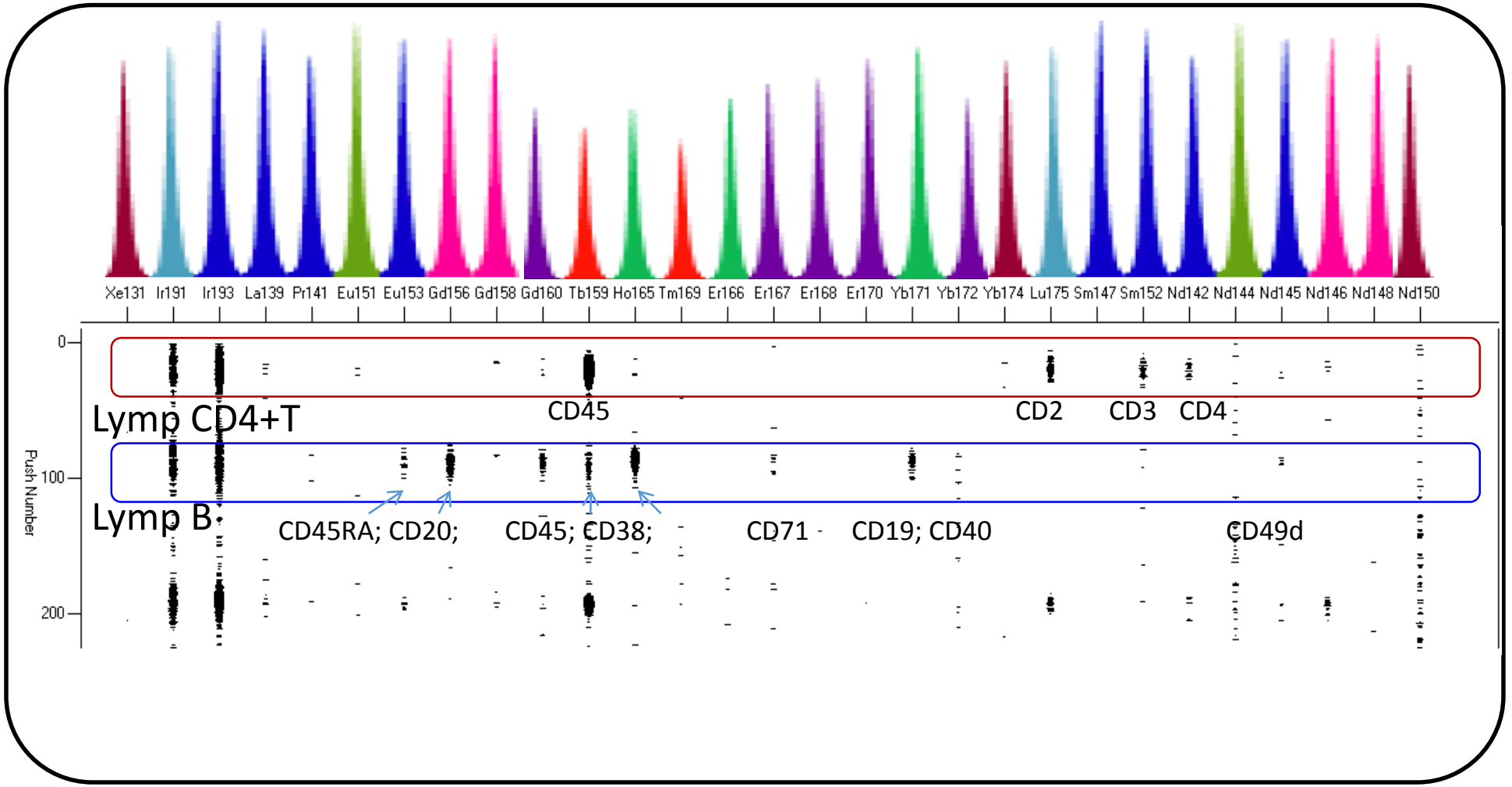
integrate from "start" to "end"

.FCS file

time (s)	Cell len.	ch. 160	ch. 170	ch. 180
10	48	1402	563	15
189	42	1212	481	36
302	51	1934	787	29
...		...	...	...

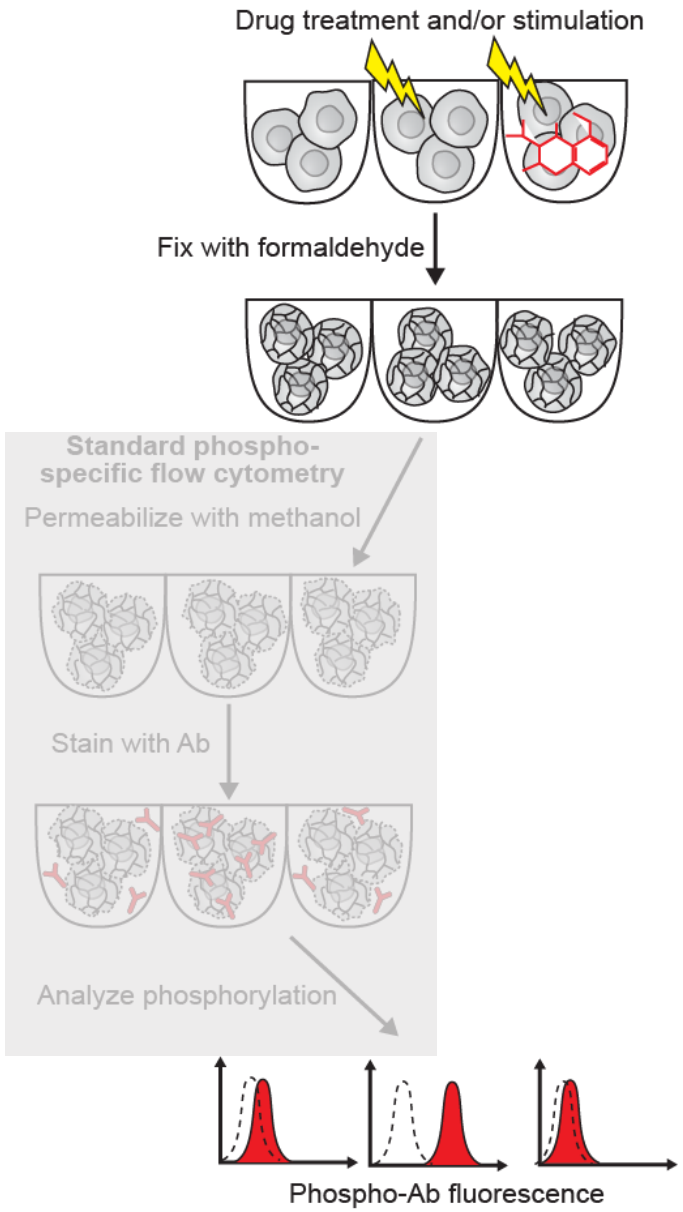
! Nolan's Lab does (still?) randomization here by shifting the axis to +100. In this way they maintain some of the the negative values.

Example reads per isotope



Adapted from S. Bendall, http://web.stanford.edu/class/immunol206a/

Cell Multiplexing/Barcoding



Krutzik PO, Nolan GP. , Nat Methods. 2006

Advantages:

1. Uniform Staining
2. Reduced Antibody Consumption
3. Reduced Acquisition Time

Adapted from R. Finck, http://web.stanford.edu/class/immunol206a/

Binary Cell Labeling Schemes for n-well MCB Multiplexing

$2^1=2$

$2^2=4$

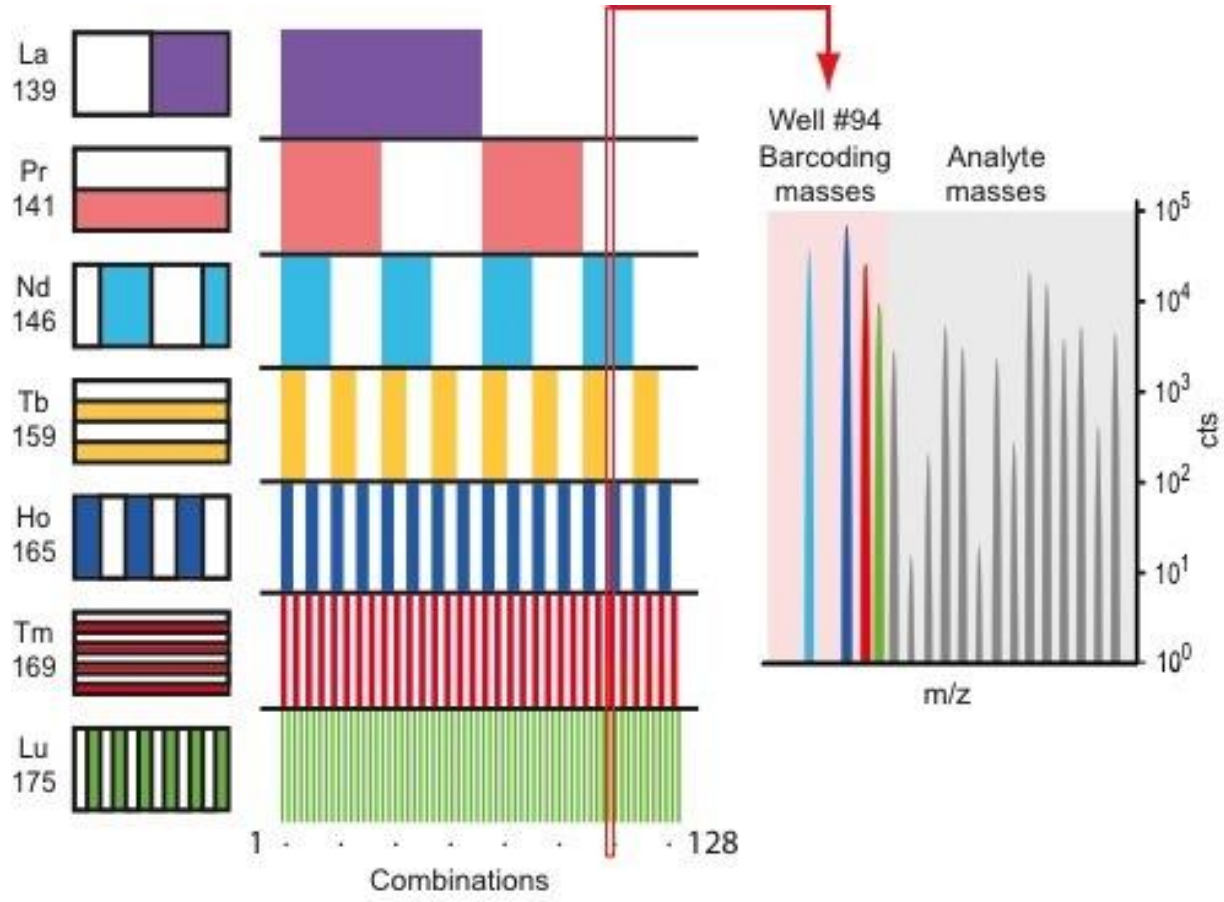
$2^3=8$

$2^4=16$

$2^5=32$

$2^6=64$

$2^7=128$



Bodenmiller, et. al., *Nature Biotech*., 2012

Before Analysis

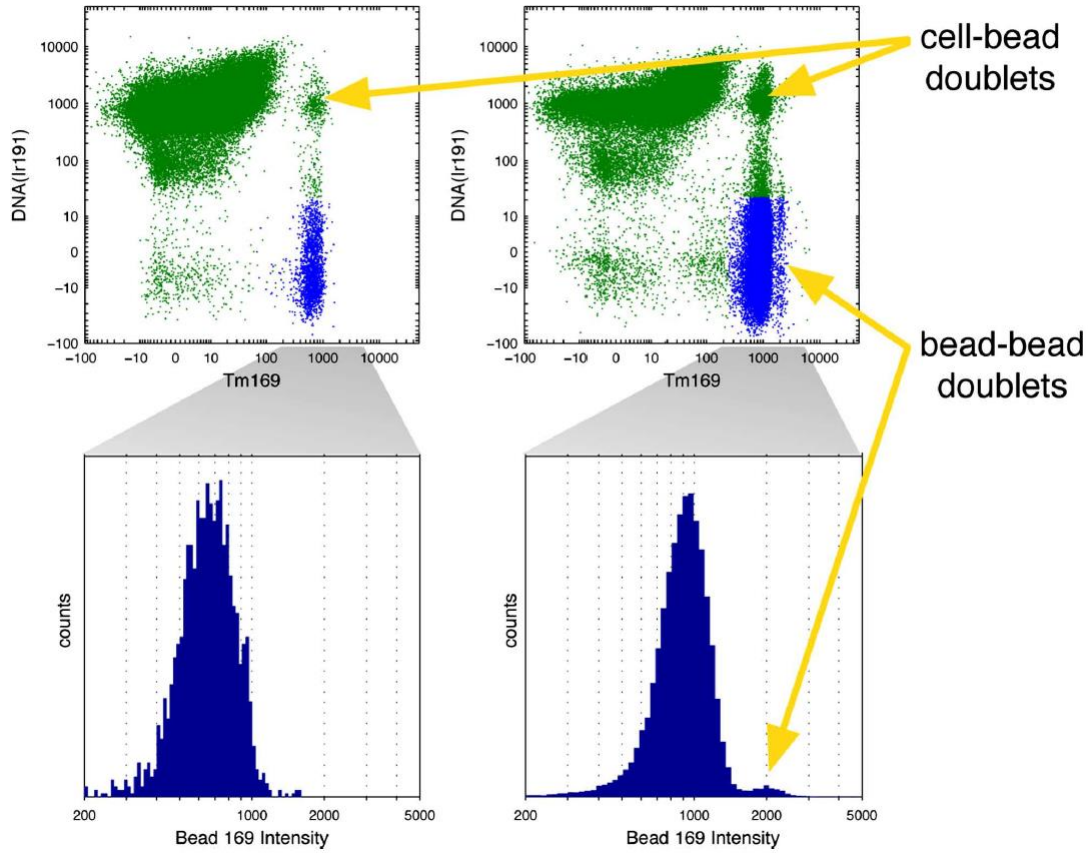
1. Doublet filtering
2. Bead normalization
3. Debarcoding (if needed)
4. Randomization (visualization)
5. Transformation(arcsinh)
6. Gating

Bead doublets

Events that result from a bead combined with

- a cell (cell-bead doublets) or
- another bead (bead-bead doublets).

To find bead doublets make a biaxial plot of the bead channel (x-axis) and of the DNA channel (y-axis) and create a bead gate.



Finck, et. al., Cytometry A, 2013

Normalization of Mass Cytometry Data

Goal: Reliably compare mass cytometry data across patients, conditions, tissues, etc.

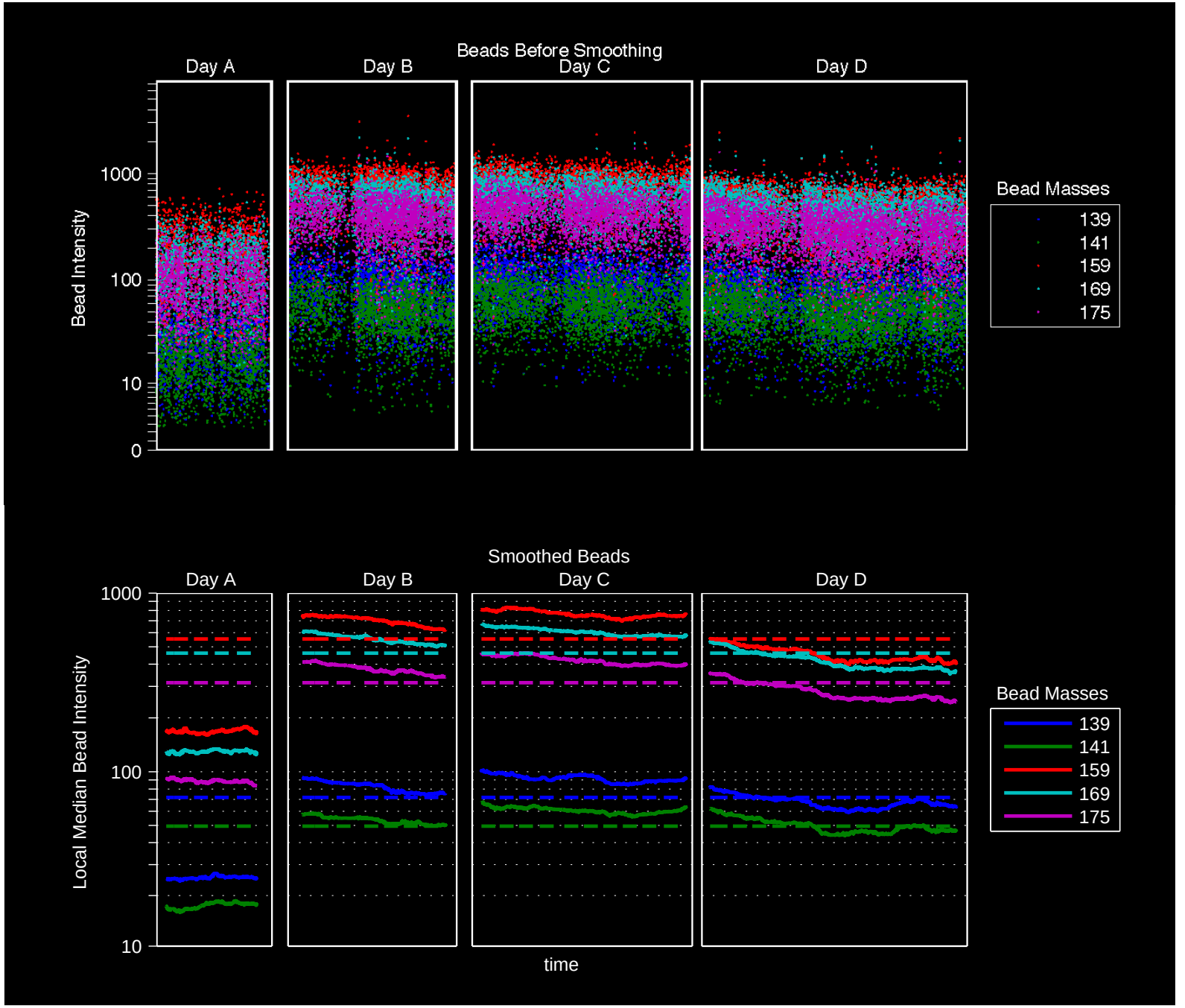
Problem: Drifts in mass cytometry instrument sensitivity over time due to cellular debris, fluctuations in plasma temperature, and calibrations.

Solution: Normalization using internal bead standards measured concurrently with cell samples.

Adapted from R. Finck, http://web.stanford.edu/class/immunol206a/

Normalization procedure: smoothing

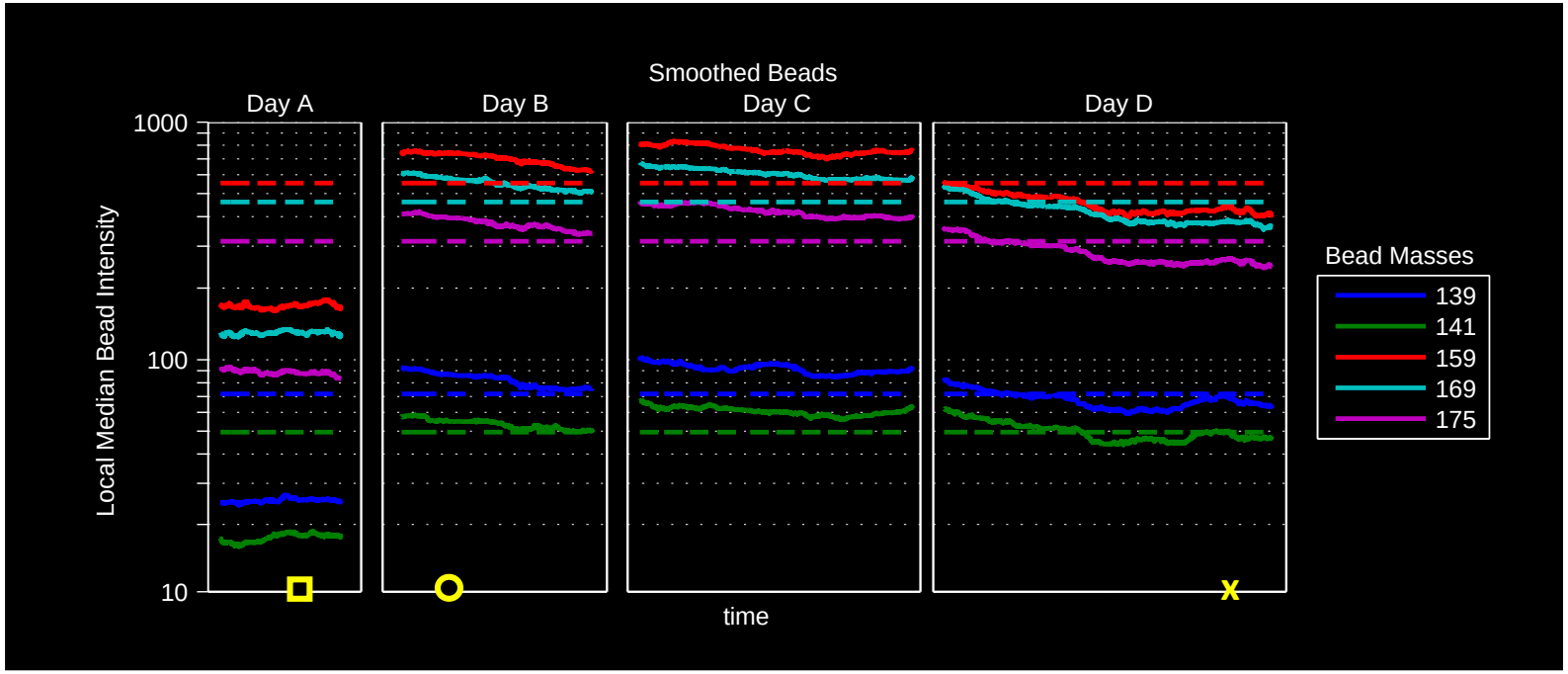
- Bead smoothing removes local variance (in a single experiment)
- use the median of a sliding window of 500 bead-associated events

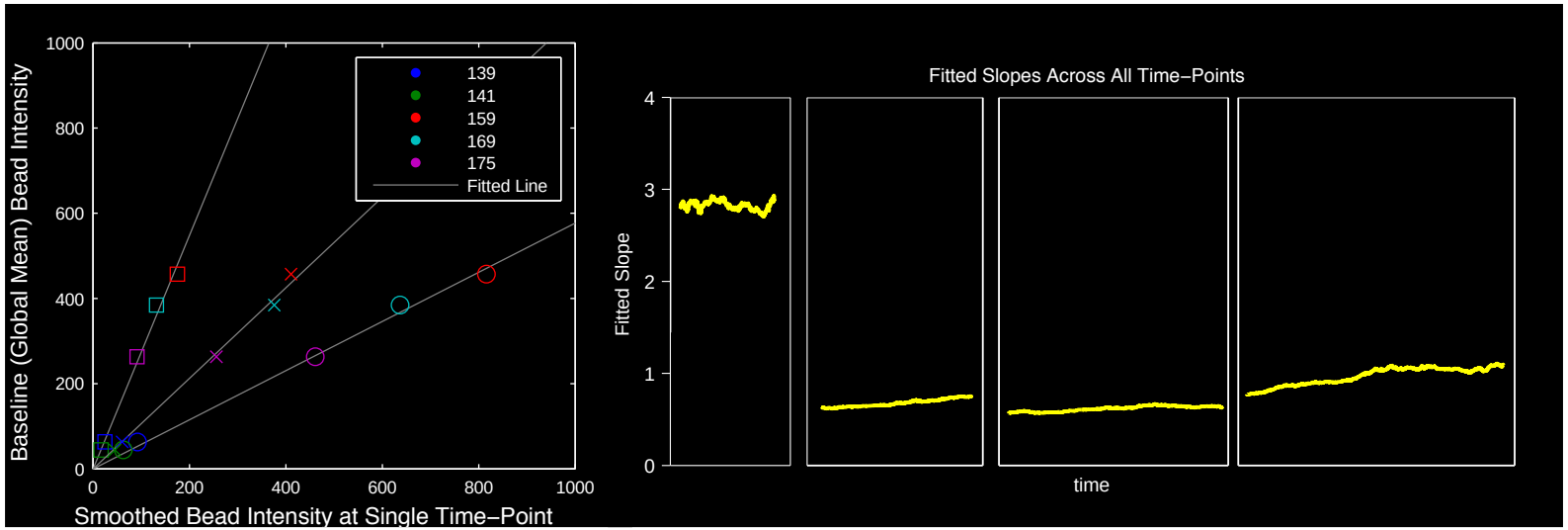


Adapted from R. Finck, http://web.stanford.edu/class/immunol206a/

Normalization: slope correction

- Fitted Slopes Define a Correction Function
- Multiple days/experiments: use the slope of the line through the origin and the point of intersection of the bead intensity at every time point and the mean smoothed bead intensities across all experiments.

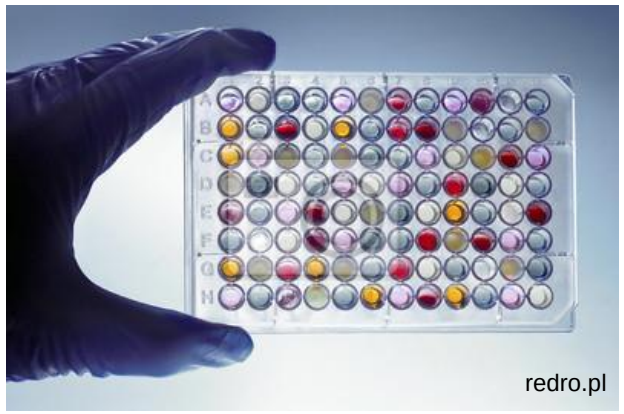


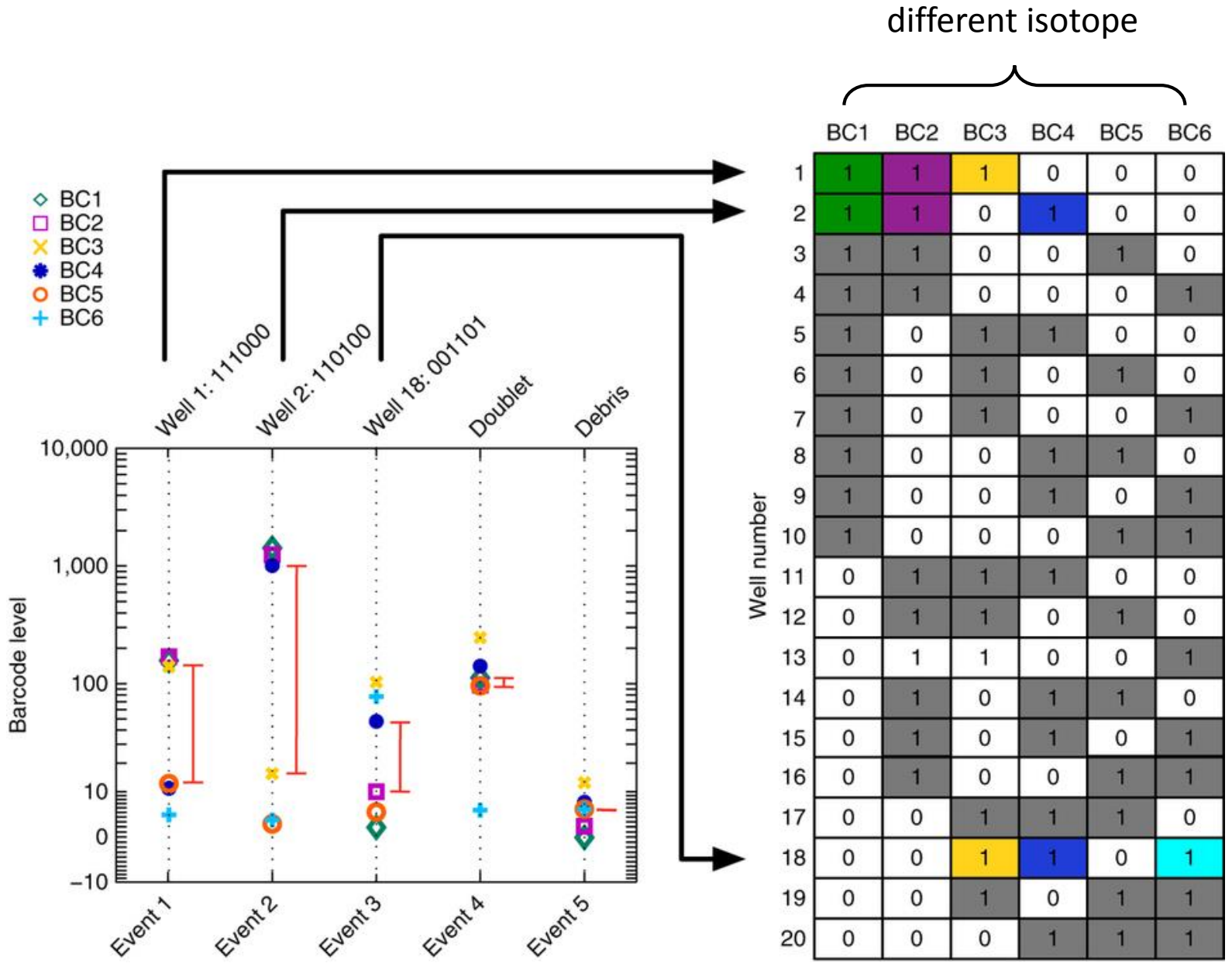


Adapted from R. Finck, http://web.stanford.edu/class/immunol206a/

Debarcoding

- Each perturbation experiment is placed in a separate well
- Each well receives a unique combination of barcodes



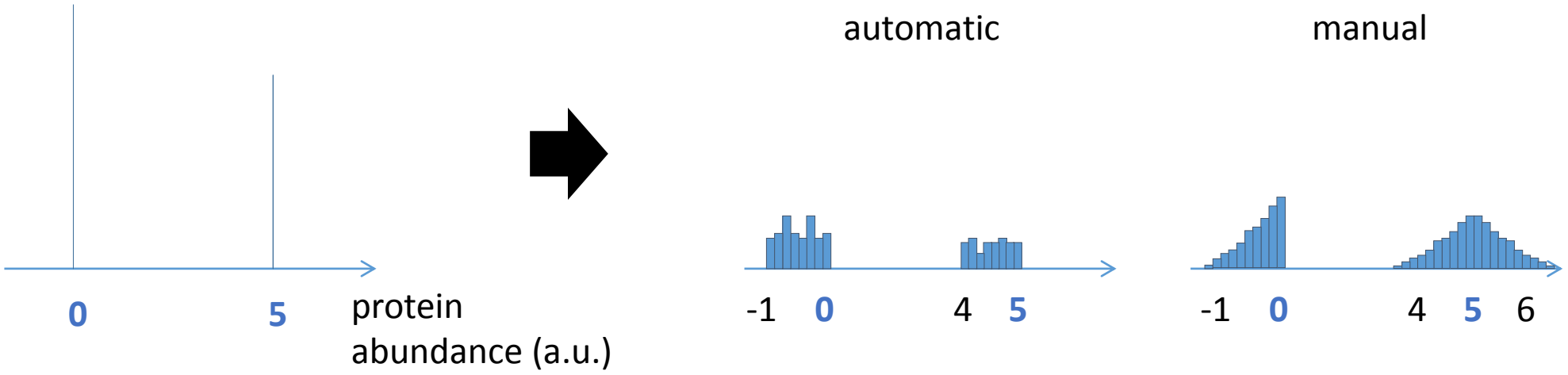


6-choose-3 MCB-multiplexing example, [Zunder et. al., Nat. Prot., 2015]

Randomization of integer values

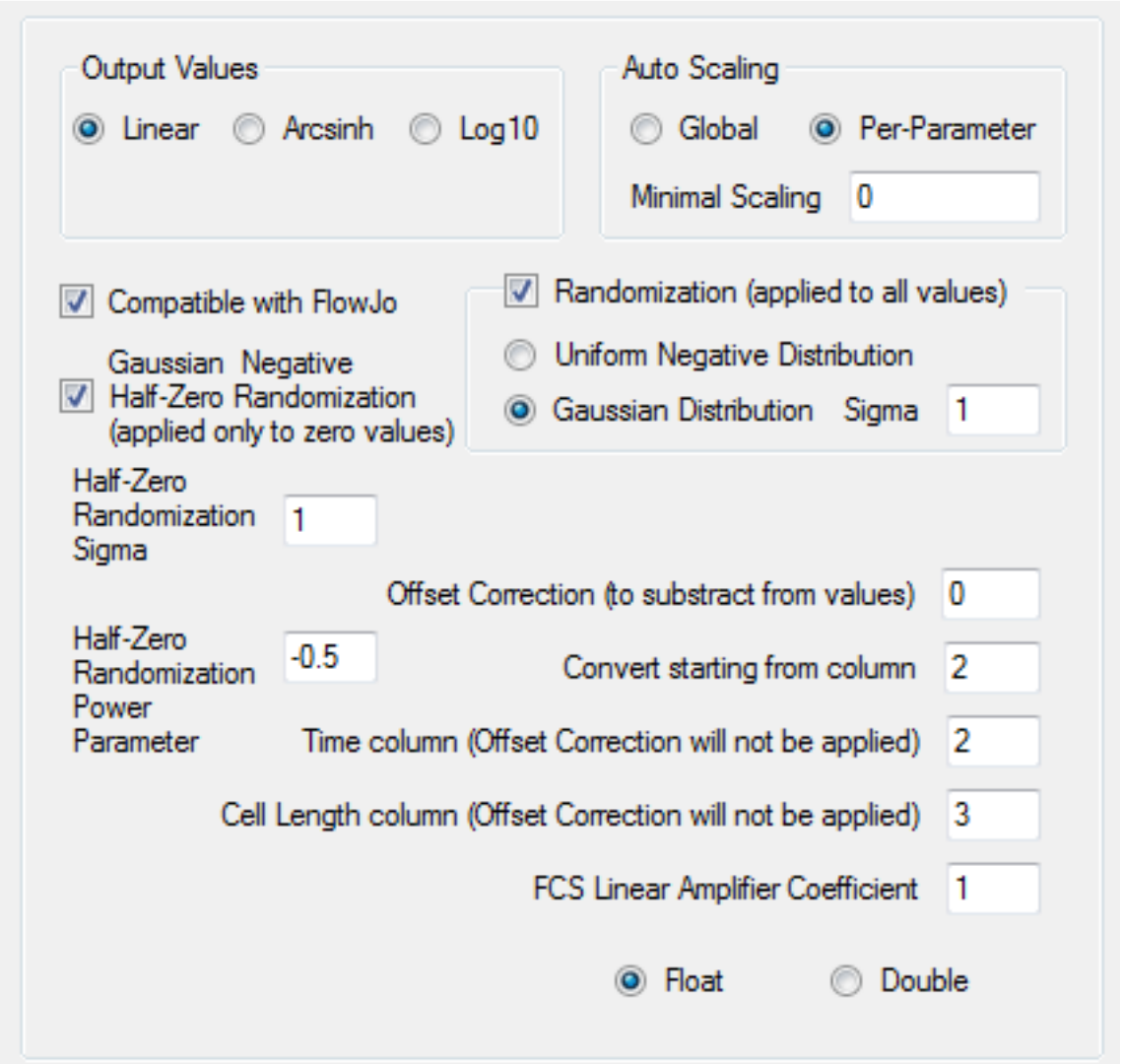
Avoid having large peaks (usually at zero) and create better scatterplot visualizations

- *Automatic*: using a negative uniform distribution
- *Manual*: add a Gaussian distributed random value (tunable variance)
  - (separate option for zero values): scatter using the negative half of a tunable Gaussian

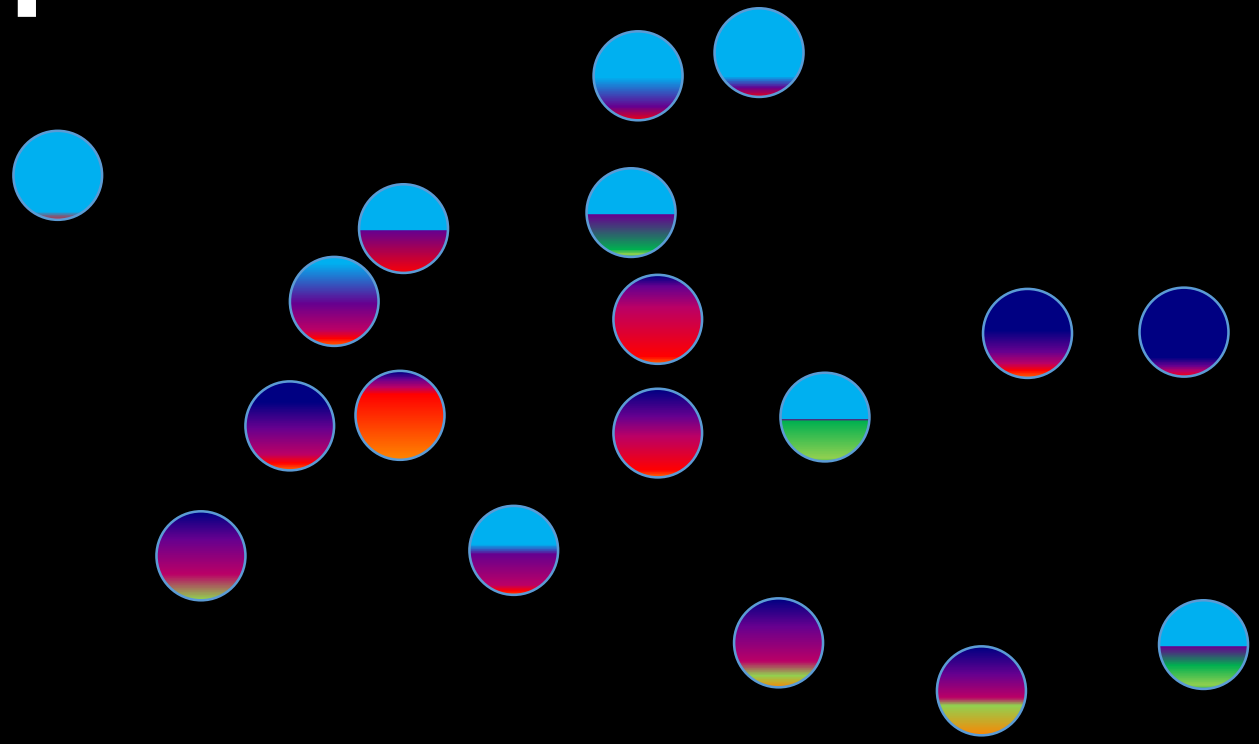


FCS conversion settings (CyTOF)

- Transformation
  - Linear, Arcsinh, Log10
- Scaling
- Randomization (only if Linear data)

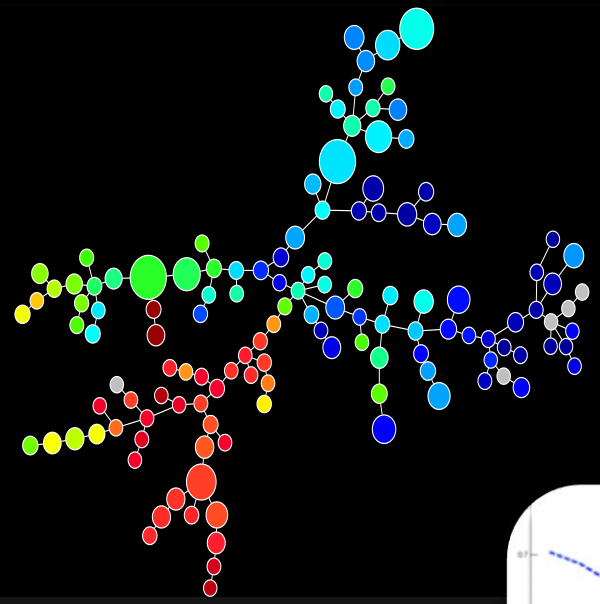


Can we create 2D maps representing higher dimensional data?



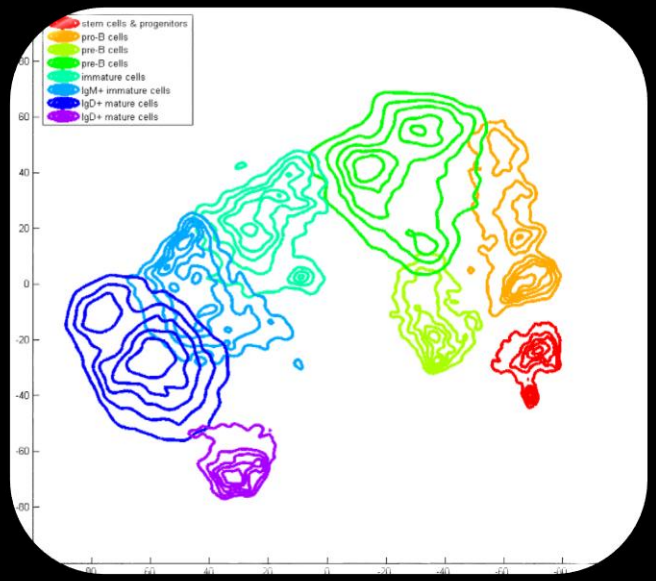
Adapted from S. Bendall, http://web.stanford.edu/class/immunol206a/

Possible Solutions



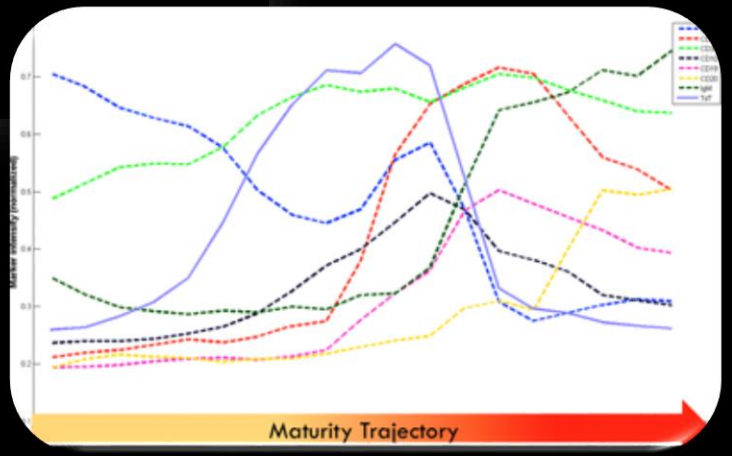
SPADE

SPADE
www.cytospade.org / www.cytobank.org
P. Qiu et al. *Nature Biotechnology, 2011*



ViSNE

ViSNE
Amir et al. *Nature Biotechnology, 2013*



Wanderlust

Wanderlust
Bendall et al. *Cell, 2014*

Adapted from S. Bendall, http://web.stanford.edu/class/immunol206a/

Causalpath


CAUSALPATH
Next Generation
Causal Analysis
... Inspired by the Induction
of Biological Pathways from
Cytometry Data

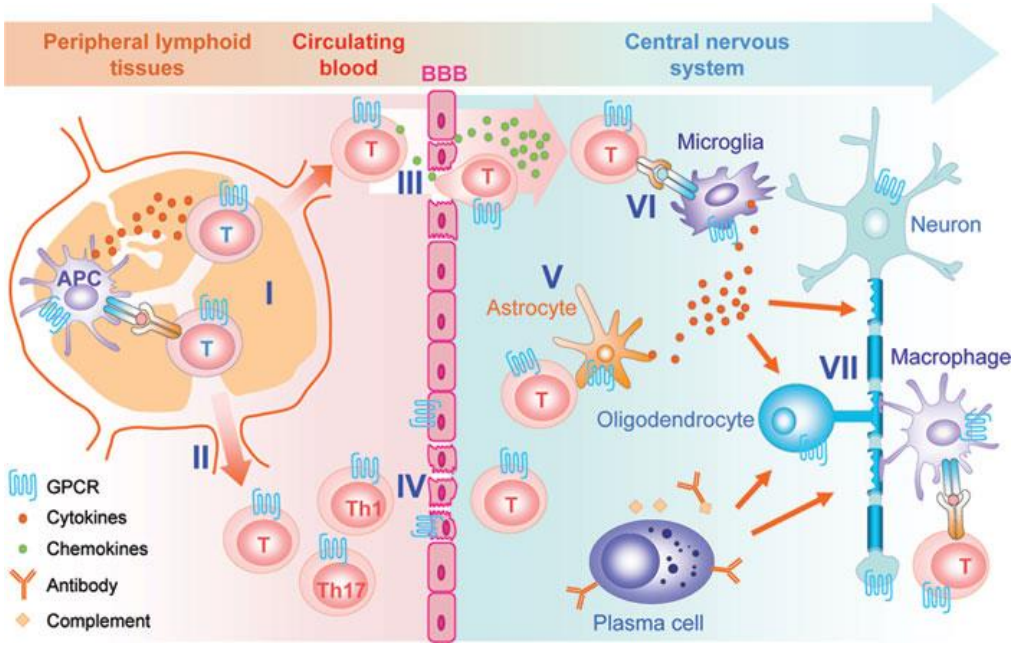

KAROLINSKA INSTITUTET
ANNO 1810
Karolinska
Institutet


ΕΥΡΩΠΗ
University of Crete
Computer Science Department


erc
European Research Council
Established by the European Commision

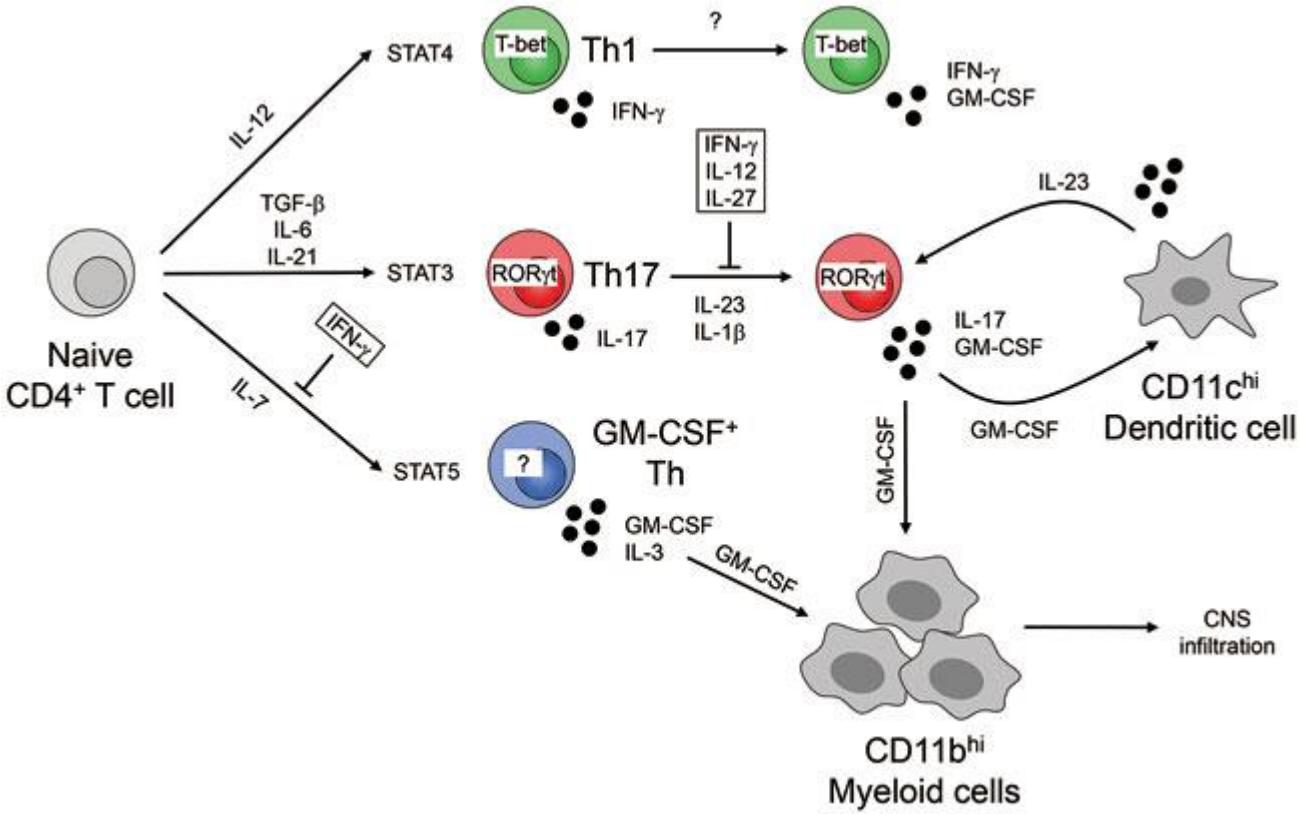
Biological Background

- Multiple Sclerosis (MS)
  - naïve T cells become T-helper (Th) cells in the blood
  - Pathogenic Th cells attack the nervous system
  - Secrete **cytokines** damaging brain cells



Du and Xie, *Cell Research* (2012), doi:10.1038/cr.2012.87

From Naïve CD4+ T cells to GM-CSF+ cells



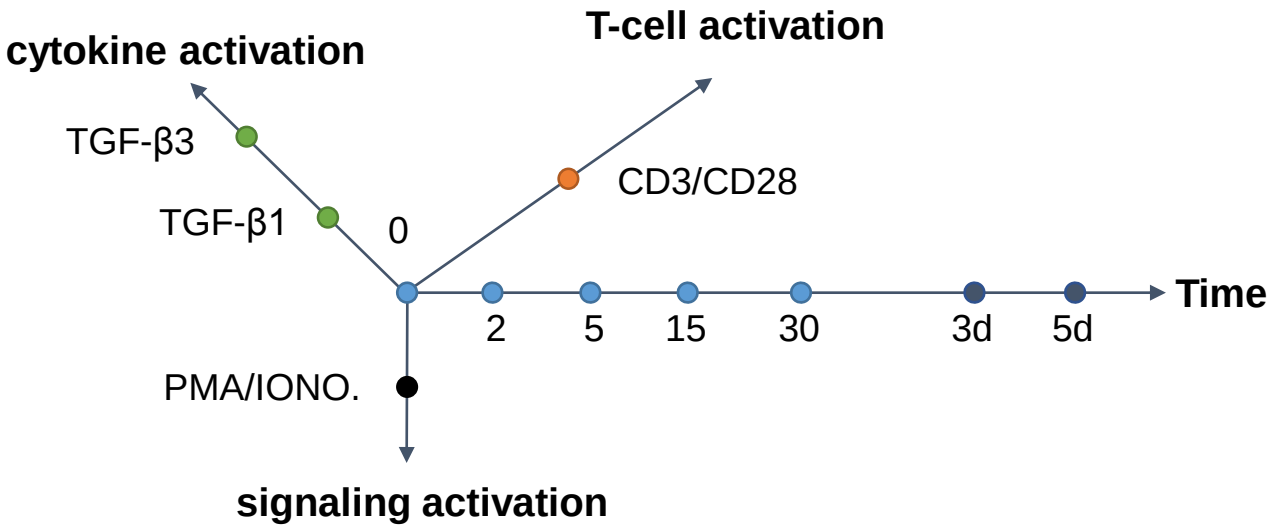
Herndler-Brandstetter, *Cell Research,* 2014

Biological questions

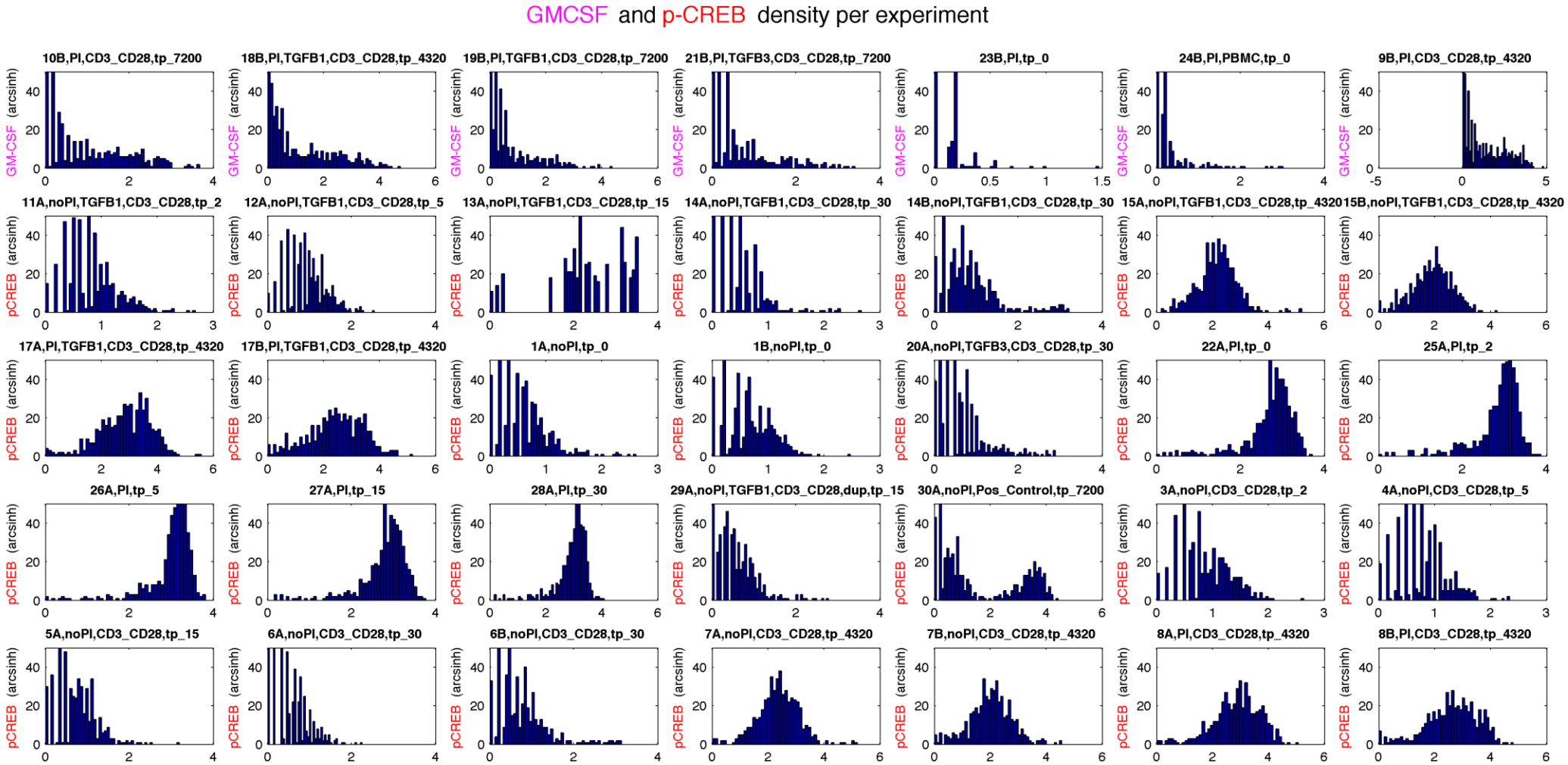
- How human CD4+ T cells differentiate to become GM-CSF+ cells?
  - Under which cytokine signals?
  - Under what stimulation conditions (signaling pathways involved)?
- Which are the cell characteristics? (co-expression of other T-cell markers (CD))
- Which are the Th subsets (Th1, Th2, Th17) present?

Experimental Design

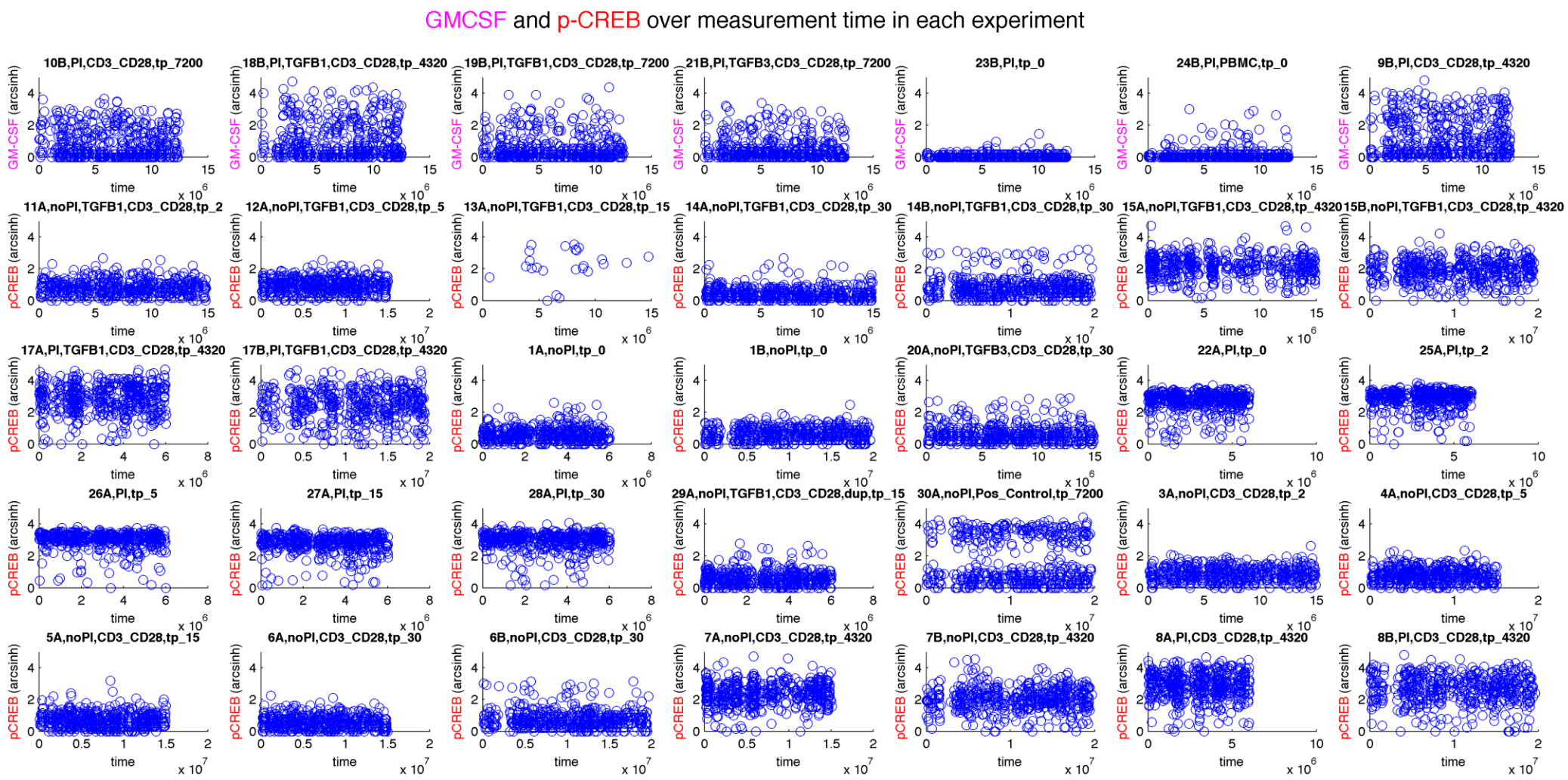
- 2 donors
- 34 distinct experiments (>100K cells/exp.)
- 48 protein markers: 27 surface, 11 signaling, 10 cytokine



Data example (density)



Data example (acquisition time vs abundance)



Thank you !