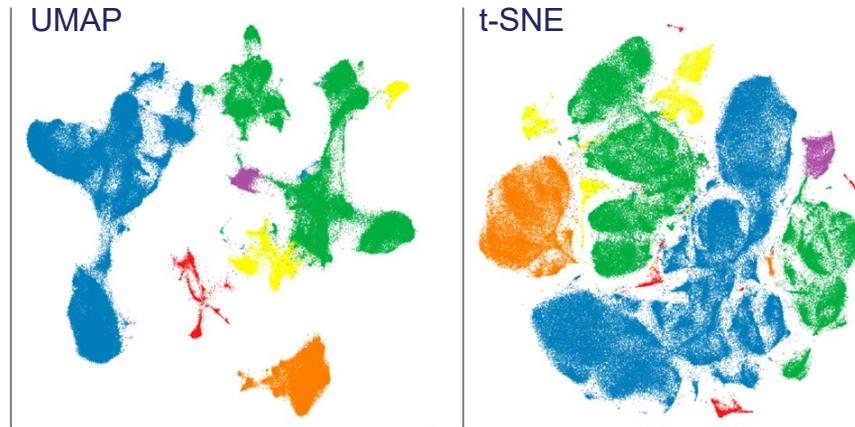


Part 1: 90 min

Analysis I: Getting Started with Cytometry Data Science



Becht et al. 2018



Jonathan Irish

Associate Professor
Cell & Developmental Biology
Pathology, Microbiology & Immunology

jonathan.irish@vanderbilt.edu



Cass Mayeda

Web Applications
Research Assistant

cass.mayeda@vanderbilt.edu



Josef Spidlen

Senior Director R&D
BD Biosciences

josef.spidlen@bd.com



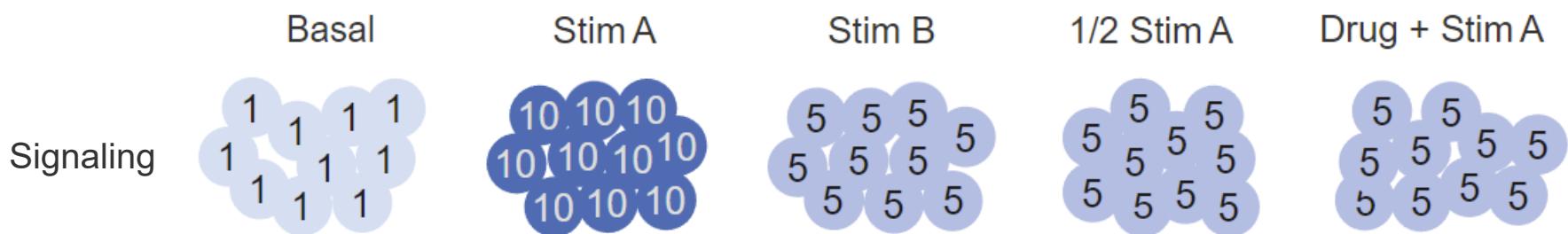
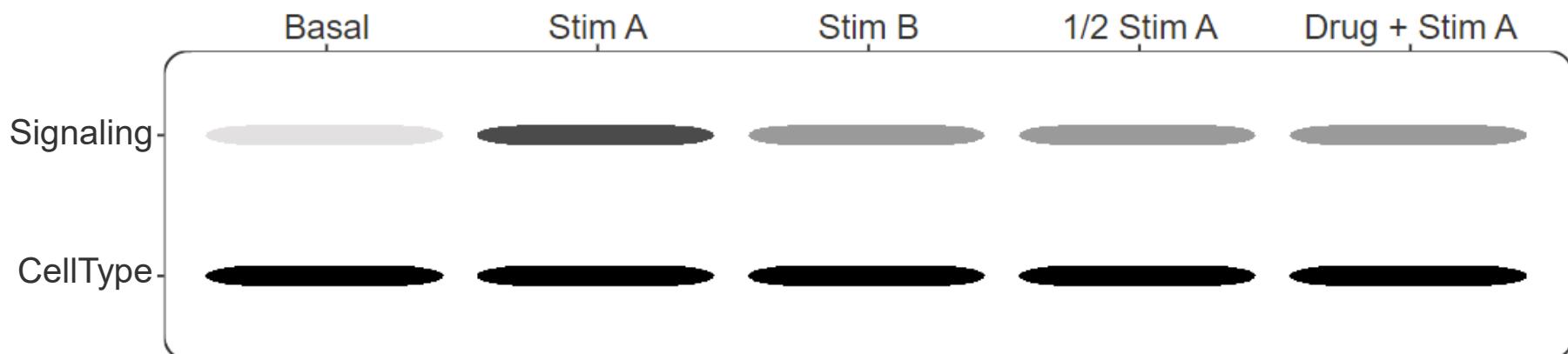
Nicolas Loof

Senior Application Scientist
Multi-omics Specialist
BD Biosciences

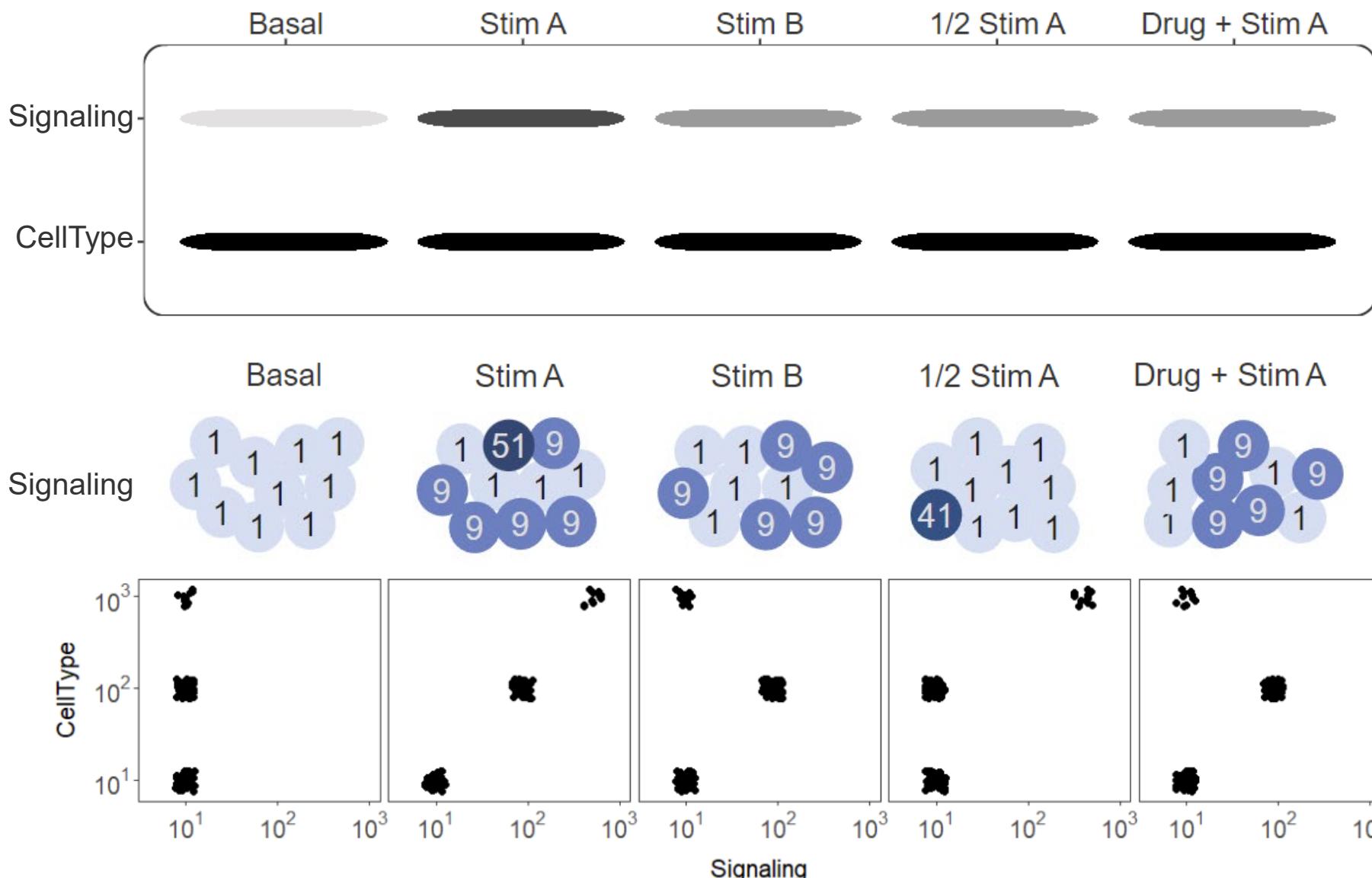
nicolas.loof@bd.com

Thinking Multidimensionally for Single Cell Cytometry

Single Cell Biology: Which Cell, How Much?

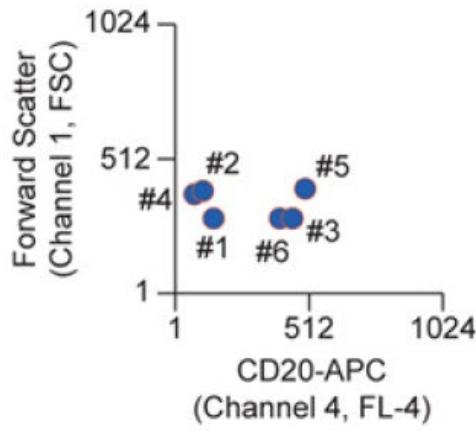
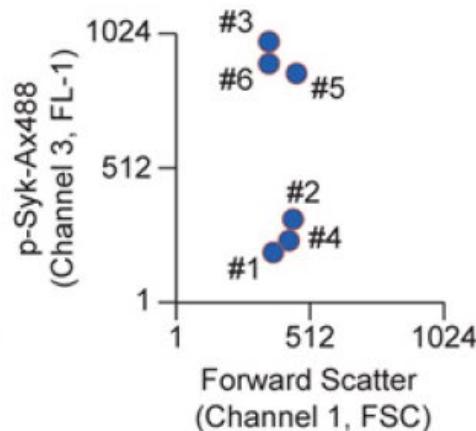
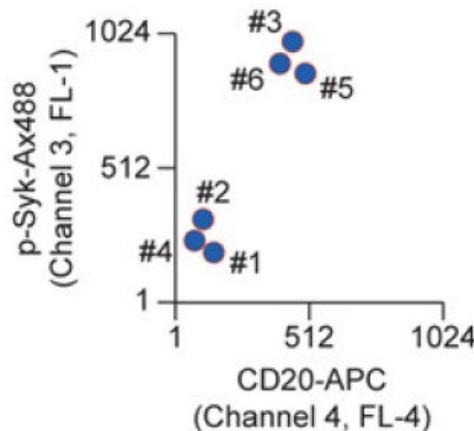


Single Cell Biology: Which Cell, How Much?



Example Four-Parameter Cytometry Data (Event List)

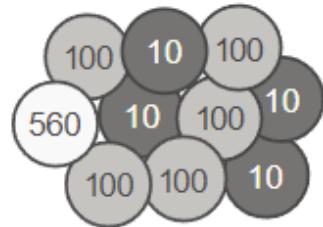
	Channel 1	Channel 2	Channel 3	Channel 4
Cell Event #1	400	290	5	50
Cell Event #2	425	301	18	45
Cell Event #3	402	292	912	503
Cell Event #4	422	303	14	40
Cell Event #5	430	310	892	510
Cell Event #6	402	282	903	499
Long Channel Name	Forward Scatter	Side Scatter	Fluorescent Channel 1 (FL-1)	Fluorescent Channel 4 (FL-4)
What biologists call this channel:	Forward Scatter (FSC)	Side Scatter (SSC)	phospho-Syk-Alexa488 (p-Syk-Ax488)	CD20-APC



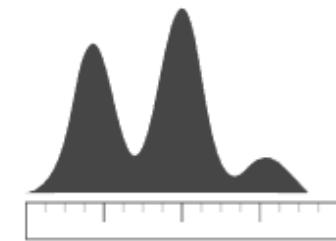
Single Cell Biology: Which Cell, How Much?

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!).



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



Why Single Cell Biology?

Explore a comparison with western blots

Cytometry Workflows

Steps in data analysis pipelines

Scaling Matters

How to prepare your data

Single Cell Biology

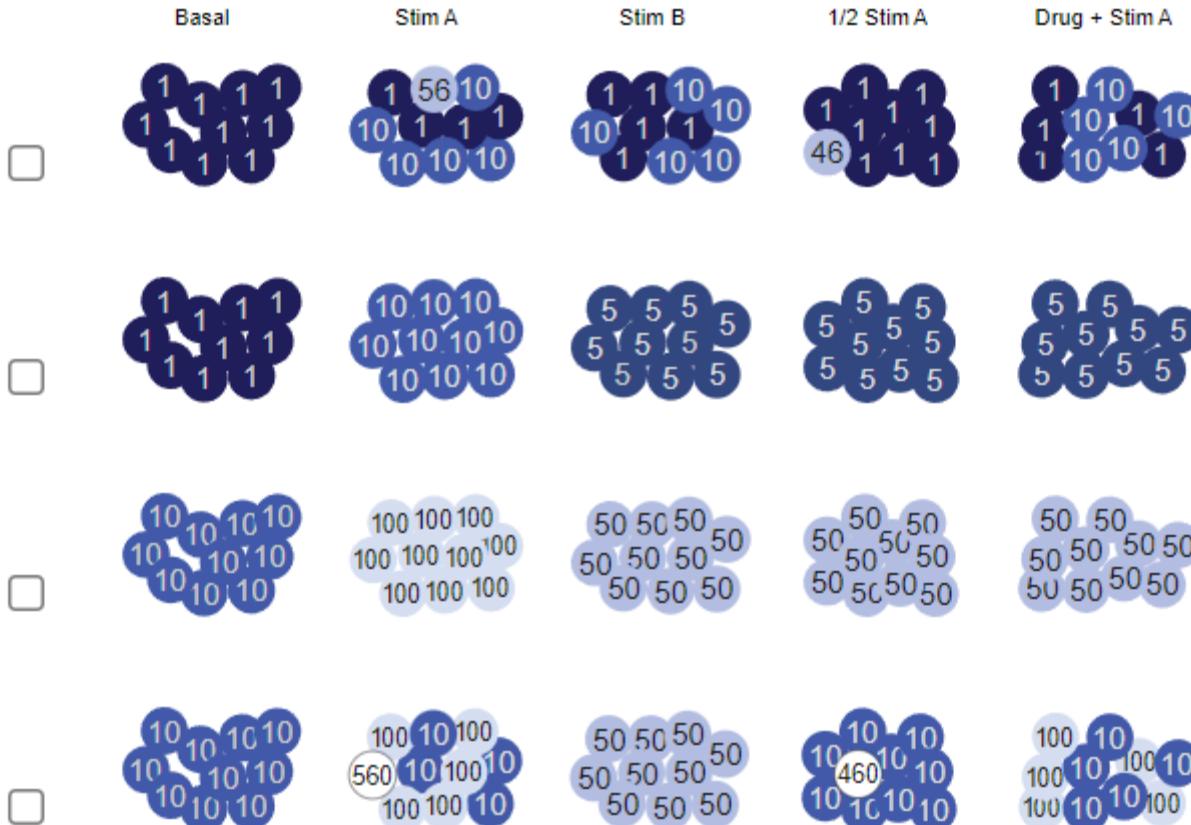
What's in a Western Blot?

If you were to perform a western blot for the signaling marker, the result would look like the following.



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.

Single Cell Biology



Single Cell Biology

What technology, which measurement?

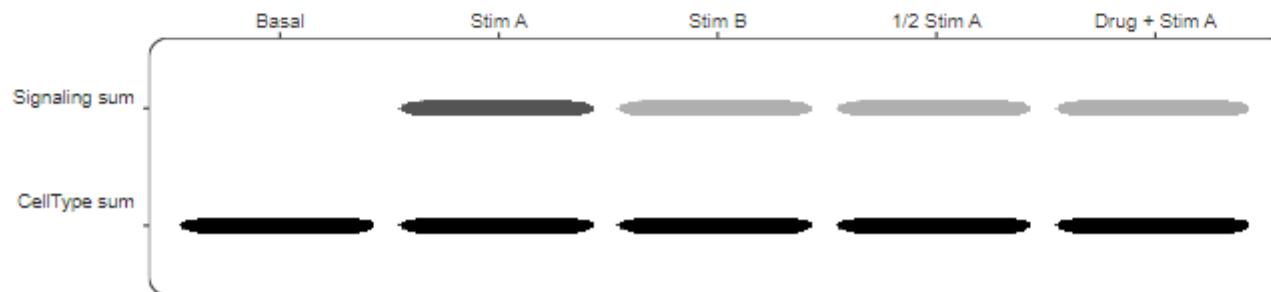
In contrast to Western Blots, Flow Cytometry experiments are usually interested in the MEDIAN signal.

Western Blot

Flow Cytometry

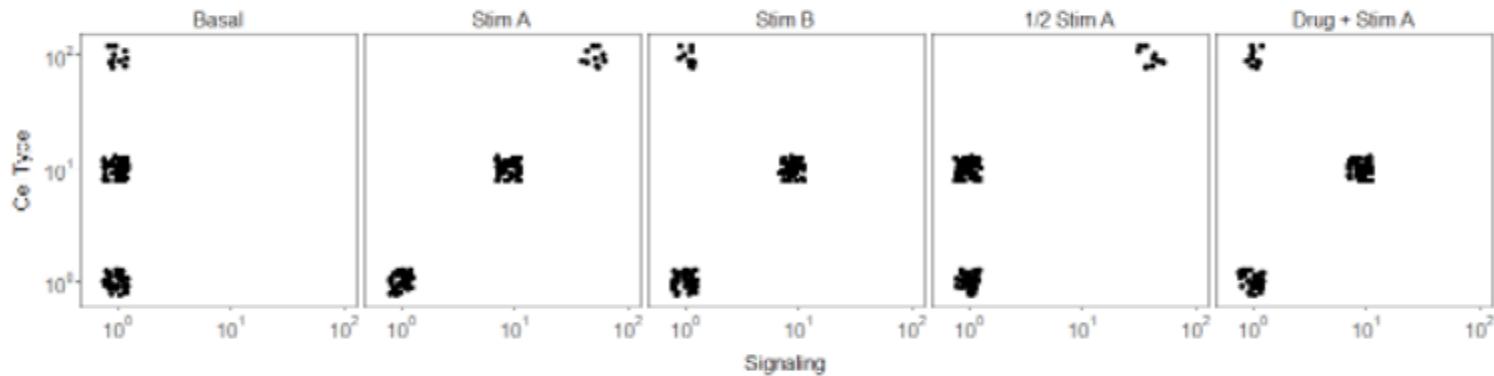
Average

Measurement type: sum



Single Cell Biology

CellType vs. Signaling



Each plot contains 100 cells.

Same data, different views

Heatmaps

Heatmaps are another common data analysis tool. With cytometry data, they can be configured to display the sum, median, or average signal values in a way that's very similar to the simulated western diagrams above.

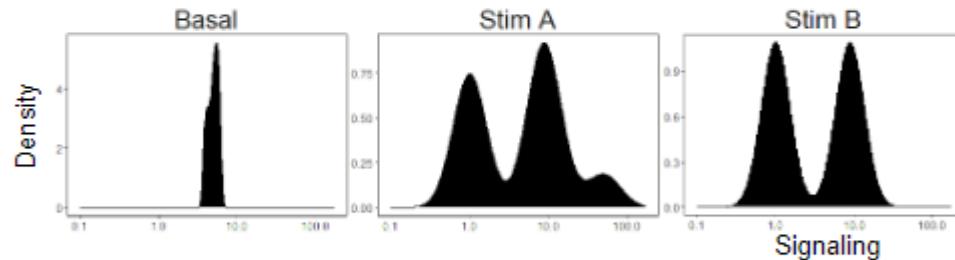


Same data, different views

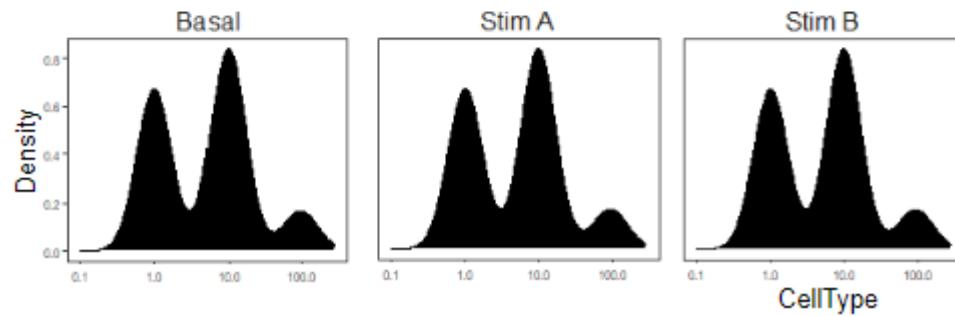
Misc

Histograms and contour plots are other methods of displaying data.

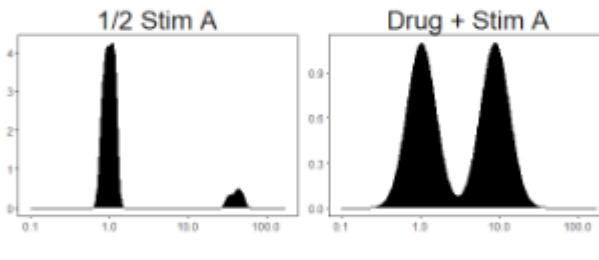
Histograms



Contours



Histograms & Contours



Same data, different views

Try for yourself

If you'd like to try plotting this dataset yourself, it's available to download below. If you don't know where or how to start, there is also an R markdown document you can use as a jumping off point for generating figures from this web page.



[Download CSV](#)

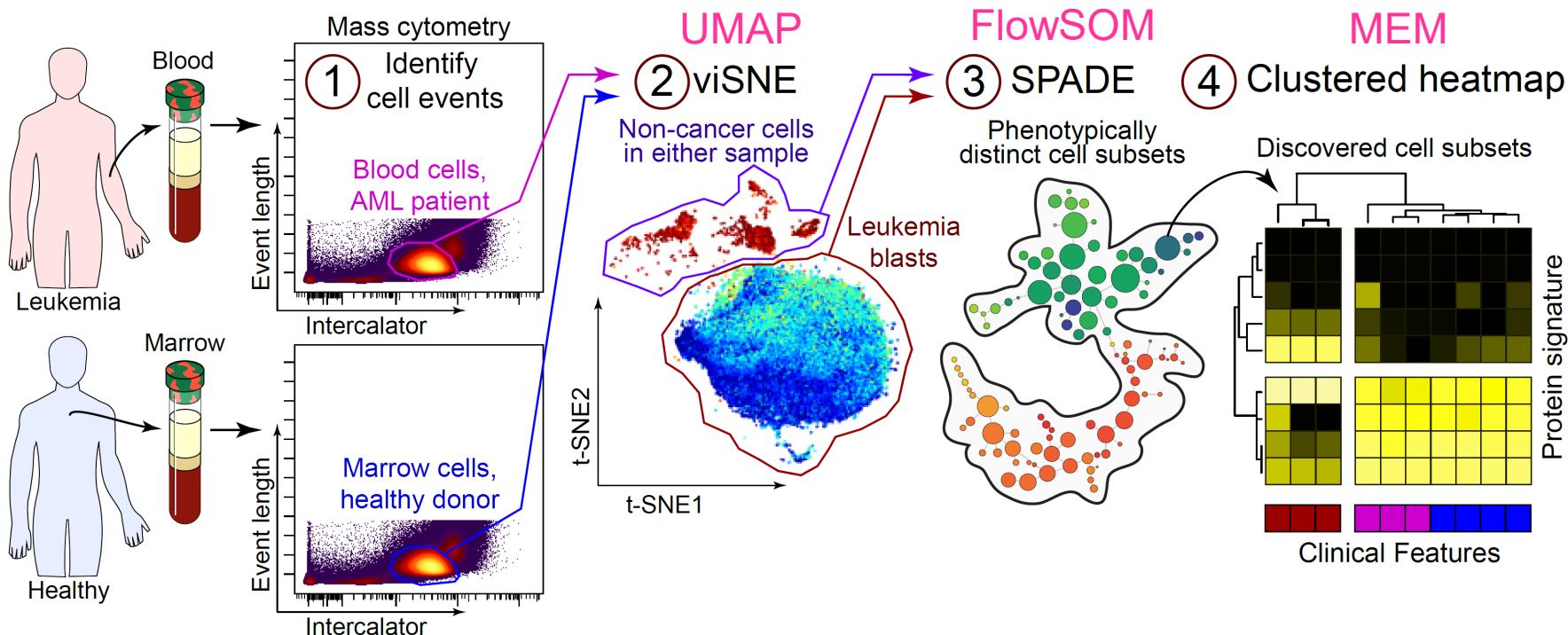


[Download FCS](#)



[Download R code](#)

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



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

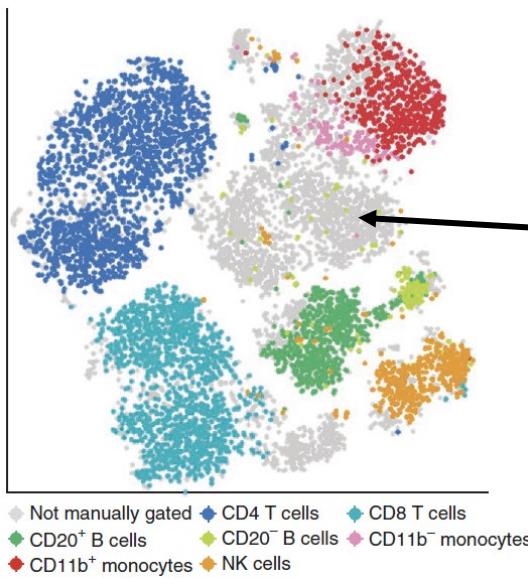
Analysis step	Traditional	Additional methods [§]		Method here
Data collection	1) Panel design	Human expert	-	-
	2) Data collection	Human expert	-	-
Data processing	3) Cell event parsing	Instrument software	Bead normalization and event parsing [31]	-
	4) Scale transformation	Human expert	Logicle [36]	-
Distinguishing initial populations	5) Live single cell gating	Biaxial gating + human expert	No event restriction, AutoGate [48]	viSNE + human expert (Figure 1) [†]
	6) Focal population gating			
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

Traditional Gating Overlooks Many Cells in Primary Samples

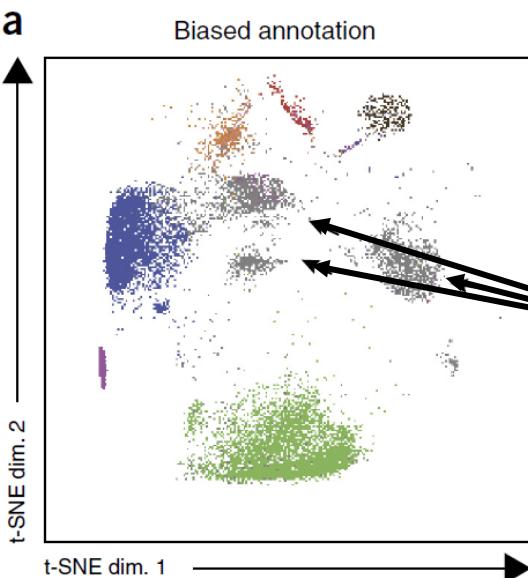


viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia

El-ad David Amir¹, 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 Krishnaswamy¹, Garry P Nolan^{2,4} & Dana Pe'er^{1,4}

nature biotechnology
2013

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**).

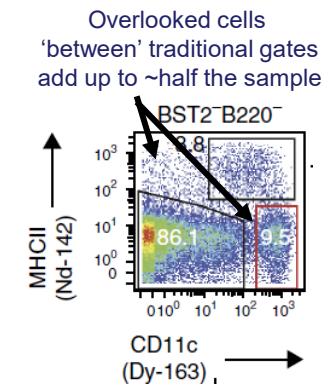


High-dimensional analysis of the murine myeloid cell system

Burkhard Becher^{1,4,5}, Andreas Schlitzer^{1,5}, Jinmiao Chen^{1,5}, Florian Mair², Hermi R Sumatoh¹, Karen Wei Weng Teng¹, Donovan Low¹, Christiane Ruedl³, Paola Riccardi-Castagnoli¹, Michael Poidinger¹, Melanie Greter², Florent Ginhoux¹ & Evan W Newell¹

nature immunology
2014

Notably, whereas traditional biased gating strategies allowed for identification of only $54.7 \pm 2.6\%$ (mean \pm 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 ($96.6 \pm 1.0\%$ (mean \pm s.e.m., $n = 3$ mice) accounted for by 14 predominant clusters).

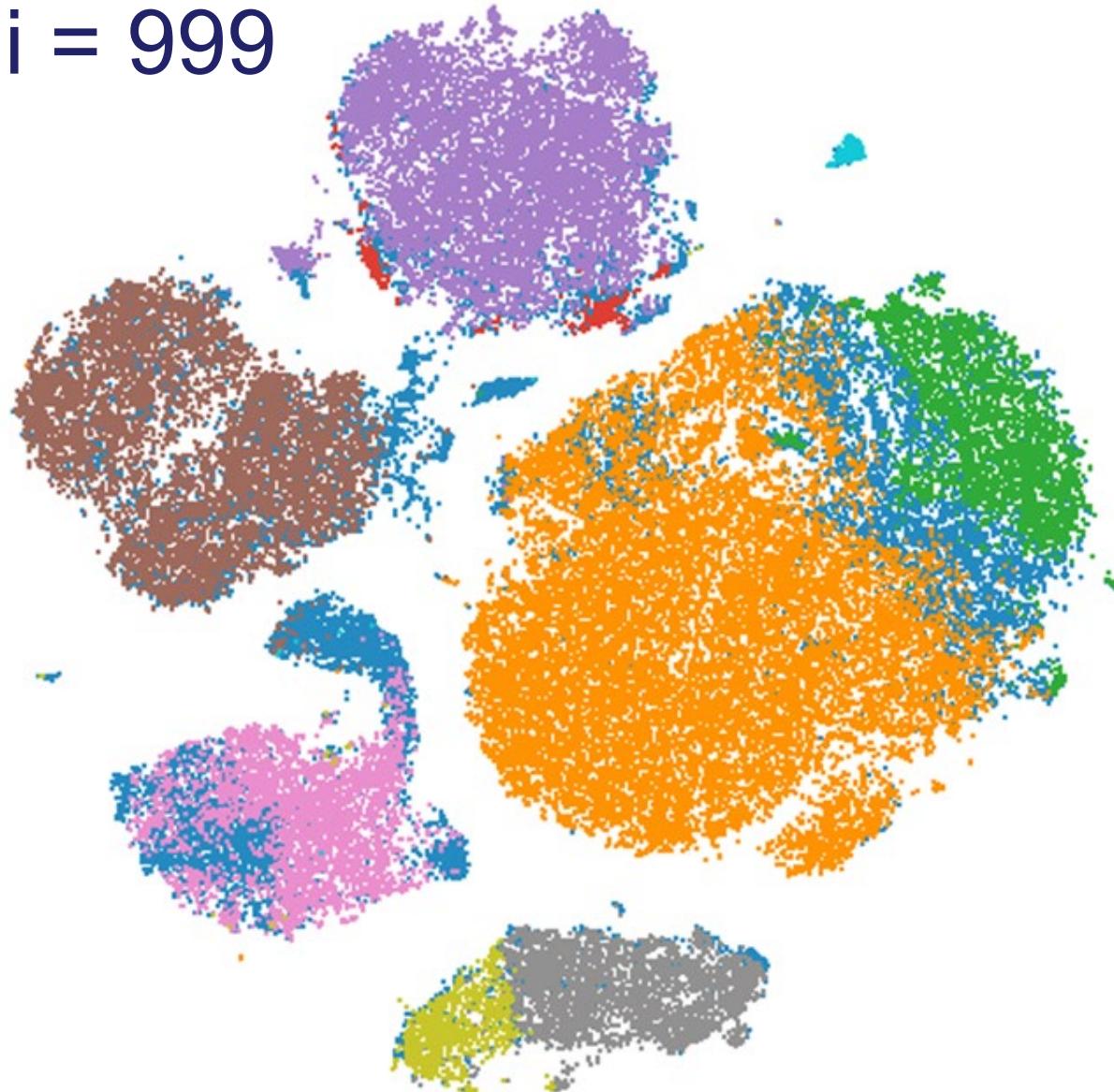


Viewing Expert Gates with viSNE Reveals Cyto Incognito

i = 999

Healthy human blood,
mass cytometry,
26D viSNE analysis

- Ungated
- CD45RA+ Naive CD4+
- CD45RA- CD4+ T cells
- CD45RA- Memory CD8+ T cells
- CD45RA+ Naive CD8+ T cells
- CD16+ NK cells
- CD14+ CD33+ Monocytes
- IgM+ B cells
- IgM- B cells
- CD123+ pDCs

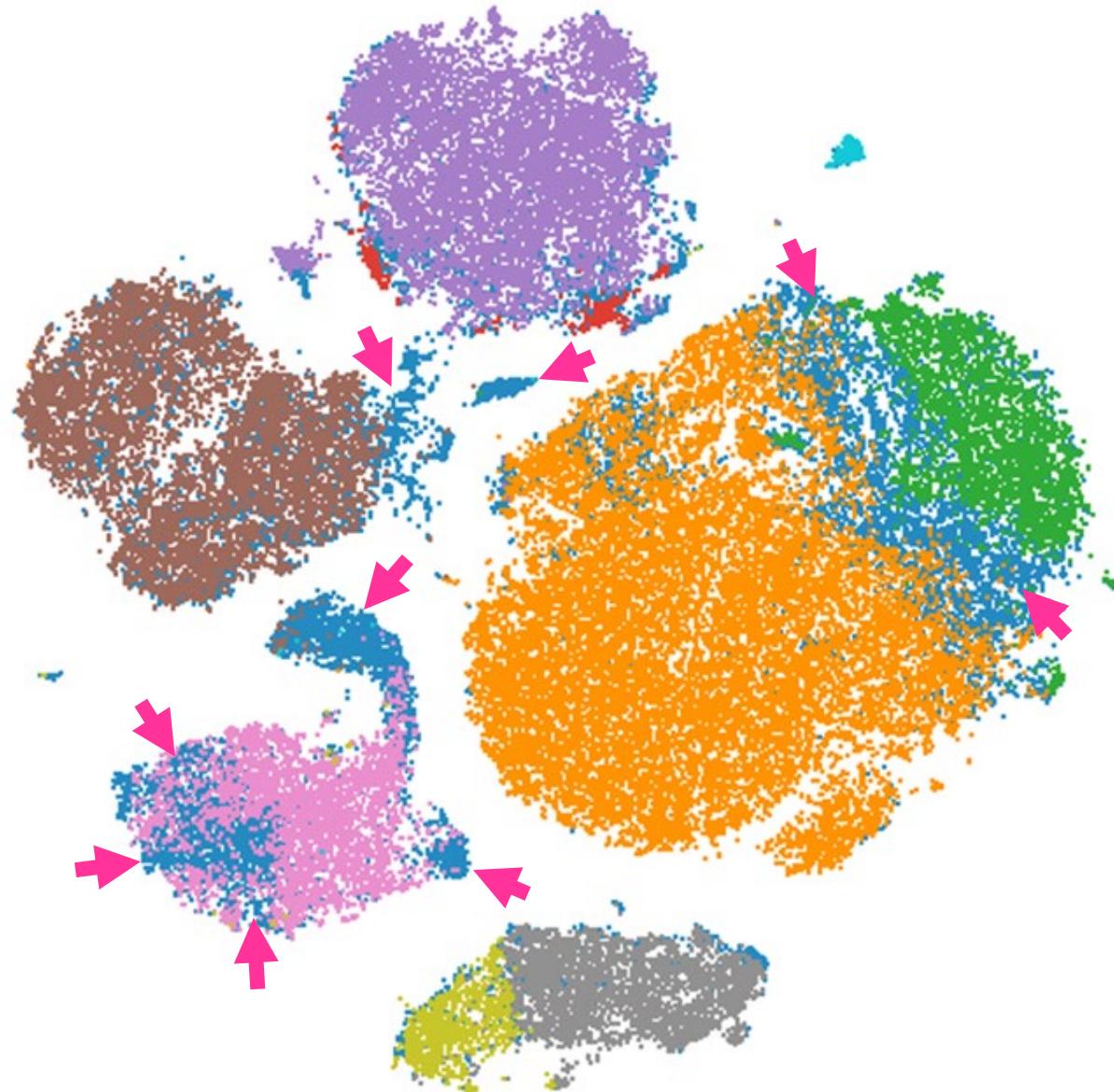


Viewing Expert Gates with viSNE Reveals Cyto Incognito

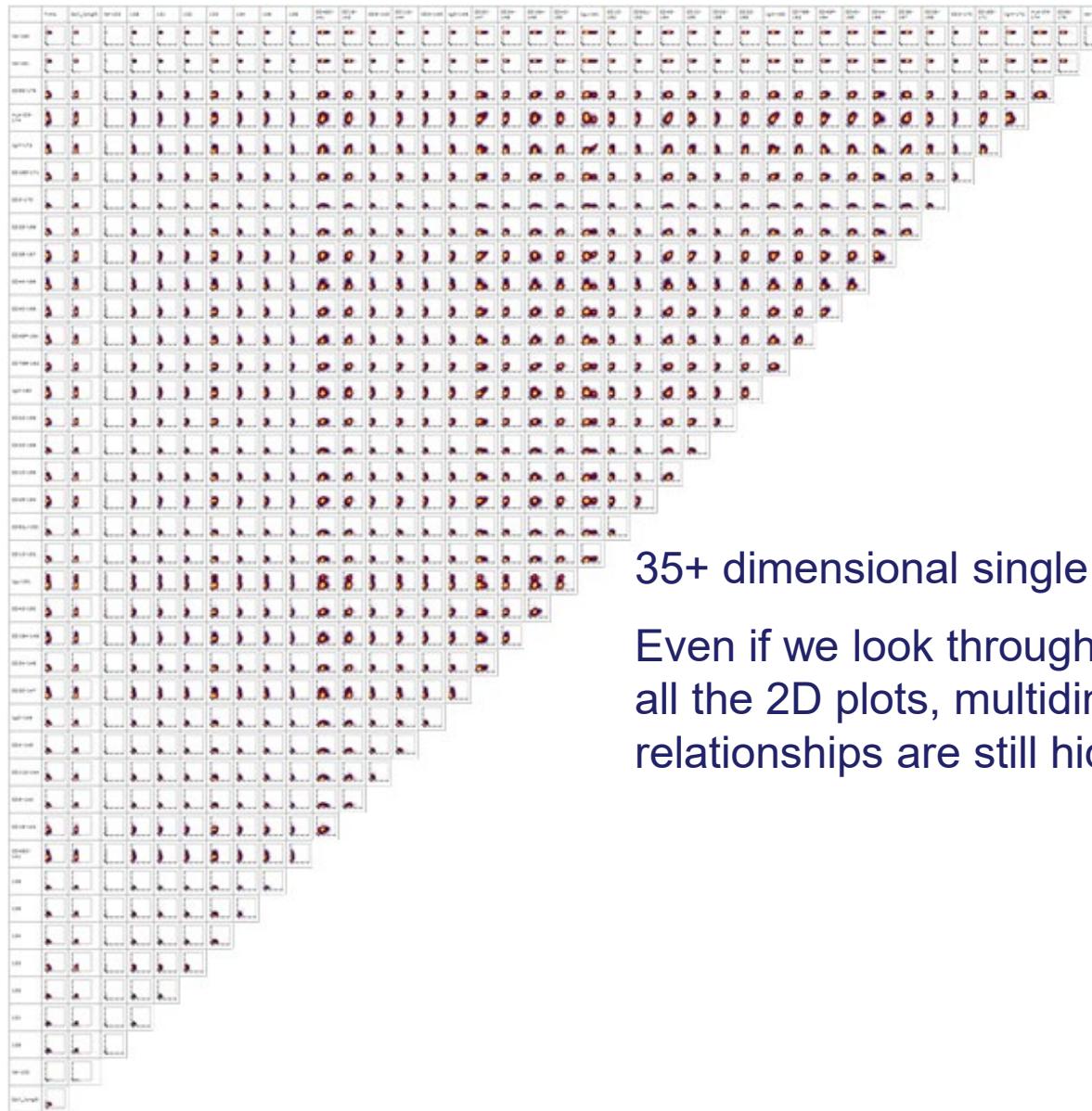
Healthy human blood,
mass cytometry,
26D viSNE analysis

- Ungated
- CD45RA+ Naive CD4+
- CD45RA- CD4+ T cells
- CD45RA- Memory CD8+ T cells
- CD45RA+ Naive CD8+ T cells
- CD16+ NK cells
- CD14+ CD33+ Monocytes
- IgM+ B cells
- IgM- B cells
- CD123+ pDCs

➡ Cyto incognito
(Cells overlooked or
hidden in expert gating)



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...

ISAC FlowRepository: Annotated Public Cytometry Data

flowrepository.org

FLOWRepository

Inbox Annotation Data Invite a User IMPC Admin Public View

Welcome, flowrepository Logout

Help

The following open access article describes how to upload and annotate flow cytometry data sets: Spidlen J, Breuer K and Brinkman R. Preparing a Minimum Information about a Flow Cytometry Experiment (MIFlowCyt) Compliant Manuscript Using the International Society for Advancement of Cytometry (ISAC) FCS File Repository (FlowRepository.org). *Current Protocols in Cytometry*. UNIT 10.18; July 2012.

We also have a [Quick start guide](#) and a [FAQ](#) section.

You may download [slides](#) from our Workshop at CYTO 2012: Publishing MIFlowCyt Compliant Data to ISAC's FlowRepository.org for Cytometry A and Other Journals

Additional links and help options are listed in our [support](#) page.

You can contact us for support regarding technical issues or by sending comments about how to make your experiment or FlowRepository in general better. Leave feedback or ask questions by filling out a [support ticket](#).

Citing FlowRepository

Please reference us by citing: Spidlen J, Breuer K, Rosenberg C, Kotekha N and Brinkman RR. FlowRepository - A Resource of Annotated Flow Cytometry Datasets Associated with Peer-reviewed Publications. *Cytometry A*. 2012 Sep; 91(9):727-31.

Supporting journal

Cytometry A

FlowRepository

FlowRepository is a database of flow cytometry experiments where you can query and download data collected and annotated according to the [MIFlowCyt standard](#). It is primarily used as a data deposition place for experimental findings published in peer-reviewed journals in the flow cytometry field. Those datasets can be queried or browsed using the form and links below. In addition, FlowRepository is hosting the primary flow cytometry data collected by the [International Mouse Phenotype Consortium](#). The goal of the IMPC is to discover functional insight for every gene by generating and systematically phenotyping 20,000 knockout mouse strains. Data generated by this project can be searched using the Search Mouse Data panel at the bottom.

Query FlowRepository

Enter a term to search all publicly available experiments:

[Show query fields](#)

Links

Browse public datasets Browse OMIP datasets Referencing FlowRepository
Browse community datasets Quick start guide FlowRepository Steering Committee & Advisory Board
Browse most popular datasets Submit data Funding

Search Mouse Data

IMPC

Search by MGI gene id or gene symbol. Use * to search for gene groups, e.g., *rik to look for all Riken genes, hox* to look for all hox genes:

<https://flowrepository.org/>



- 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

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.

Defining Your System

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.

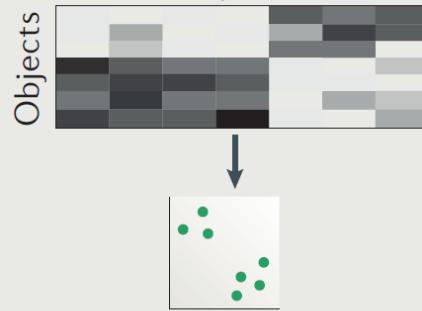
- ▶ Organization of elements (groups, order, etc.), feature relationships.
- ▶ Supervised analysis explicitly uses prior knowledge.
- ▶ Unsupervised analysis looks for patterns without prior knowledge.

There Are Many Ways to Analyze Modern Datasets

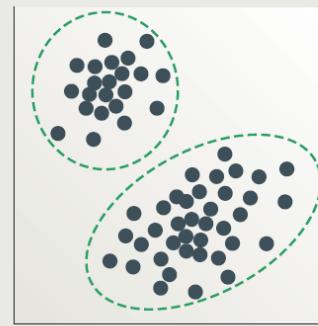
a Unsupervised machine learning: learning structures

Dimensionality reduction

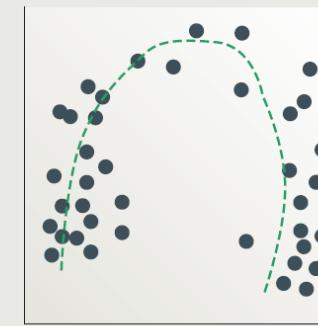
Properties



Clustering

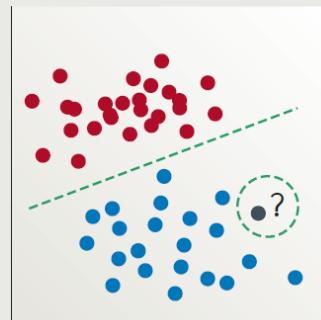


Seriation

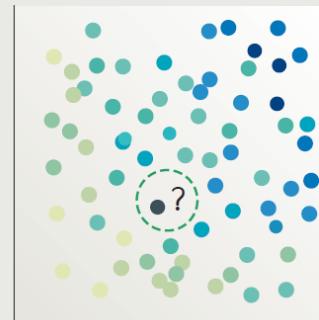


b Supervised machine learning: learning from examples

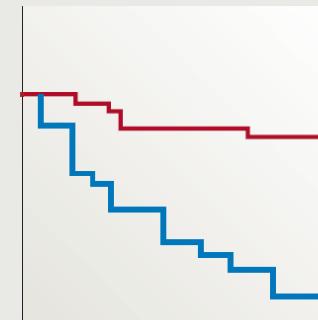
Classification



Regression



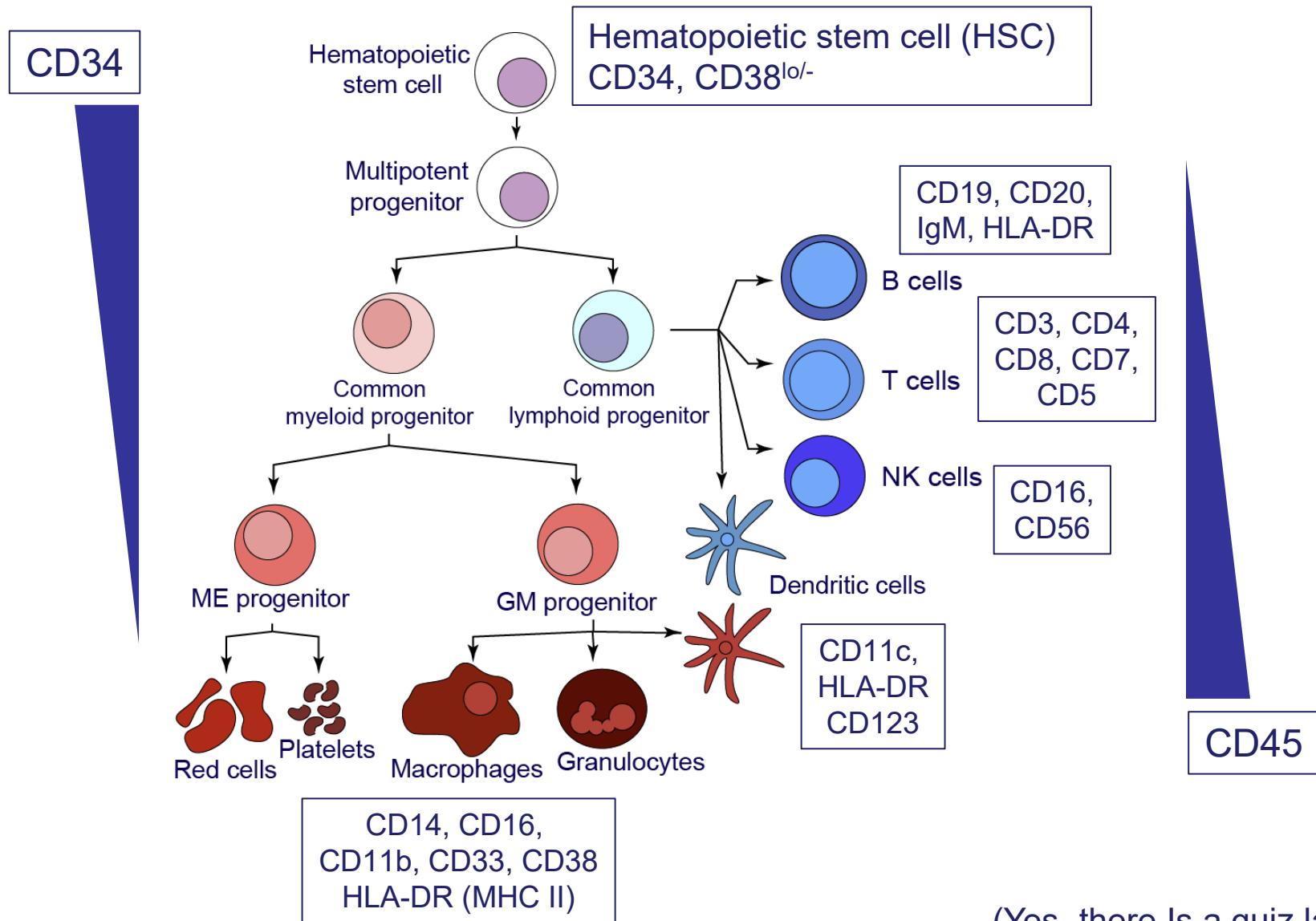
Survival analysis



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



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

Diggins et al., *Methods* 2015

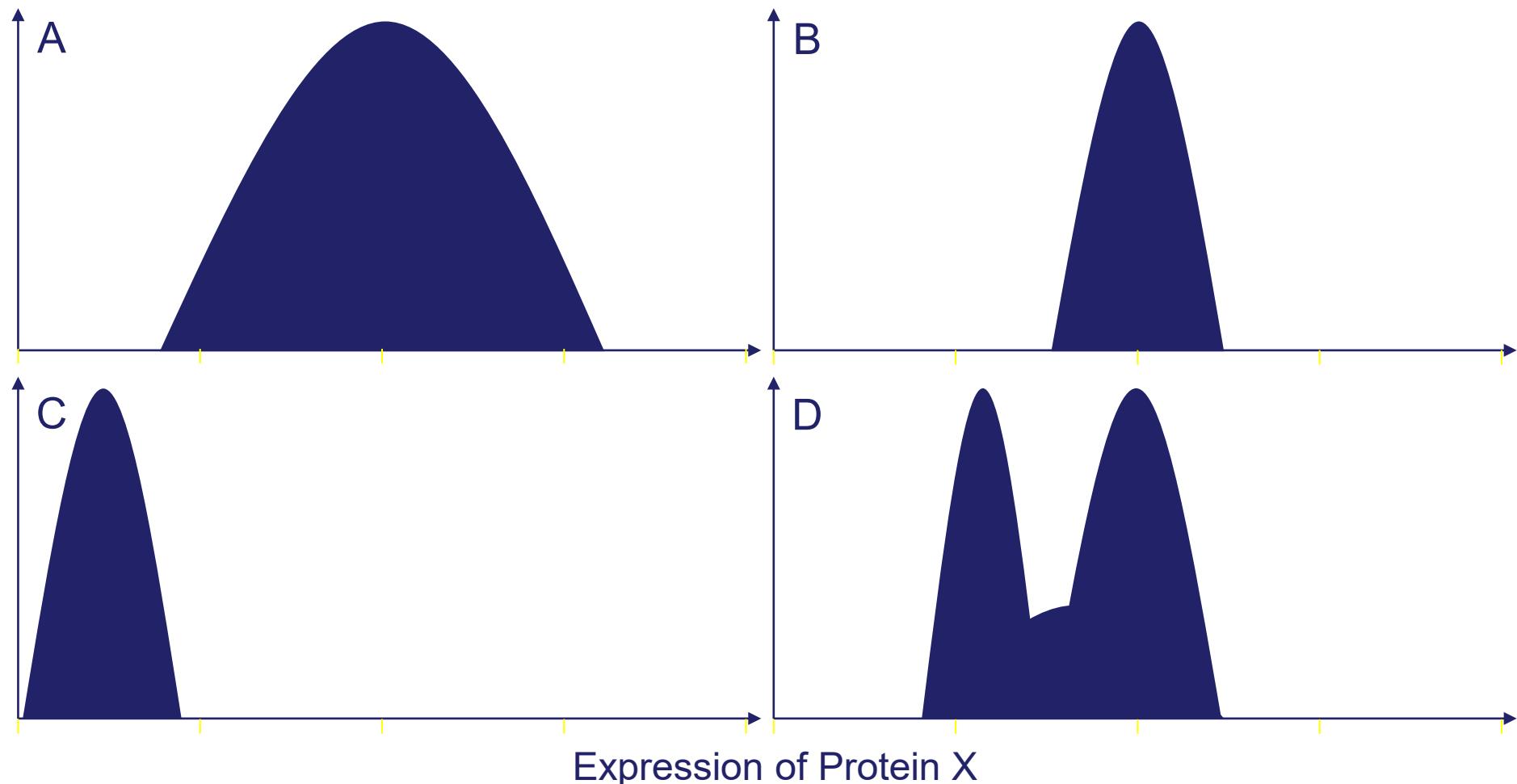
Populations are often labeled by metaphors of function
 (“cancer stem cells”, “central memory T cells”)
 or incomplete labels based on a few features (e.g. “PD-1+ CD8 T cells”).

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

Enrichment Tracks Feature Exclusivity In a Subset

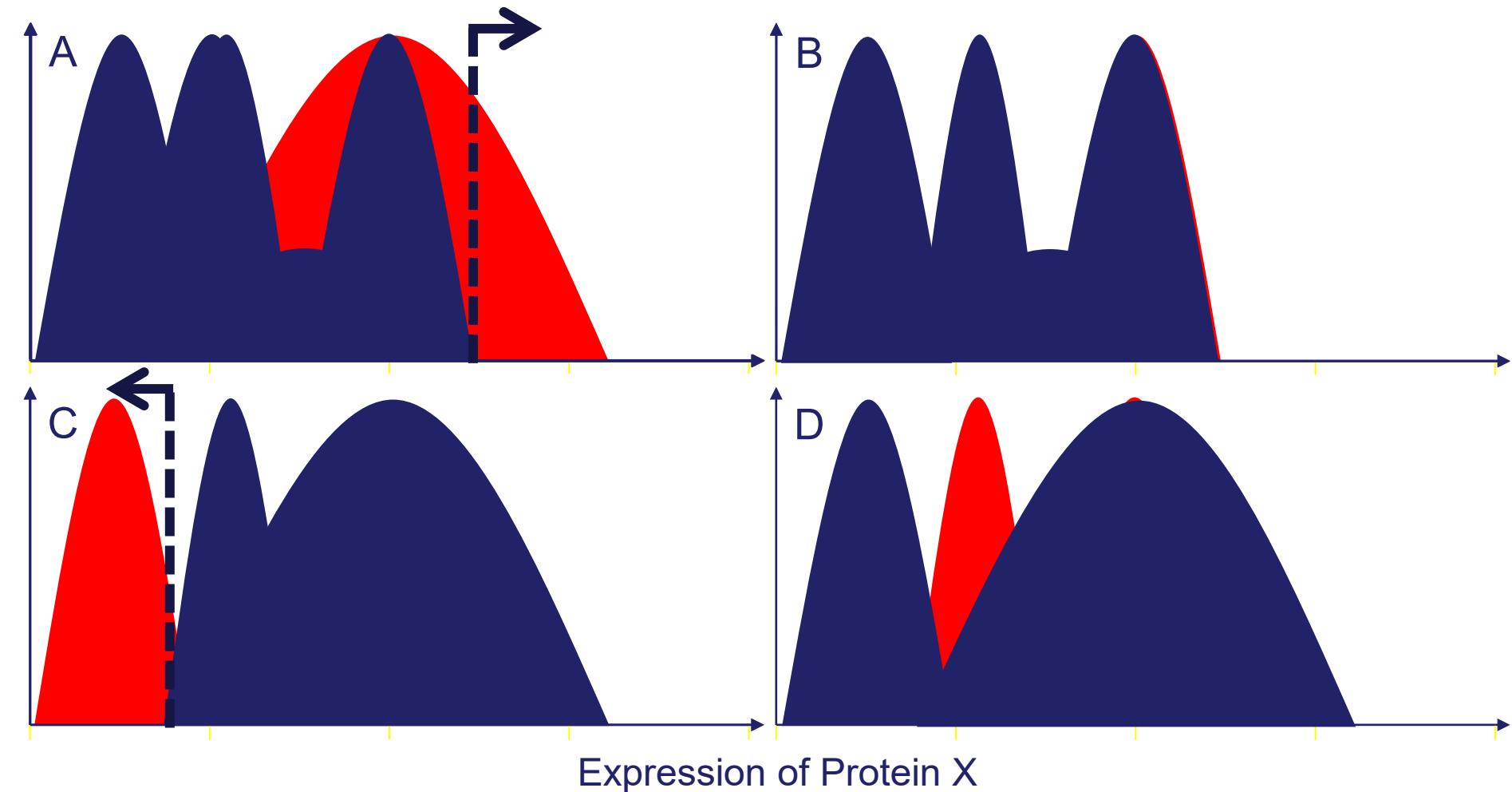
A, B, C, and D are 4 cell types within the same sample and they are each 25% of the sample. If I were to offer to pay you based on the purity of a sort based only on protein X, which population would you want to isolate?

I'll pay you \$1 per correct cell & take away \$1 for each incorrect cell and \$1 for each cell you miss.

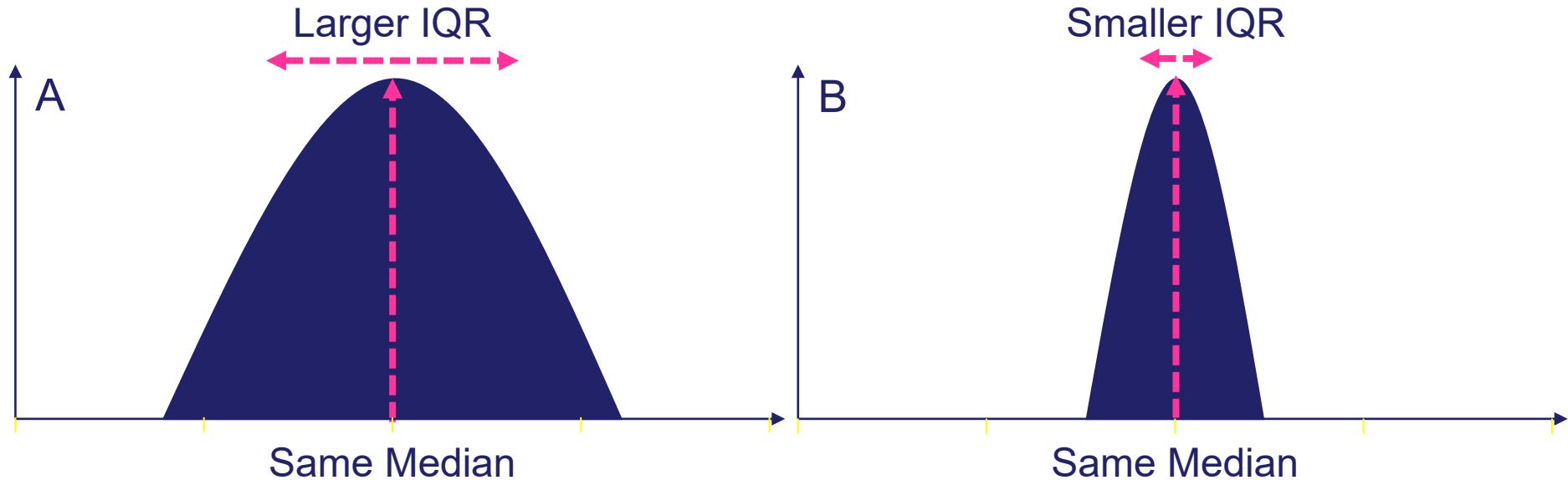


Enrichment Tracks Feature Exclusivity In a Subset

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.)

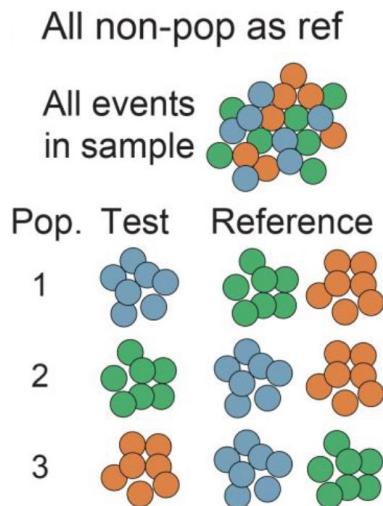
MEM Quantifies Relative Enrichment By Combining Magnitude & Interquartile Range

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



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

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

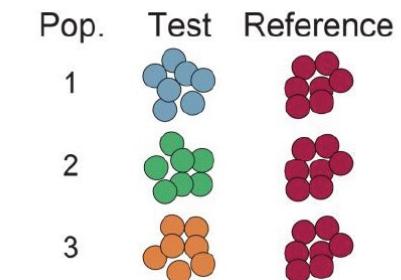


MEM label (CD19⁺ cells)

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

▼ CD7⁻²

Standard control reference pop



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%

Natural killer cells

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

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%

CD8⁺ T cells

▲ CD8⁺⁸ CD7⁺⁵ CD3⁺³
CD45RA⁺³ CXCR4⁺² 9.25%

Early myeloid cells

▲ MHCII⁺⁹ CD33⁺⁸ CD38⁺⁵
CD4⁺³ CD15⁺² CD14⁺²
▼ CD45⁻² CD7⁻² 0.02%

CD4⁺ T cells

▲ CD4⁺⁷ CD7⁺⁵ CD3⁺⁵
CD47⁺² CD45RA⁺² 8.12%

Monocytes

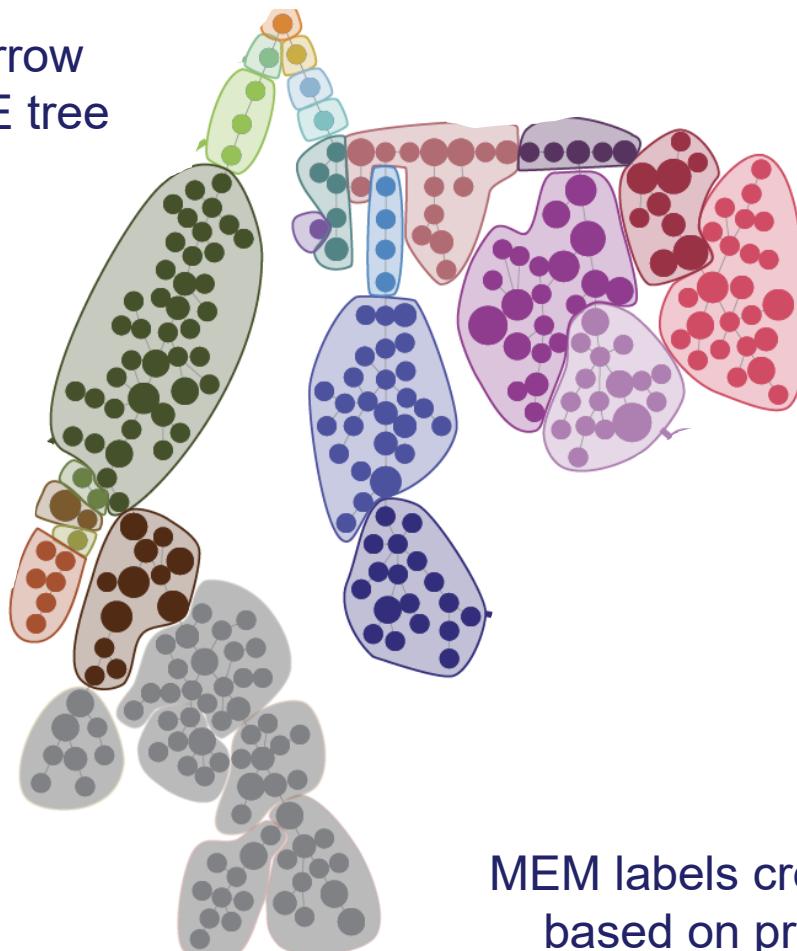
▲ CD33⁺¹⁰ CD14⁺⁸ CD11b⁺⁷
MHCII⁺⁵ CD4⁺⁴ CD11c⁺⁴
CD38⁺⁴ CD13⁺³
▼ CXCR4⁻² CD47⁻² 10.57%

B cells

▲ MHCII⁺¹⁰ CD20⁺⁹ CD19⁺⁷ IgM⁺⁵ CD34⁺³
CD45RA⁺³ CXCR4⁺² CD47⁺² CD33⁺²
▼ CD7⁻² 2.44%

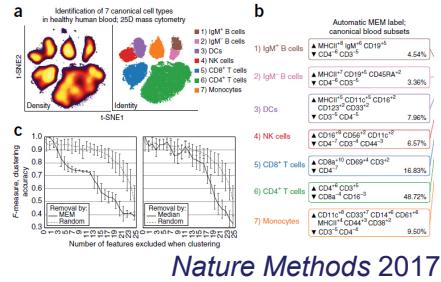
Marker Enrichment Modeling Automatically Labels Cell Types in Human Bone Marrow Using -10 to +10 Enrichment Values

Cells from bone marrow grouped in a SPADE tree



MEM labels created automatically based on protein enrichment

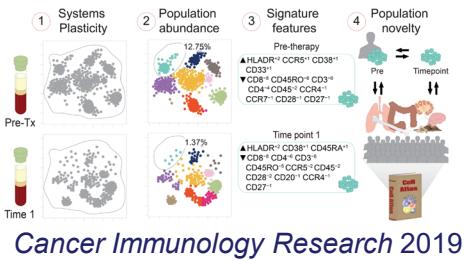
Tools for Automated Cell Discovery & Characterization



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



Greenplate et al., PMC6318034

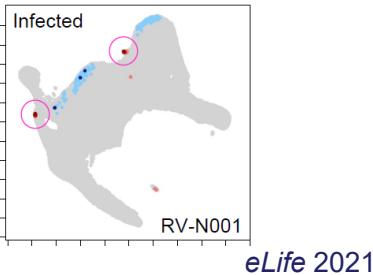
- Set of tools for **longitudinal** single cell tumor immunology.
- Revealed abnormal immune cells in multiple tumor types.
- Includes datasets (AML, melanoma) used by T-REX
- Greenplate+ vs. COVID-19 (*Science* 2020 PMC7263500)

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 2020



T-REX

Barone, Paul, Muehling et al., PMC8370768

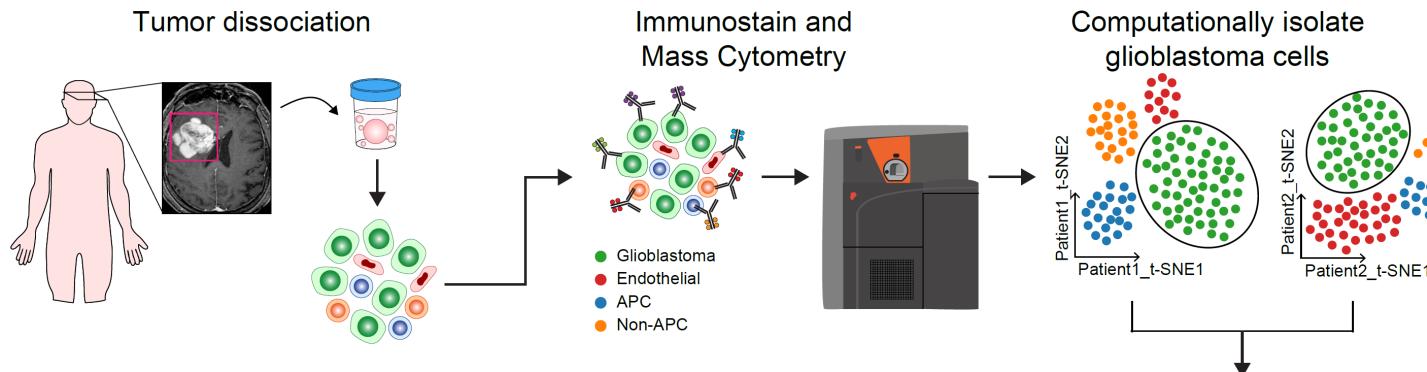
- T-REX: compares a pair of samples (e.g. pre- and post-)
- Revealed rhinovirus-specific cells based on rapid expansion
- Revealed identity of memory T cells in SARS-CoV-2 vaccine response (*bioRxiv* preprint PMC8328055)



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

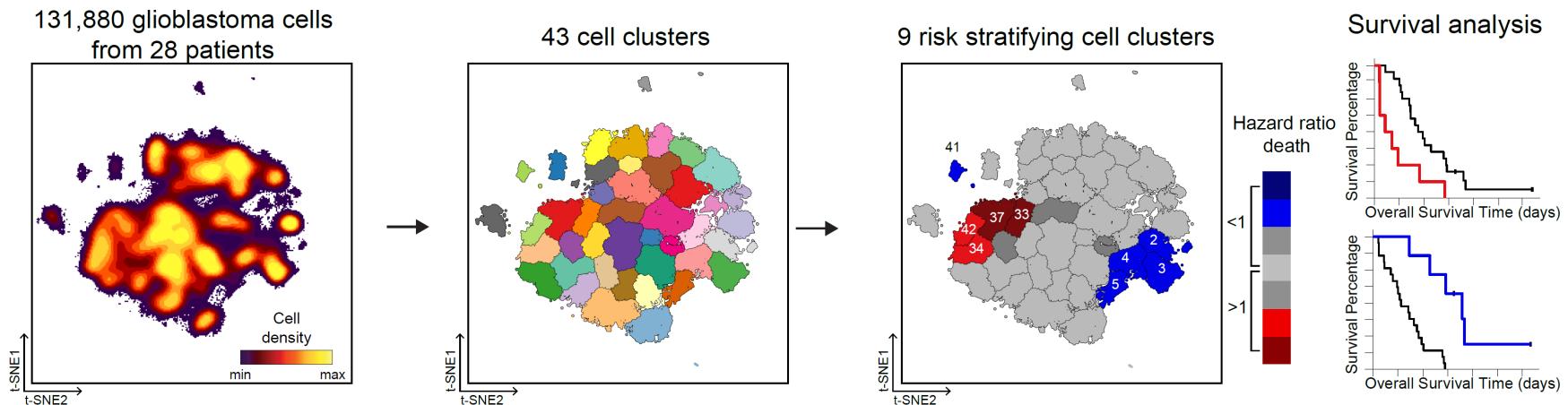
a

Tumor Preparation and Mass Cytometry Dataset Generation



b

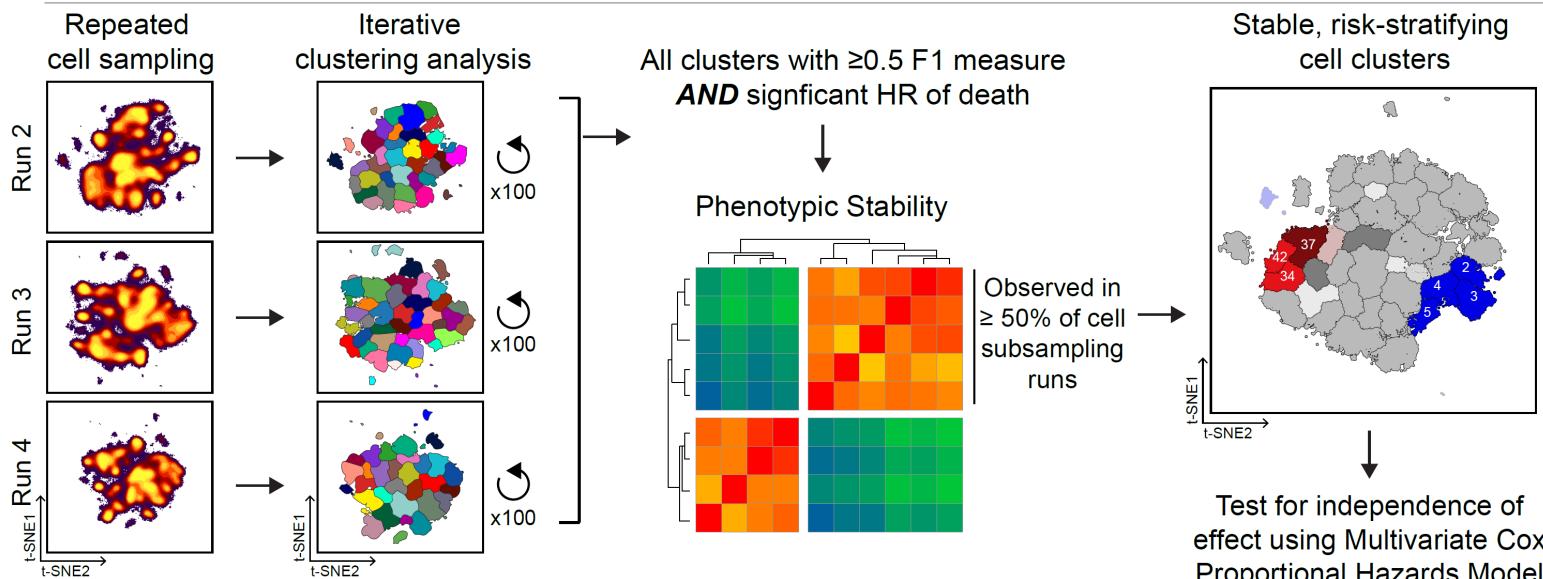
Risk Assessment Population IDentification (RAPID)



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

C

Cluster and Phenotypic Stability Testing



D

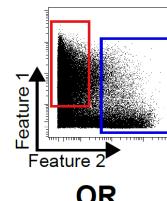
Biological Validation

Feature selection for validation

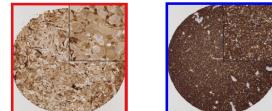
- ▲ Feature1⁺⁵ FeatureX⁺³
- ▼ Feature2⁻⁷ FeatureY⁻²

- ▲ Feature2⁺⁵ FeatureY⁺³
- ▼ Feature1⁻⁷ FeatureX⁻²

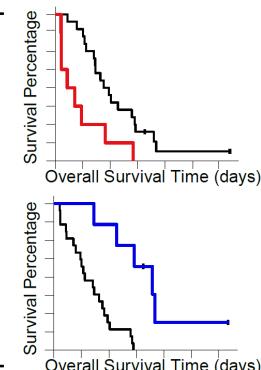
Low dimensional gating strategy



Immunohistochemistry staining



Survival analysis



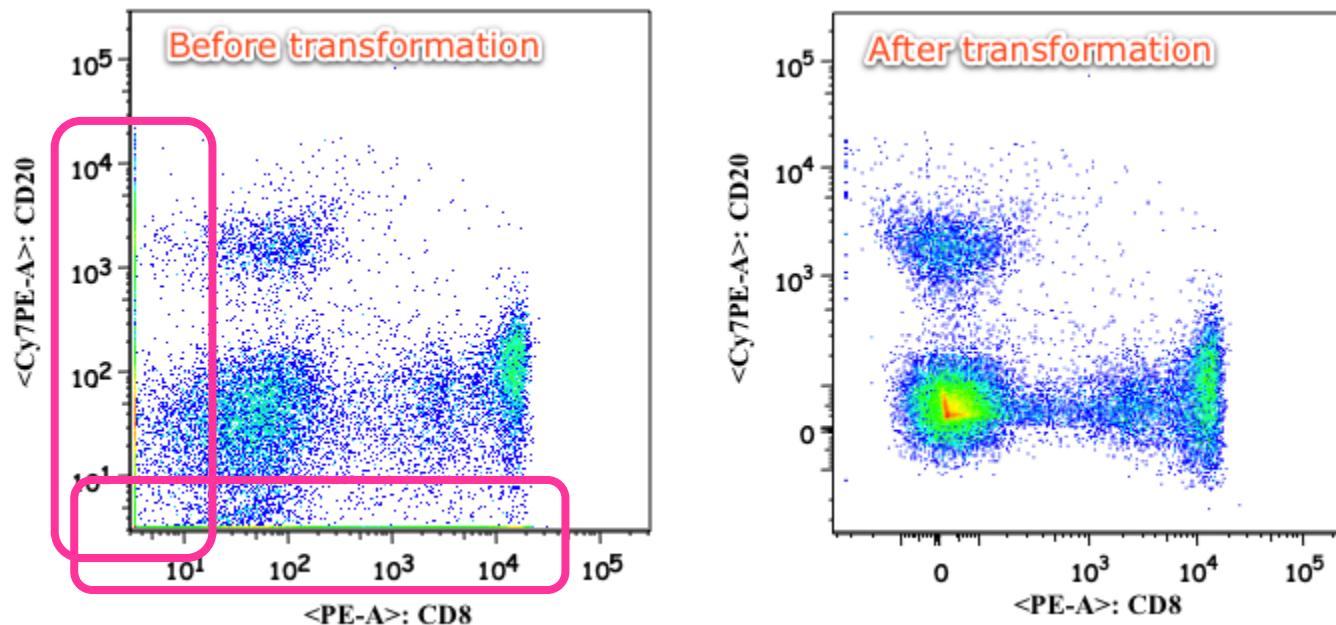
The Importance of Scales

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 the cells that are biologically 100% negative for a marker?



<http://www.flowjo.com/v76/en/displaytransformwhy.html>

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

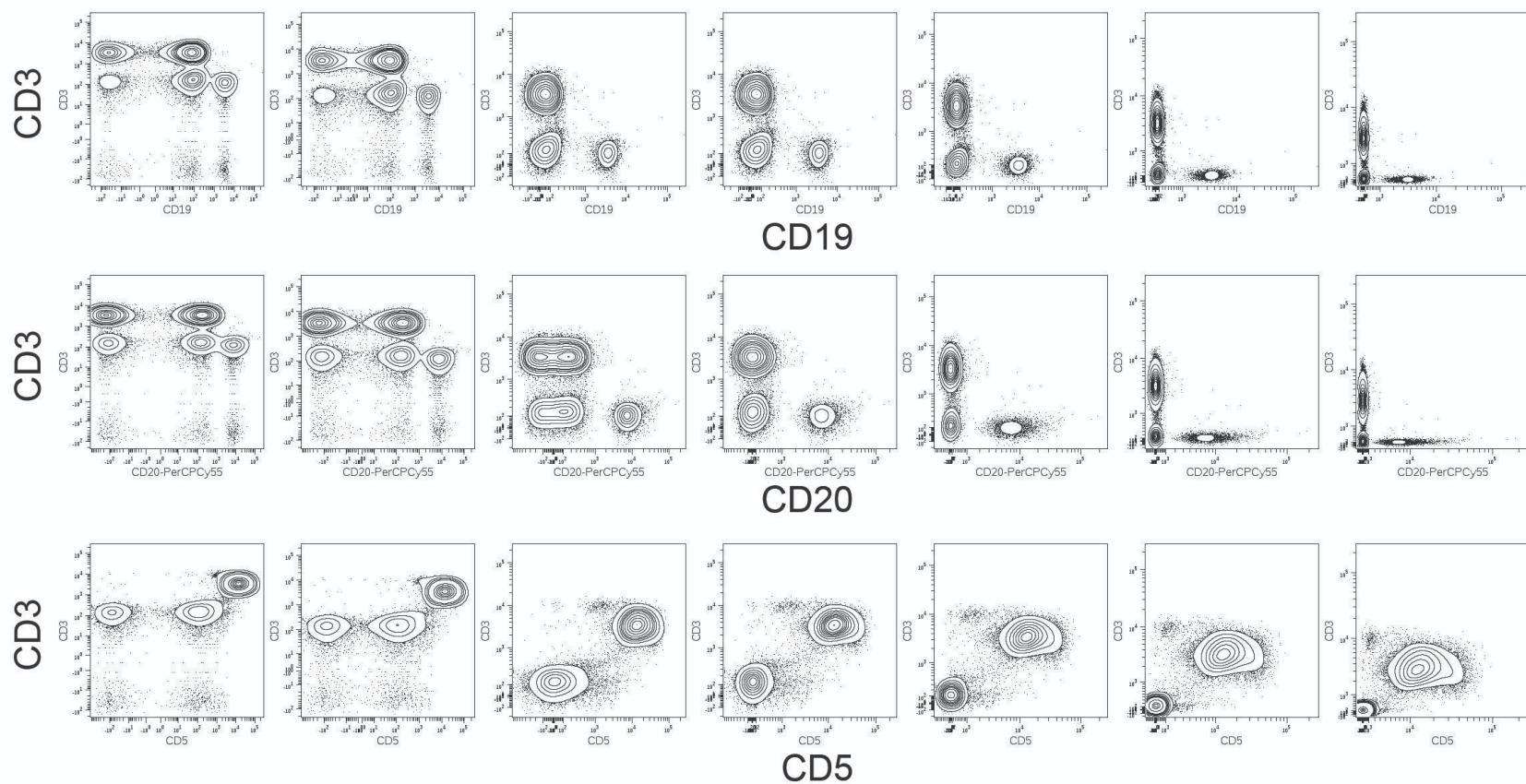
Scaling is important in both mass and fluorescence cytometry.

Scaling Matters for Measuring Distance (Fluorescence Flow)

Healthy human PBMC, intact cells gate

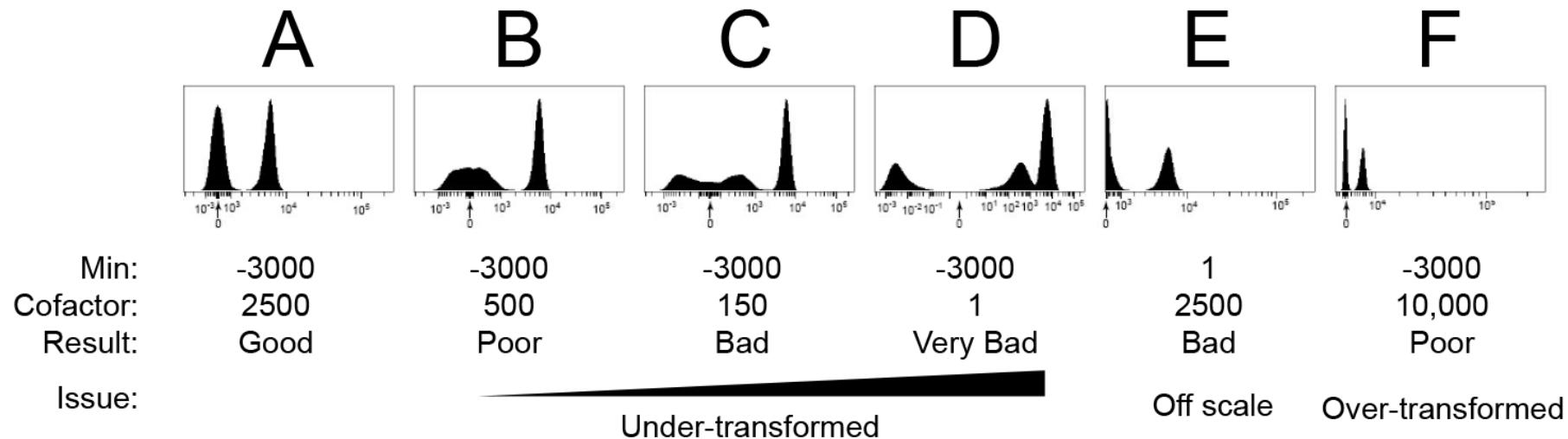
arcsinh(intensity / argument)

CD3 arg	1	5	150	150	300	750	1500
CD19 arg	1	5	150	150	300	750	1500
CD20 arg	1	5	150	500	1500	3000	6000
CD5 arg	1	5	150	500	1500	3000	6000



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.



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

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

Since computational analysis techniques compare distance similar to what a person does when looking at a plot, these techniques can identify artificial populations near zero (see C and D) if data are not appropriately transformed prior to analysis.

Examples of Four Common Data Scales

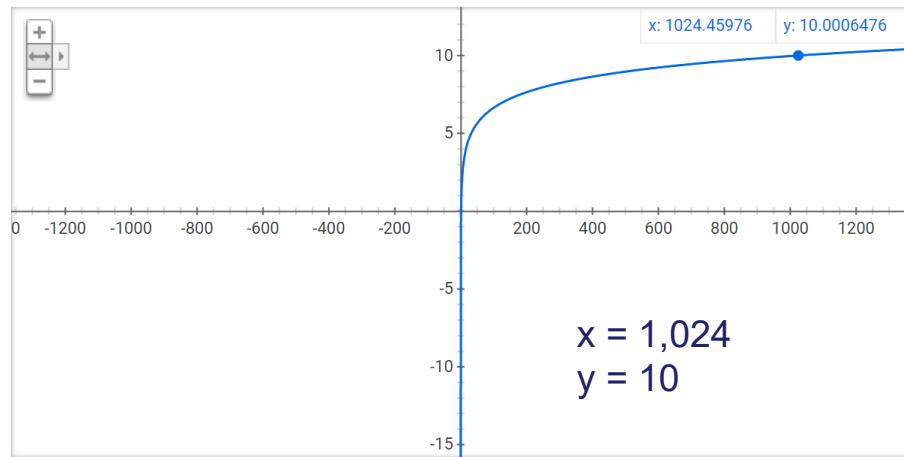
$\log_2(x)$

$\log_{10}(x)$

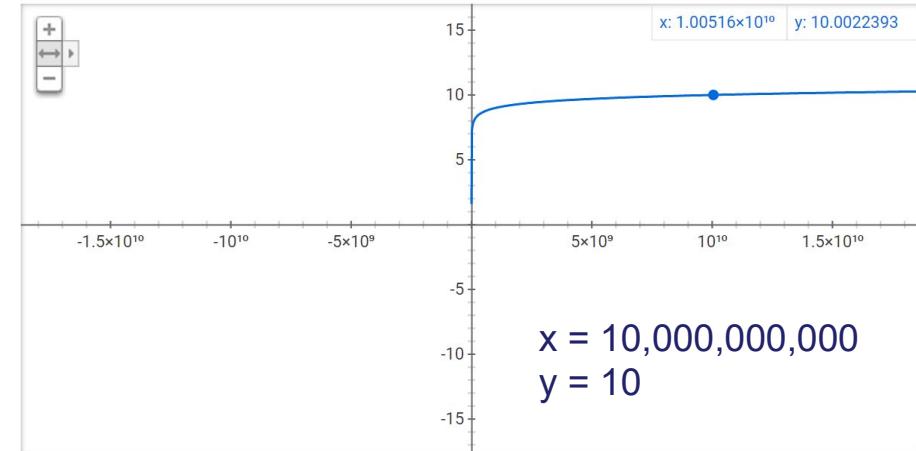
$\text{arcsinh}(x/5)$

$\text{arcsinh}(x/150)$

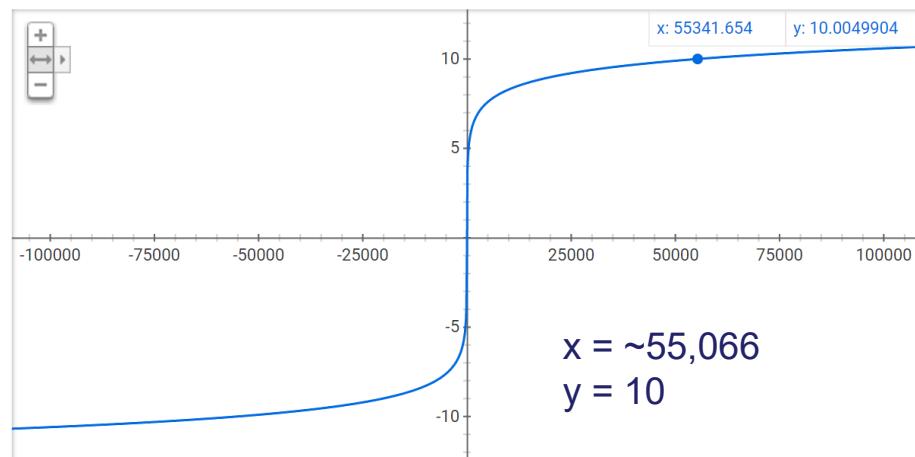
Graph for $\ln(x)/\ln(2)$



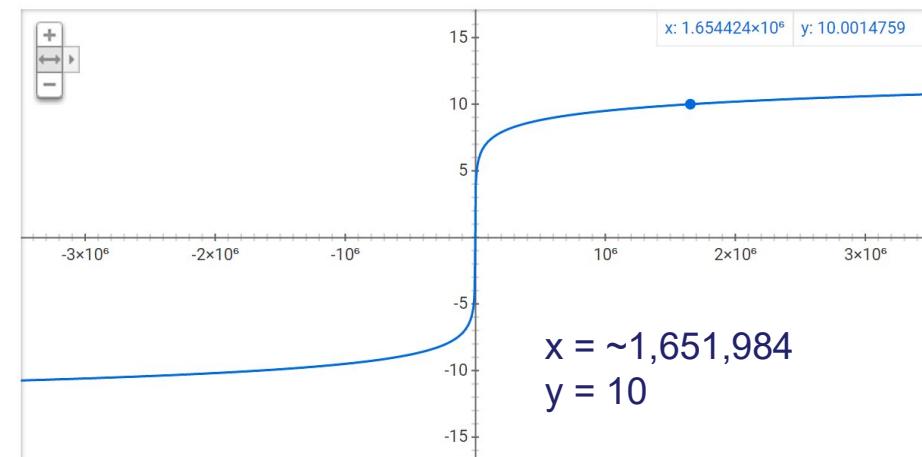
Graph for $\ln(x)/\ln(10)$



Graph for $\ln(x/5+\sqrt{1+(x/5)^2}) = \text{arcsinh}(x/5)$



Graph for $\ln(x/150+\sqrt{1+(x/150)^2}) = \text{arcsinh}(x/150)$



Linear: Do I consider change from 1 to 10 equivalent to from 1001 to 1010? Or is it equivalent to 1000 to 10,000?

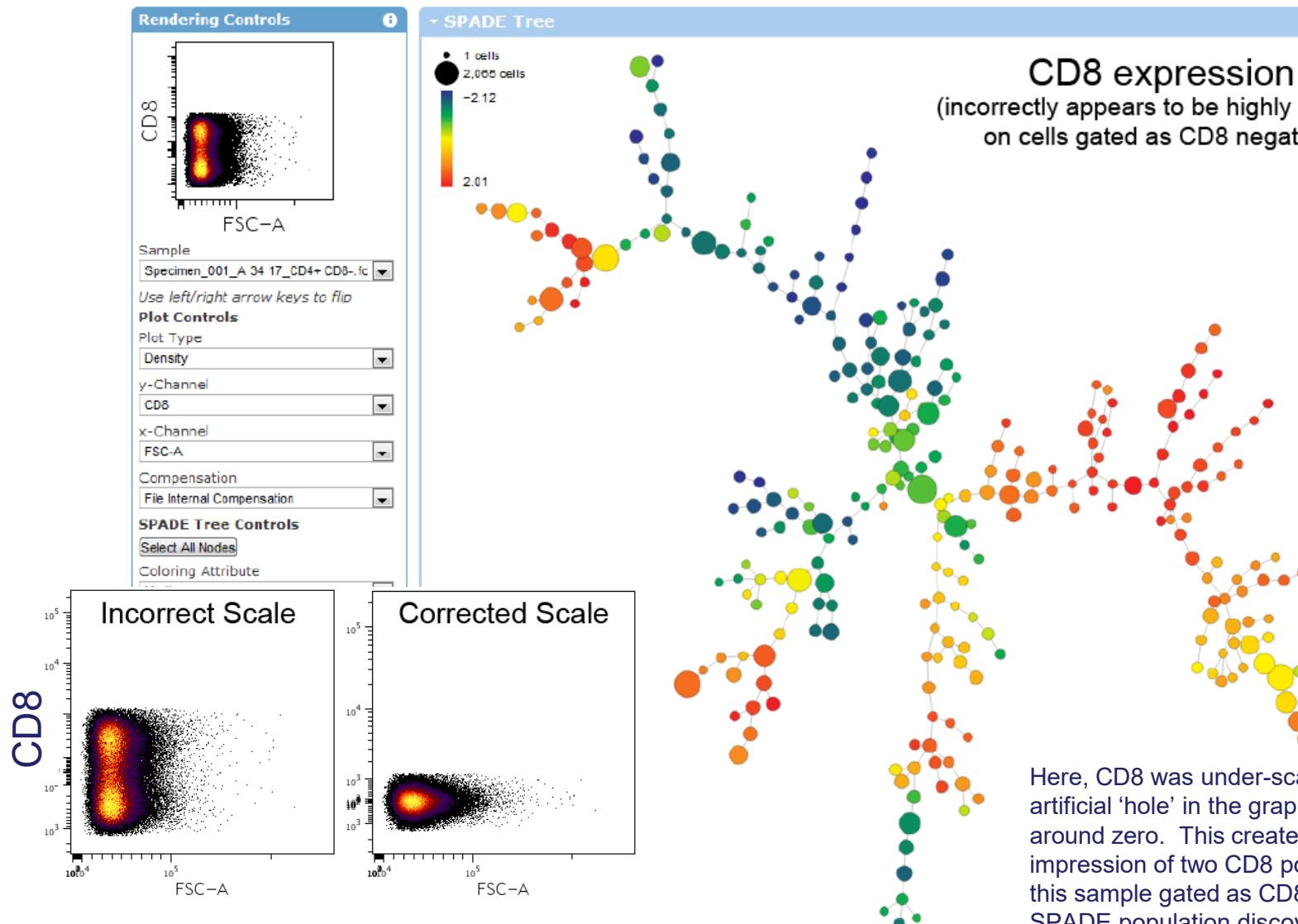
Log10: If $x = 1024$, $y = 3.0$

arcsinh5: If $x = 1024$, $y = 6.0$

Log2: If $x = 1024$, $y = 10$

Continuous: If $x = 0$, what is y ? Can I plot a negative number? Symmetrical: Is -10 to -1 the same change as from 1 to 10?

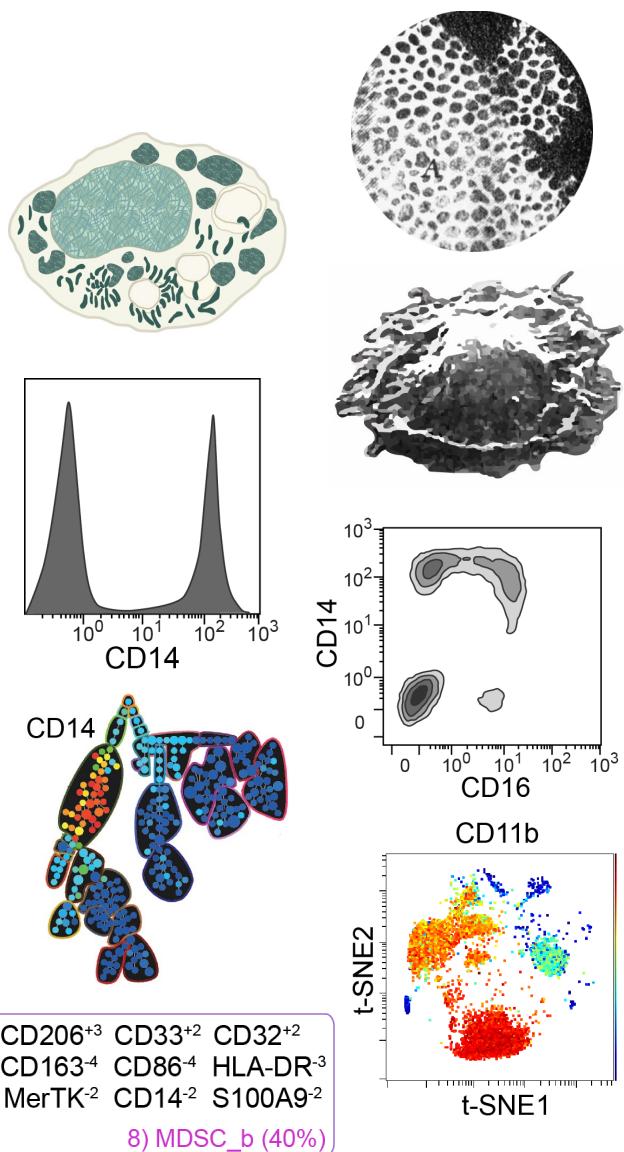
Inappropriate Scaling Can Lead to False Population Discovery



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

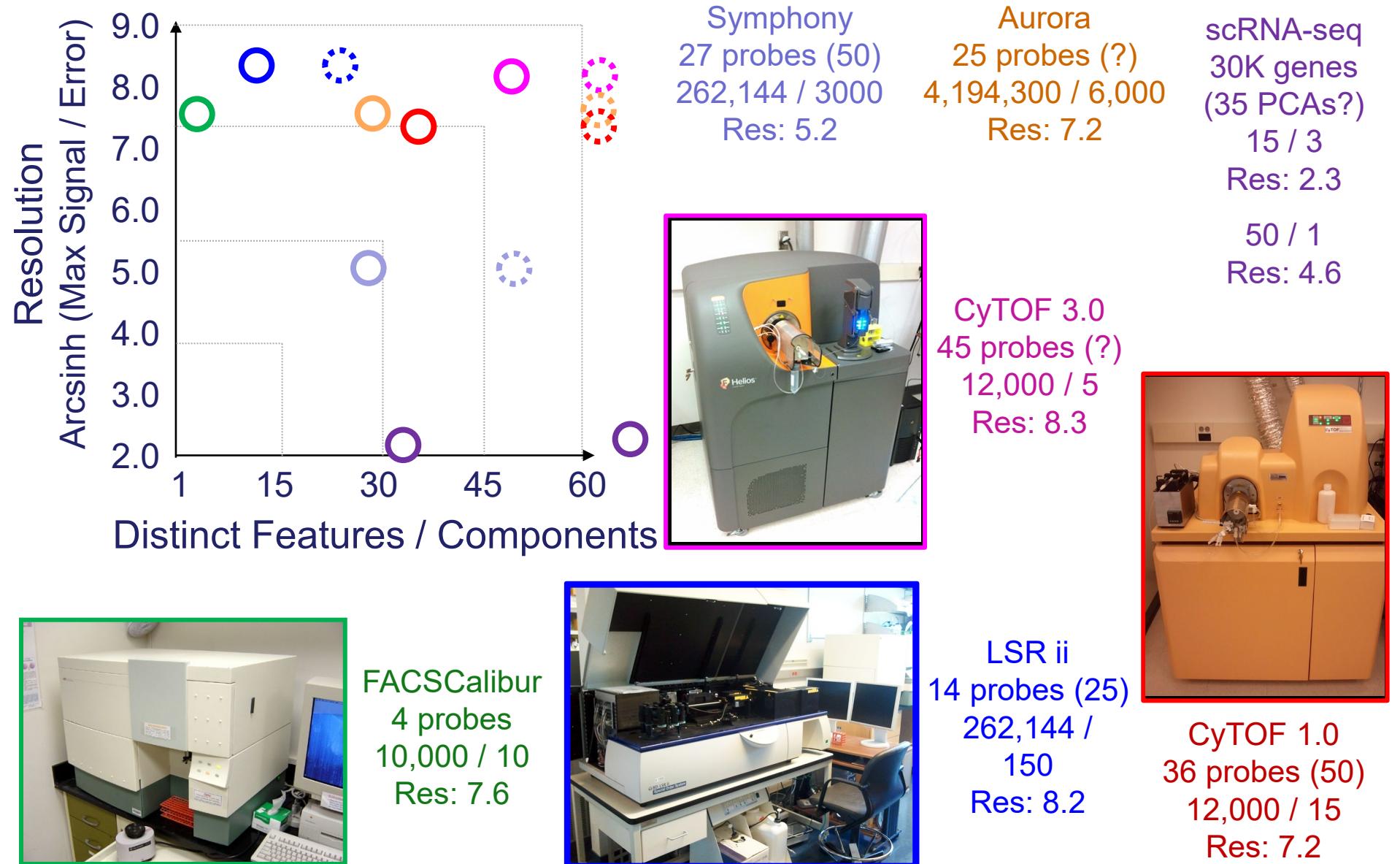


* Robert Hooke describes 'cells' in *Micrographia: or Some Physiological Descriptions of Miniature Bodies Made by Magnifying Glasses*

** 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

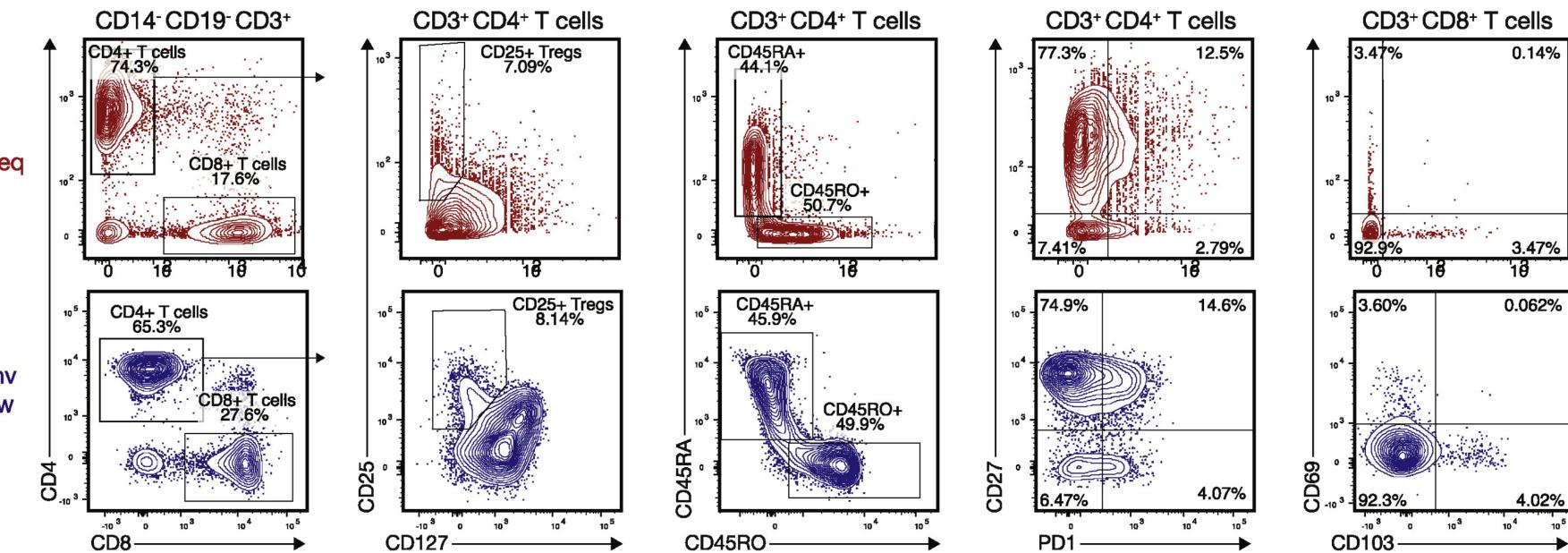
Cytometry Probes & Instruments Keep Improving (ca. 2019)



Scaling Matters for Measuring Distance

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

D



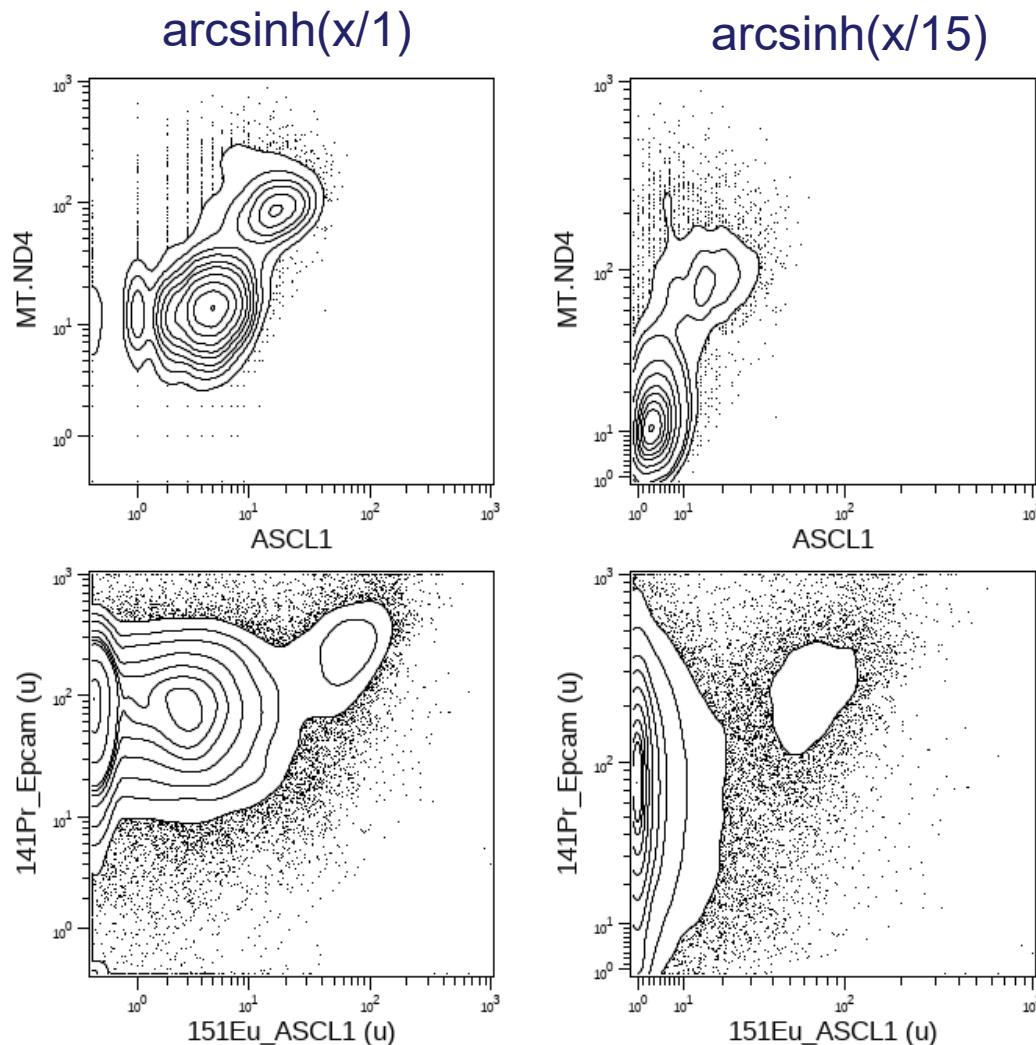
Mair et al., Cell Reports 2020

Example Challenge, Scaling to Compare scRNA-seq vs. CyTOF (Here with one example cell line, DMS454, and ASCL1 on the x-axis)

DMS454
SCLC cell line

scRNA-seq

CyTOF



Qualitative: Is there a subset of cells that is ASCL1⁺?

Quantitative: How much ASCL1 is in each cell? How much does ASCL1 change after Tx?