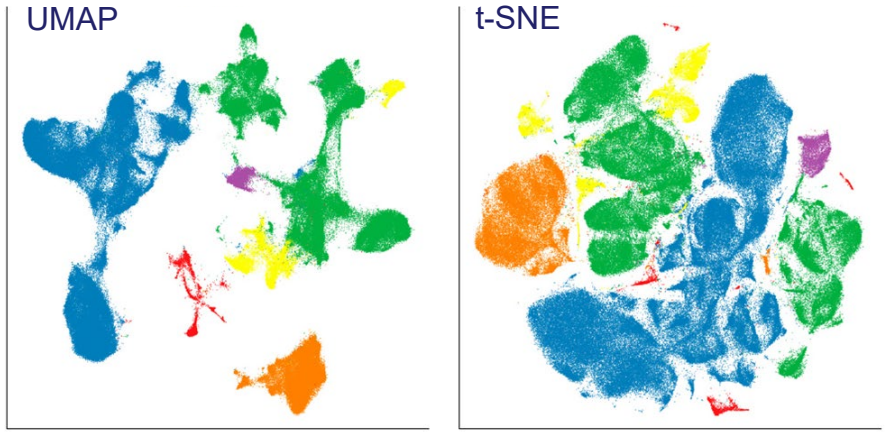
Analysis I: Getting Started with Cytometry Data Science (90 min)

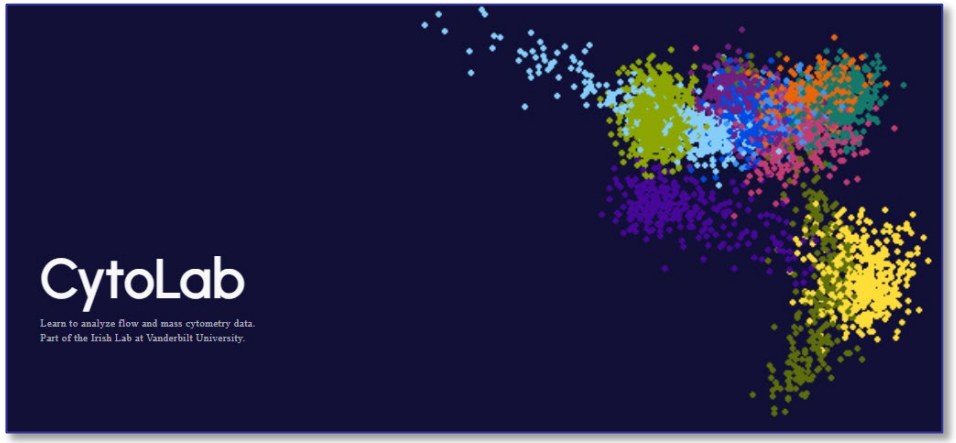
Starting Poll: https://www.menti.com/

Poll Code: 5033 7474

Analysis I: Getting Started with Cytometry Data Science



Becht et al. 2018



Course slides & webapps on CytoLab: https://cytolab.github.io/

Jonathan Irish

Associate Professor
Cell & Developmental Biology
Vanderbilt University

jonathan.irish@vanderbilt.edu

Cass Mayeda

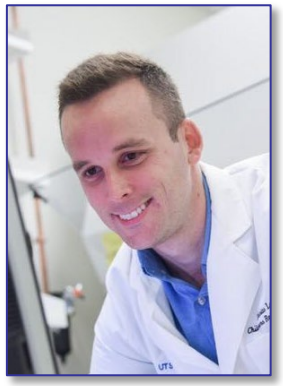
Web Applications
Research Assistant
Vanderbilt University

cass.mayeda@vanderbilt.edu

Josef Spidlen

Senior Director R&D
BD Biosciences

josef.spidlen@bd.com



Nicolas Loof

Senior Application Scientist
Multi-omics Specialist
BD Biosciences

nicolas.loof@bd.com

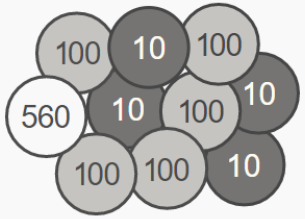
Thinking Multidimensionally for Single Cell Cytometry

We'll Start with the Single Cell Biology Webapp at CytoLab

Introduction to

Cytometry

Looking for a crashcourse in cytometry data analysis? You've come to the right place. Below you'll find a series of tutorial applications designed to introduce cytometry data analysis fundamentals for first-time users (or experienced scientists looking to brush up!).



Why Single Cell Biology?

Explore a comparison with western blots

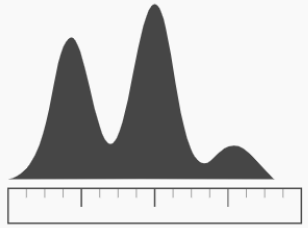
VISIT

1. Gate	Human
2. Reduce Dimensions	t-SNE
3. Cluster	FlowSOM

Cytometry Workflows

Steps in data analysis pipelines

COMING SOON



Scaling Matters

How to prepare your data

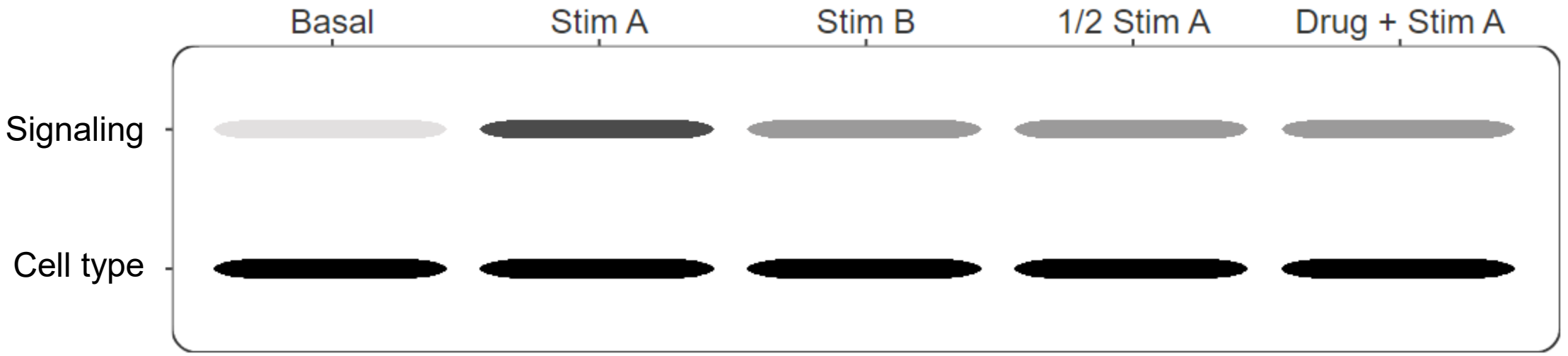
COMING SOON

Webapp: https://cytolab.shinyapps.io/SCBIO/

Reviews: Doxie & Irish, *Current Topics in Microbiology & Immunology* 2014; Krutzik et al., *Clinical Immunology* 2004

Single Cell Biology Asks: Which Cell, How Much?

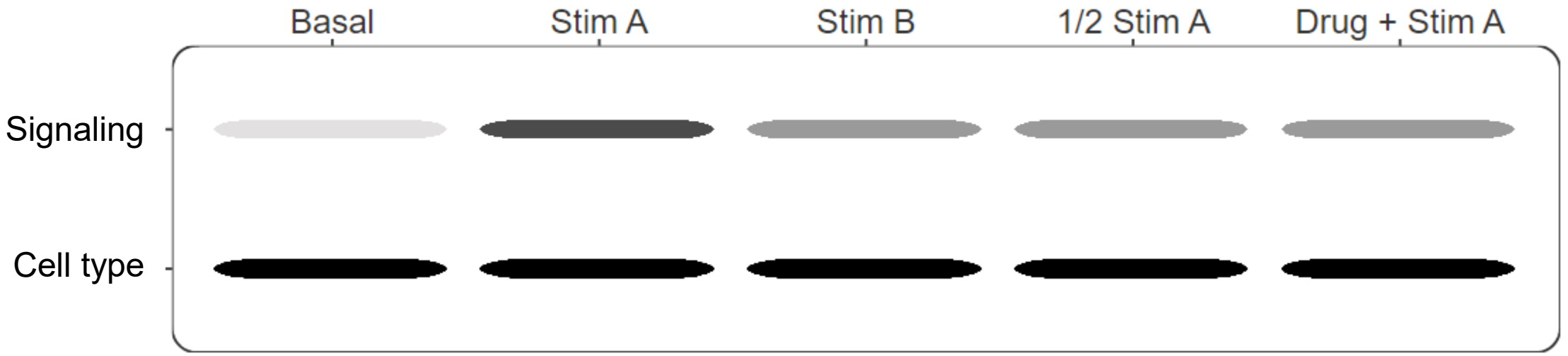
Imagine Western Blot results for 2 proteins across 5 signaling conditions…



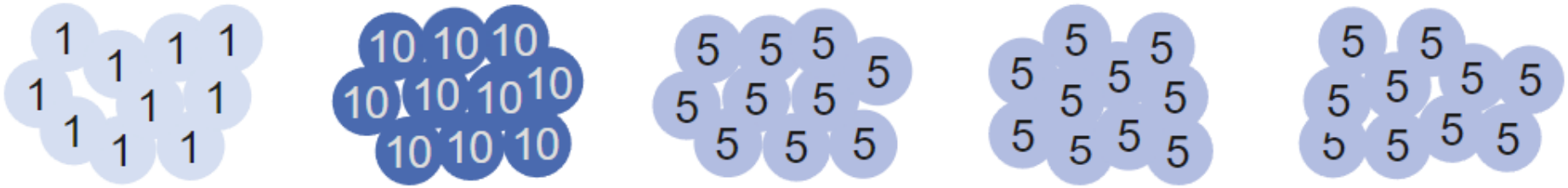
Webapp: https://cytolab.shinyapps.io/SCBIO/

Reviews: Doxie & Irish, *Current Topics in Microbiology & Immunology* 2014; Krutzik et al., *Clinical Immunology* 2004

Single Cell Biology Asks: Which Cell, How Much?



Could this level of signaling protein per cell match the Western?

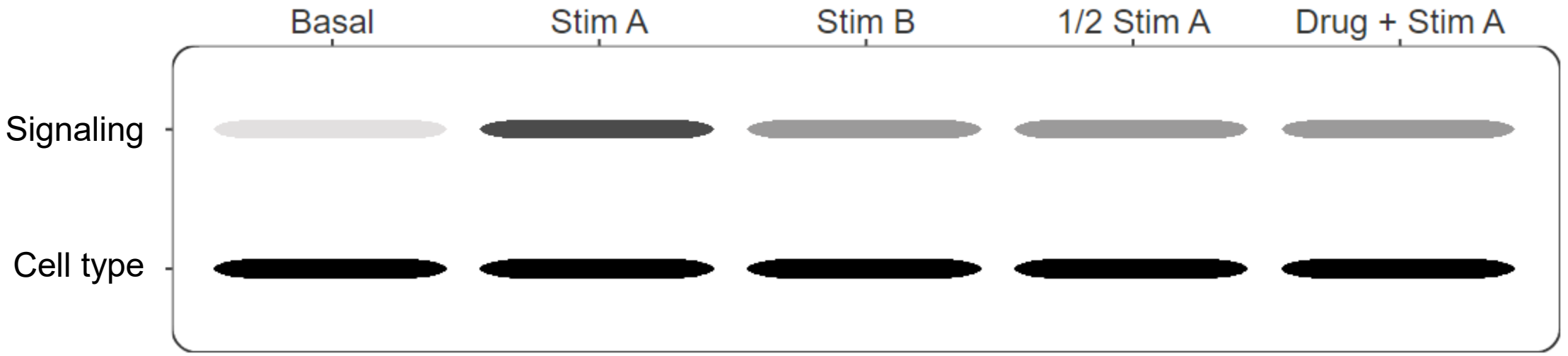


Sum:	10	100	50	50	50
Average:	1	10	5	5	5

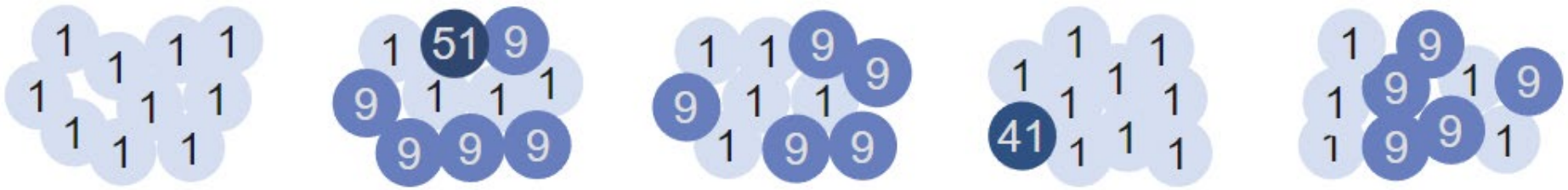
Webapp: https://cytolab.shinyapps.io/SCBIO/

Reviews: Doxie & Irish, *Current Topics in Microbiology & Immunology* 2014; Krutzik et al., *Clinical Immunology* 2004

Single Cell Biology Asks: Which Cell, How Much?



What about this level of signaling?



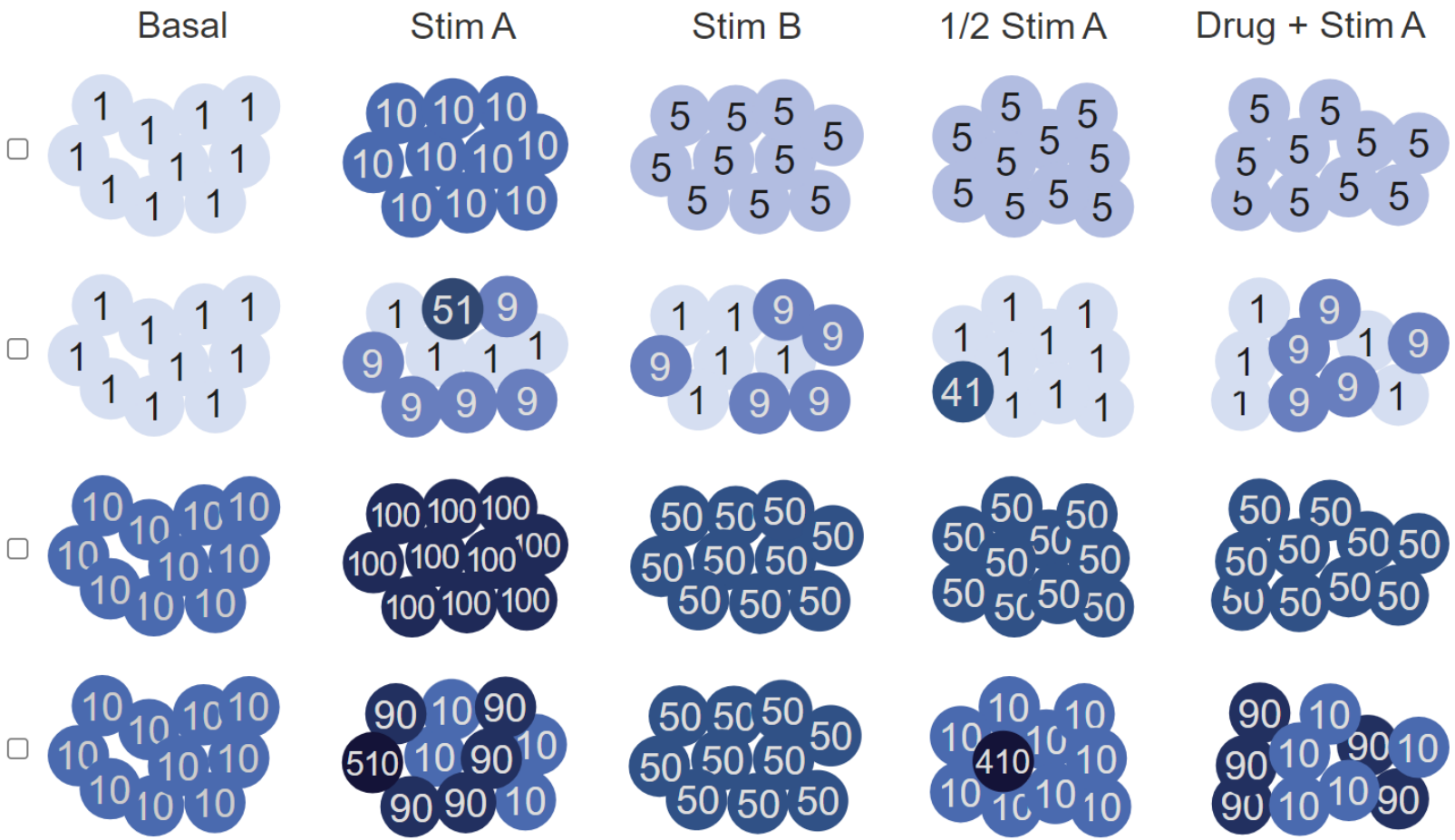
Sum:	10	100	50	50	50
Average:	1	10	5	5	5

Webapp: https://cytolab.shinyapps.io/SCBIO/

Reviews: Doxie & Irish, *Current Topics in Microbiology & Immunology* 2014; Krutzik et al., *Clinical Immunology* 2004

Are Any of The Columns Different at the Sample Level?

Based on the simulated western blot above, and assuming different potential exposure times, select all of the following that could represent measurements of signal at the cell level.

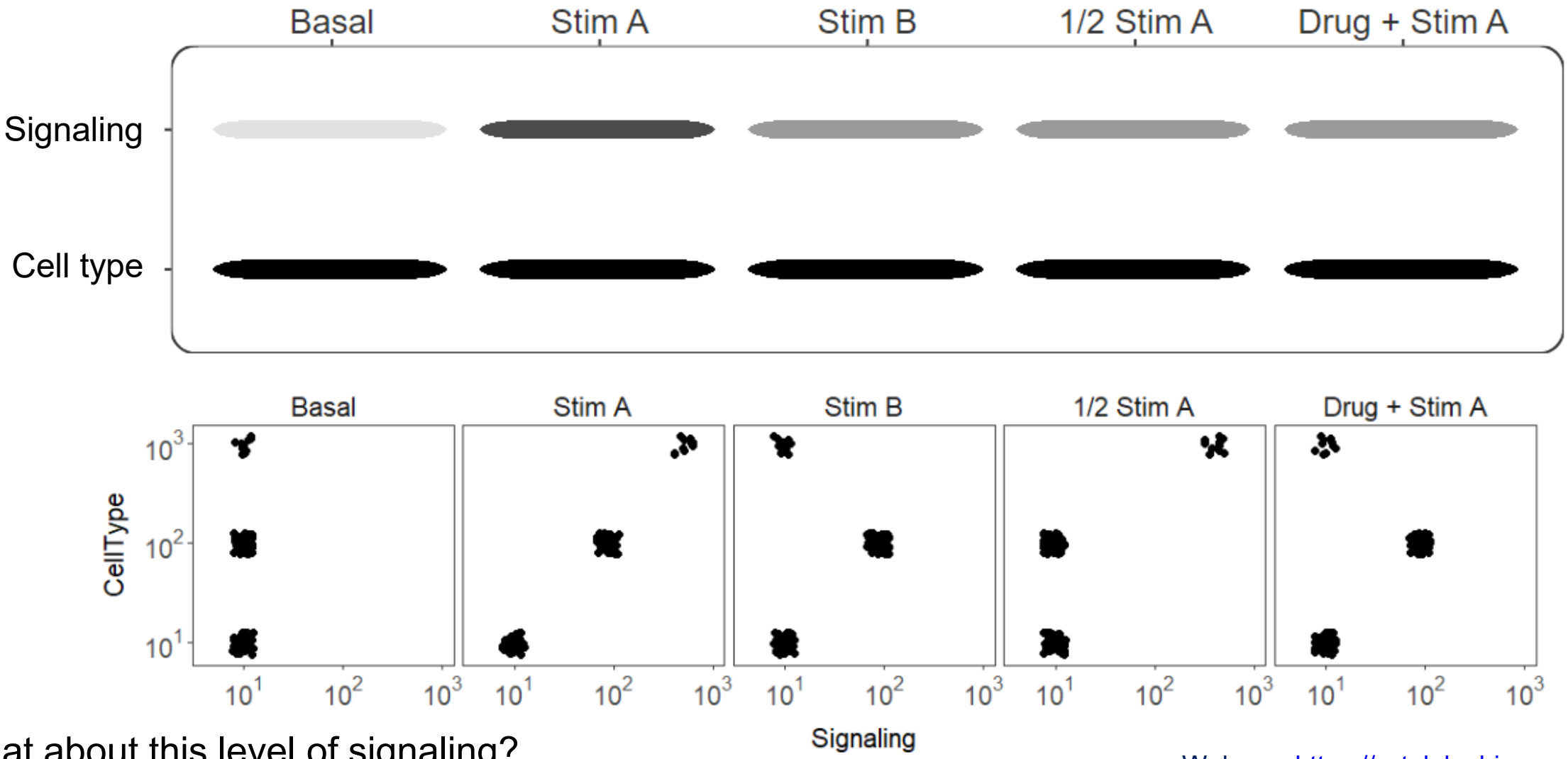


Submit answers

Webapp: https://cytolab.shinyapps.io/SCBIO/

Reviews: Doxie & Irish, *Current Topics in Microbiology & Immunology* 2014; Krutzik et al., *Clinical Immunology* 2004

Another Advantage of Single Cell Biology Is Multidimensionality

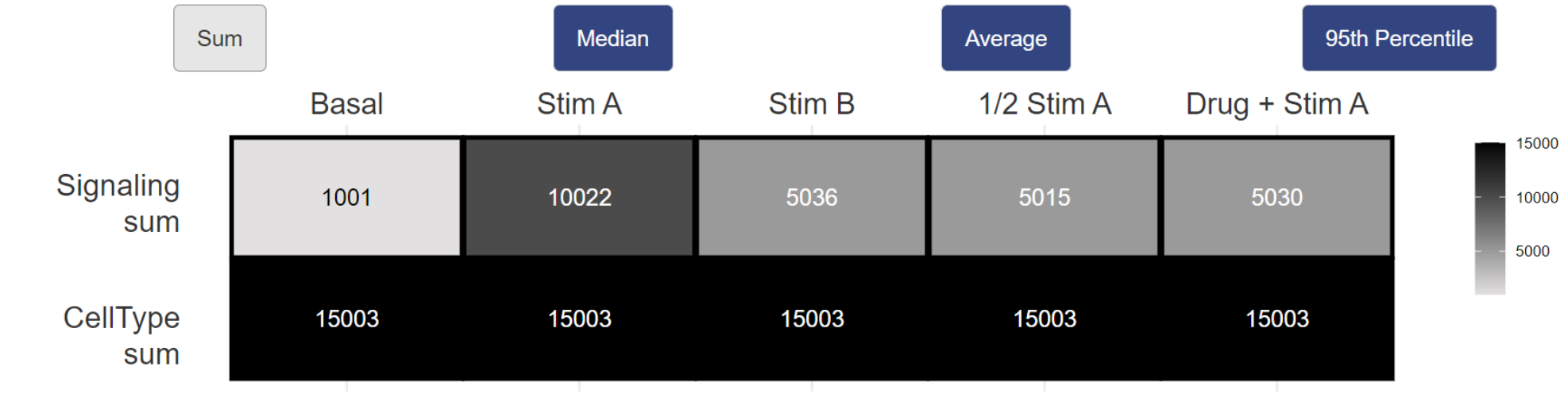


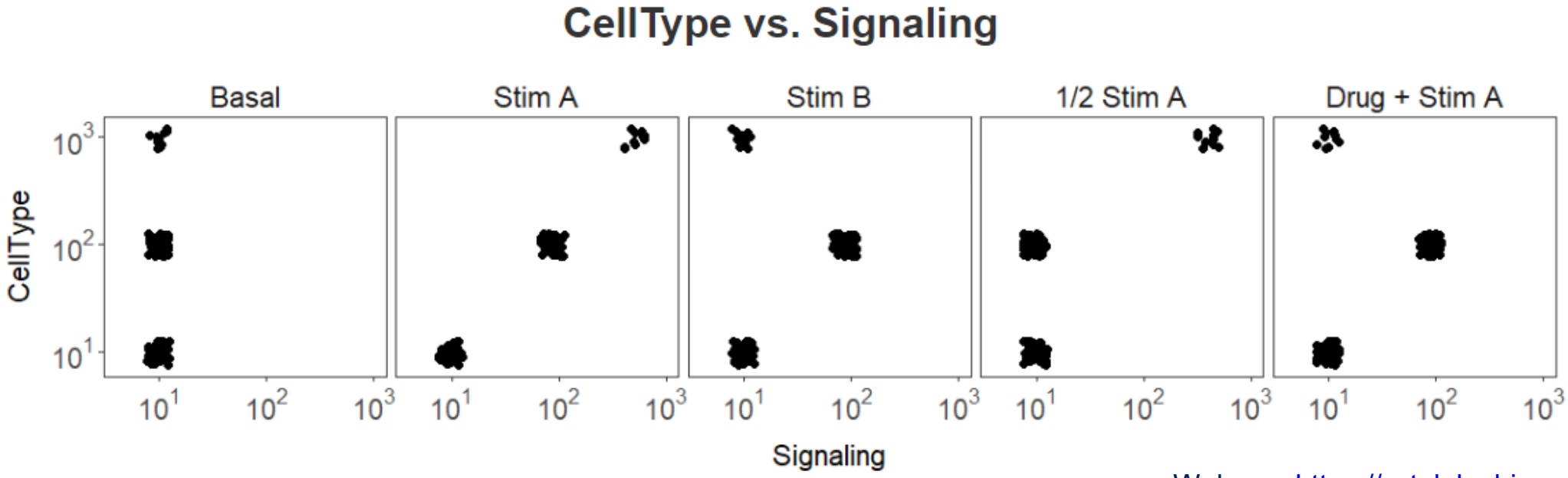
What about this level of signaling?

Webapp: https://cytolab.shinyapps.io/SCBIO/

Reviews: Doxie & Irish, *Current Topics in Microbiology & Immunology* 2014; Krutzik et al., *Clinical Immunology* 2004

Sums Can Be Shifted By Small Populations

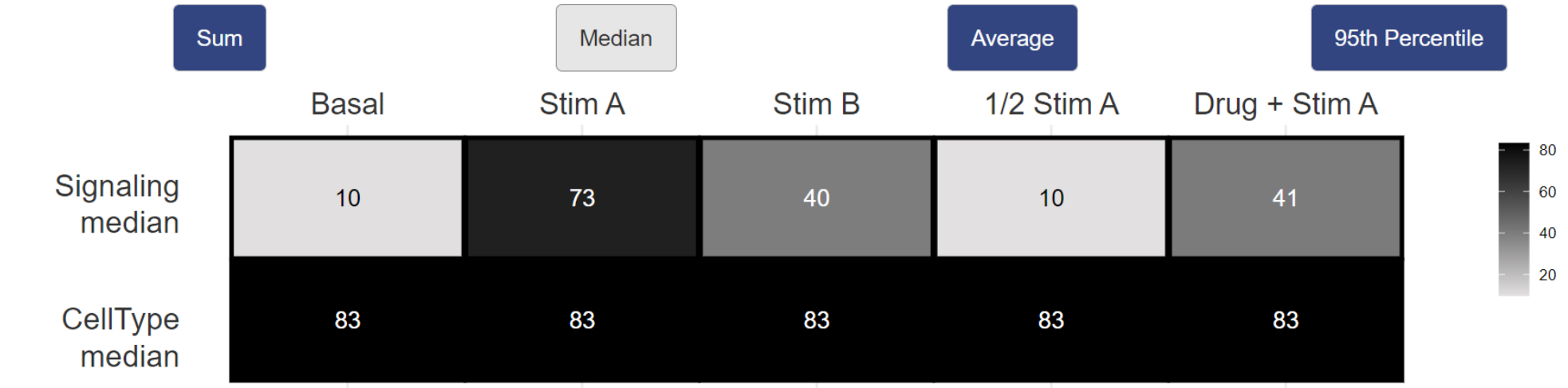


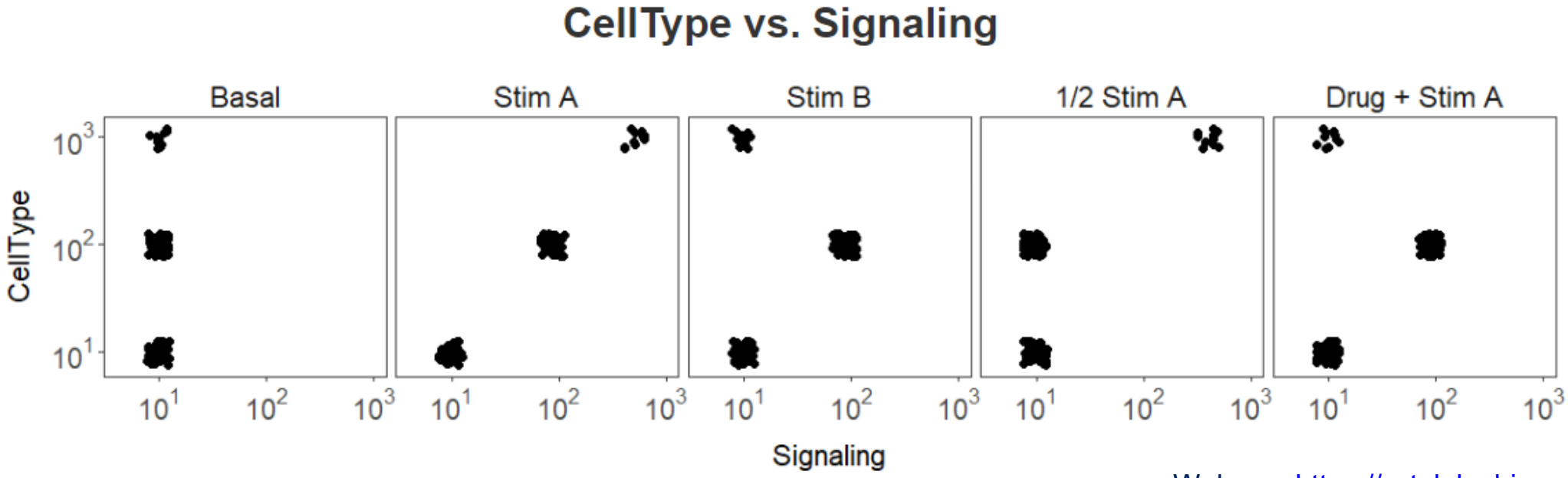


Webapp: https://cytolab.shinyapps.io/SCBIO/

Reviews: Doxie & Irish, *Current Topics in Microbiology & Immunology* 2014; Krutzik et al., *Clinical Immunology* 2004

Medians May Overlook Small Subsets

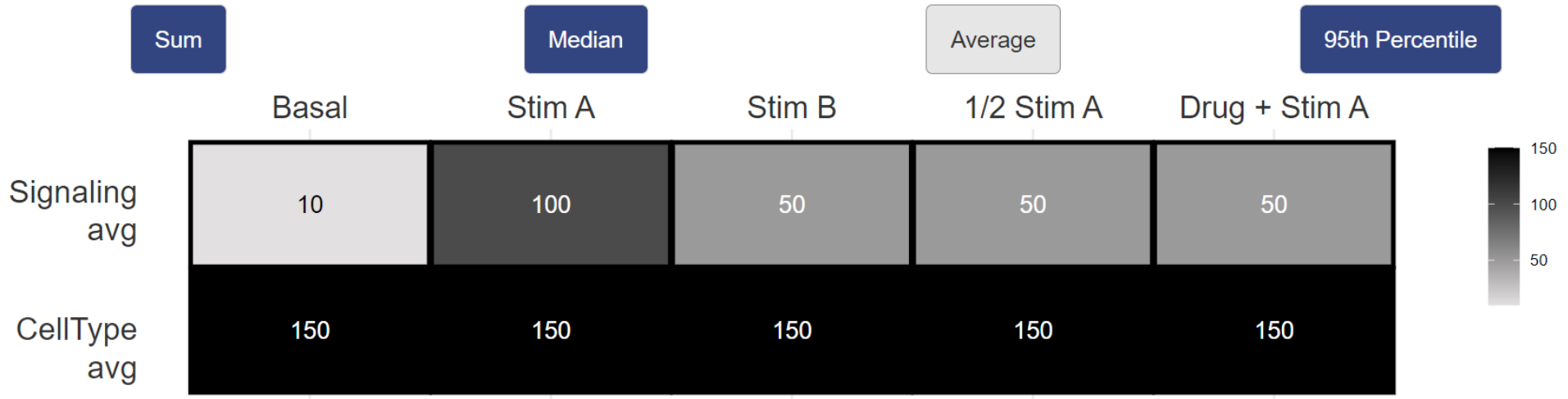


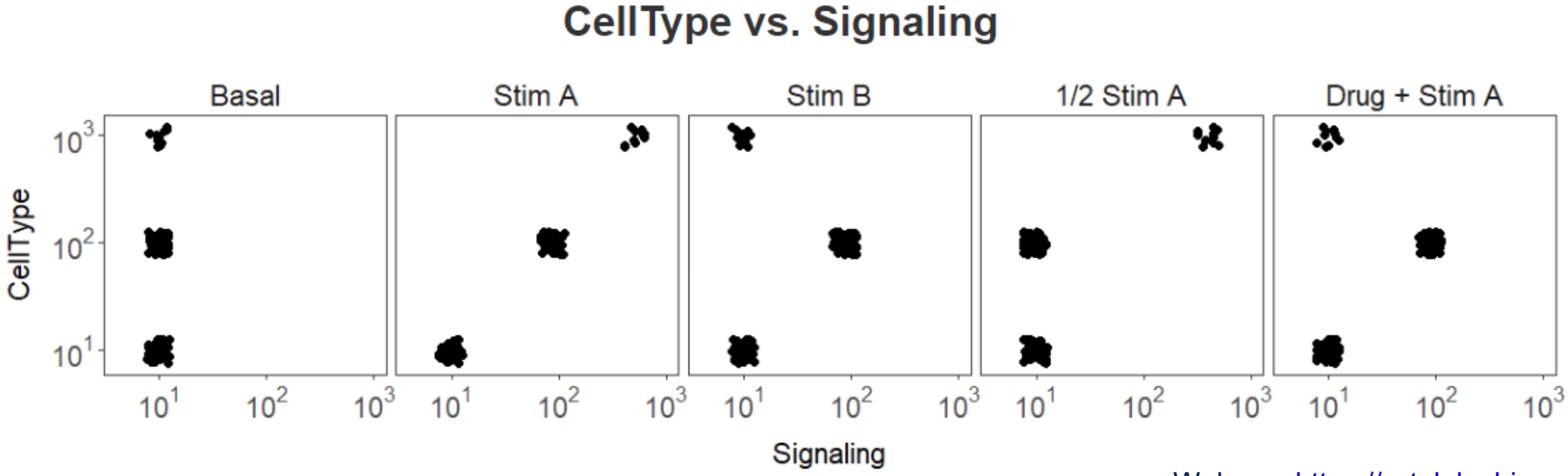


Webapp: https://cytolab.shinyapps.io/SCBIO/

Reviews: Doxie & Irish, *Current Topics in Microbiology & Immunology* 2014; Krutzik et al., *Clinical Immunology* 2004

Averages Are Impacted By Small Populations (Like Sums)

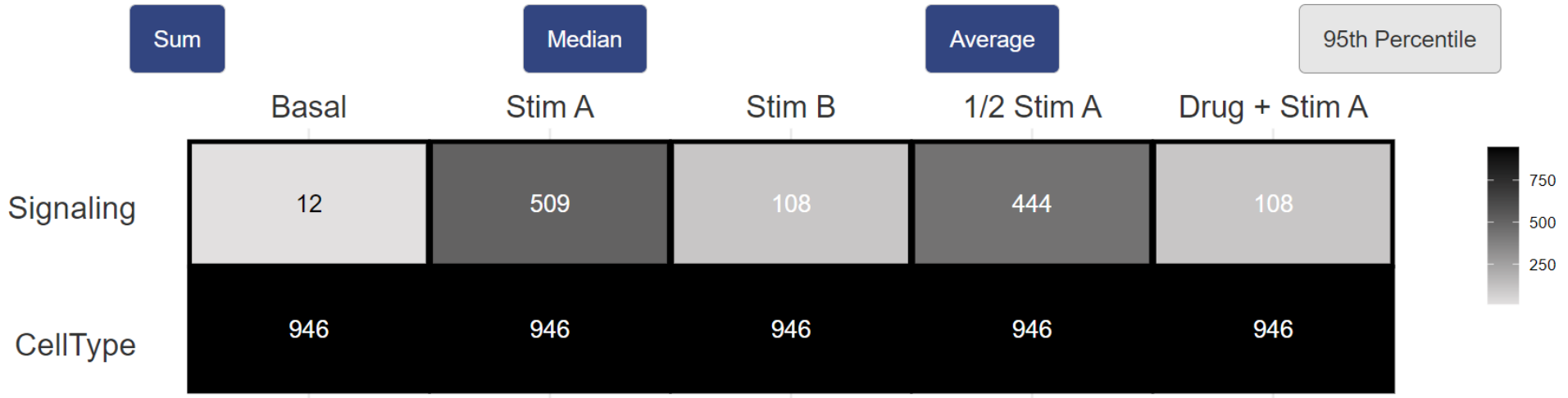


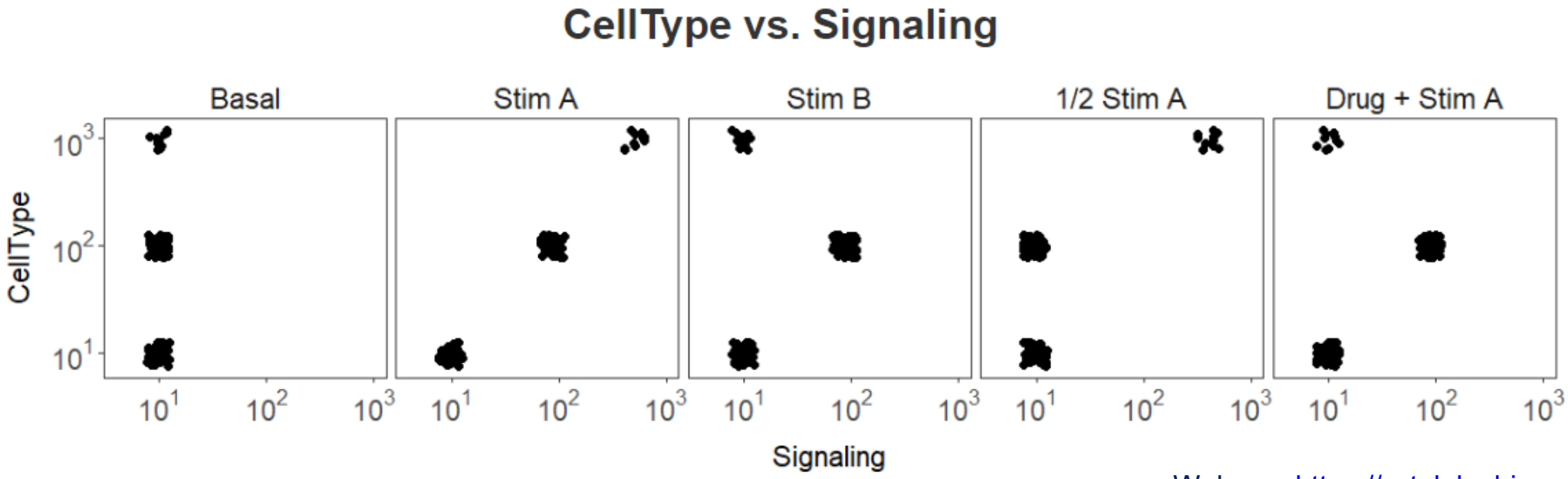


Webapp: https://cytolab.shinyapps.io/SCBIO/

Reviews: Doxie & Irish, *Current Topics in Microbiology & Immunology* 2014; Krutzik et al., *Clinical Immunology* 2004

The 95th Percentile Tracks the Leading Edge of a Distribution

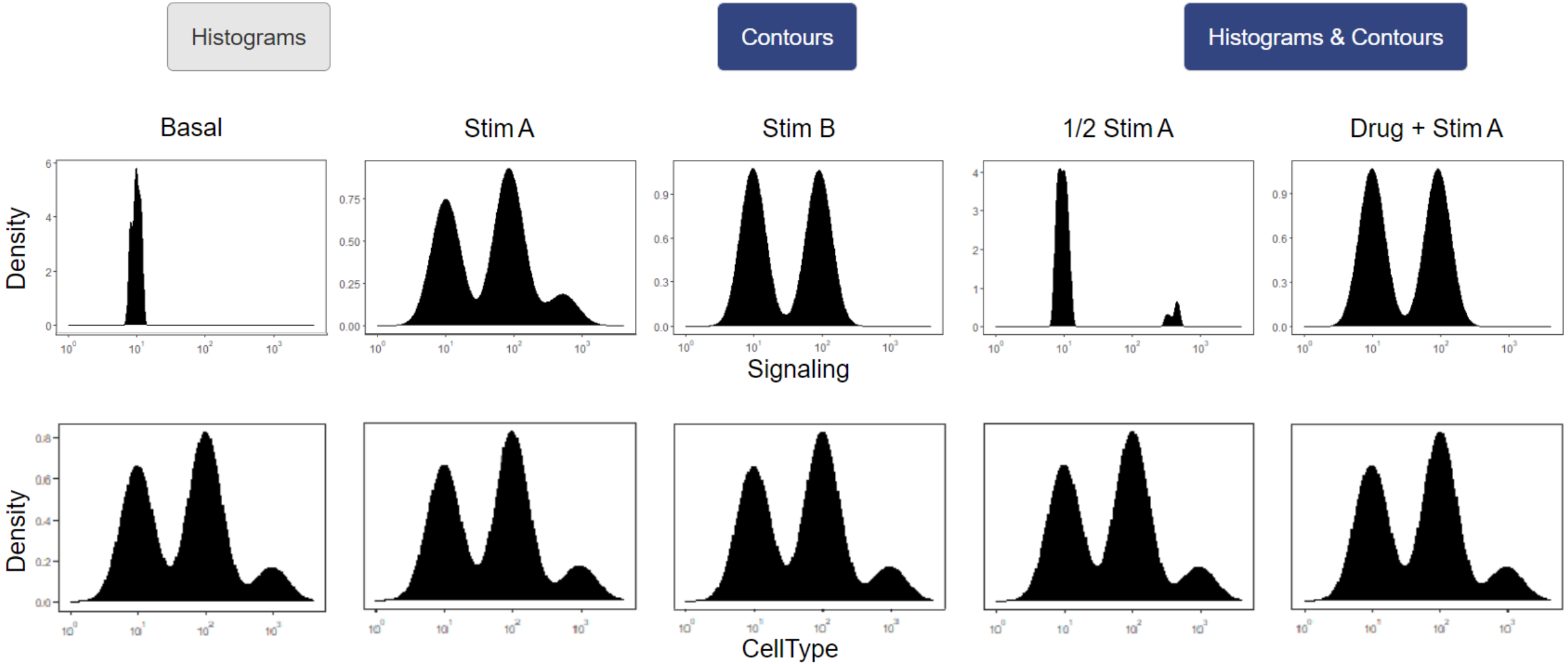




Webapp: https://cytolab.shinyapps.io/SCBIO/

Reviews: Doxie & Irish, *Current Topics in Microbiology & Immunology* 2014; Krutzik et al., *Clinical Immunology* 2004

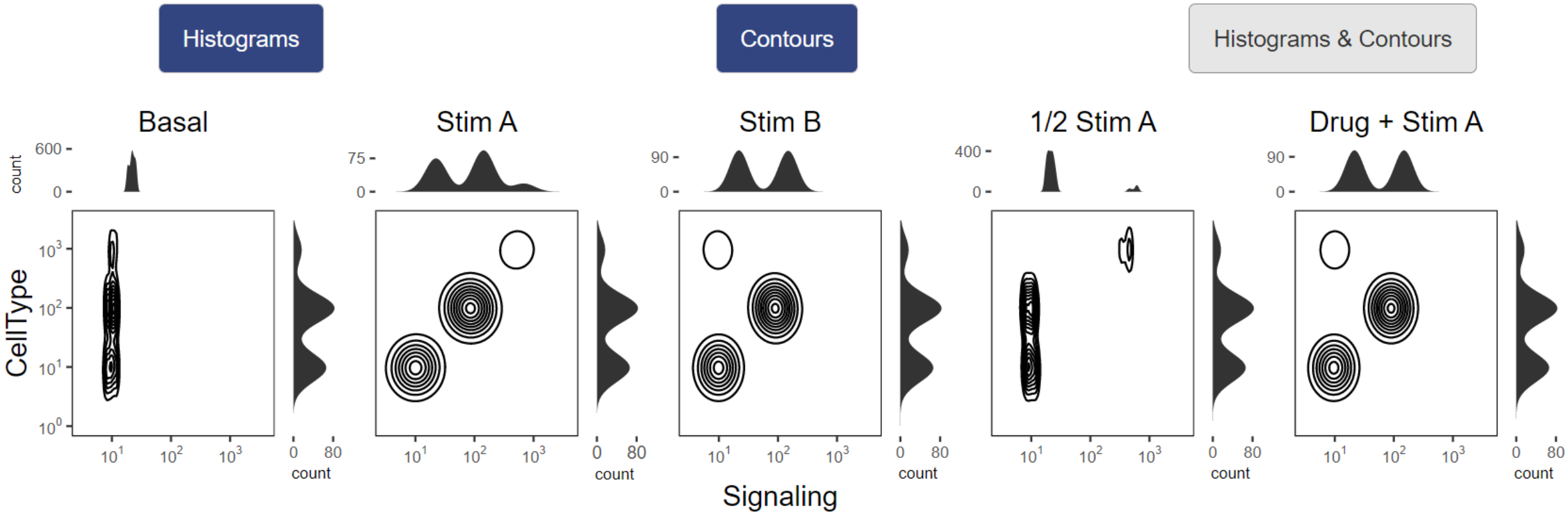
1D Histograms Emphasize Shifts in Each Measured Feature



Webapp: https://cytolab.shinyapps.io/SCBIO/

Reviews: Doxie & Irish, *Current Topics in Microbiology & Immunology* 2014; Krutzik et al., *Clinical Immunology* 2004

Use Different Analysis Tools to Test Data Science Hypotheses

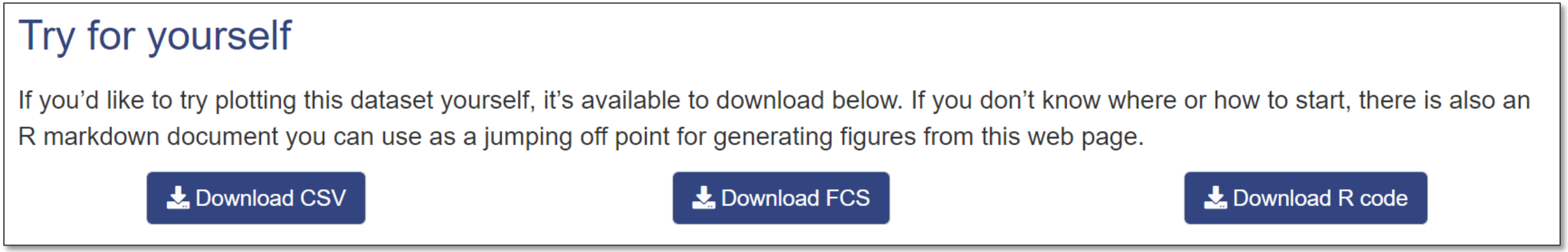


Example Hypothesis: In ½ Stim A, the small subset of cells that is at 1000 for Signaling is the subset of cells that is at 1000 Cell Type marker.

Webapp: https://cytolab.shinyapps.io/SCBIO/

Reviews: Doxie & Irish, *Current Topics in Microbiology & Immunology* 2014; Krutzik et al., *Clinical Immunology* 2004

Access the Data & View Across Analysis Platforms (CSV as Text Table, FCS File in Cytometry Software, Code in R)

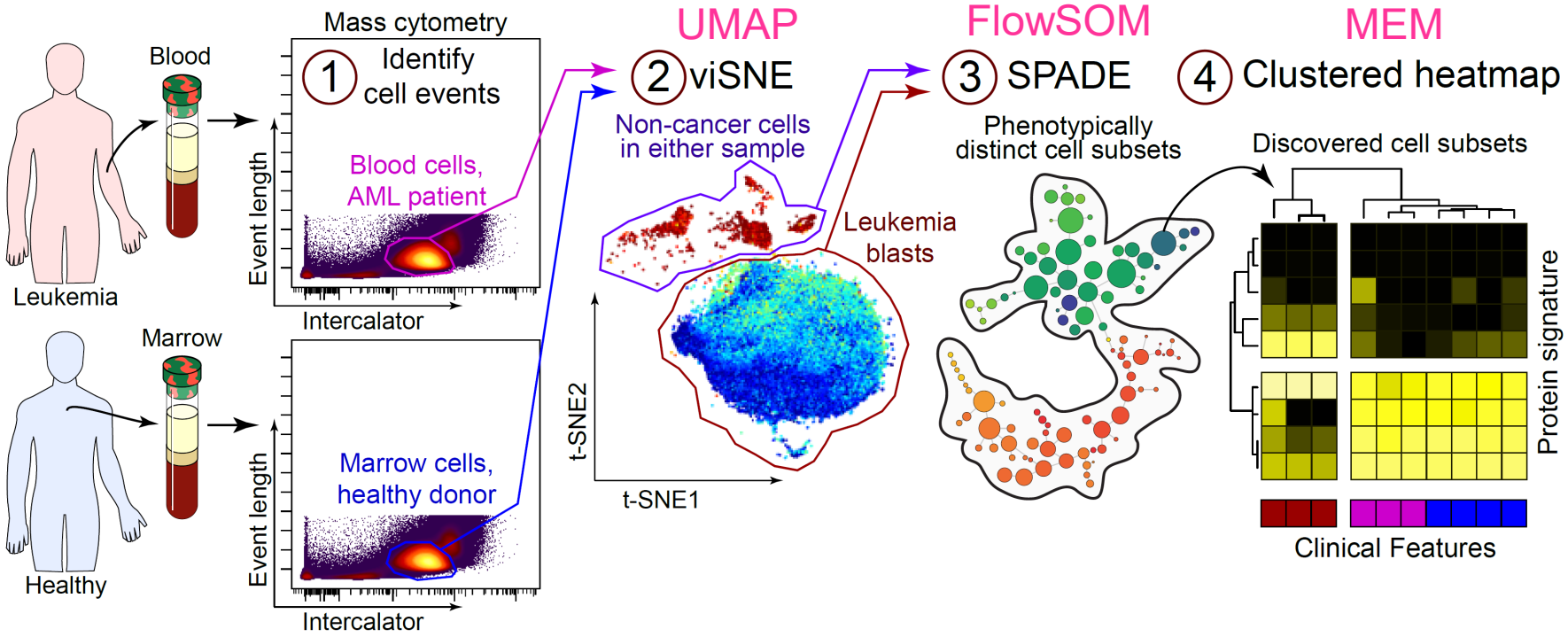


Webapp: https://cytolab.shinyapps.io/SCBIO/

Reviews: Doxie & Irish, *Current Topics in Microbiology & Immunology* 2014; Krutzik et al., *Clinical Immunology* 2004

‘Cyborg Learning’:
Humans & Machines Working Together

Machine Learning Is a Key Skillset for Biologists (And the Tools Are Rapidly Evolving)



Protocol: Diggins et al., *Current Protocols in Cytometry* 2017
Data files: https://flowrepository.org/id/FR-FCM-ZZKZ
More great tools: Saeys et al, *Nature Reviews Immunology* 2016

Original research:
Diggins et al., *Methods* 2015
Diggins et al., *Nature Methods* 2017

Typical workflow and goal: learn & label cytotypes (cell identities), reveal and assess unexpected & abnormal cells

Need: human reference data (more examples) with annotations

Which Parts Are Machines Good At?

Table 1 – A modular machine learning workflow for unsupervised high-dimensional single cell data analysis

Analysis step		Traditional	Additional methods§	Method here
Data collection	1) Panel design	Human expert	-	-
	2) Data collection	Human expert	-	-
Data processing	3) Cell event parsing	Instrument software	Bead normalization and event parsing [31]	-
	4) Scale transformation	Human expert	Logicle [36]	-
Distinguishing initial populations	5) Live single cell gating 6) Focal population gating	Biaxial gating + human expert	No event restriction, AutoGate [48]	viSNE + human expert (Figure 1)†
Revealing cell subsets	7) Select features	Human expert	Statistical threshold [40]	Human expert†
	8) Reduce dimensions or transform data	N/A	Heat plots [49], SPADE [12], t-SNE [50], viSNE [9], ISOMAP [23], LLE [25], PCA in R/flowCore [51]	SPADE†, viSNE
	9) Identify clusters of cells	Human expert	SPADE, k-medians, R/flowCore, flowSOM [52], Misty Mountain [13], JCM [26], Citrus [14], ACCSENSE [53], DensVM [24], AutoGate	SPADE (Figure 2)†, viSNE + human expert (Figure 1)
	10) Cluster refinement	Human expert	Citrus, DensVM, R/flowCore	-
Characterizing cell subsets	11) Feature comparison	Select biaxial single cell views	viSNE, SPADE, Heatmaps [34, 40], Histogram overlays [34, 40], Violin or box and whiskers plots [51]	Heatmaps (Figure 3A)†, viSNE (Figure 1C), SPADE (Figure 2C)
	12) Model populations	N/A	JCM, PCA	-
	13) Learn cell identity	Human expert	-	Human expert† (Figure 1B, Figure 2B, and Figure 3B)
	14) Statistical testing	Prism, Excel	R/flowCore	-

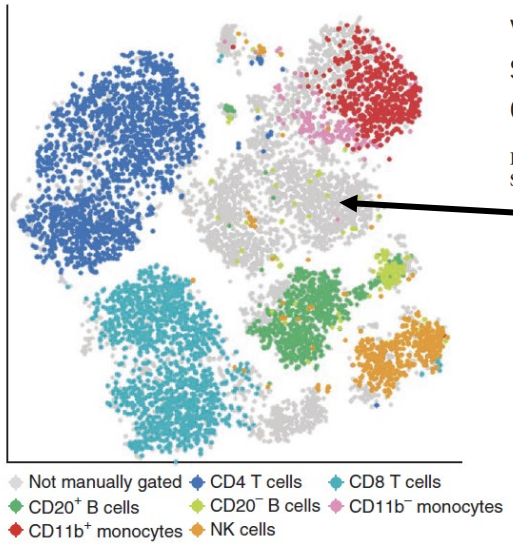
§Methods with broad application (e.g. R/flowCore) are listed minimally at select steps based on particular strengths or published applications.

†Denotes the primary approach used at each step in the sequential analysis workflow shown here.

A major gap in the field
is in true learning of cell identity

Diggins et al., *Methods* 2015

Traditional Gating Overlooks Many Cells in Primary Samples

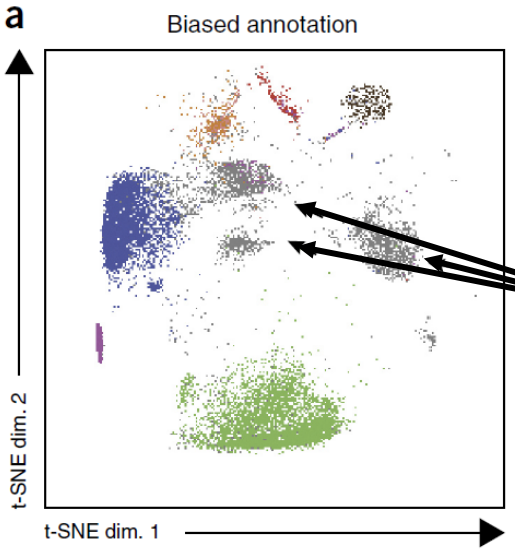


viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia

nature biotechnology
2013

El-ad David Amir[1], Kara L Davis[2,3], Michelle D Tadmor[1,3], Erin F Simonds[2,3], Jacob H Levine[1,3], Sean C Bendall[2,3], Daniel K Shenfeld[1,3], Smita Krishnaswamy[1], Garry P Nolan[2,4] & Dana Pe'er[1,4]

In all cases, the viSNE gate included cells that were not classified by the expert manually gated biaxial plots; these cells are labeled in gray in the viSNE map. Examination of the marker expression of these cells reveals that they are typically just beyond the threshold of one marker, but the viSNE classification is strongly supported based on the expression of all other markers. For example, in **Figure 1d**, wherein cells are colored for CD11b marker expression, the cells in the gated region express the canonical monocyte marker CD33 (**Supplementary Fig. 1b**). However, only 47% of these cells were classified as monocytes by the manual gating (**Fig. 1b**).



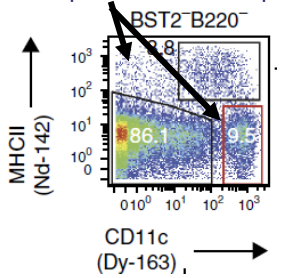
High-dimensional analysis of the murine myeloid cell system

nature immunology
2014

Burkhard Becher[1,4,5], Andreas Schlitzer[1,5], Jinmiao Chen[1,5], Florian Mair[2], Hermi R Sumatoh[1], Karen Wei Weng Teng[1], Donovan Low[1], Christiane Ruedl[3], Paola Riccardi-Castagnoli[1], Michael Poidinger[1], Melanie Greter[2], Florent Ginhoux[1] & Evan W Newell[1]

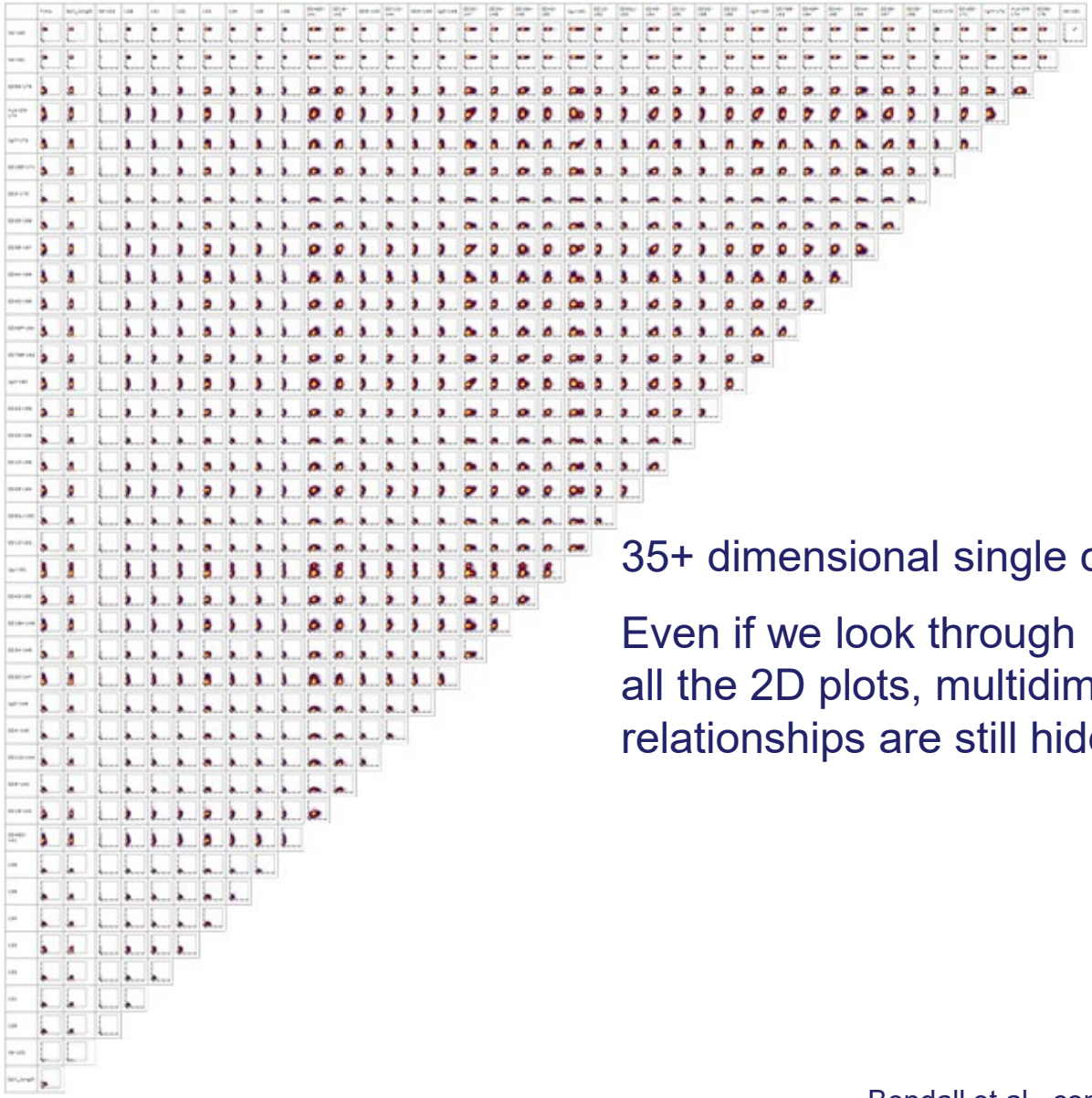
Notably, whereas traditional biased gating strategies allowed for identification of only 54.7 ± 2.6% (mean ± s.e.m., n = 3 mice) of lung myeloid cells (different DC subsets, macrophages, monocytes, neutrophils), the automatic, computational approach identified nearly 100% of the cells (96.6 ± 1.0% (mean ± s.e.m., n = 3 mice) accounted for by 14 predominant clusters).

Overlooked cells 'between' traditional gates add up to ~half the sample



Amir et al., *Nature Biotechnology* 2013; Becher et al., *Nature Immunology* 2014

We Now Make Billions of Multi-D Single Cell Measurements
=> Need for Machine Learning Tools & Human Readable Views



35+ dimensional single cell data:

Even if we look through
all the 2D plots, multidimensional
relationships are still hidden…

Bendall et al., concept plot of 35 features plotted pairwise, 2011

But first: what is data science?

Irish lab view of data science:

Systematically varying analytical elements in order to test a hypothesis

(Varied analytical elements might be different data types, data sub-samples, different initial assumptions, contrasting analytical tools, input parameters, etc.)

It's relatively new that datasets are robust enough to enable mining & exploration.

Rumsfeldian Data Science

Known knowns: What do you know about your system?

Known unknowns: What do you know remains to be learned?

Unknown unknowns: What don’t you know you don’t know?

Donald Rumsfeld (Feb 12, 2002): Reports that say that something hasn't happened are always interesting to me, because as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones.

Socratic Data Science

Known knowns:	What do you know about your system?
Known unknowns:	What do you know remains to be learned?
Unknown unknowns:	What don't you know you don't know?
Unknown knowns:	What don't you know, but think you do? i.e. Which 'priors' are incorrect?

If you fear incorrect priors, unsupervised analysis may be able to help.

Socrates according to Plato's *Apology*: I am wiser than this man, for neither of us appears to know anything great and good; but he fancies he knows something, although he knows nothing; whereas I, as I do not know anything, do not fancy I do. In this trifling particular, then, I appear to be wiser than he, because I do not fancy I know what I do not know.

Define Your System At the Start of Planning a Study

1) Elements, the studied units of the system.

- Patients, cells, images, pixels, transcripts, genomes, peptides.
- We will envision elements as “rows” in a spreadsheet.

2) Features, the things measured for each element.

- Clinical outcomes, phospho-proteins, pixel density, nucleotides.
- We will envision features as “columns” in a spreadsheet.
- Feature selection may rely on hypotheses, rules, or prior knowledge.

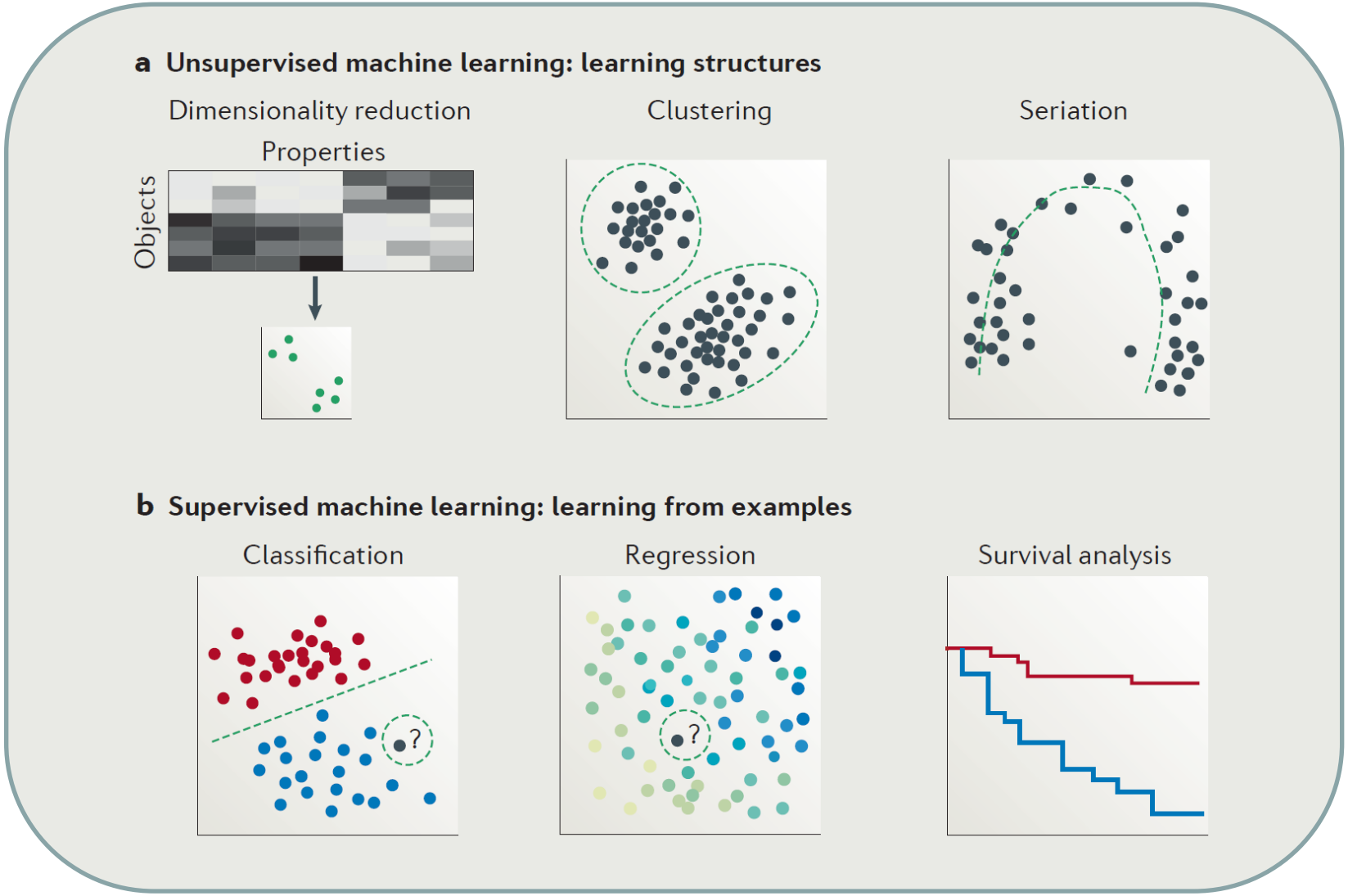
3) Scales, the type & range of the measurements for each feature.

- Categorical, linear, log & base, arcsinh & cofactor.
- -150 to 262,144; 1 to 10,000; 0 to 50; 1 to 100; 0 to 1; NR, PR, CR.
- Will largely explore the data without units until we create reports.

4) Prior knowledge, the things assumed to be known for the system.

- Organization of elements (groups, order, etc.), feature relationships.
- Supervised analysis explicitly uses prior knowledge.
- Unsupervised analysis looks for patterns without prior knowledge.

There Are Many Ways to Analyze Modern Datasets



Saeys Y, Van Gassen S, and Lambrecht BN, *Nature Reviews Immunology* 2016

ISAC FlowRepository: Share Annotated Cytometry Data Provides a Quantitative Reference & Can Reproduce An Analysis

https://flowrepository.org/

ISAC
INTERNATIONAL SOCIETY FOR ADVANCEMENT OF CYTOMETRY

- Initially sponsored by ISAC and Wallace H Coulter Foundation
- Maintained through volunteer efforts; use required for *Cytometry A* journal
- +1,000 users in the last year (+36%), >3500 total
- >100 users uploading and annotating data per month
- +922 experiments in just last year (+40%); >3000 total; downloads have doubled.
- Total data volume: ~4.0 TB in >1,380 public data sets

Over to Nicholas Loof…

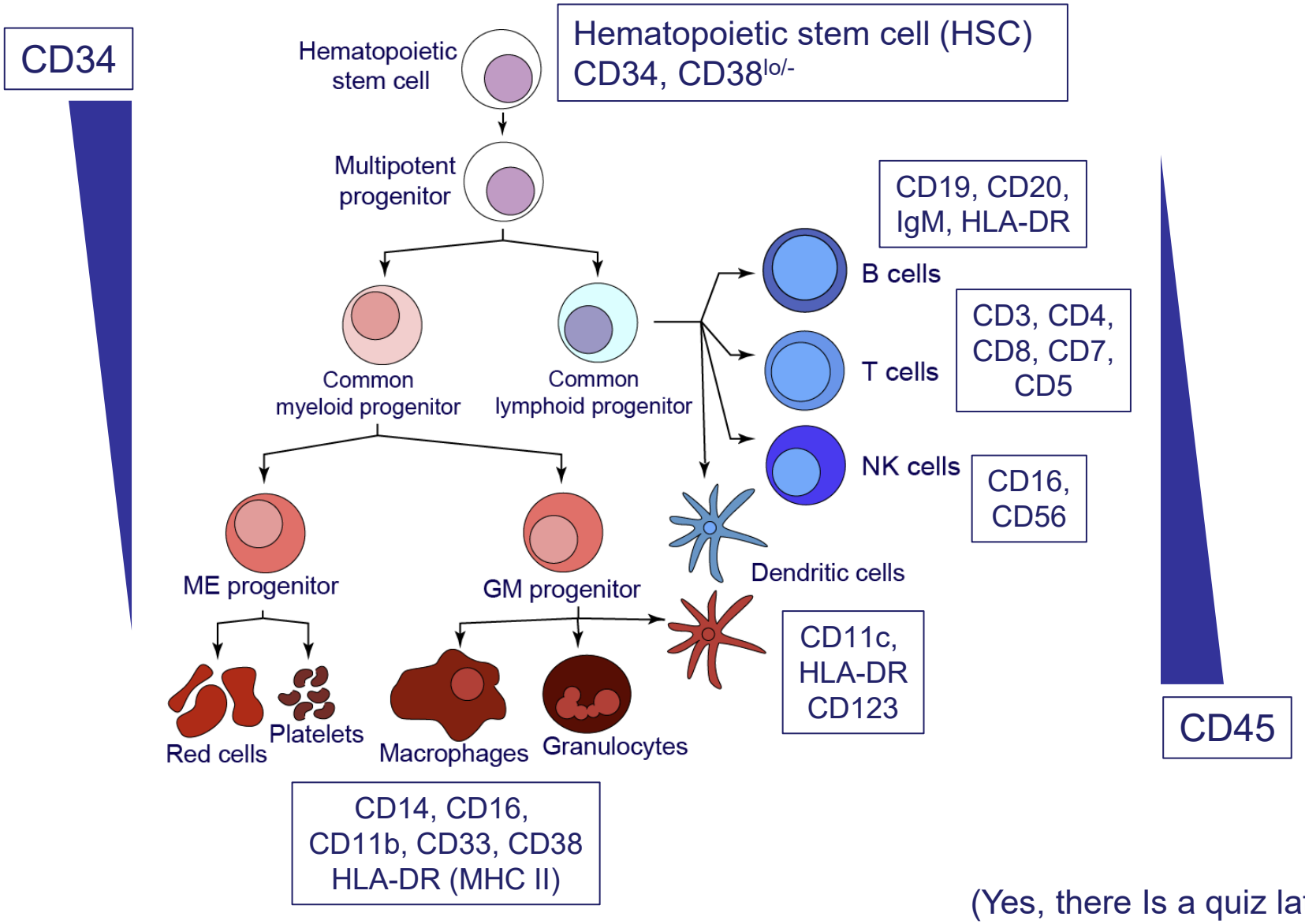
Identifying Cell Populations (MEM, RAPID, and Statistical Tools)

MEM summarizes a population's special features
and is used in workflows "at the end"
(in place of box and whisker plots or heatmaps)

[So MEM complements tools from other steps, including

t-SNE, SPADE, Citrus, FlowSOM, SCAFFOLD, Phenograph]

Human Bone Marrow Hematopoiesis & “Famous” Cell Identity Markers



(Yes, there Is a quiz later)

Despite advances, no computational tools learn & label cell identity,
a human must “stare and compare” using expert knowledge

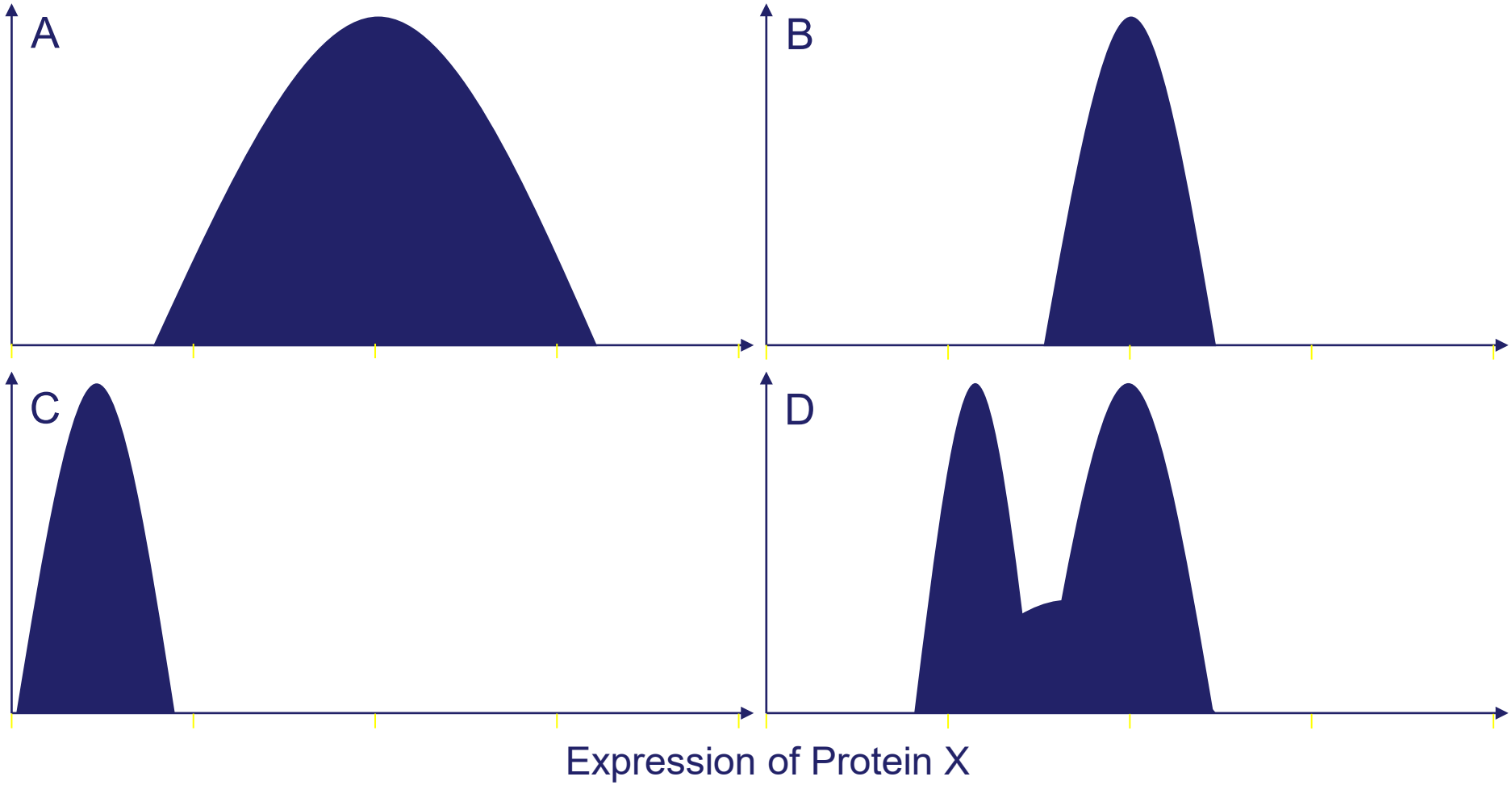
Diggins et al., *Methods* 2015

Populations are often labeled by metaphors of function
(“cancer stem cells”, “central memory T cells”)
or incomplete labels based on a few features (e.g. “PD-1+ CD8 T cells”).

We need an unbiased way to label & identify cells
(regardless of how they are found)

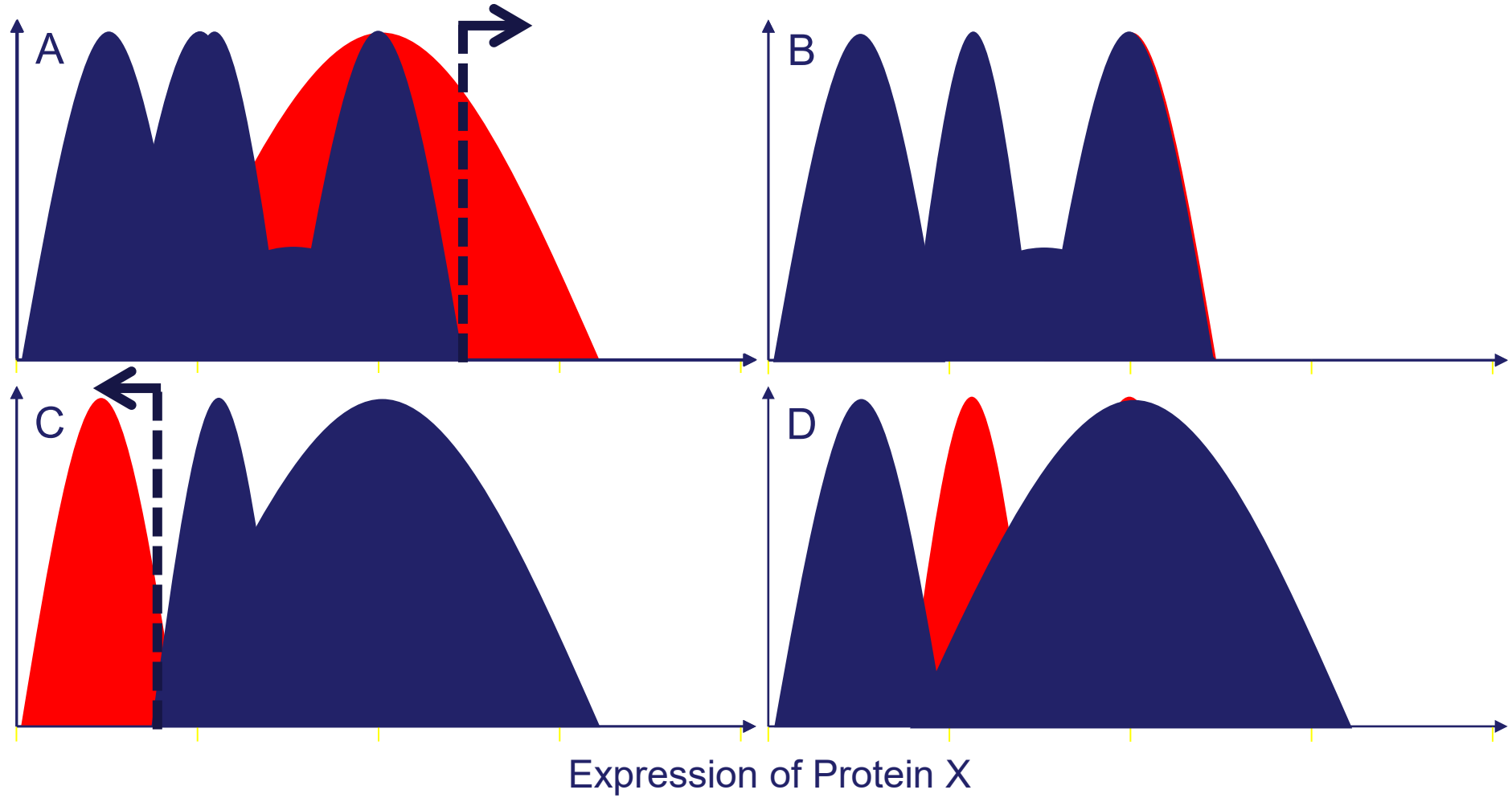
Enrichment Tracks Feature Exclusivity In a Subset

A, B, C, and D are 4 cell types within the same sample and they are each 25% of the sample. If I were to offer to pay you based on the purity of a sort based only on protein X, which population would you want to isolate? I'll pay you $1 per correct cell & take away $1 for each incorrect cell and $1 for each cell you miss.

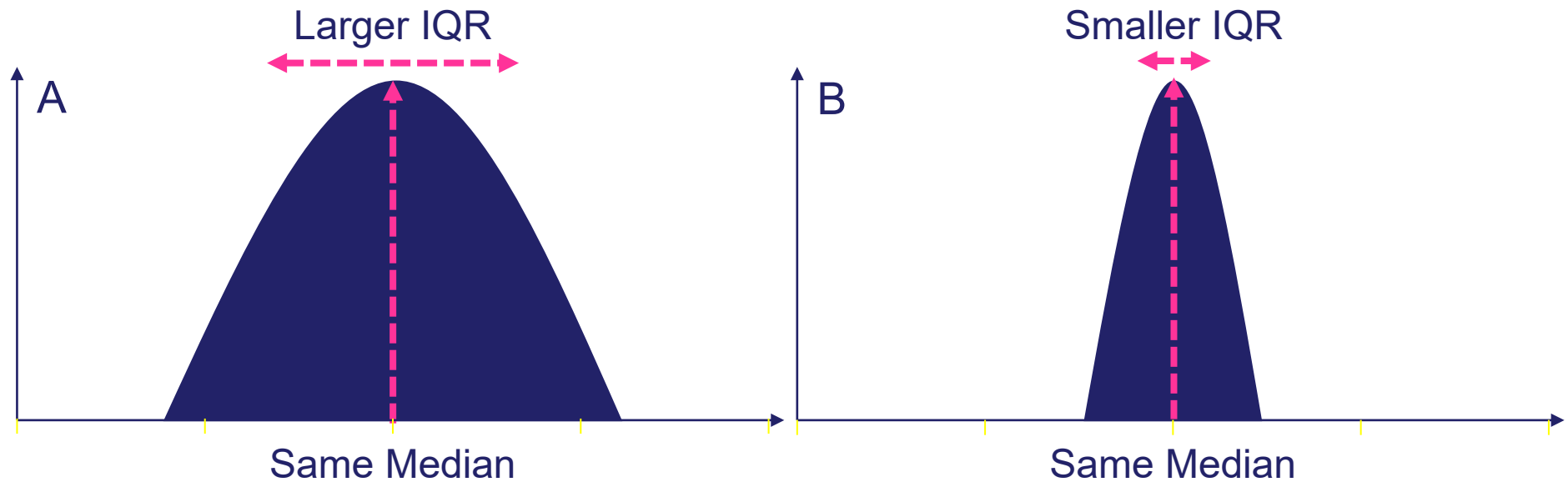


Enrichment Tracks Feature Exclusivity In a Subset

A, B, C, and D are 4 subsets where Protein X was measured.
In which subset is Protein X most distinct? (Which would be easiest to gate?)



Median (50%) and Interquartile Range (25%-75%) Represent Key Features of Distributions



Core idea in MEM: given two protein distributions with equal medians, a smaller interquartile range (IQR) indicates greater enrichment

Not captured by median & IQR are other elements of shape (skewness, symmetry, # peaks, outliers, etc.)

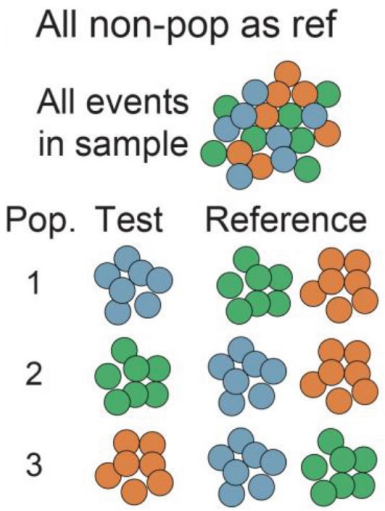
Diggins et al., *Nature Methods* 2017

MEM Quantifies Relative Enrichment By Combining Magnitude & Interquartile Range

$$MEM = |MAG_{test} - MAG_{ref}| + \frac{IQR_{ref}}{IQR_{test}} - 1$$

Linear transformation to -10 to +10
(d20 scale, cause that's how we roll)

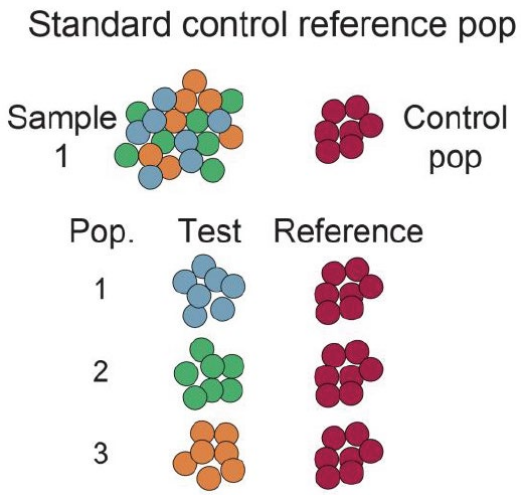
If $MAG_{test} - MAG_{ref} < 0$, MEM = -MEM



MEM label ($CD19^{+}$ cells)

▲$HLADR^{+10}$ $CD20^{+9}$ $CD19^{+7}$ IgM^{+5} C
$CD45RA^{+3}$ $CXCR4^{+2}$ $CD47^{+2}$ CD33

▼$CD7^{-2}$



Diggins et al., *Nature Methods* 2017

Quiz Time: What Are These Cell Subsets & What Is This Tissue?

Stem cells (HSCs)

▲$CD34^{+6}$ $CD33^{+4}$ $CD15^{+3}$ $CD38^{+3}$ $MHCII^{+3}$ $CXCR4^{+2}$
▼$CD44^{-5}$ $CD45^{-5}$ $CD7^{-3}$ 0.07%

Progenitors

▲$MHCII^{+10}$ $CD33^{+7}$ $CD38^{+5}$ $CD123^{+3}$ $CD117^{+3}$ $CD19^{+2}$ $CD34^{+2}$ $CD13^{+2}$ $CD14^{+2}$ $CXCR4^{+2}$
▼$CD45^{-3}$ $CD15^{-2}$ 0.002%

Early myeloid cells

▲$MHCII^{+9}$ $CD33^{+8}$ $CD38^{+5}$ $CD4^{+3}$ $CD15^{+2}$ $CD14^{+2}$
▼ $CD45^{-2}$ $CD7^{-2}$ 0.02%

Monocytes

▲$CD33^{+10}$ $CD14^{+8}$ $CD11b^{+7}$ $MHCII^{+5}$ $CD4^{+4}$ $CD11c^{+4}$ $CD38^{+4}$ $CD13^{+3}$
▼ $CXCR4^{-2}$ $CD47^{-2}$ 10.57%

Natural killer cells

▲$CD16^{+9}$ $CD7^{+6}$ $CD38^{+5}$ $CD56^{+4}$ $CD161^{+4}$ $CD45RA^{+3}$ $CD8^{+2}$ $CD11b^{+2}$ $CD47^{+2}$ 5.27%

$CD8^+$ T cells

▲$CD8^{+8}$ $CD7^{+5}$ $CD3^{+3}$ $CD45RA^{+3}$ $CXCR4^{+2}$ 9.25%

$CD4^+$ T cells

▲$CD4^{+7}$ $CD7^{+5}$ $CD3^{+5}$ $CD47^{+2}$ $CD45RA^{+2}$ 8.12%

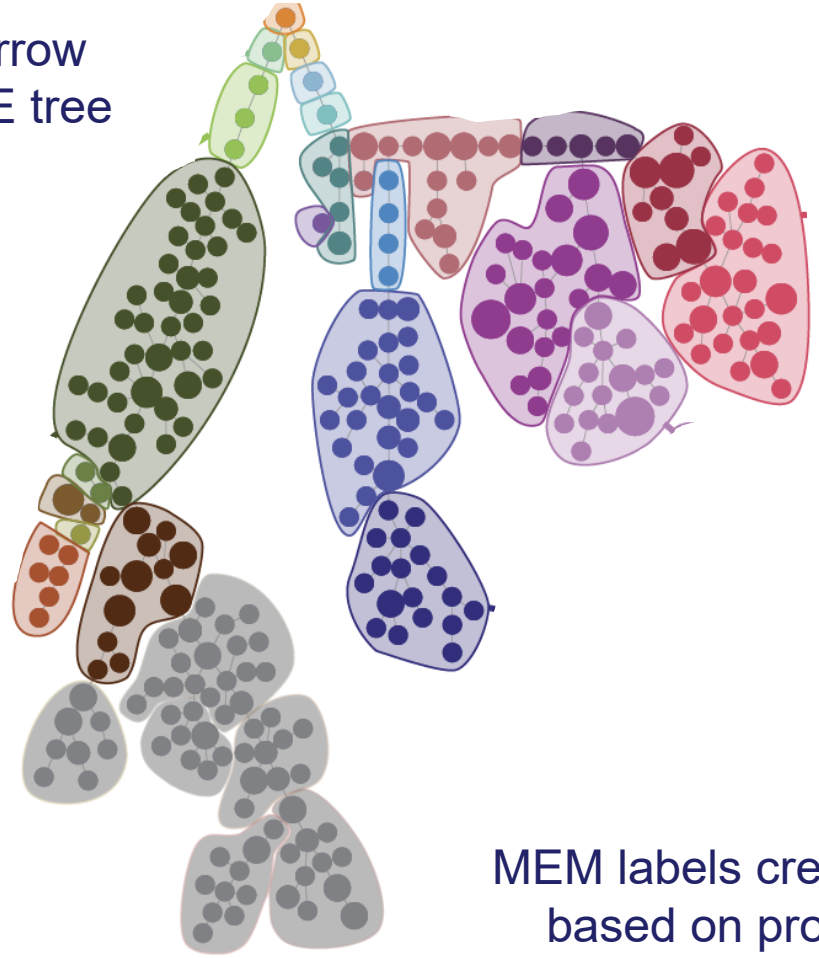
B cells

▲$MHCII^{+10}$ $CD20^{+9}$ $CD19^{+7}$ IgM^{+5} $CD34^{+3}$ $CD45RA^{+3}$ $CXCR4^{+2}$ $CD47^{+2}$ $CD33^{+2}$
▼$CD7^{-2}$ 2.44%

Data from healthy human bone marrow, Bendall et al., *Science* 2011
Diggins et al., *Nature Methods* 2017

Marker Enrichment Modeling Automatically Labels Cell Types in Human Bone Marrow Using -10 to +10 Enrichment Values

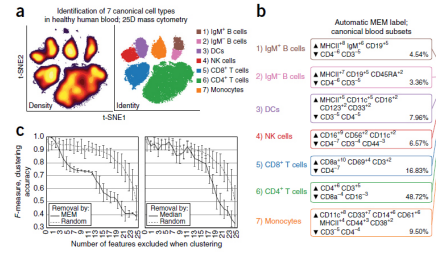
Cells from bone marrow grouped in a SPADE tree



MEM labels created automatically based on protein enrichment

Diggins et al., *Nature Methods* 2017
Data from Bendall et al., *Science* 2011

Tools for Automated Cell Discovery & Characterization



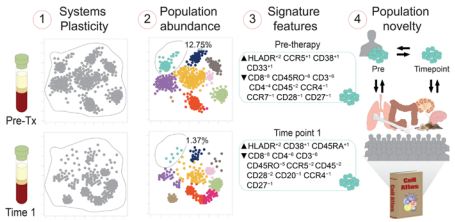
Nature Methods 2017

MEM

Diggins et al., PMC5330853

- MEM: machine labeling & identification of cell type clusters
- Enabled comparison across single cell platforms
- E.g., memory CD4+ T cells, 10-point enrichment scale:

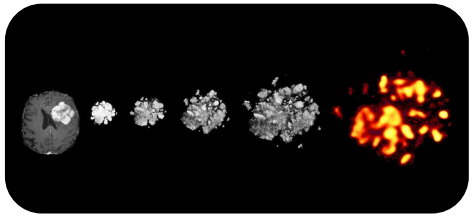
  $ICOS^{+8}$ $CD38^{+8}$ $CD4^{+7}$ $CD45R0^{+6}$ $CD3^{+5}$ $Ki\text{-}67^{+4}$



Cancer Immunology Research 2019

Greenplate et al., PMC6318034

- Set of tools for longitudinal single cell tumor immunology.
- Revealed abnormal immune cells in multiple tumor types.
- Includes datasets (AML, melanoma) used by T-REX
- Greenplate+ vs. COVID-19 (*Science* 2020 PMC7263500)

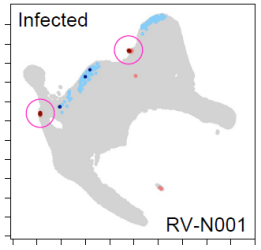


eLife 2020

RAPID

Leelatian, Sinnaeve et al., PMC7340505

- RAPID: probabilistic clinical outcomes on t-SNE & UMAP
- Reveals associations with extreme clinical outcomes
- Revealed JAK + AKT cooperation in glioblastoma cells



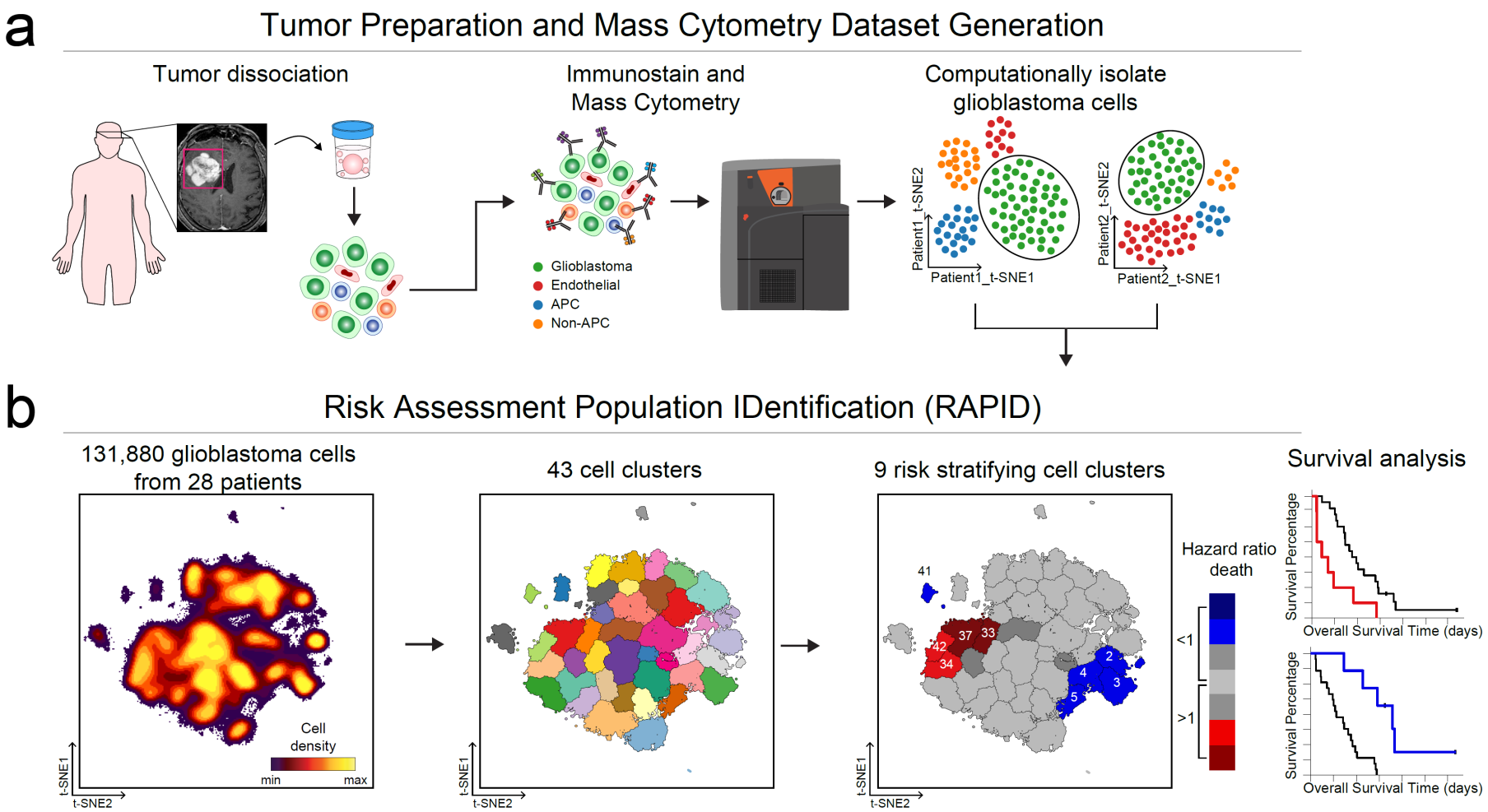
eLife 2021

T-REX

Barone, Paul, Muehling et al., PMC8370768

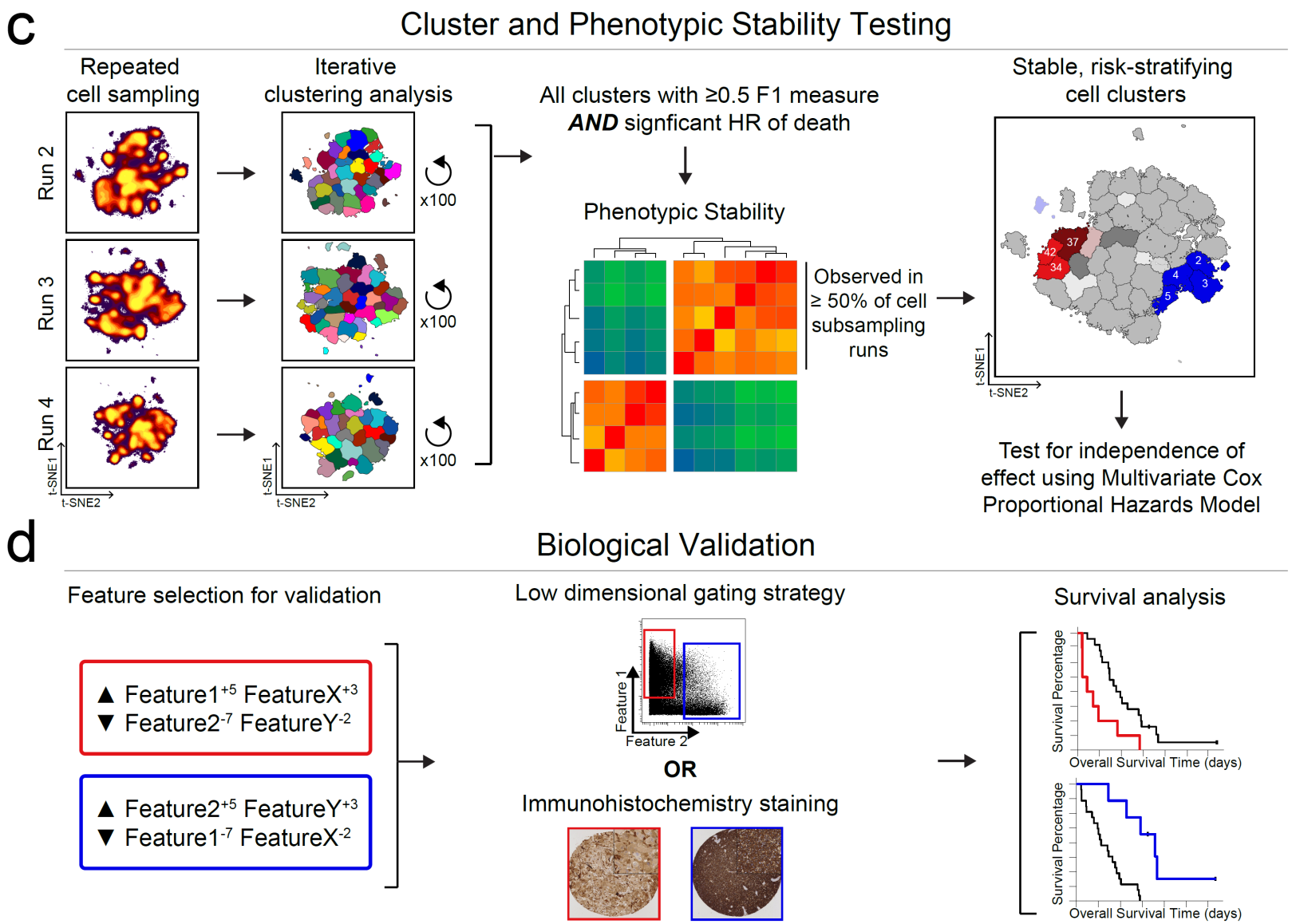
- T-REX: compares a pair of samples (e.g. pre- and post-)
- Revealed rhinovirus-specific cells based on rapid expansion
- Revealed identity of memory T cells in SARS-CoV-2 vaccine response (*bioRxiv* preprint PMC8328055)

Considering a Recent Algorithm: RAPID is Designed for Unsupervised Analysis of Survival



Leelatian, Sinnaeve et al., *eLife 2020*

Statistical & Biological Validation Should Be Designed In & Will Be Essential During Peer Review



Leelatian, Sinnaeve et al., *eLife 2020*

The Importance of Scales (time permitting)

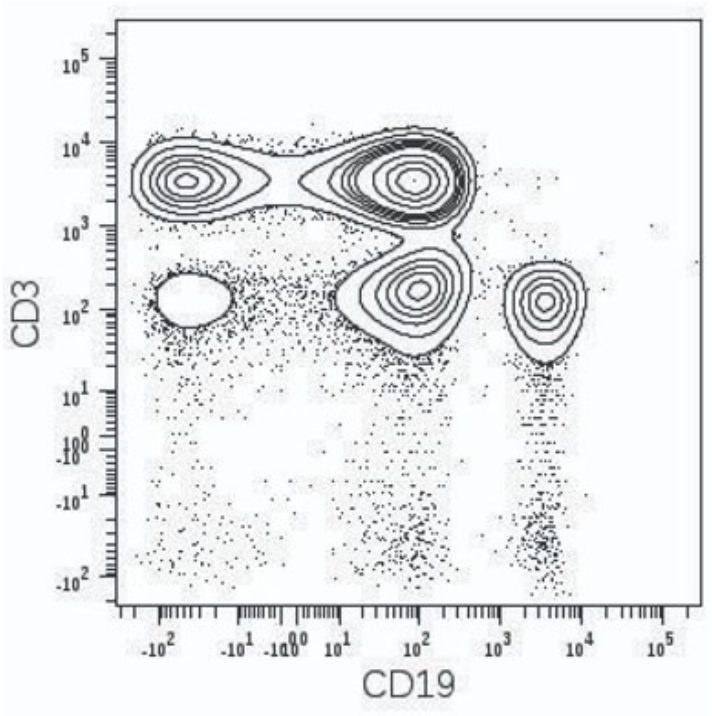
Before we get to ‘dessert’, some math ‘veggies’:

Scales matter: poorly or variably scaled data can destroy an analysis, most issues arise near zero

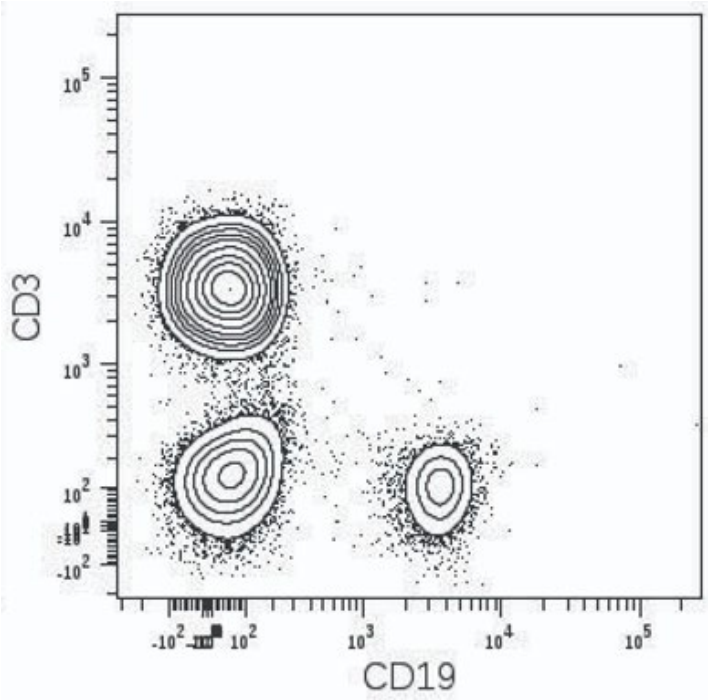
(pre-processing & normalization can also be critical)

Have you ever noticed two peaks within a cell subset that is biologically 100% negative for a marker?

Incorrect scales



Better scales

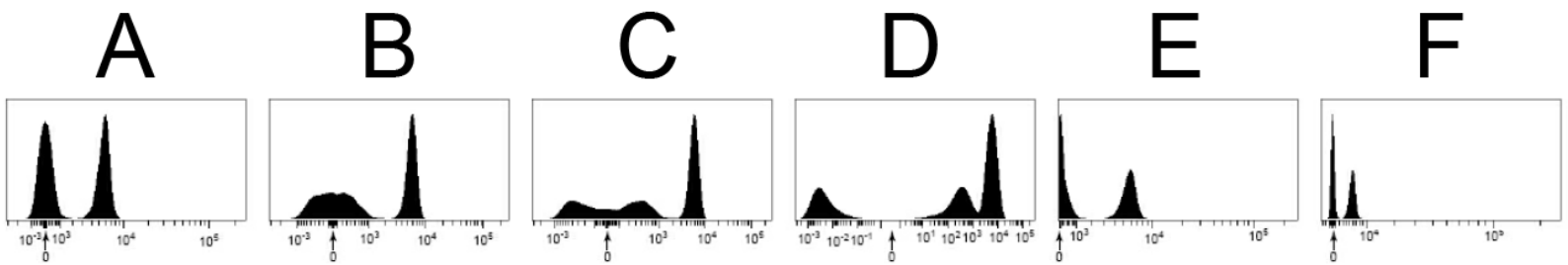


Results from bad scaling (poor transformation)
and it can be an issue for computational analysis.

Scaling is important in both mass and fluorescence cytometry.

Scaling Matters for Measuring Distance (Compensation Beads)

A 50:50 mix of + and - events stained only for PerCP-Cy5.5 is shown using different scales.



	A	B	C	D	E	F
Min:	-3000	-3000	-3000	-3000	1	-3000
Cofactor:	2500	500	150	1	2500	10,000

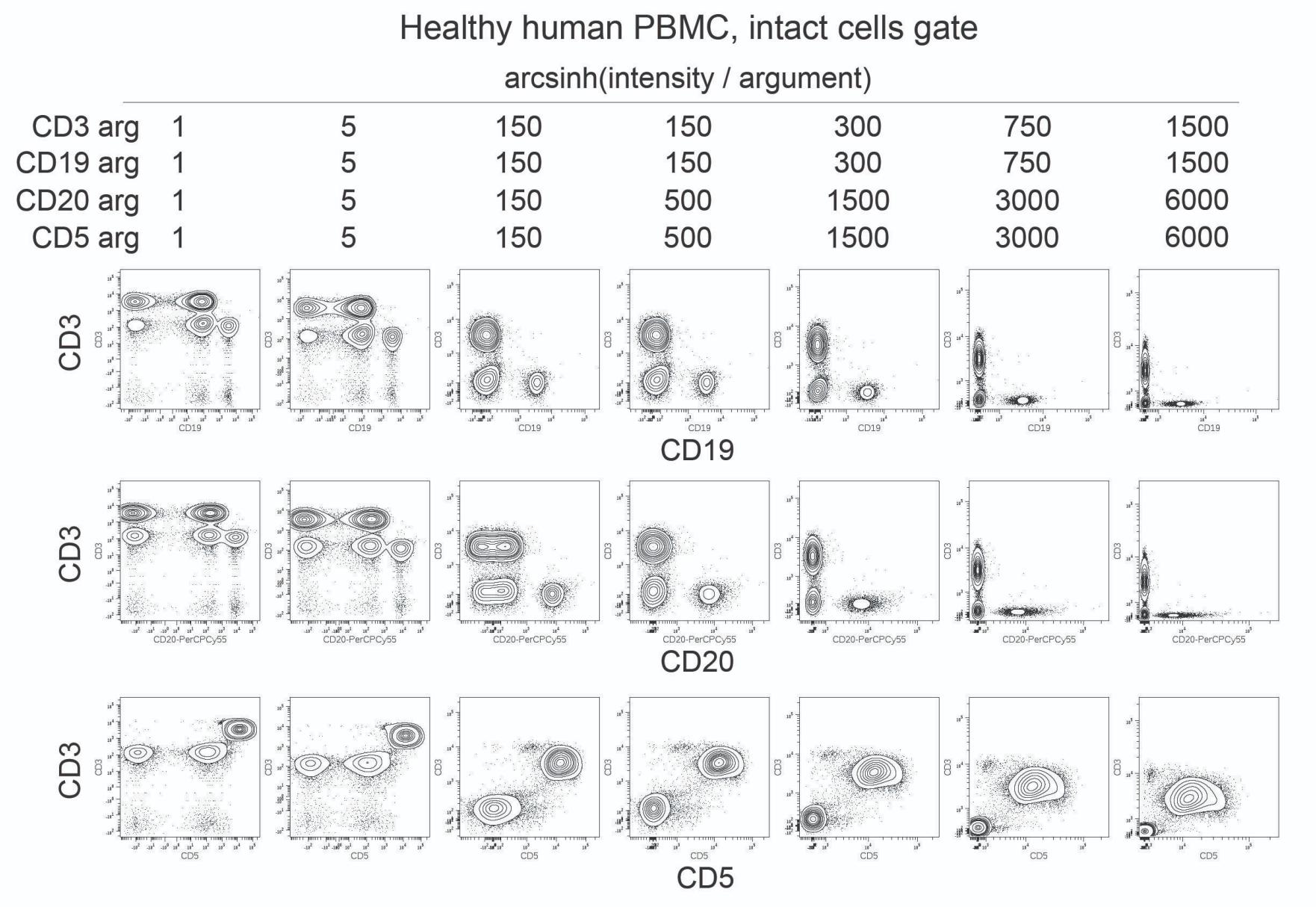
$$\text{arcsinh(x) with cofactor c} = \ln\left(\frac{x}{c} + \sqrt{1 + \left(\frac{x}{c}\right)^2}\right)$$

For fluorescence flow cytometry data: a biexponential or arcsinh transformation corrects the scale near zero.

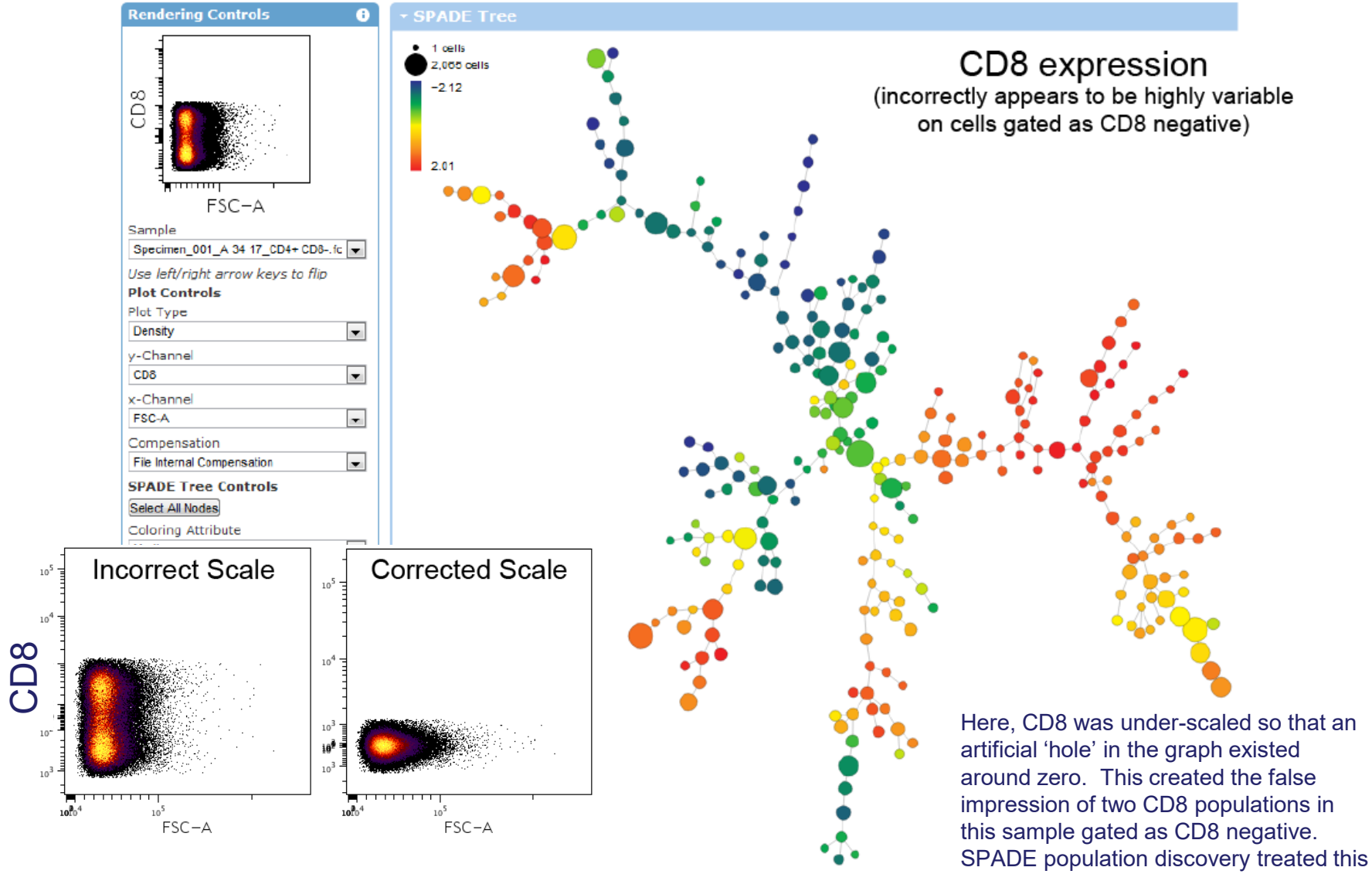
Since computational analysis techniques compare distance similarly to what a person does when looking at a plot, poorly scaled data can lead to identification of artificial populations near zero (see C and D) if data are not appropriately transformed prior to analysis.

More information: https://my.vanderbilt.edu/irishlab/protocols/scales-and-transformation/
FlowJo webpage on scales & transformation

Scaling Matters for Measuring Distance (Fluorescence Flow)



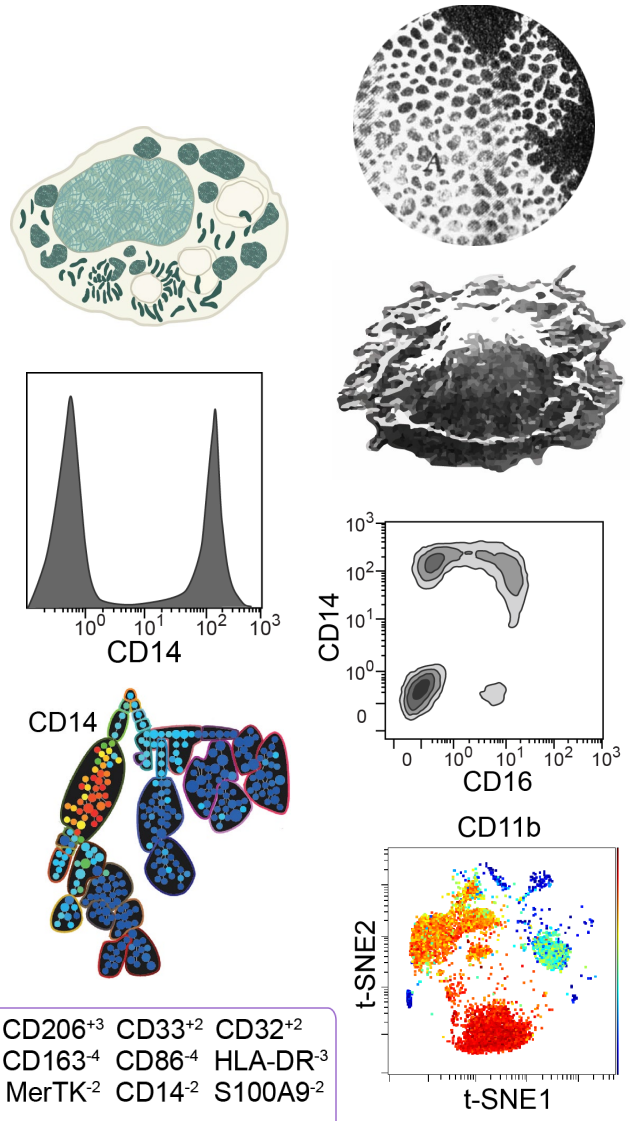
Inappropriate Scaling Can Lead to False Population Discovery



Here, CD8 was under-scaled so that an artificial 'hole' in the graph existed around zero. This created the false impression of two CD8 populations in this sample gated as CD8 negative. SPADE population discovery treated this as significant.

New Technology Reveals & Characterizes New Cells

Date	Approach	Dimensions (D) Per Cell & Speed	
1665*	Light microscopy	Low	Low
1908**	Light microscopy	Low	Low
1946	Scanning EM	Low	Low
1989	Flow cytometry identification	Low	1K cells/s
2001	Flow cytometry subsetting	4D	2 – 50K cell/s
2011	Mass cytometry + SPADE	32D	500 cell/s
2014	Mass cytometry + t-SNE / viSNE	38D	500 cell/s
(now)	Flow or Imaging MC + UMAP, FlowSOM, MEM	50D	500 cell/s



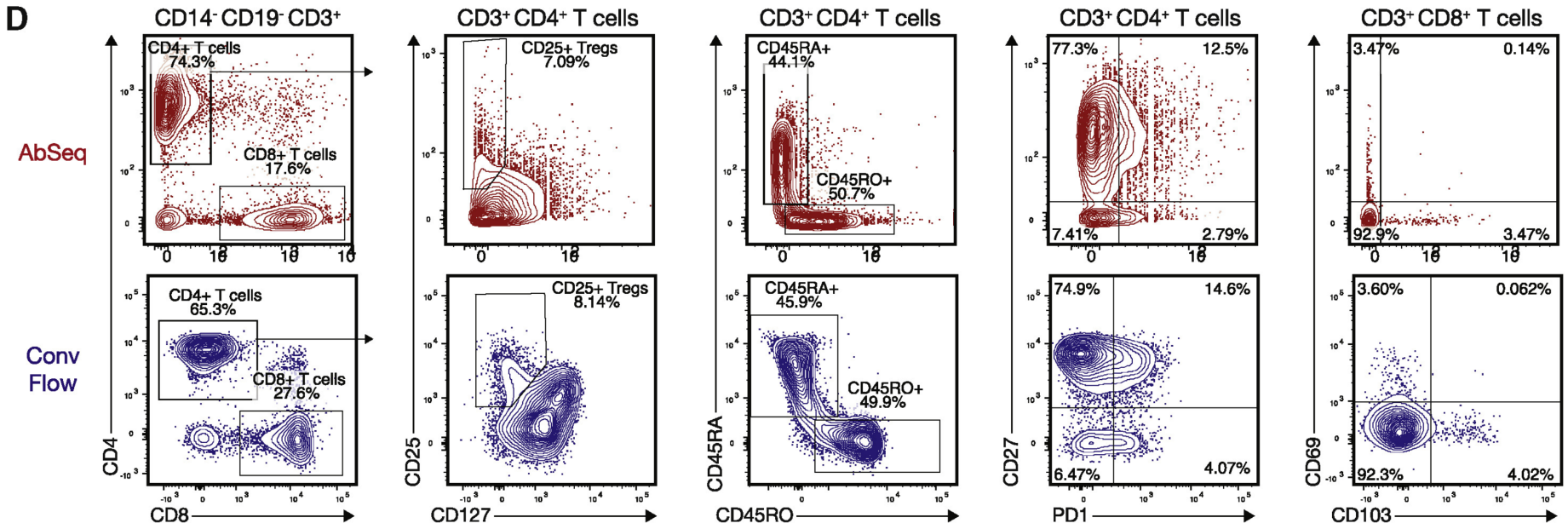
* Robert Hooke describes 'cells' in *Micrographia: or Some Physiological Descriptions of Miniature Bodies Made by Magnifying Glasses*

** Élie Metchnikoff characterizes mononuclear phagocytes: Lectures on the Comparative Pathology of Inflammation, Pasteur Institute in 1891, Nobel Prize in 1908 w/ Ehrlich.

Adapted from
Roussel et al., *Human Innate Immunity* 2016

Different Data Types May Need Different Scale Types or Cofactors

(Measuring Antibodies by AbSeq scRNA-seq vs. Fluorescence Flow)



Mair et al., *Cell Reports* 2020