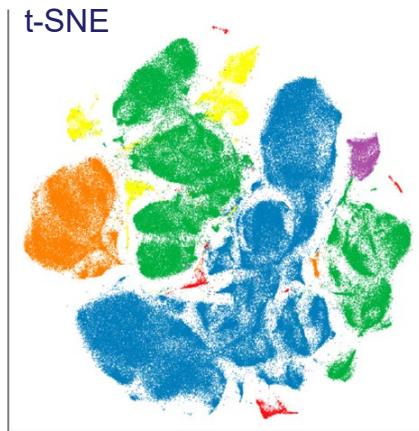
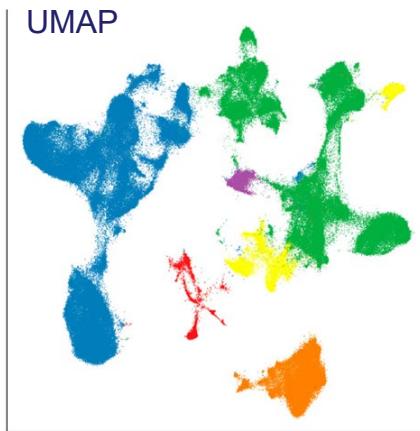


# Analysis II: Tailoring Cytometry Data Science Workflows (90 min)

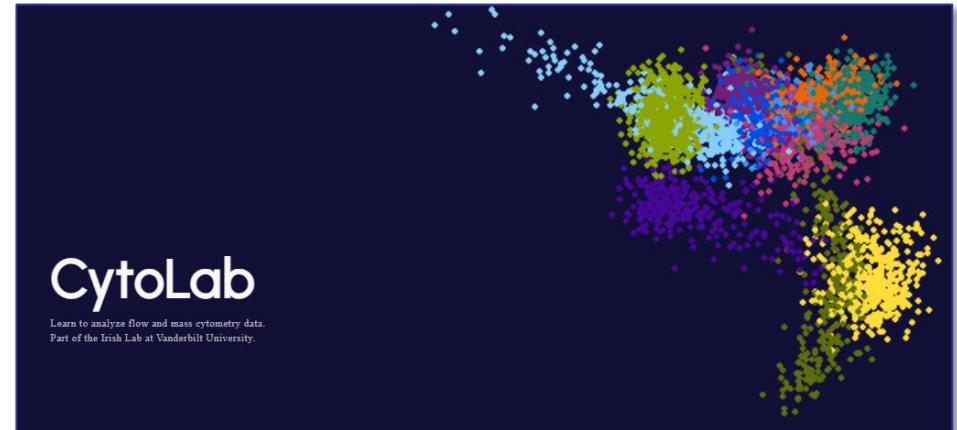
Starting Poll: <https://www.menti.com/>

Poll Code: 7939 8162

# Analysis II: Tailoring Cytometry Data Science Workflows



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](mailto:jonathan.irish@vanderbilt.edu)



**Cass Mayeda**

Web Applications  
Research Assistant  
Vanderbilt University

[cass.mayeda@vanderbilt.edu](mailto:cass.mayeda@vanderbilt.edu)



**Josef Spidlen**

Senior Director R&D  
BD Biosciences

[josef.spidlen@bd.com](mailto:josef.spidlen@bd.com)



**Nicolas Loof**

Senior Application Scientist  
Multi-omics Specialist  
BD Biosciences

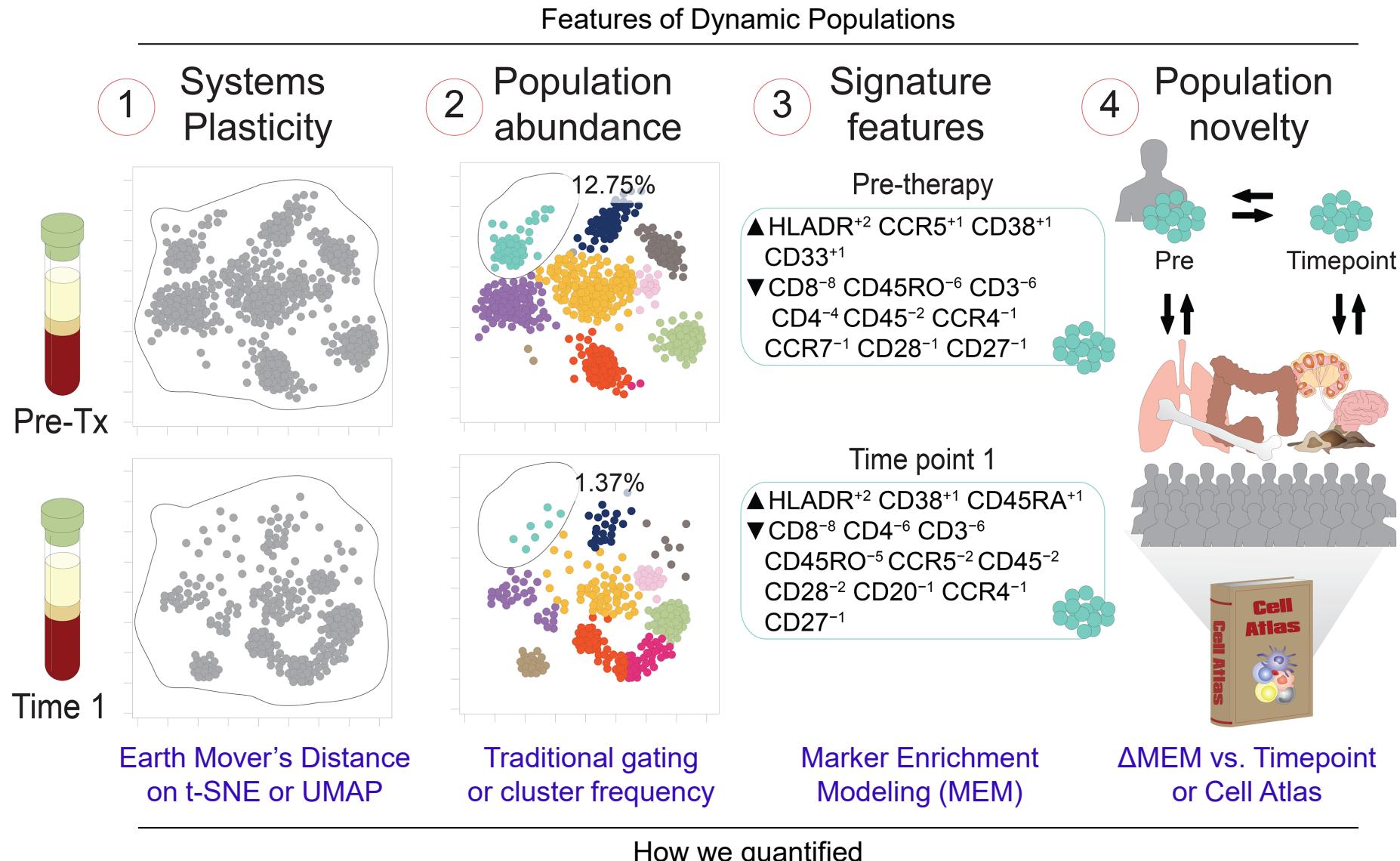
[nicolas.loof@bd.com](mailto:nicolas.loof@bd.com)

# Systems Immune Monitoring & Tailoring Workflows

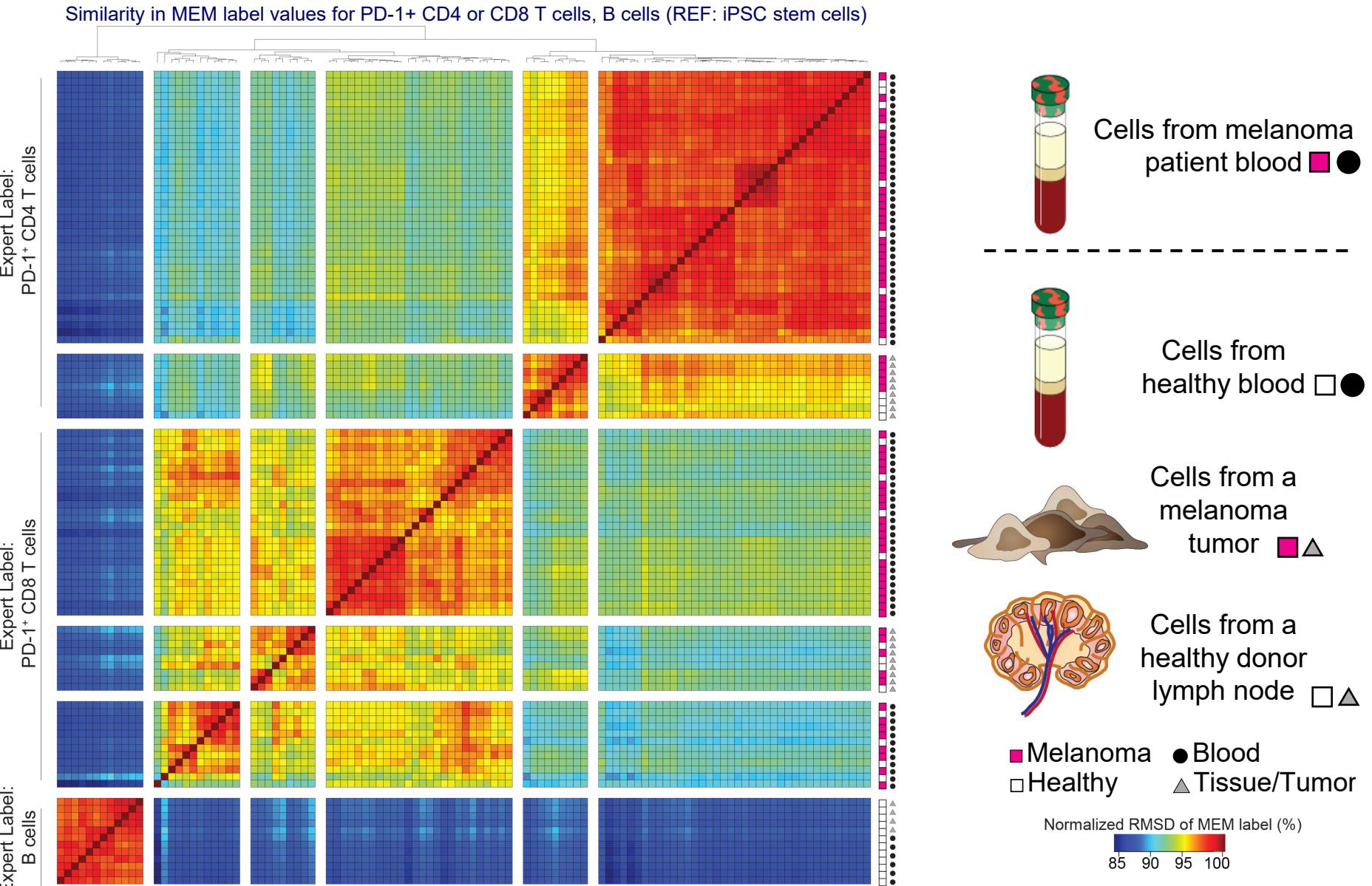
# Samples Over Time Reveal Immune System Dynamics

Comparisons with Earth Mover's Distance,  
Root Mean Square Deviation (RMSD),  
and Change in MEM label ( $\Delta$ MEM)

# Clinical Trial Monitoring: What Do We Need to Know? Automate Four Key Readouts vs. Clinical Outcomes



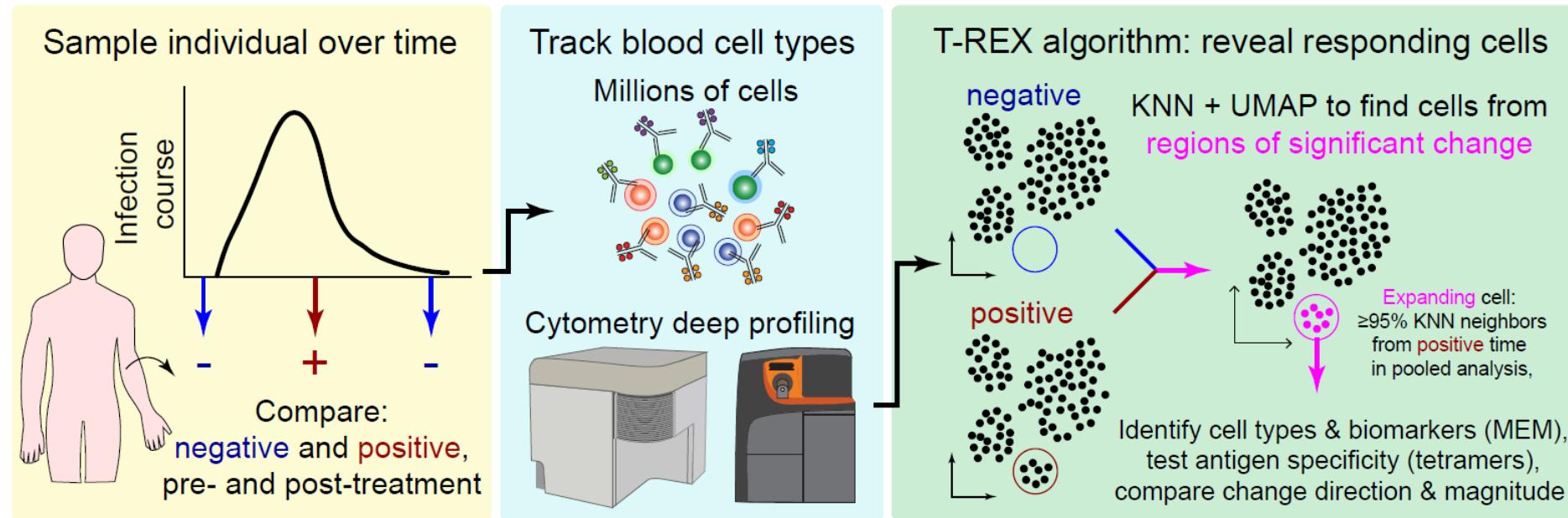
# Distinct Phenotypes of PD-1<sup>+</sup> CD8<sup>+</sup> T cells in Melanoma Tumors Revealed by Quantitatively Comparing MEM Text Labels



Data files: <http://flowrepository.org/id/FR-FCM-ZYCC>

Greenplate et al., *Cancer Immunology Research* 2019  
Methods: Diggins et al., *Nature Methods* 2017; *Curr Prot Cyt* 2018

# RAPID & T-REX Are Both Unsupervised, RAPID: Continuous Outcomes vs. T-REX: Categorical Groups



T-REX (Tracking Responders EXpanding) identifies phenotypic hotspots undergoing great change between conditions (e.g., +/- infection)

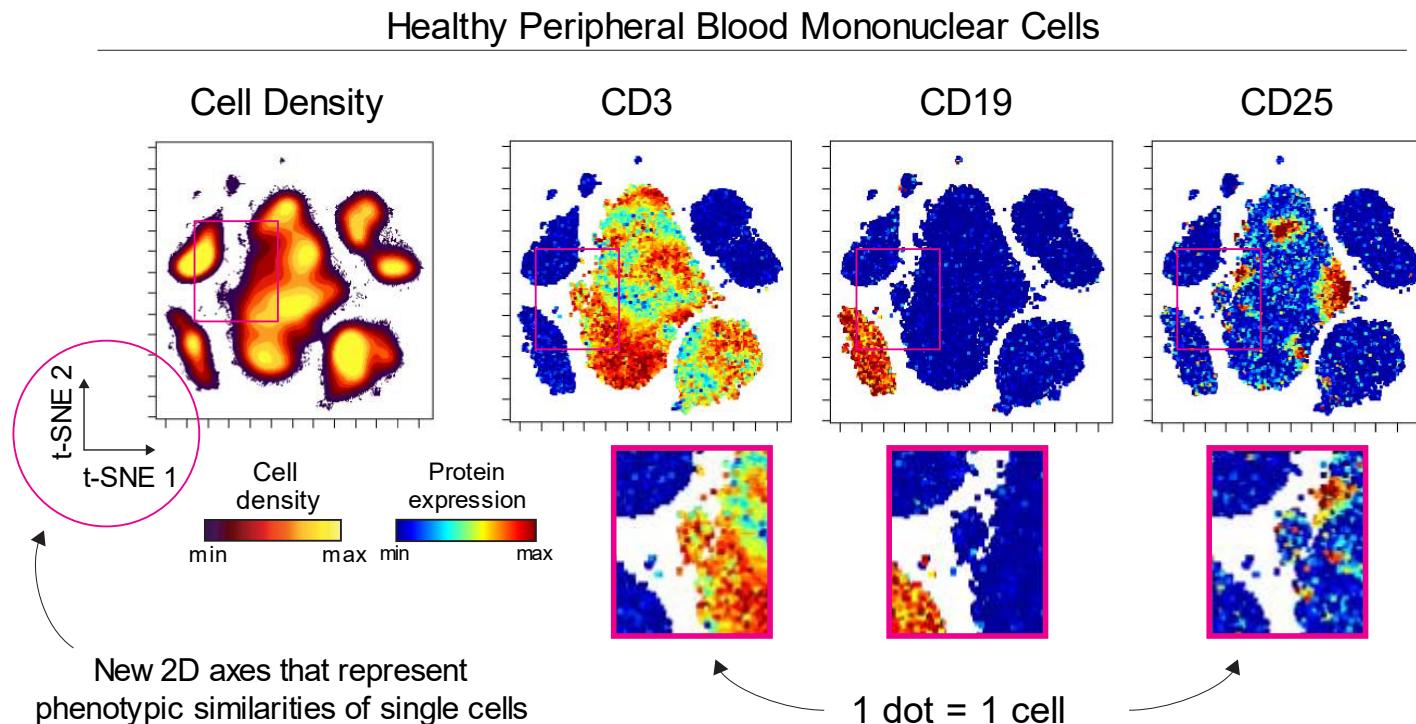
Code: <https://github.com/cytolab/t-rex>

Manuscript: <https://elifesciences.org/articles/64653>



# Running the Workflow on PBMC

Dots = 50,000 cells  
t-SNE = 25 measured protein features (25D)  
Identification of 7 canonical cell types (CD4+ T cells, CD8+ T cells, NK cells, Monocytes, Dendritic Cells, IgM+ B cells, IgM- B cells)



# Let's Analyze PBMC Data!

---

<https://cytolab.shinyapps.io/PBMC/>

This web app is running R code live.

## Data Science Tutorial on Human Blood Cells

Welcome to a data science tutorial on healthy human peripheral blood mononuclear cells (PBMCs). Here you will apply t-SNE, FlowSOM, and MEM algorithms on the data, and learn how changing different settings impacts your results.

The dataset is from [Diggins et al., Nature Methods 2017](#), and contains around 50,000 cells each measured for 25 different proteins. Viewing the first few cells in spreadsheet form, the data looks like the following:

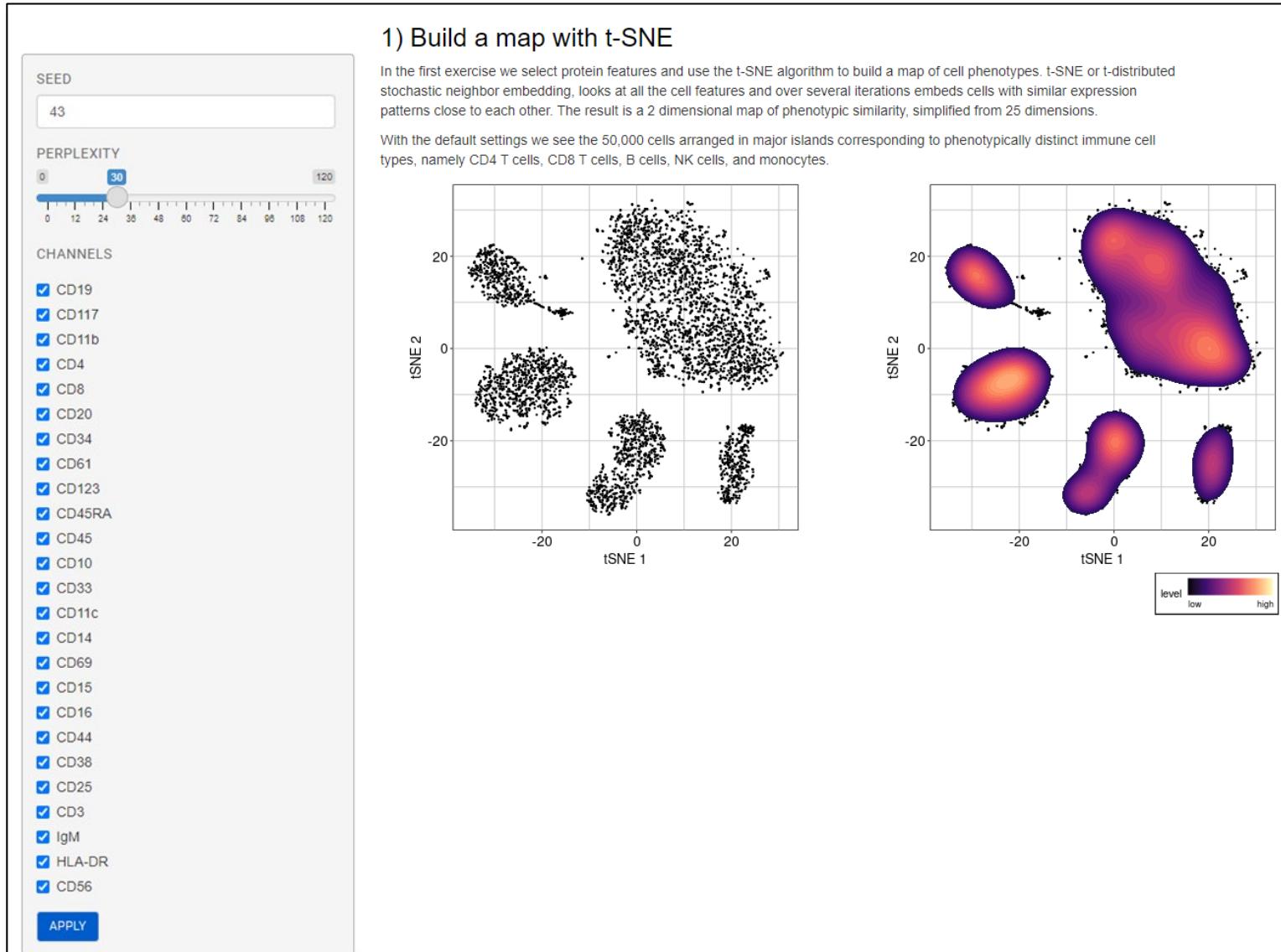
	CD19	CD117	CD11b	CD4	CD8	CD20	CD34	CD61	CD123	CD45RA	CD45	CD10	CD33	CD11c	CD14	CD69	CD15	CD16	CD44	CD38	CD25	CD3	IgM	HLA-DR	CD56
cell 1	-.3	-.6	11	132	12	-.8	-.8	-.7	-.5	101	284	-.2	-.8	2	-.06	-.8	.7	-.6	46	-.1	10	71	-.9	10	-.09
cell 2	-.3	-.4	-.6	204	4.6	-.2	-.6	-.03	-.8	-.1	400	-.7	-.7	-.6	1	-.1	-.8	-.2	222	10	3	99	-.5	-.9	-.04
cell 3	1.4	-.5	-.5	145	2.4	-.2	-.4	-.5	-.6	25	360	-.6	5	-.08	-.8	-.3	-.2	-.4	320	24	18	50	-.6	-.8	5

For this tutorial, we've taken a random sample of 5,000 cells from the 50,000 to run analyses on. If you'd like a larger or smaller sample size, you have the option to change that in the following menu. Alternatively if you'd like to reset your session, you can use the clear session button.

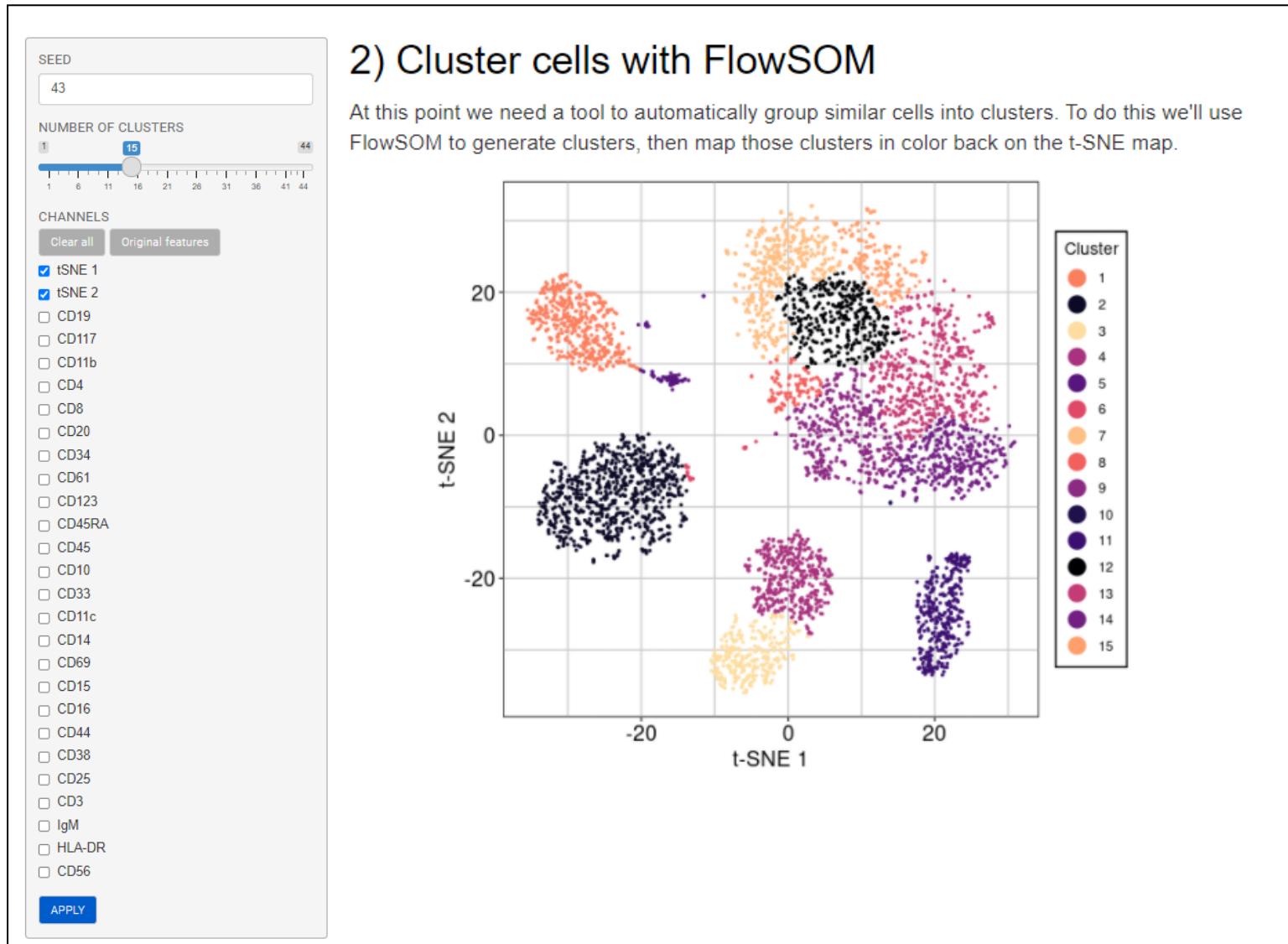
SAMPLE SIZE

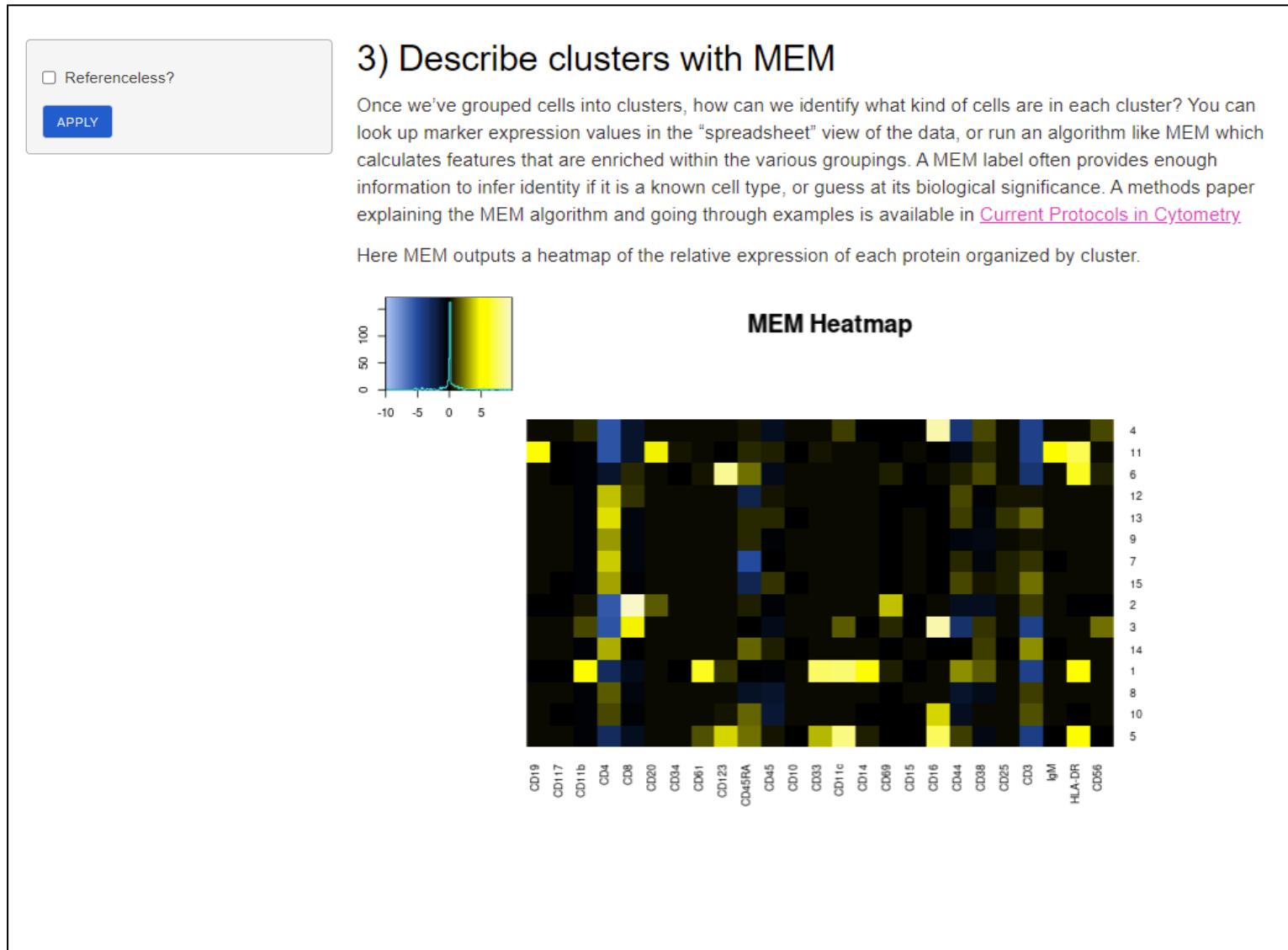
CLEAR SESSION

APPLY

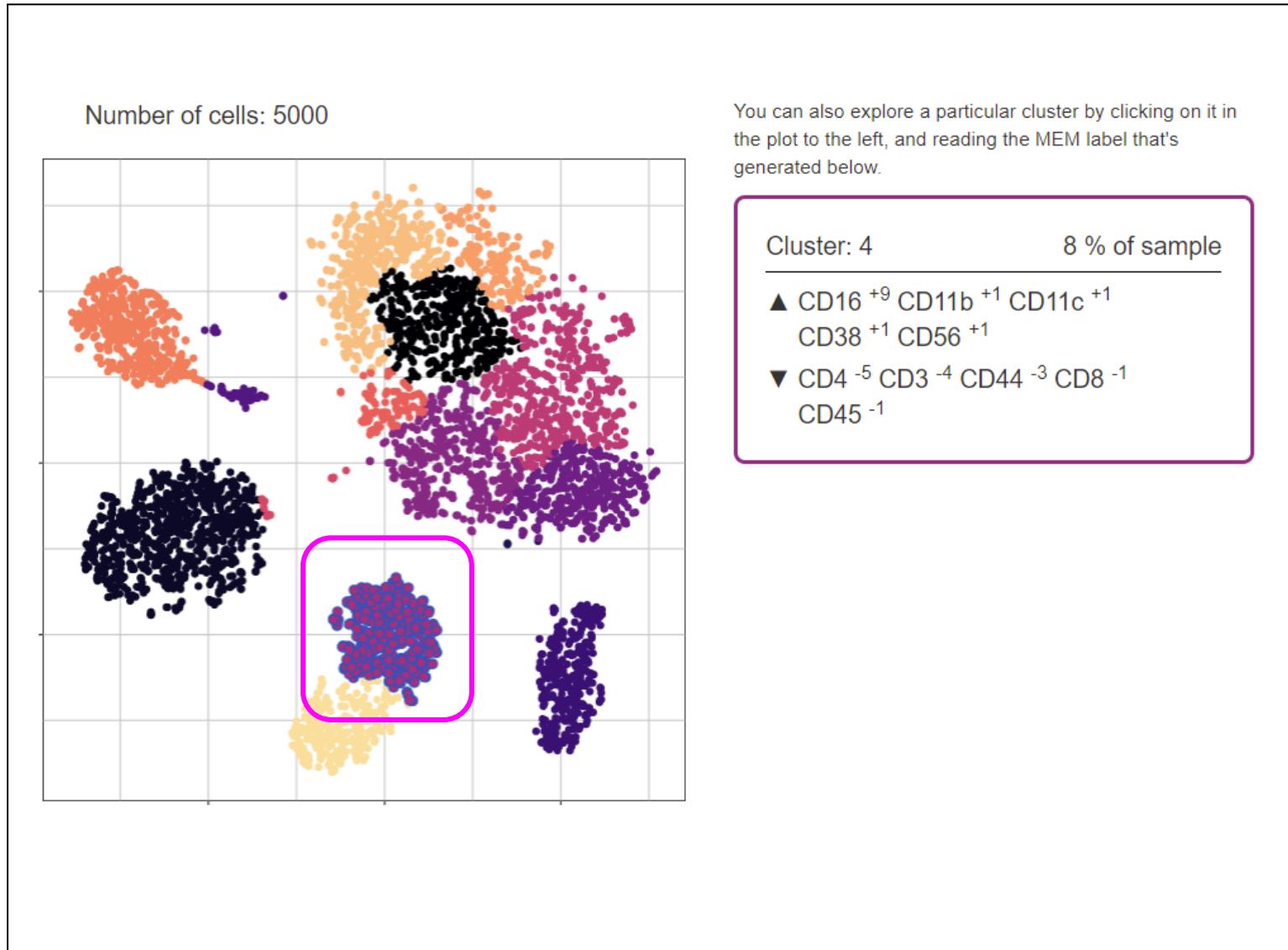


<https://cytolab.shinyapps.io/PBMC/>

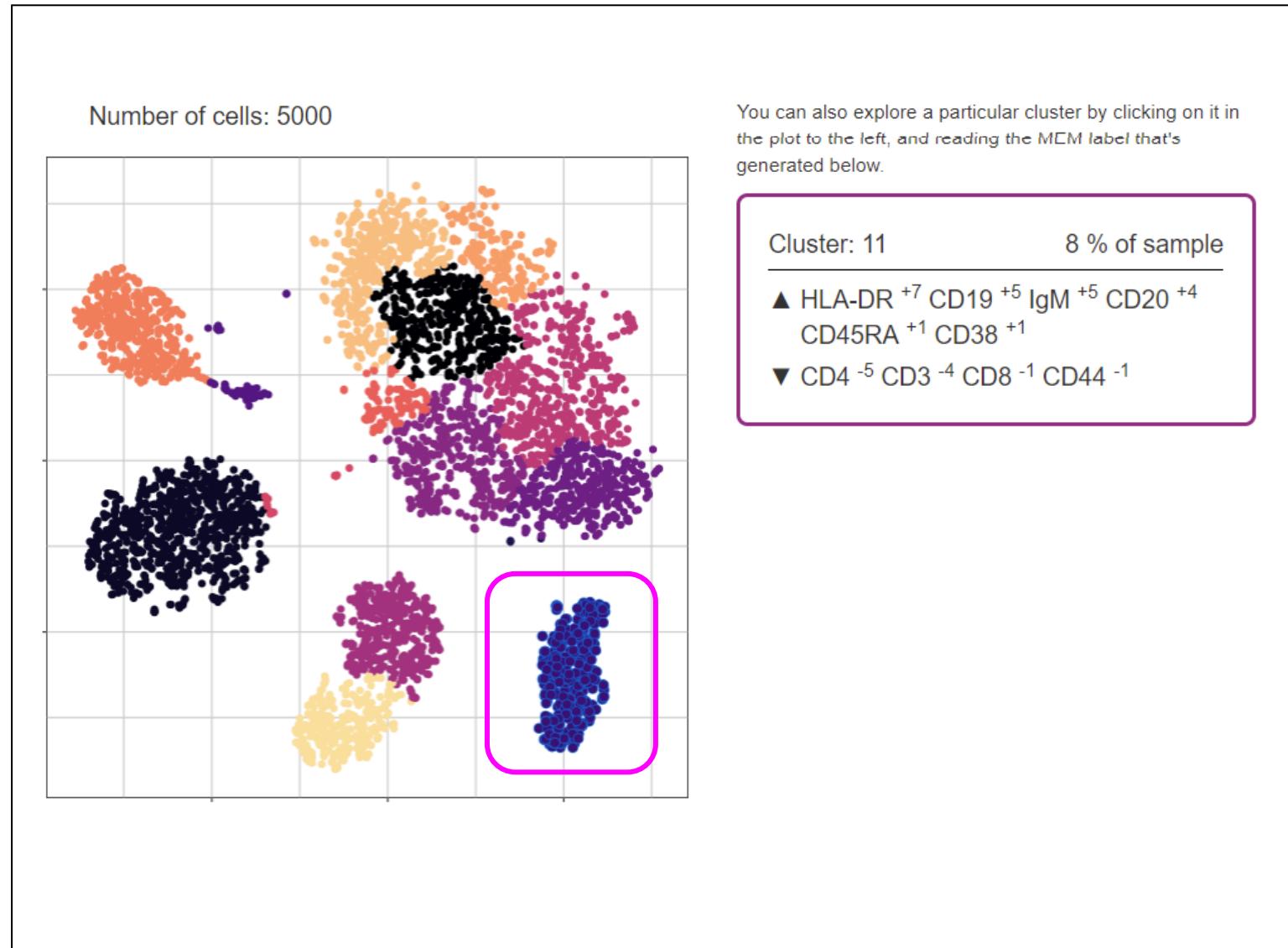


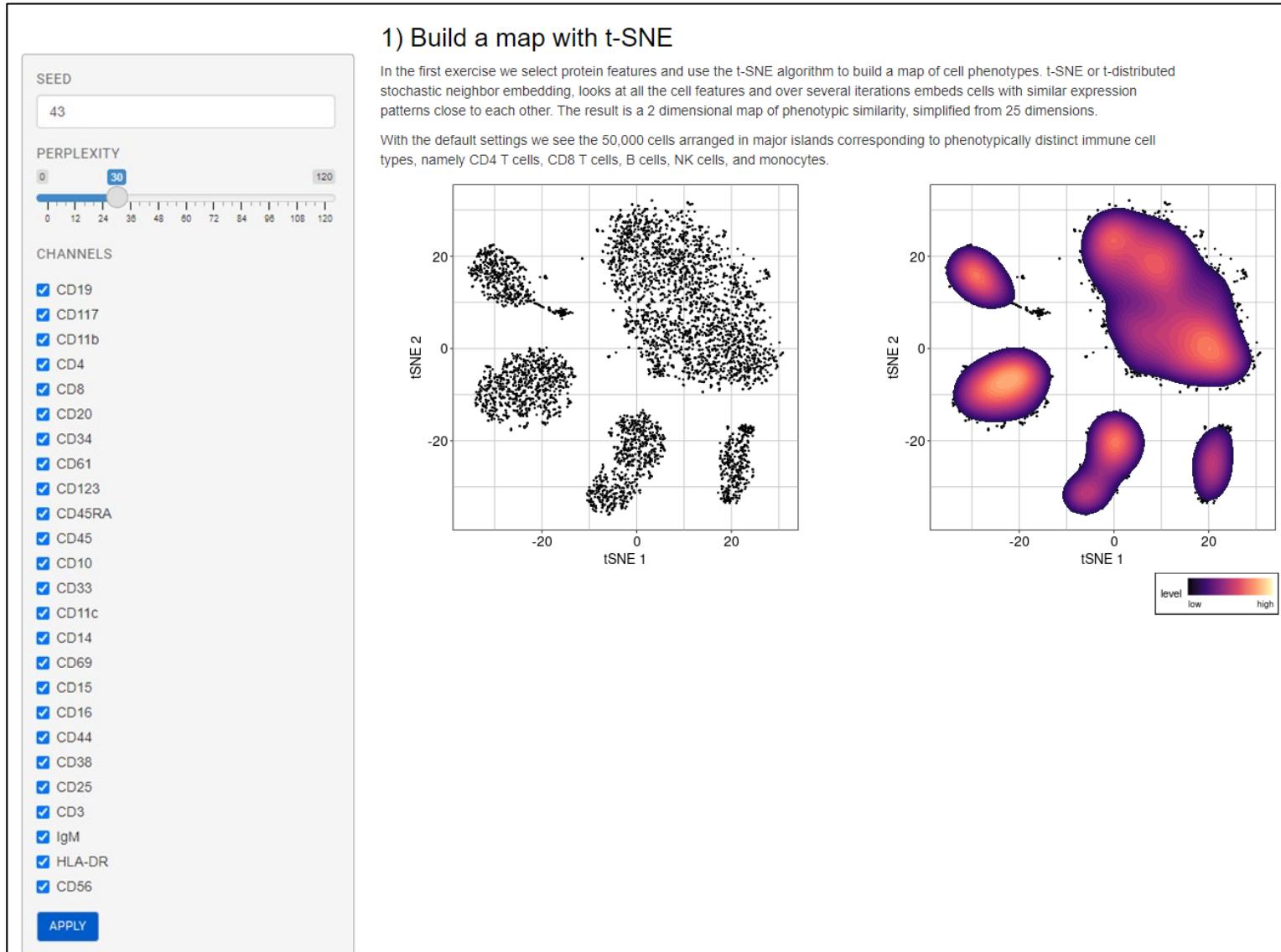


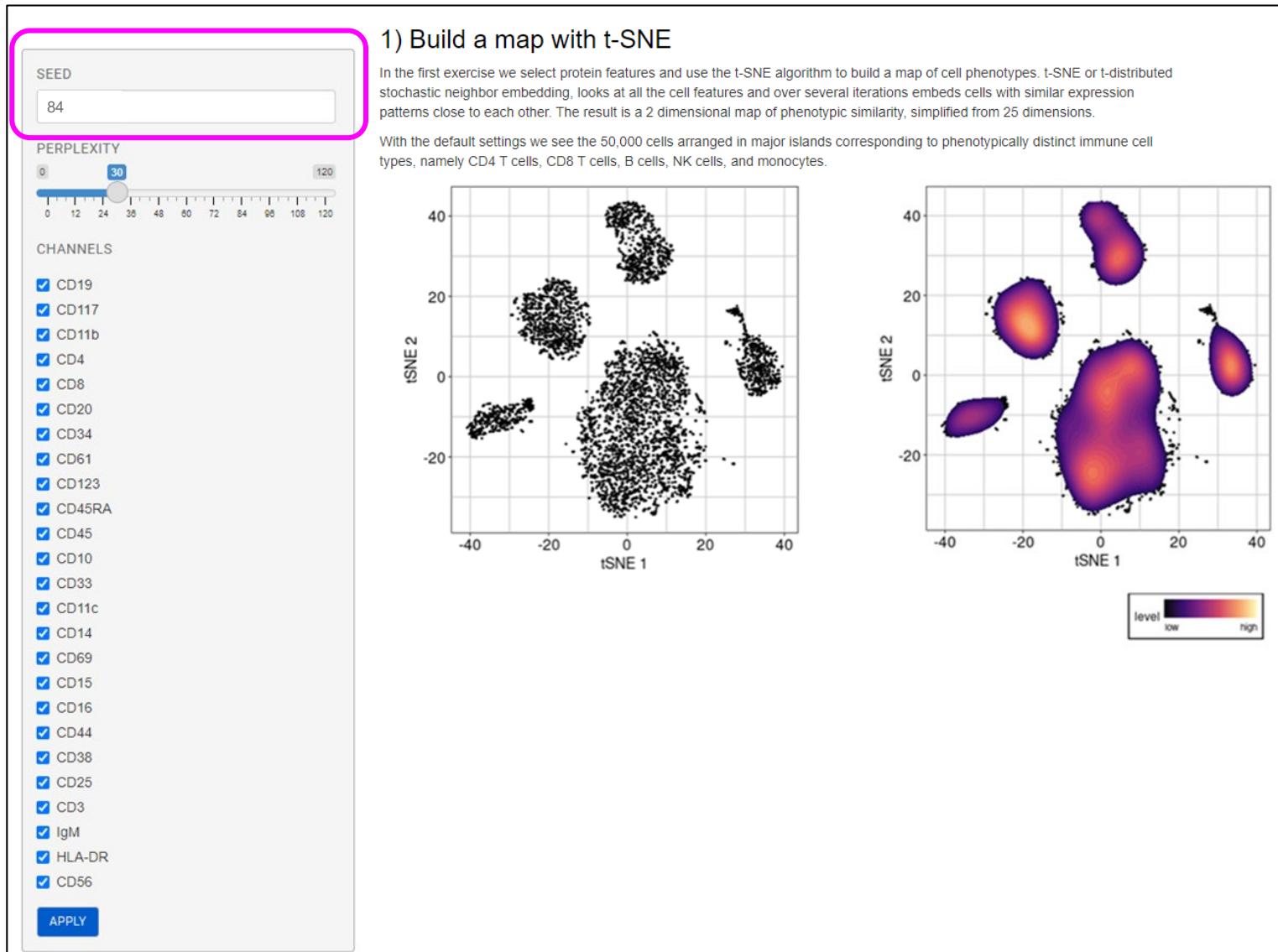
<https://cytolab.shinyapps.io/PBMC/>



<https://cytolab.shinyapps.io/PBMC/>







SEED  
43

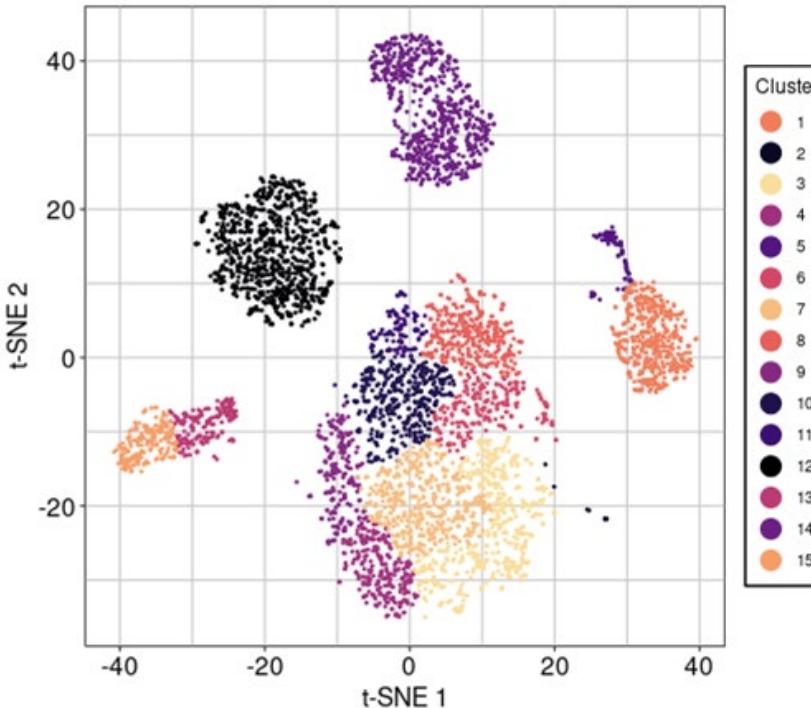
NUMBER OF CLUSTERS  
1 15 44

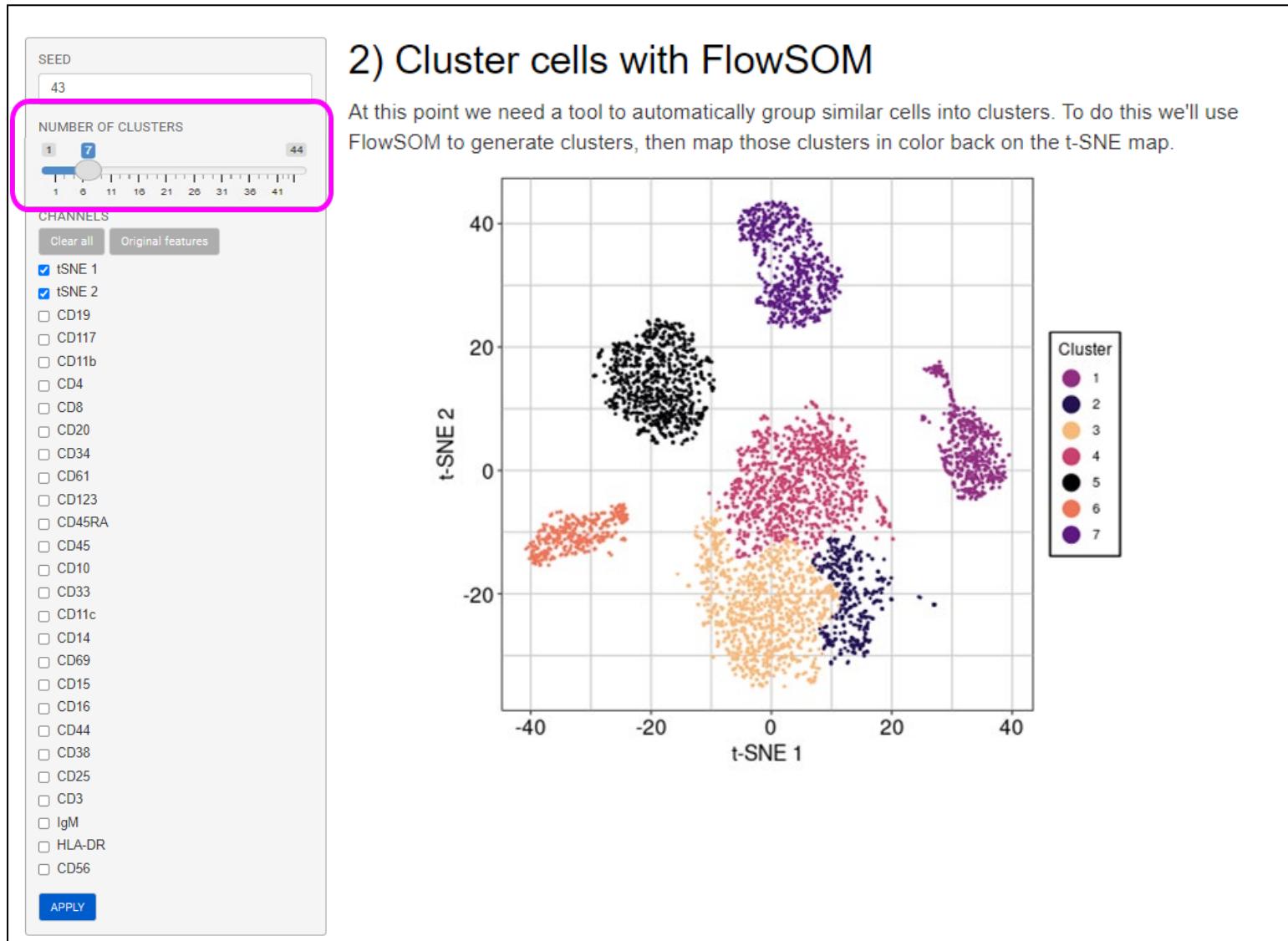
CHANNELS

tSNE 1  
 tSNE 2  
 CD19  
 CD117  
 CD11b  
 CD4  
 CD8  
 CD20  
 CD34  
 CD61  
 CD123  
 CD45RA  
 CD45  
 CD10  
 CD33  
 CD11c  
 CD14  
 CD69  
 CD15  
 CD16  
 CD44  
 CD38  
 CD25  
 CD3  
 IgM  
 HLA-DR  
 CD56

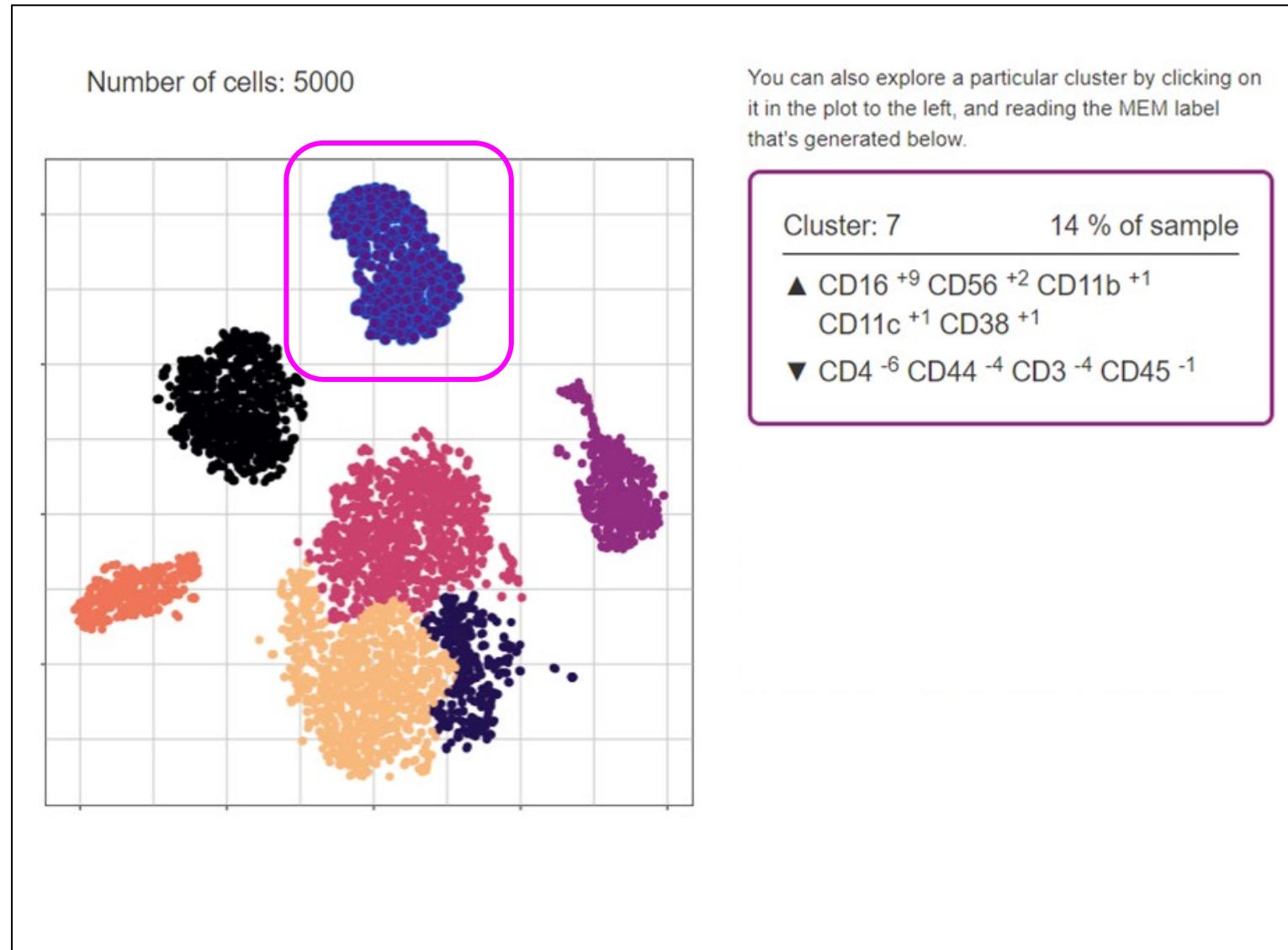
## 2) Cluster cells with FlowSOM

At this point we need a tool to automatically group similar cells into clusters. To do this we'll use FlowSOM to generate clusters, then map those clusters in color back on the t-SNE map.

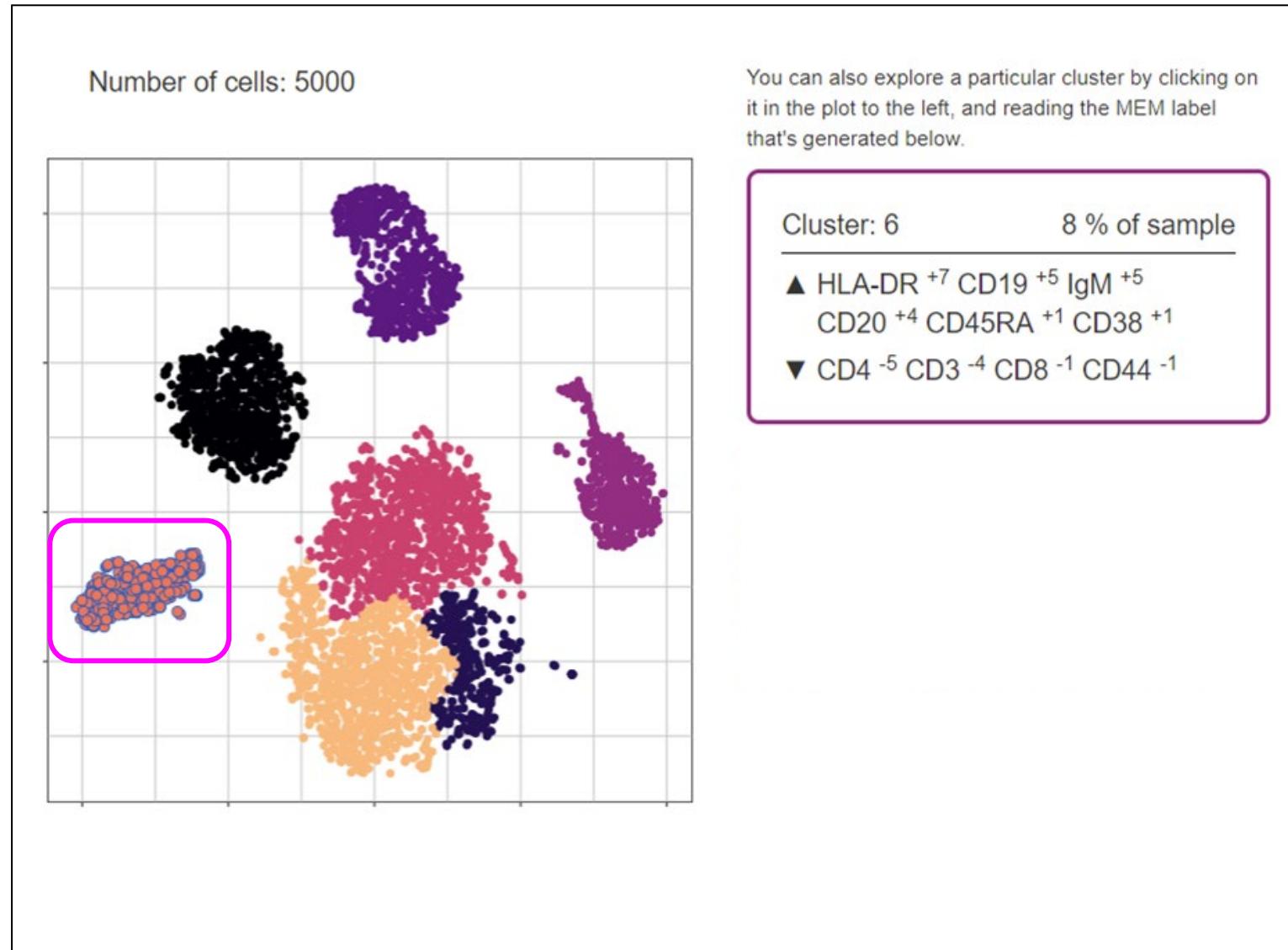


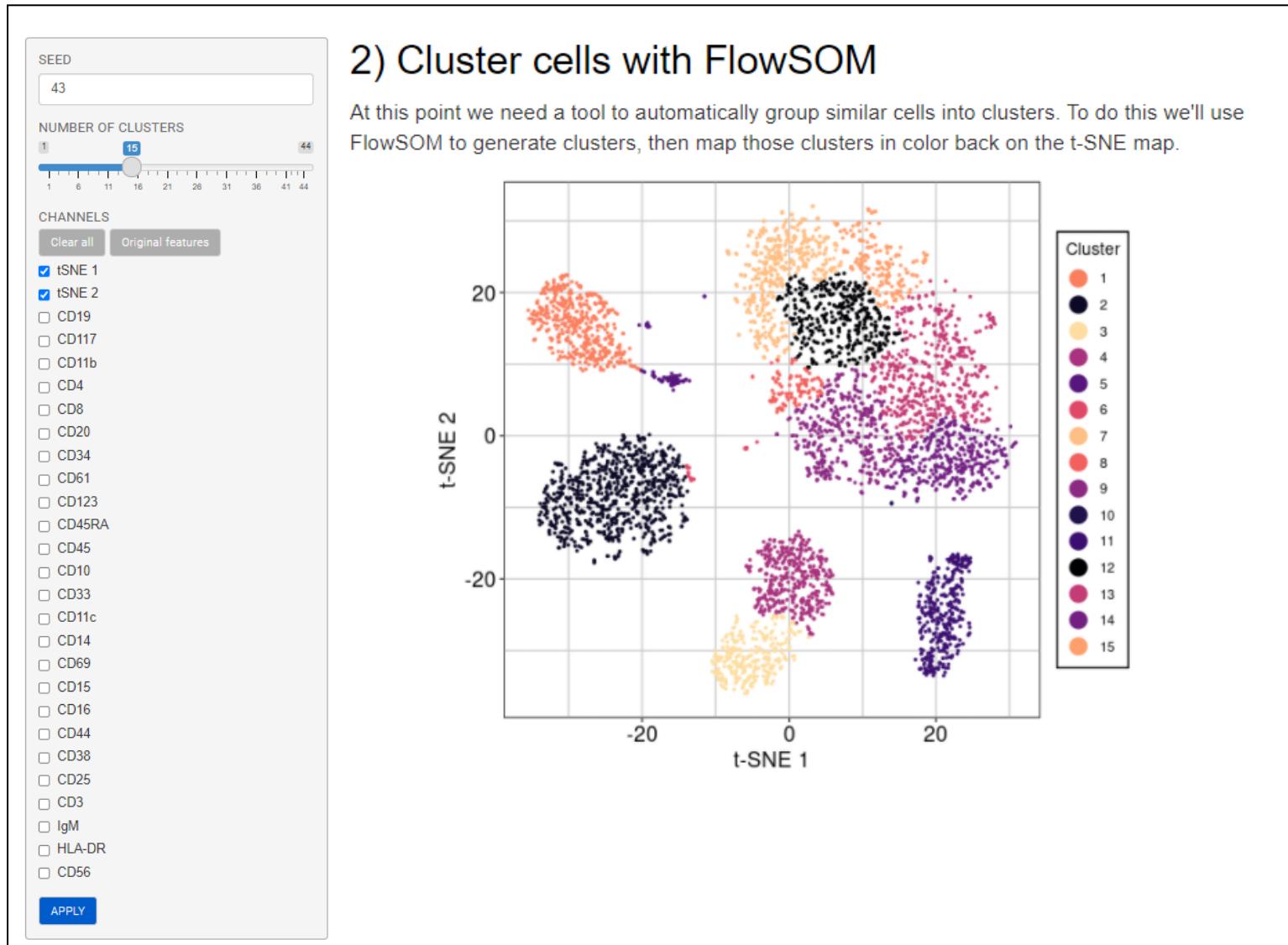


<https://cytolab.shinyapps.io/PBMC/>



<https://cytolab.shinyapps.io/PBMC/>





SEED  
43

NUMBER OF CLUSTERS  
1 15 44

CHANNELS

tSNE 1

tSNE 2

CD19

CD117

CD11b

CD4

CD8

CD20

CD34

CD61

CD123

CD45RA

CD45

CD10

CD33

CD11c

CD14

CD69

CD15

CD16

CD44

CD38

CD25

CD3

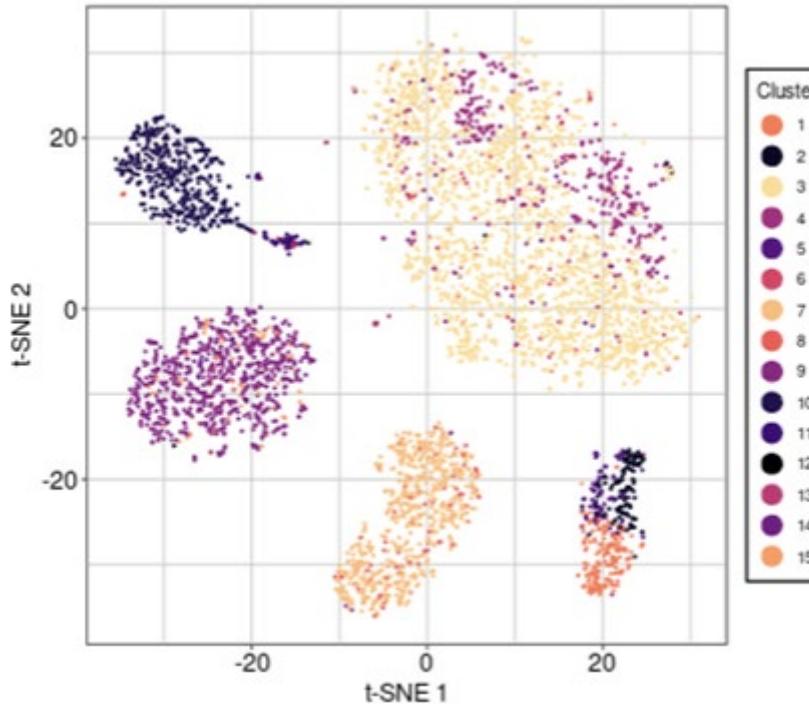
IgM

HLA-DR

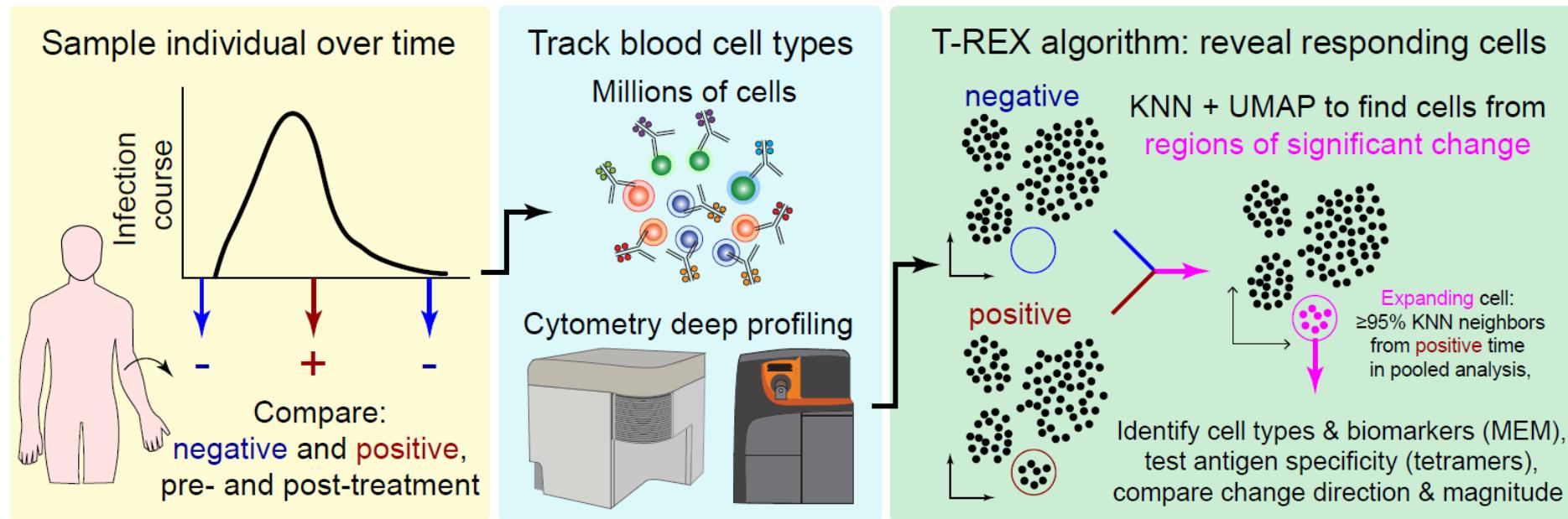
CD56

## 2) Cluster cells with FlowSOM

At this point we need a tool to automatically group similar cells into clusters. To do this we'll use FlowSOM to generate clusters, then map those clusters in color back on the t-SNE map.



# T-REX: Compare Two Samples to Identify Things Enriched in Either One; e.g., Reveal Rare, Virus-Specific Immune Cells



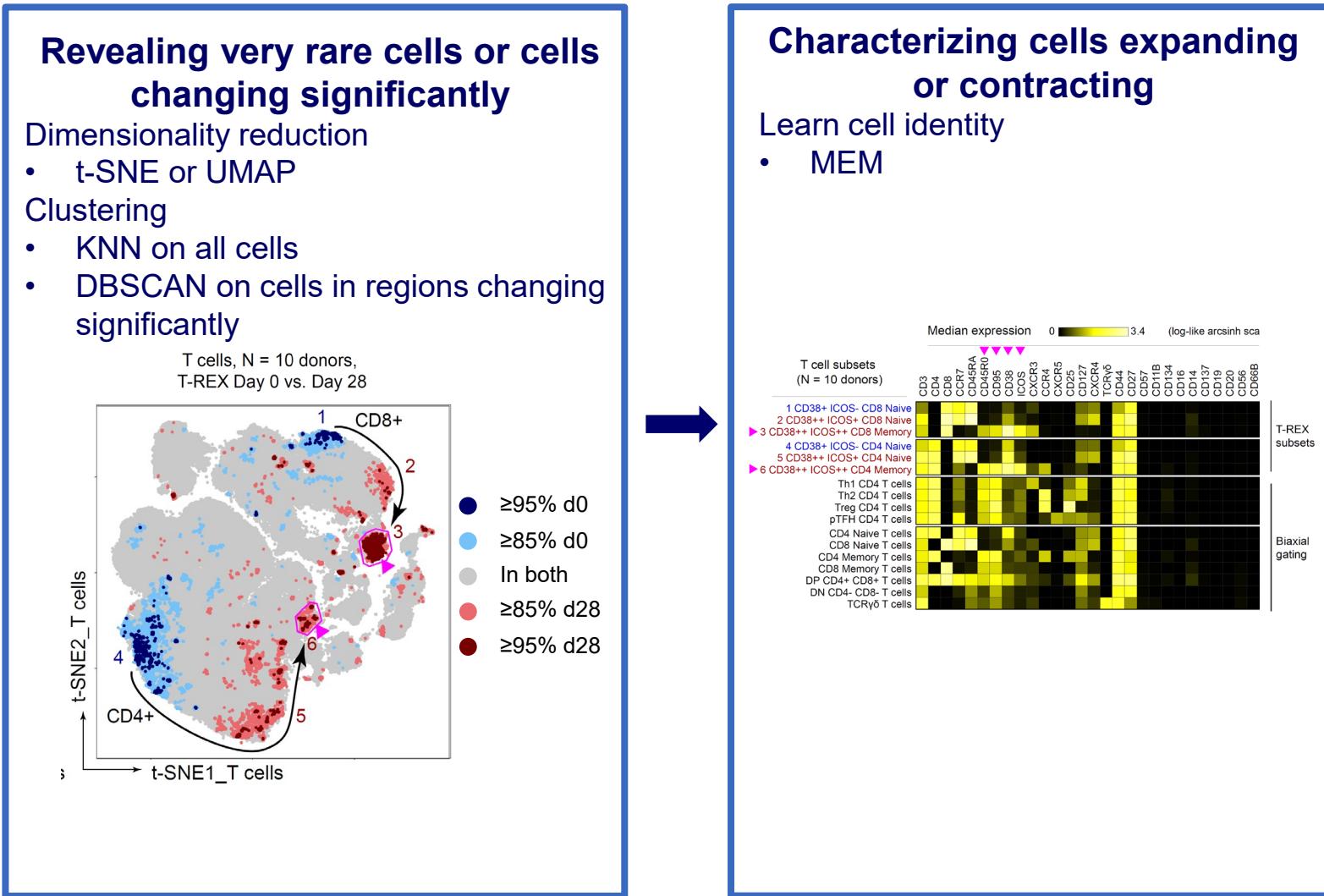
New algorithm: T-REX (Tracking Responders EXpanding)

Code: <https://github.com/cytolab/t-rex>

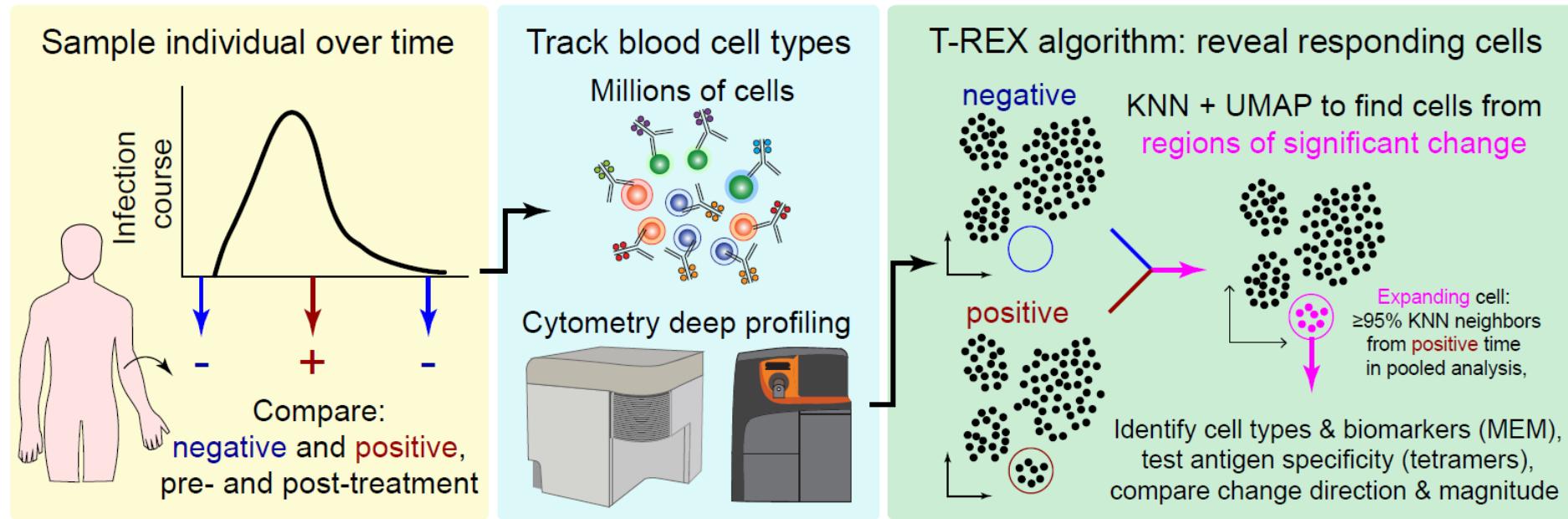
Manuscript: <https://elifesciences.org/articles/64653>



# Data Science Workflow Using T-REX

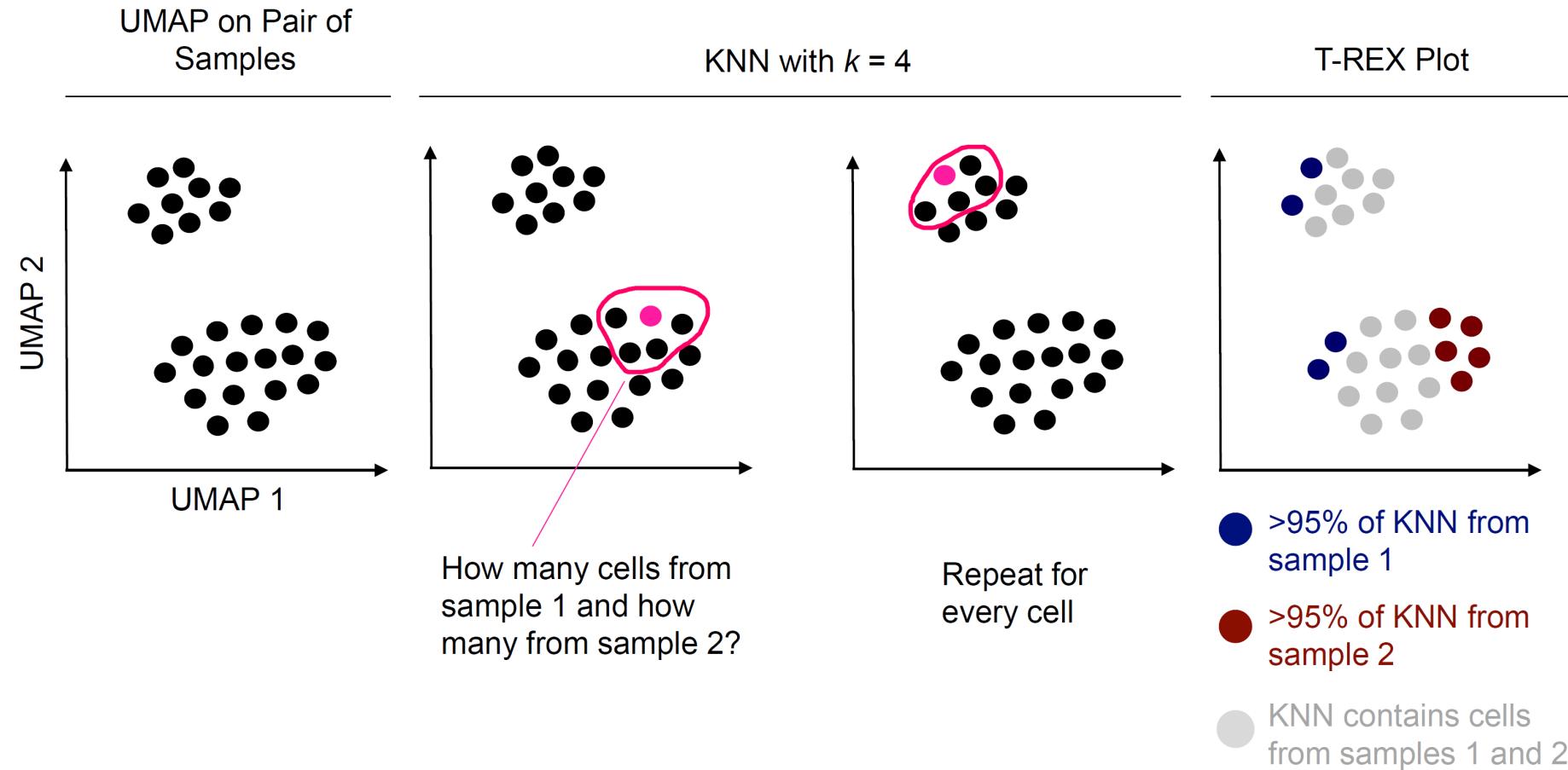


# Key Ideas & Findings in Today's Talk

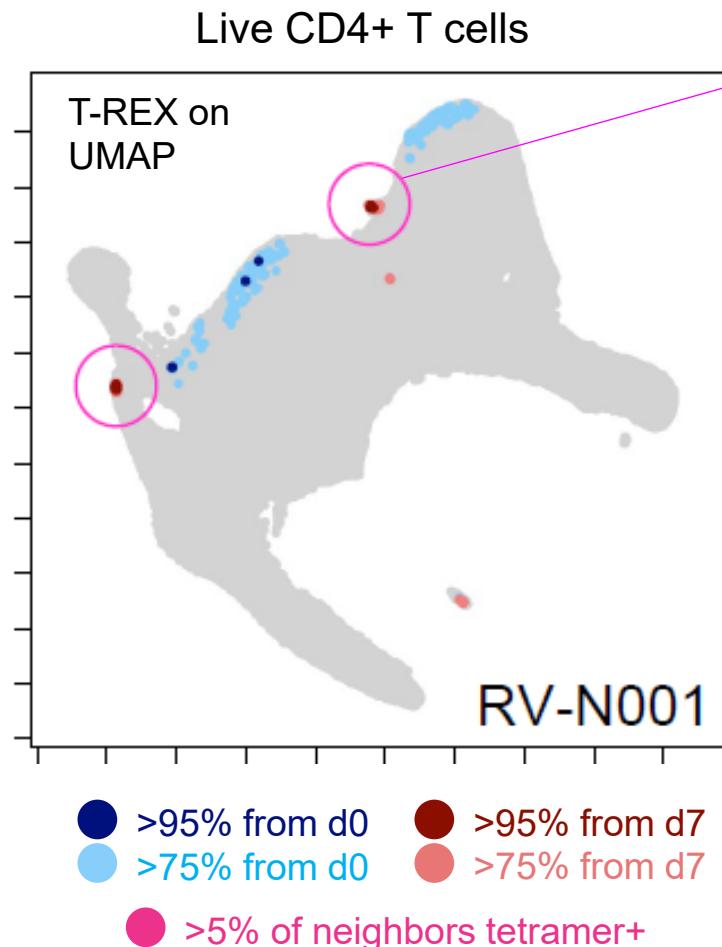


- Idea 1: T-REX automatically reveals virus-specific T cells in rhinovirus & SARS-CoV-2 vaccine response (without the need for tetramers, sorting, or sequencing)
- Idea 2: Approach focuses on extreme change & can summarize disease, therapy, or perturbation response (direction & magnitude of change; rhinovirus, COVID-19, cancer therapy, compound screening)
- Finding: Mass cytometry + T-REX characterized SARS-CoV-2 vaccine-induced memory CD4 and CD8 T cells (phenotype: CD38++ ICOS++ CD45R0+ PD-1+ Ki-67+ CXCR5-)
- Finding: Phenotype of SARS-CoV-2 vaccine responding T cells closely matched rhinovirus-specific T cells

# T-REX Algorithm Uses K-Nearest Neighbors (KNN) to Characterize Each Cell's Immediate Phenotypic Neighborhood



# T-REX: Tracking Responders EXPanding, Every Cell Is Characterized in a Search for Hotspots of Change



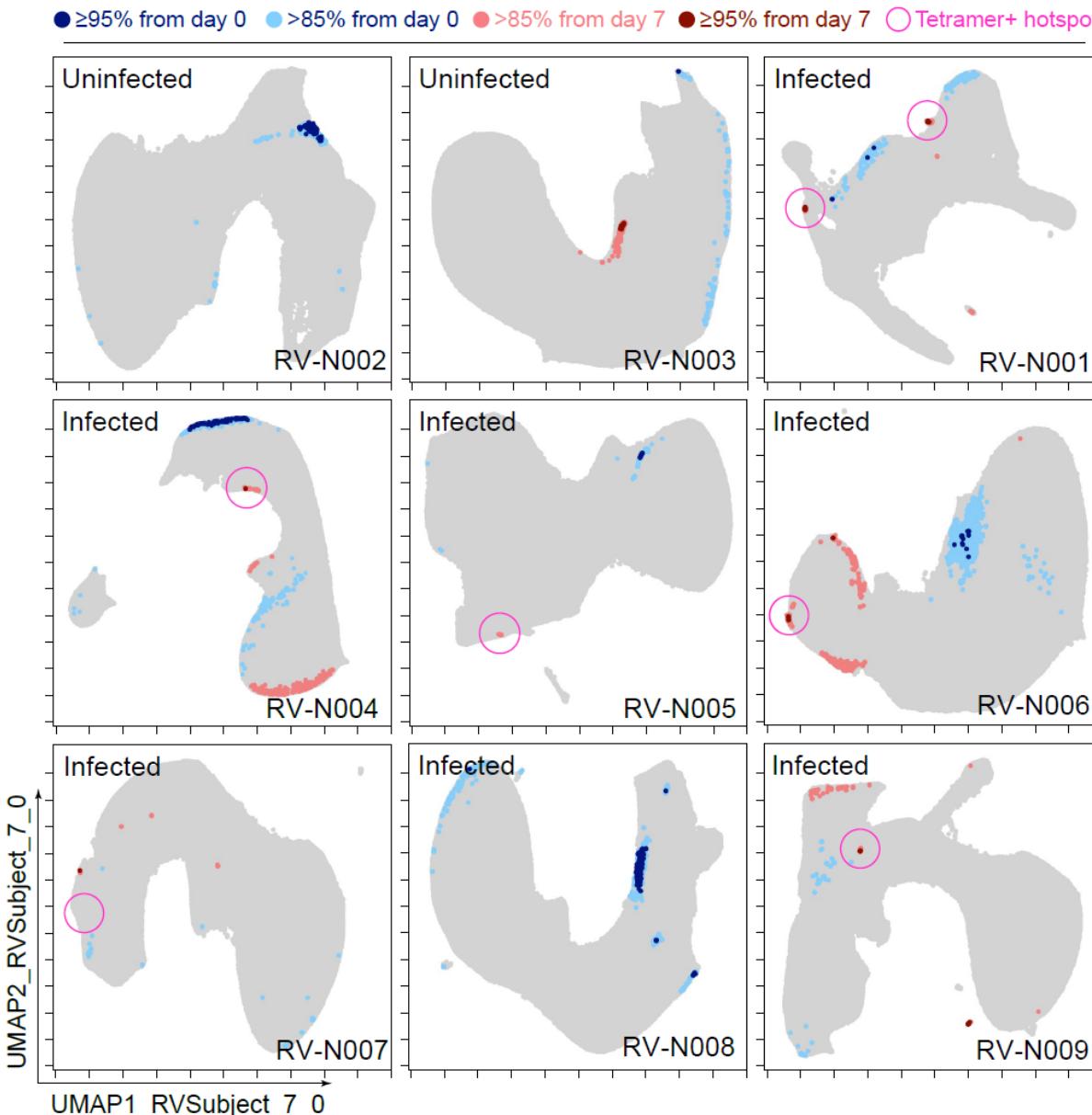
MHCII tetramers marking rhinovirus specific CD4 T cells were not used to make the UMAP, instead used to show: Change hotspots were enriched for virus-specific T cells

Color: cells in that phenotypic neighborhood are mostly from one sample

Dark red = cells mostly from day 7 (expanding) 

CD4 T cells, Day 0 vs. Day 7,  
individual infected with rhinovirus (RV-N001)  
no cell enrichment, Aurora data,  $\sim 3 \times 10^6$  cells

# In Analysis of a Rhinovirus Challenge Cohort, T-REX Revealed Virus-Specific Cell Phenotypes



CD4 T cells, Day 0 vs. Day 7,  
individuals infected with rhinovirus  
no cell enrichment, Cytek Aurora data

In 5 of 7 infected individuals, **expansion hotspots** were enriched for **virus-specific cells**



The phenotype of rhinovirus-specific memory  
CD4+ T cells calculated by MEM:  
CCR5+ ICOS+ CD38+ PD-1+ CXCR5-

Gating based on this MEM phenotype =>  
enriched for tetramer+ cells  
(without gating on tetramers):

Indicated we could sort cells (FACS)  
based on T-REX MEM labels

## T-REX revealed virus-specific T cells without tetramers

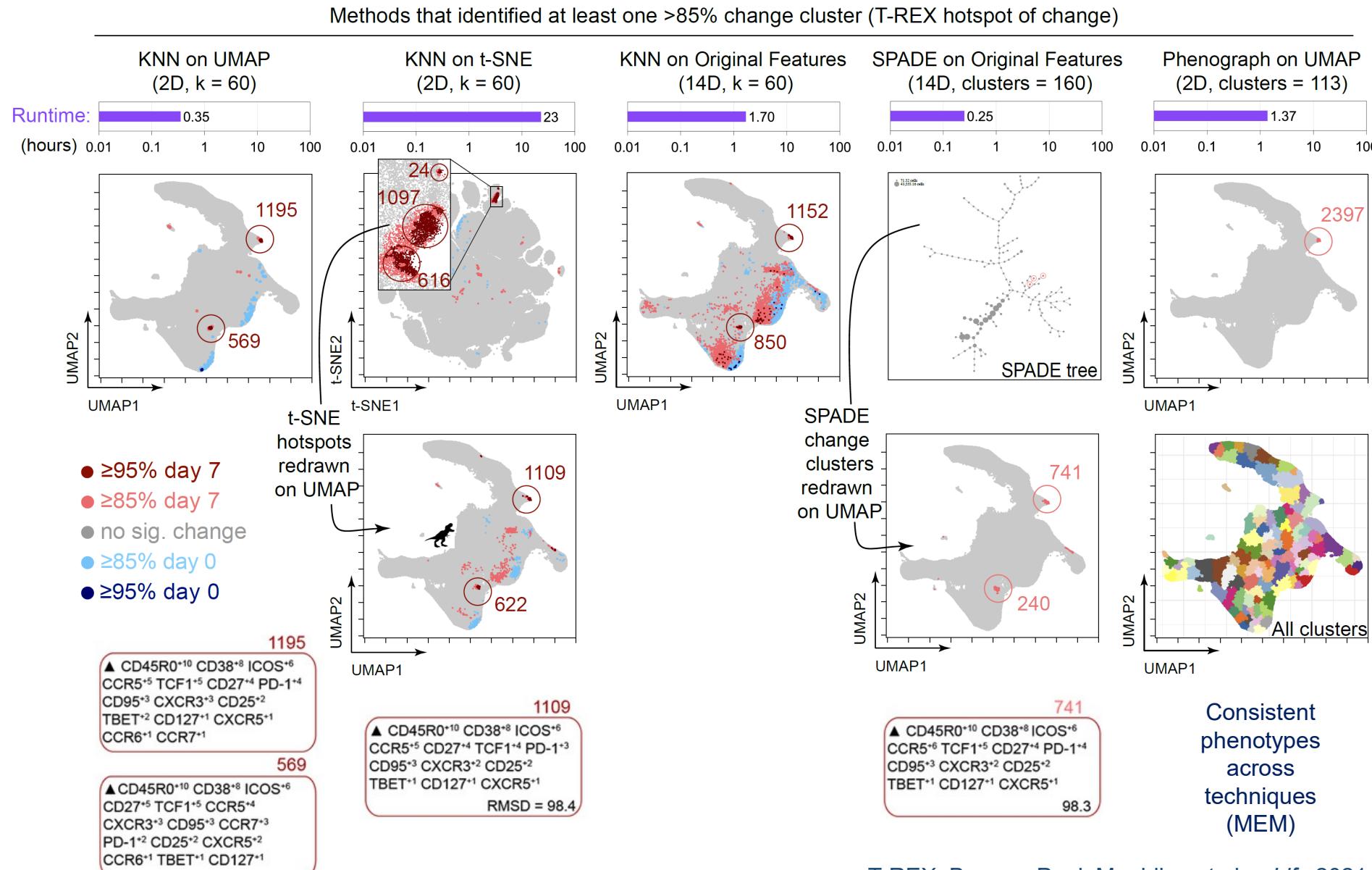


Would this approach work with other clustering algorithms?

Is it 'OK' to do KNN on UMAP axes as parameters?

(Perhaps: all embeddings are wrong, but some are useful...)

# T-REX Worked with Other Algorithms to Identify Comparable Cells, But KNN on UMAP or t-SNE Outperformed KNN on Original Features



## T-REX revealed virus-specific T cells without tetramers

Also found to work for:

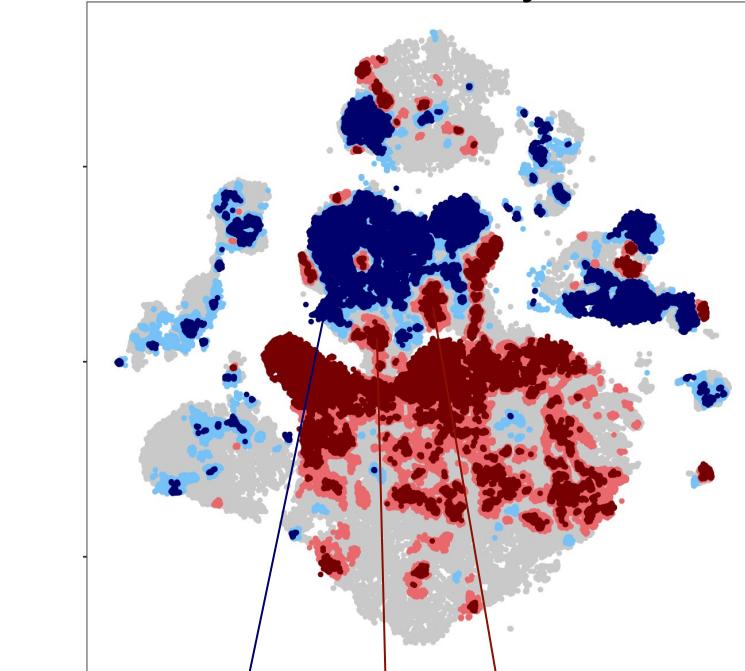
- a range of k-values ( $k = 60$  was optimal)
- post-infection as the comparison point to day 7
- data from a range of cytometers, studies, and labs
- COVID-19, melanoma immunotherapy response, AML

(see the manuscript for this & more!)



# Massive Immune Change, Common Shifts in Expanding Cell Subsets Observed Between Day 0 and Day 7 in COVID-19

COV-994535 Day 0 vs. 7



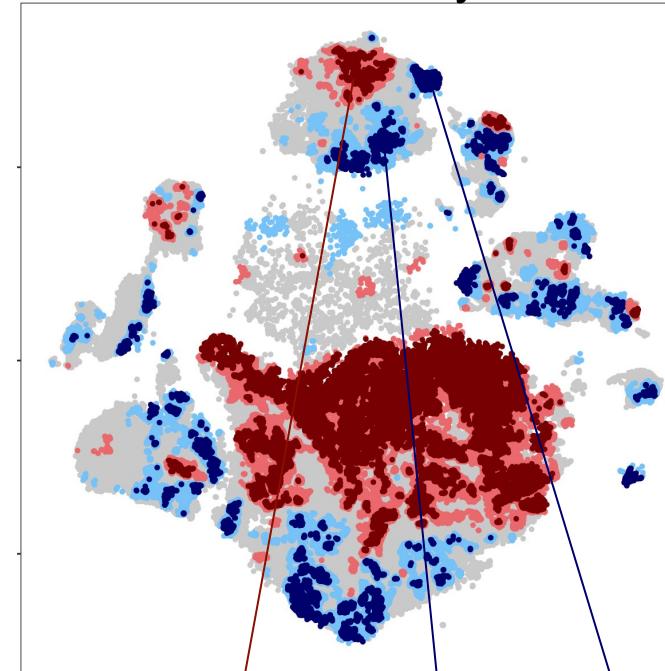
CM

ICOS<sup>+5</sup> CD27<sup>+4</sup> CD95<sup>+3</sup> CCR7<sup>+3</sup> T-bet<sup>+3</sup> CXCR5<sup>+2</sup>  
CD3<sup>+2</sup>

CX3CR1<sup>+7</sup> T-bet<sup>+7</sup> CXCR5<sup>+4</sup> TCF-1<sup>+3</sup> CD95<sup>+3</sup> Ki-67<sup>+3</sup>  
CD39<sup>+3</sup> CD16<sup>+2</sup> Eomes<sup>+2</sup>

T-bet<sup>+8</sup> CX3CR1<sup>+7</sup> CXCR5<sup>+4</sup> TCF-1<sup>+3</sup> CD95<sup>+3</sup> Ki-67<sup>+3</sup>  
CD39<sup>+3</sup>

COV-994536 Day 0 vs. 7



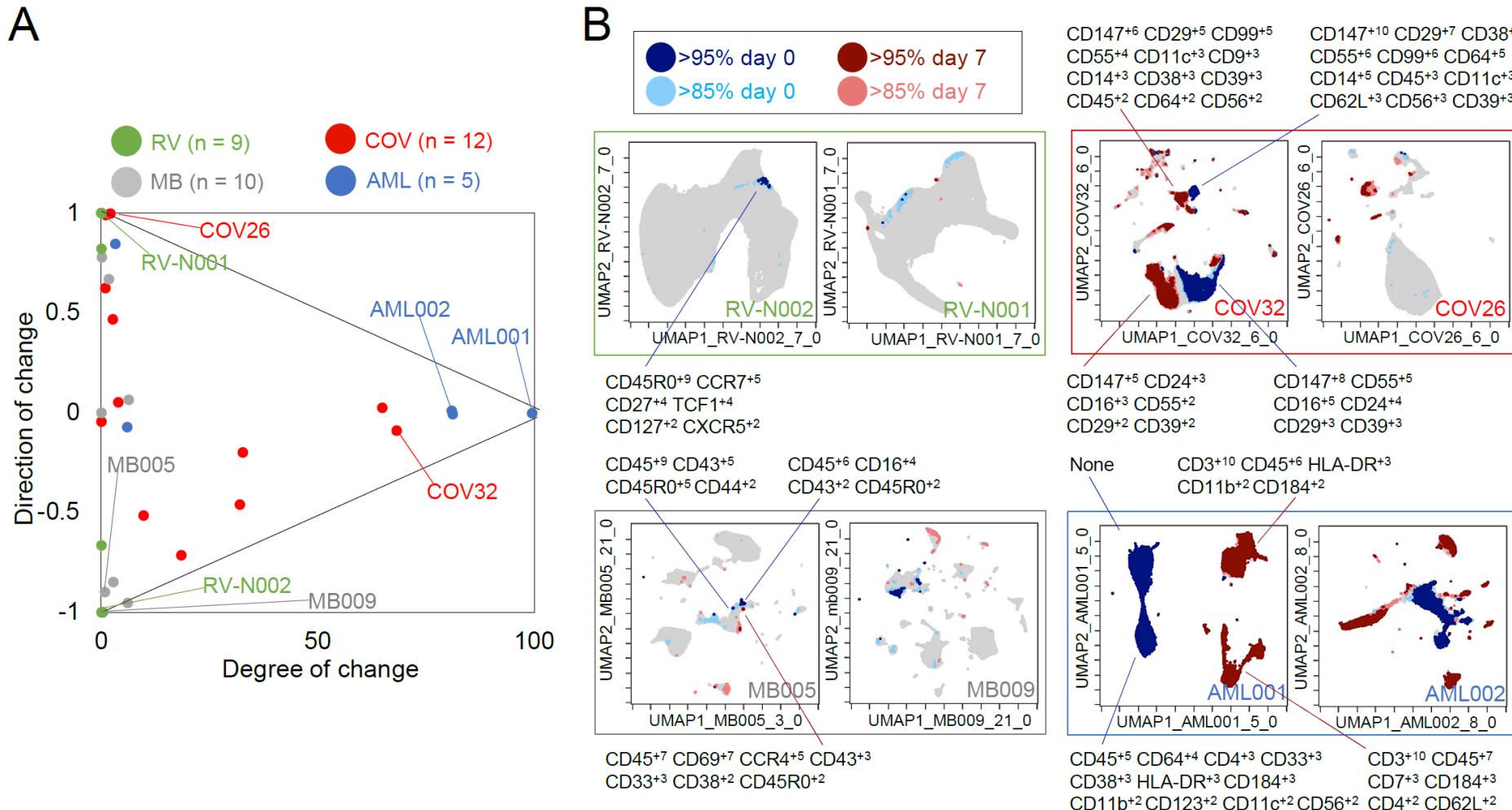
Naive

ICOS<sup>+6</sup> T-bet<sup>+5</sup> CD95<sup>+4</sup> CD27<sup>+4</sup> CXCR5<sup>+2</sup> CD3<sup>+2</sup>  
CCR7<sup>+2</sup>

CD45RA<sup>+5</sup> CD27<sup>+5</sup> T-bet<sup>+5</sup> ICOS<sup>+4</sup> CD3<sup>+3</sup> CCR7<sup>+3</sup>  
CXCR5<sup>+2</sup>

CX3CR1<sup>+6</sup> T-bet<sup>+6</sup> HLA-DR<sup>+6</sup> CD95<sup>+4</sup> CD39<sup>+4</sup>  
CXCR5<sup>+3</sup> CD38<sup>+3</sup>

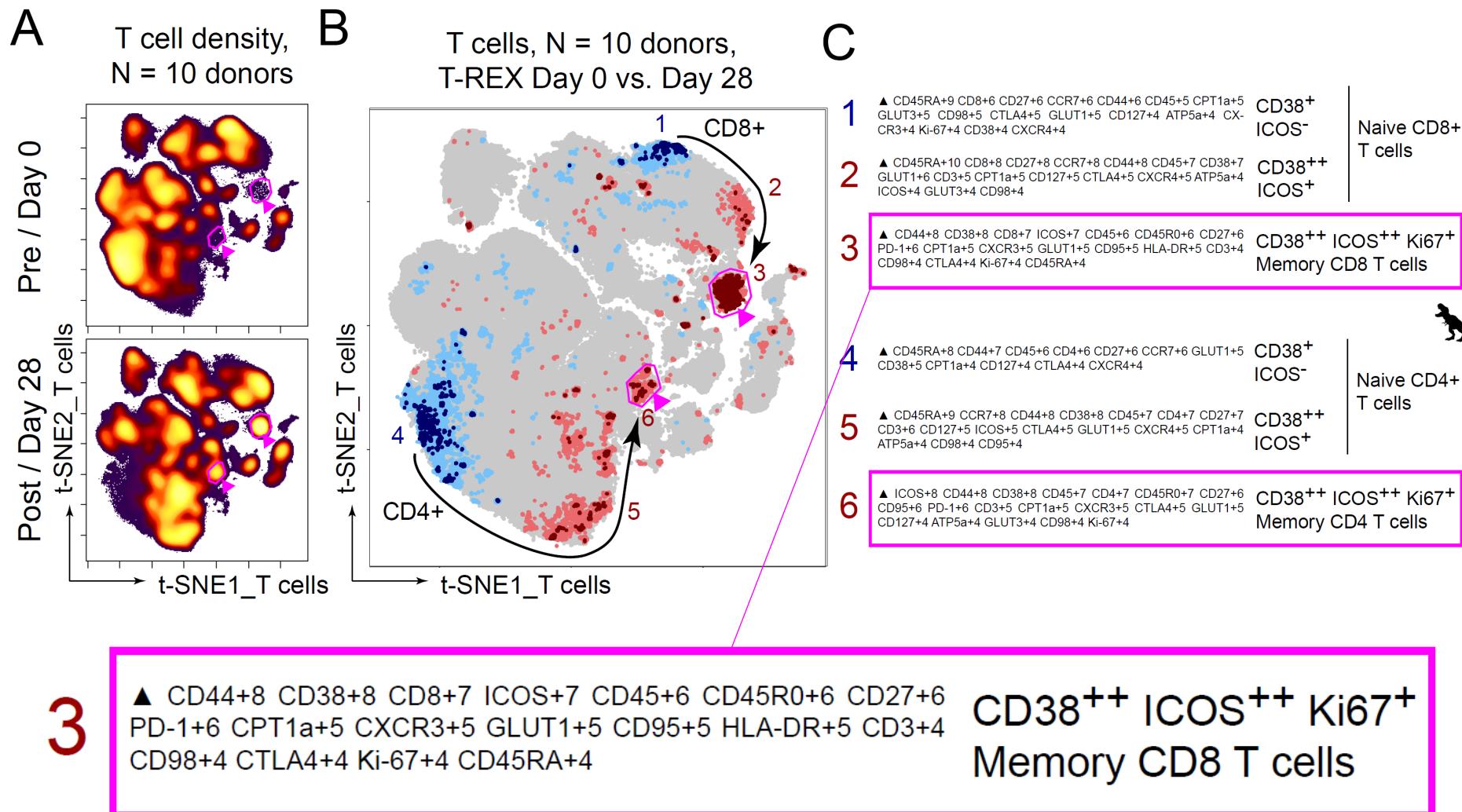
# Half of COVID-19 Patients Displayed Immune Changes Comparable to AML Patients with a Complete Response to Chemotherapy



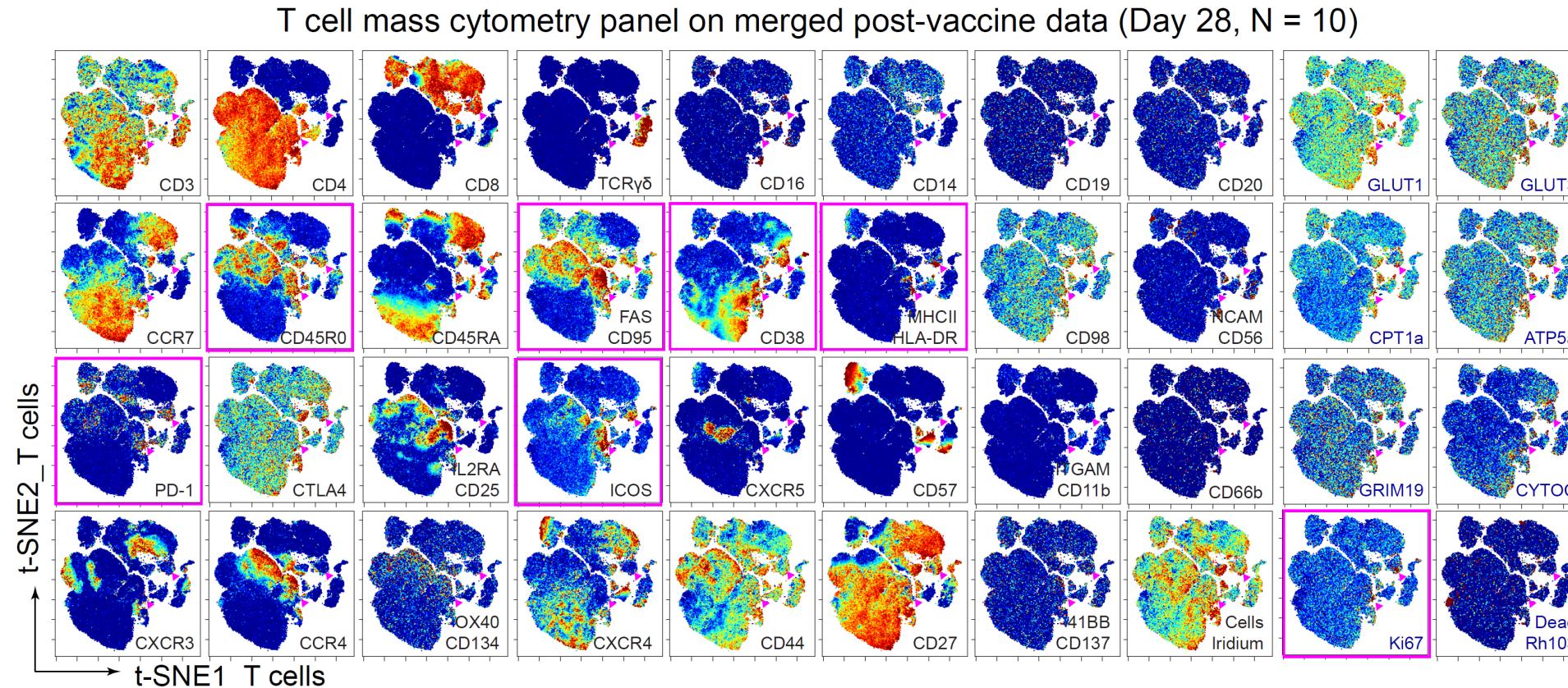
T-REX revealed virus-specific T cells without tetramers  
& characterized massive immune changes in COVID-19

Would it also work to characterize SARS-CoV-2 vaccine response? 

# T-REX Reveals Memory CD4 & CD8 T Cell Phenotypes Expanding following BNT162b2 SARS-CoV-2 RNA Vaccine



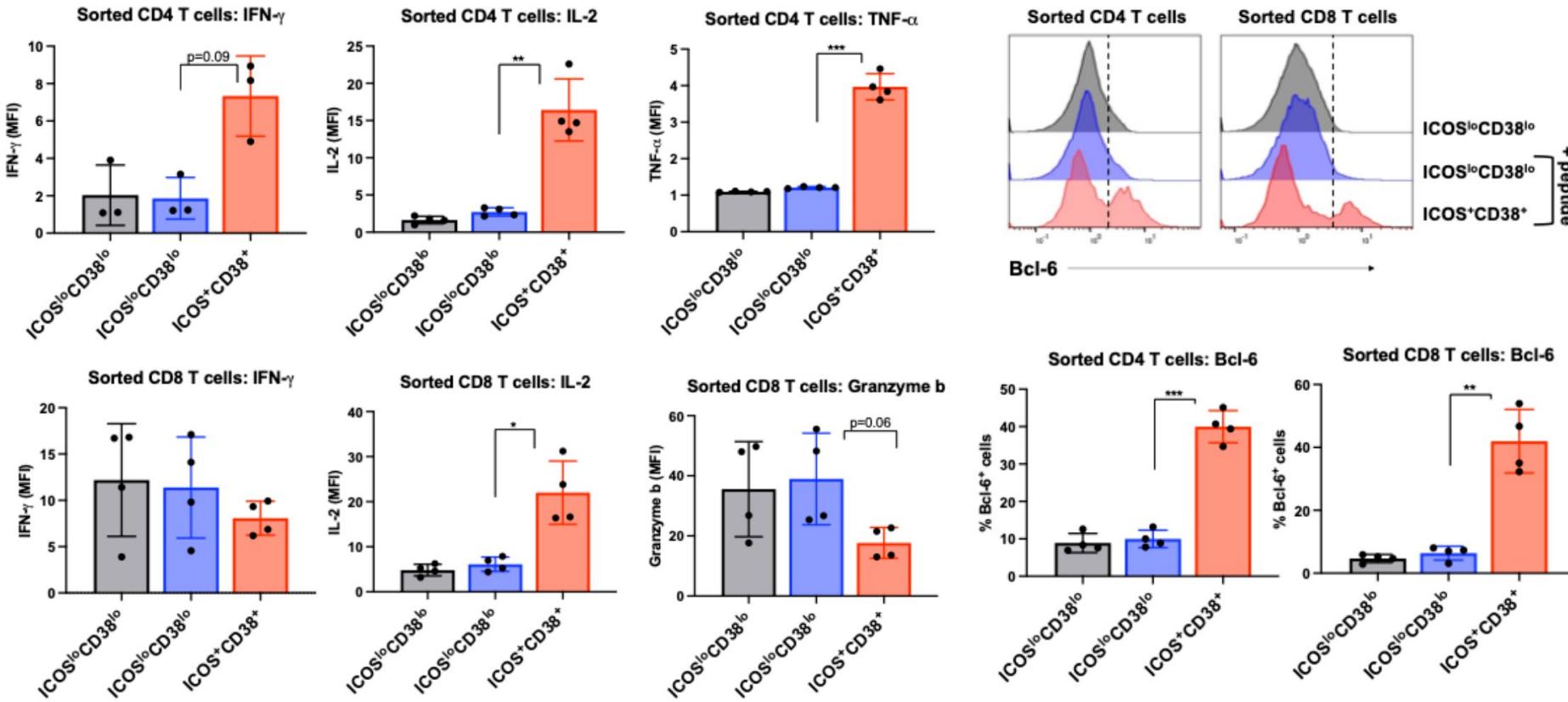
# Mass Cytometry Phenotyping of ICOS+ CD38+ PD-1+ Ki-67+ CXCR5- Memory CD4 & CD8 T Cells following SARS-CoV-2 Vaccination



# Mass Cytometry Phenotyping of ICOS+ CD38+ PD-1+ CXCR5- Memory CD4 & CD8 T Cells following SARS-CoV-2 Vaccination



# Sorting T cells on T-REX MEM Phenotype (ICOS<sup>++</sup> CD38<sup>++</sup>) Confirms Specific SARS-CoV-2 Spike Peptide Reactivity



$T_{FH}/T_{FC}$ ? Only half of these cells were BCL-6+, and the cells from T-REX were CXCR5-



T-REX revealed virus-specific T cells without tetramers,  
characterized massive immune changes in COVID-19,  
& identified a SARS-CoV-2 reactive non-canonical  
memory T cell that expands by day 28 following RNA vaccination

Check out the pre-print for more, including plasmablasts, B cell LIBRA-seq,  
and a breakthrough case who did NOT generate the ICOS+ CD38+ T cells.



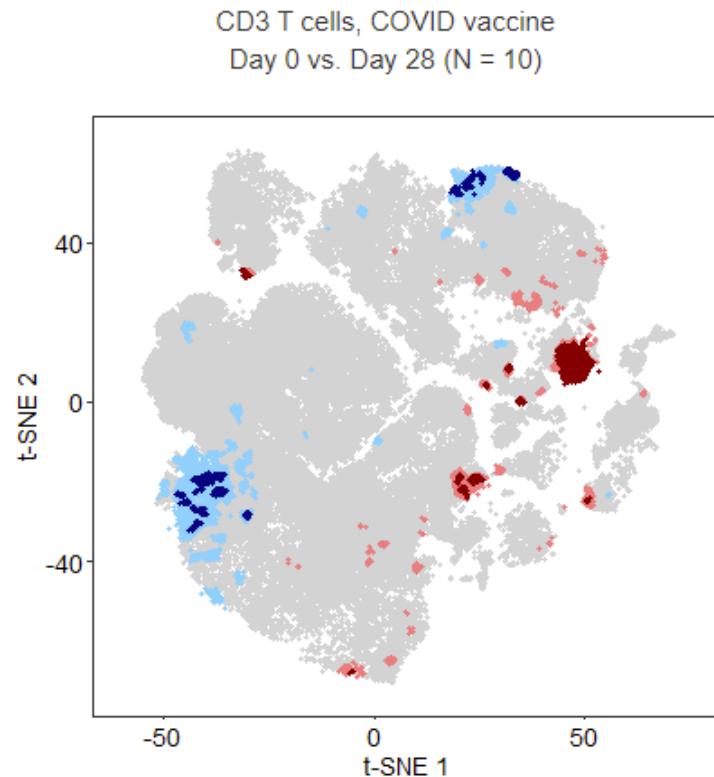
# Let's Analyze Using T-REX!

---

<https://cytolab.shinyapps.io/TREX/>

This web app is running R code live.

1) Identify populations of expansion and contraction with T-REX



- CUTOFF
- $\geq 95\%$  from day 0
  - 85-95% from day 0
  - from day 0 and 28
  - 85-95% from day 28
  - $\geq 95\%$  from day 28

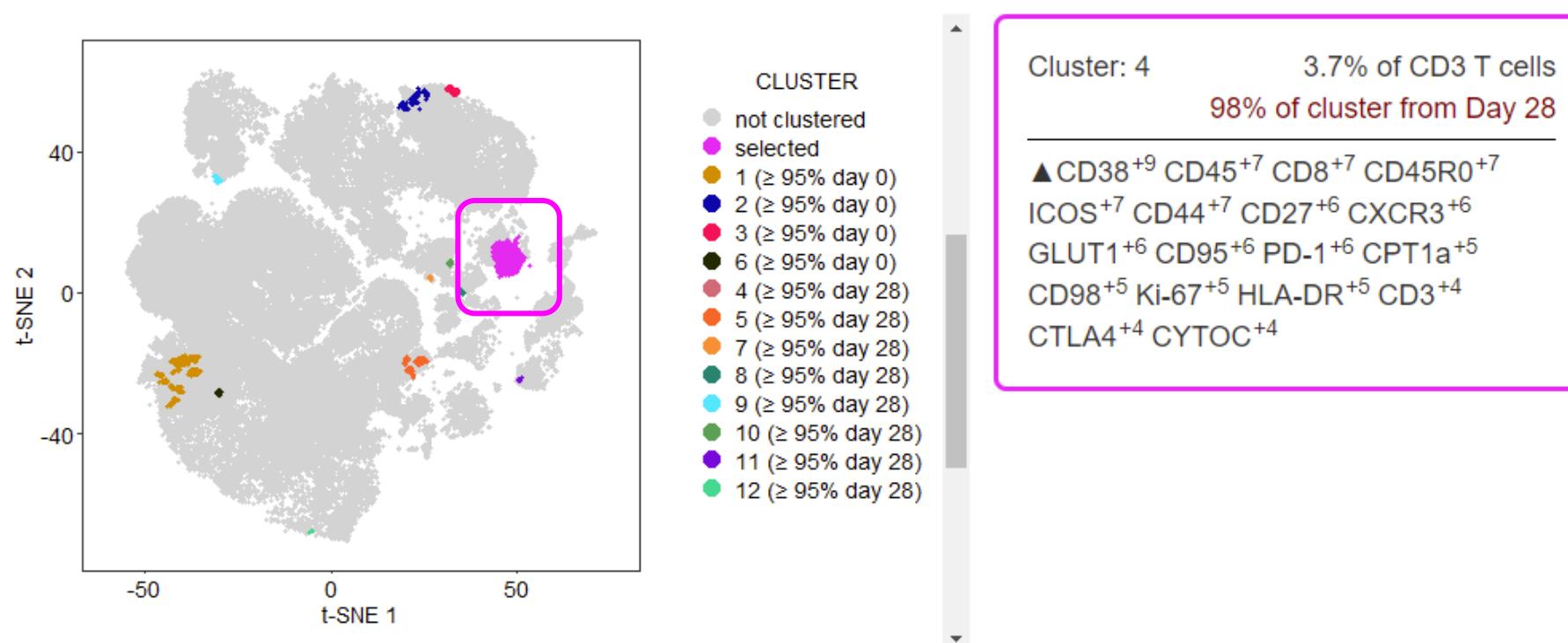
K-VALUE (# of nearest neighbors)

2      60      300

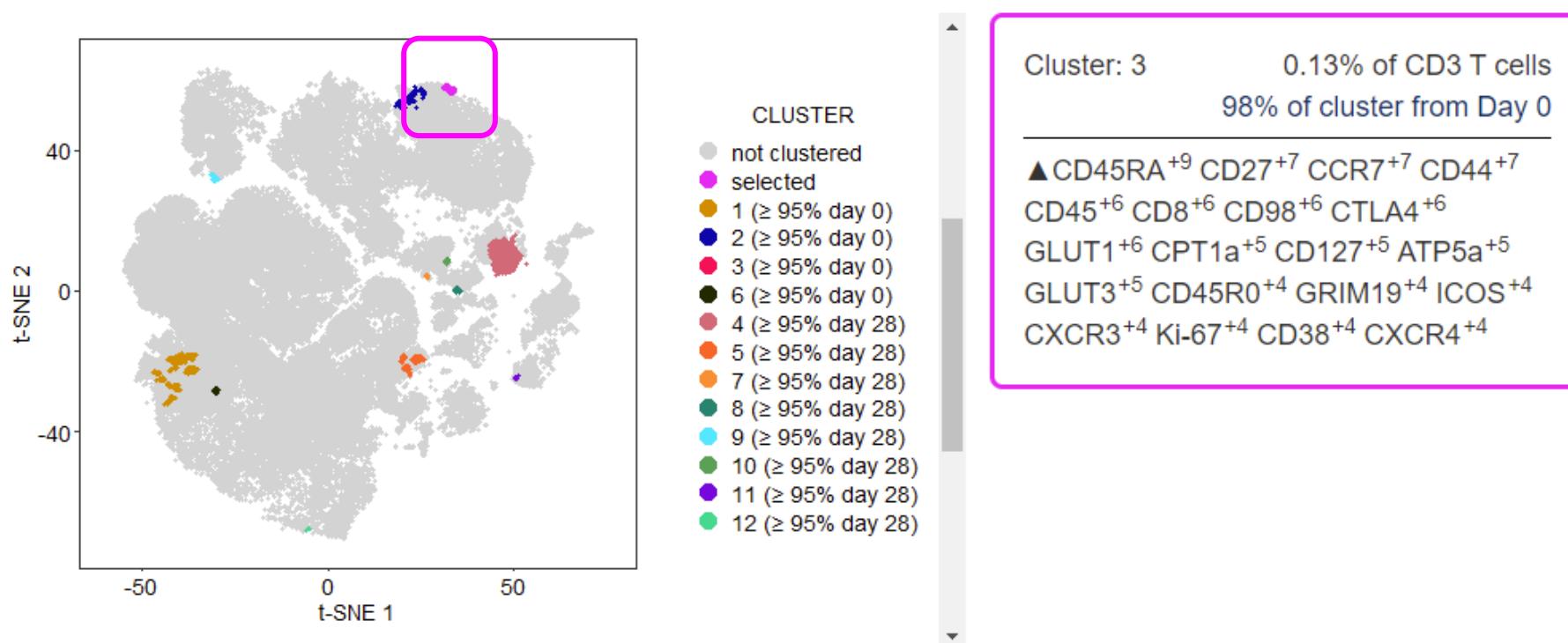
2 32 62 92 122 152 182 212 242 272 300

APPLY

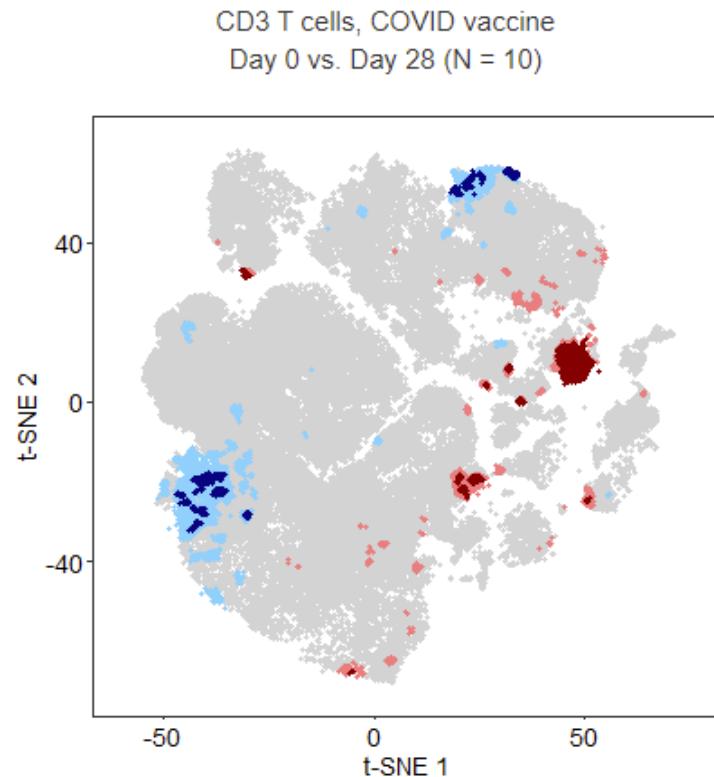
2) Cluster with DBSCAN, and examine MEM labels



2) Cluster with DBSCAN, and examine MEM labels



1) Identify populations of expansion and contraction with T-REX



- CUTOFF
- $\geq 95\%$  from day 0
  - 85-95% from day 0
  - from day 0 and 28
  - 85-95% from day 28
  - $\geq 95\%$  from day 28

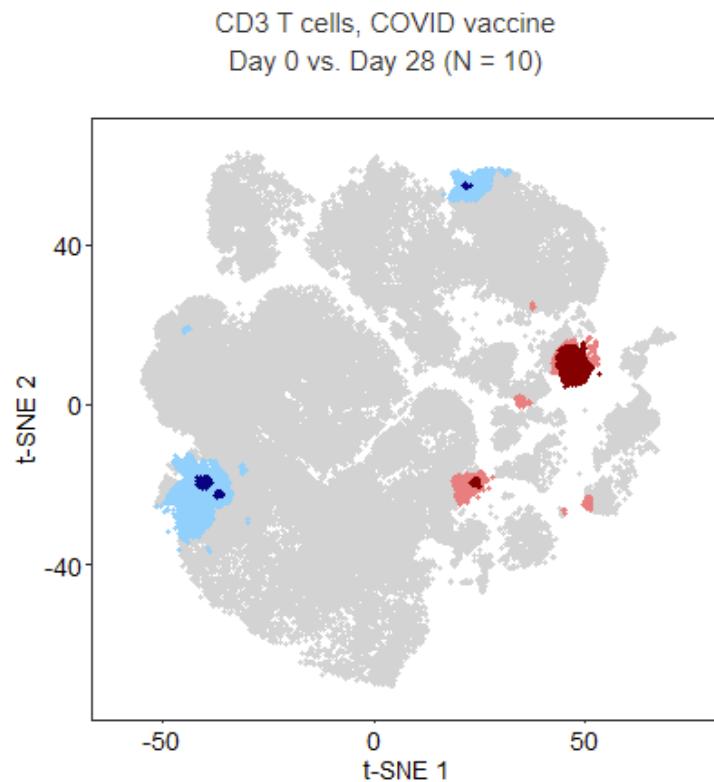
K-VALUE (# of nearest neighbors)

2      60      300

2 32 62 92 122 152 182 212 242 272 300

APPLY

1) Identify populations of expansion and contraction with T-REX



- CUTOFF
- $\geq 95\%$  from day 0
  - 85-95% from day 0
  - from day 0 and 28
  - 85-95% from day 28
  - $\geq 95\%$  from day 28

K-VALUE (# of nearest neighbors)

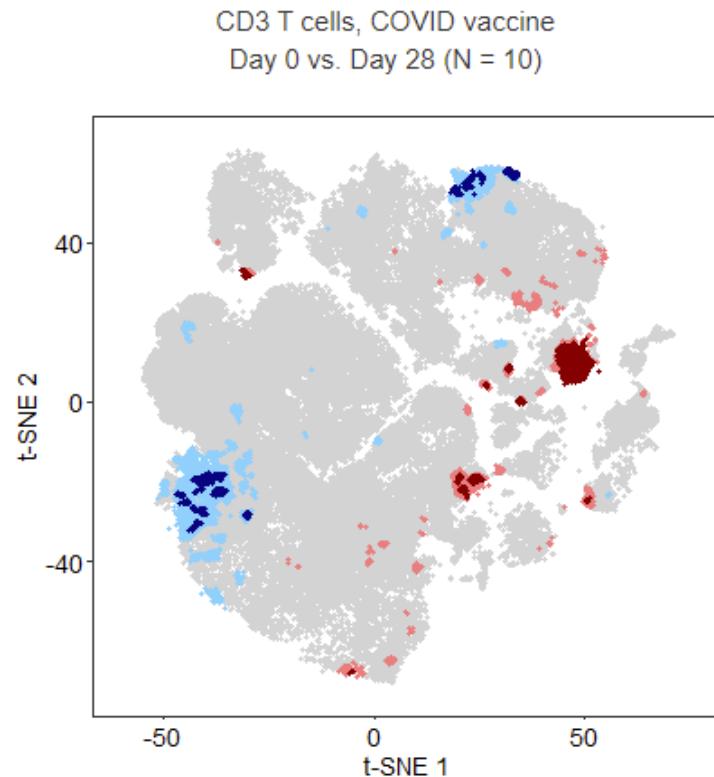
2      240      300

2 32 62 92 122 152 182 212 242 272 300

APPLY

A slider interface for selecting the K-value (number of nearest neighbors). The slider is set to 240, which is highlighted with a pink rectangle. The range of the slider is from 2 to 300, with major tick marks every 32 units. A blue button labeled 'APPLY' is located below the slider.

1) Identify populations of expansion and contraction with T-REX



- CUTOFF
- $\geq 95\%$  from day 0
  - 85-95% from day 0
  - from day 0 and 28
  - 85-95% from day 28
  - $\geq 95\%$  from day 28

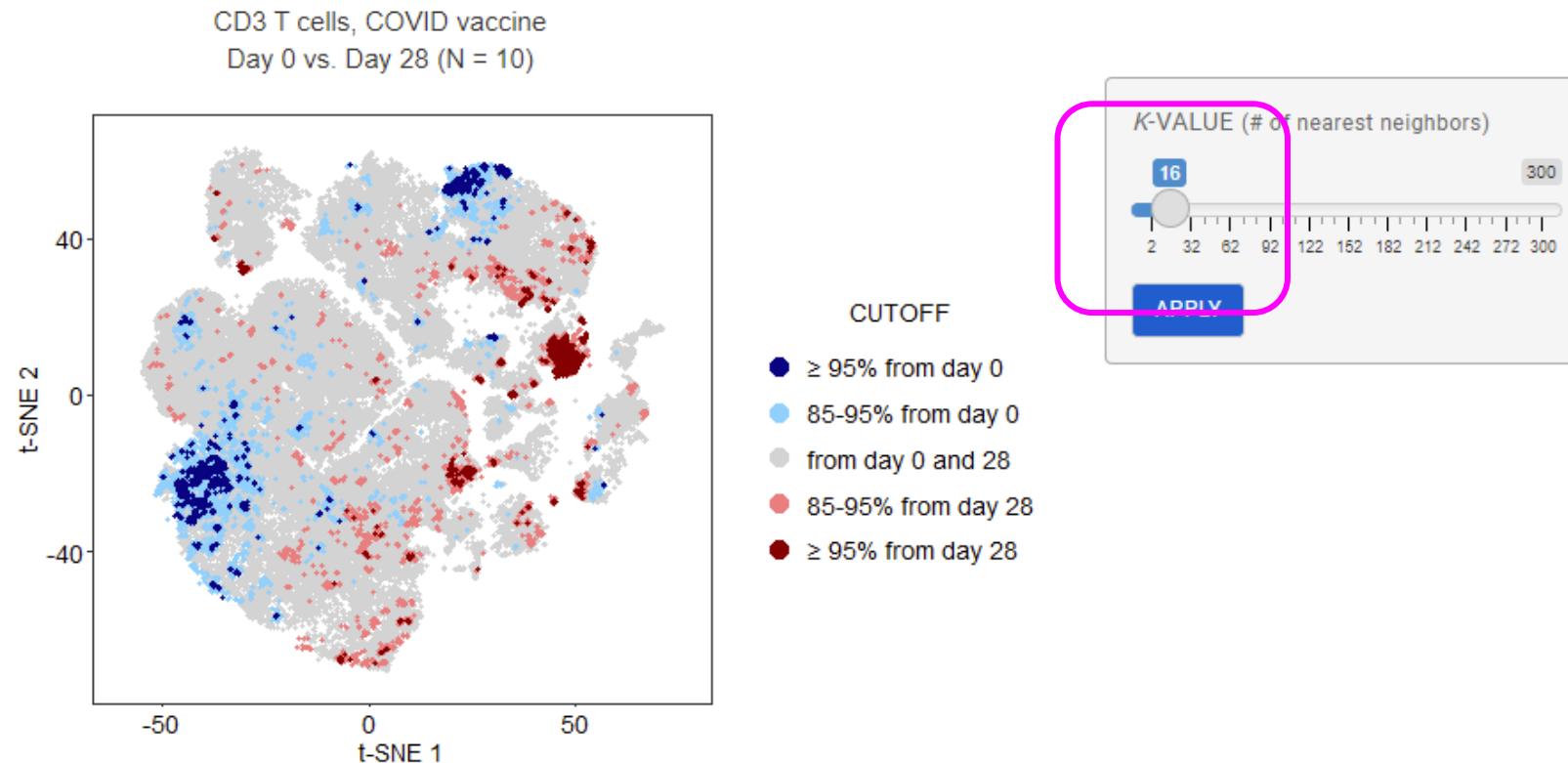
K-VALUE (# of nearest neighbors)

2      60      300

2 32 62 92 122 152 182 212 242 272 300

APPLY

1) Identify populations of expansion and contraction with T-REX



# T-REX & COVID-19 Acknowledgements

## Irish Lab at Vanderbilt University + Cancer & Immunology Core



Stephanie Medina  
PhD Student



Amanda Kouaho  
VU Undergraduate



Sierra Barone Lima  
Data Science  
Program Coordinator



Todd Bartkowiak  
PhD Postdoc  
K00 Fellow



Caroline Roe  
CIC/MCCE  
Senior Research  
Specialist



Madeline Hayes  
Lab Development  
Program  
Coordinator

## University of Virginia Collaborators (Rhinovirus T-REX)



Alberta Paul



Lyndsey Muehling



Judith Woodfolk



Kevin Kramer



Erin Wilfong



Kelsey Voss



Rachel Bonami



Ivelin Georgiev



Jeff Rathmell

Southeastern Brain Tumor Foundation (Ihrie & Irish), NIH/NCI : R00 CA143231 (Irish), R01 CA226833 (Bachmann & Irish), **HIDI TIPS (Vanderbilt)**  
U01 CA196405 (Massion), CSBC U54 CA217450 (Quaranta), **U01 AI125056 (Woodfolk)**, **R01 HL136664 (Rathmell)**, F31 CA199993 (Greenplate),  
T32 CA009592 & R25 GM062459 (Doxie), R25 CA136440 (Diggins), K12 CA090625 (Ferrell), P30 CA68485 (Vanderbilt-Ingram Cancer Center)

# Acknowledgements & Thank You!

## Past Lab Members

### Grad & Med Students

Deon Doxie (Emory)  
Cara Wogsland (UiB & BergenBio)  
Kirsten Diggins (Benaroya Institute)  
Allison Greenplate (U Penn)  
Nalin Leelatian (Yale & Vanderbilt)  
Jocelyn Gandelman (UCSF)

### Postdocs

Kanutte Huse (Oslo University)  
Mikael Roussel (CHU Rennes)  
P. Brent Ferrell, Jr. (Vanderbilt)  
Ashley Wu (Vanderbilt)

### Undergrads & Staff

Sierra Barone Lima (Data Science)  
Hannah Polikowsky (Vanderbilt)  
Alejandra Rosario-Crespo (U Puerto Rico)  
Daniel McClanahan (Twitter & LBL)  
Daniel Liu (Univ Wisc., Madison)  
Nathan Wasserman (Univ Florida, Miami)

## Visiting Scholars

Shahram Kordasti (King's College London)  
Faustine L'homme (CHU Rennes)  
Monica Hellesøy (Univ. Bergen)  
Aïda Meghraoui-Kheddar (Paris & Nice)  
Eleni Syrimi (Univ. Birmingham, UK)  
Laura Ferrer Font (Malaghan, NZ)

## Irish Lab, Vanderbilt University



Stephanie Medina  
PhD Student  
Cell & Dev Biology

Claire Cross  
PhD Student  
Chem & Phys Bio

Hannah Thirman  
PhD Student  
Chem & Phys Bio

Niraj Rama  
Vanderbilt  
Undergraduate

Amanda Kouaho  
Vanderbilt  
Undergraduate



Todd Bartkowiak  
PhD Postdoc  
K00 Fellow

Caroline Roe  
Director  
CIC Core

Madeline Hayes  
Program Coordinator

Cass Mayeda  
Web Applications Research Assistant

Jonathan Irish  
Principal Investigator

## Collaborations

### Glioblastoma & Neural Stem Cells

Rebecca Ihrie  
Vivian Gama  
Asa Brockman  
Bret Mobley  
Lola Chambliss  
Reid Thompson

### Chemical Biology

Brian Bachmann &  
Lab for Biosynthetic Studies

John Porco (BU)  
Lauren Brown (BU)

### Human Immunology

Discovery Initiative (HIDI)

Jeff Rathmell  
Jim Connelly  
Saara Kavany

### Viral Immunology

Judith Woodfolk (UVA)  
Lyndsey Muehling (UVA)  
Glenda Canderan (UVA)

### Quantitative & Systems Biology Center

Vito Quaranta & many more  
(U54 CA217450)

### Center for Extracellular Vesicles

Alissa Weaver  
Heather Pua  
Andries Zijlstra

Michael David Greene Brain Cancer Fund (Thompson, Ihrie, Irish), Southeastern Brain Tumor Foundation (Ihrie & Irish),  
Ivy Foundation (Ihrie & Irish), R01 NS118580 (Ihrie & Ess), R00 CA143231 (Irish), R01 CA226833 (Bachmann & Irish),  
U01 CA196405 (Massion), CSBC U54 CA217450 (Quaranta), R01 HL136664 (Rathmell), U01 AI125056 (Woodfolk), U01 TR002625 (Porco),  
Cancer & Immunology Core, Flow Cytometry Core, HIDI TIPS (Vanderbilt), P30 CA68485 (Vanderbilt-Ingram Cancer Center)



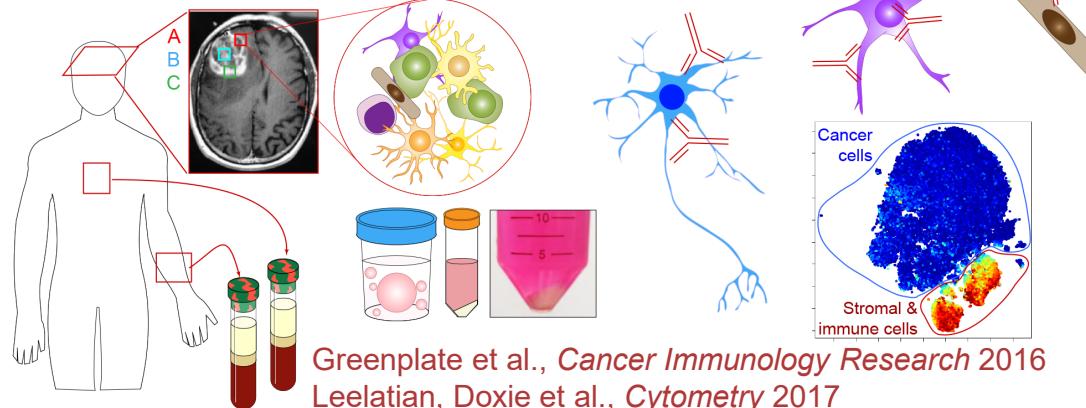
# Irish Lab @ Vanderbilt University

## Single cell biology for precision medicine

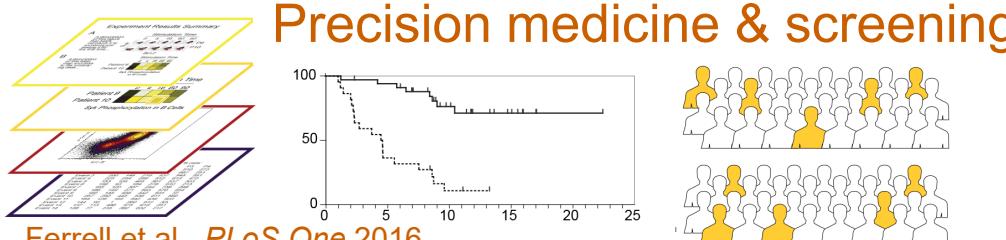
Irish lab website:



### Human immune & tumor tissues

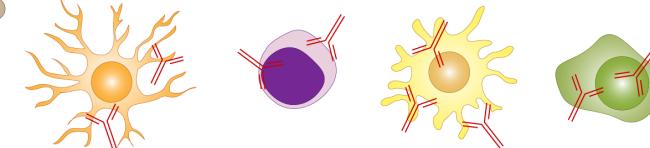


Greenplate et al., *Cancer Immunology Research* 2016  
Leelatian, Doxie et al., *Cytometry* 2017  
Doxie et al., *Pigment Cell & Melanoma Research* 2018  
Leelatian, Sinnaeve et al., *eLife* 2020 ([RAPID](#))



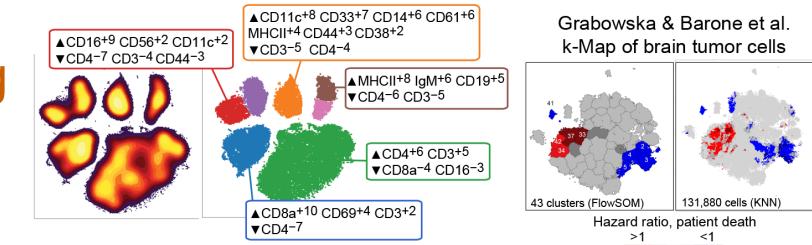
Ferrell et al., *PLoS One* 2016  
Earl, Ferrell et al., *Nature Communications* 2018  
Greenplate et al., *Cancer Immunology Research* 2019  
Kramer, Wilfong, Voss et al., *bioRxiv* 2021 (COVID vaccine response)

### Cell signaling networks & immune development



Polikowsky et al., *J Immunology* 2015  
Roussel et al., *J Leukocyte Biology* 2017  
Huse et al., *Cytometry* 2018  
Bartkowiak et al., 2021 (in prep)

### AI & machine learning



Diggins et al., *Methods* 2015  
Diggins et al., *Nature Methods* 2017 ([MEM](#))  
Gandelman et al., *Hematologica* 2018  
Barone, Paul, Muehling et al., *eLife* 2020 ([T-REX](#))

### Cell & Developmental Biology

#### Cancer Biology

### Chemical & Physical Biology

#### Molecular Pathology & Immunology

# Vanderbilt University in Nashville, TN



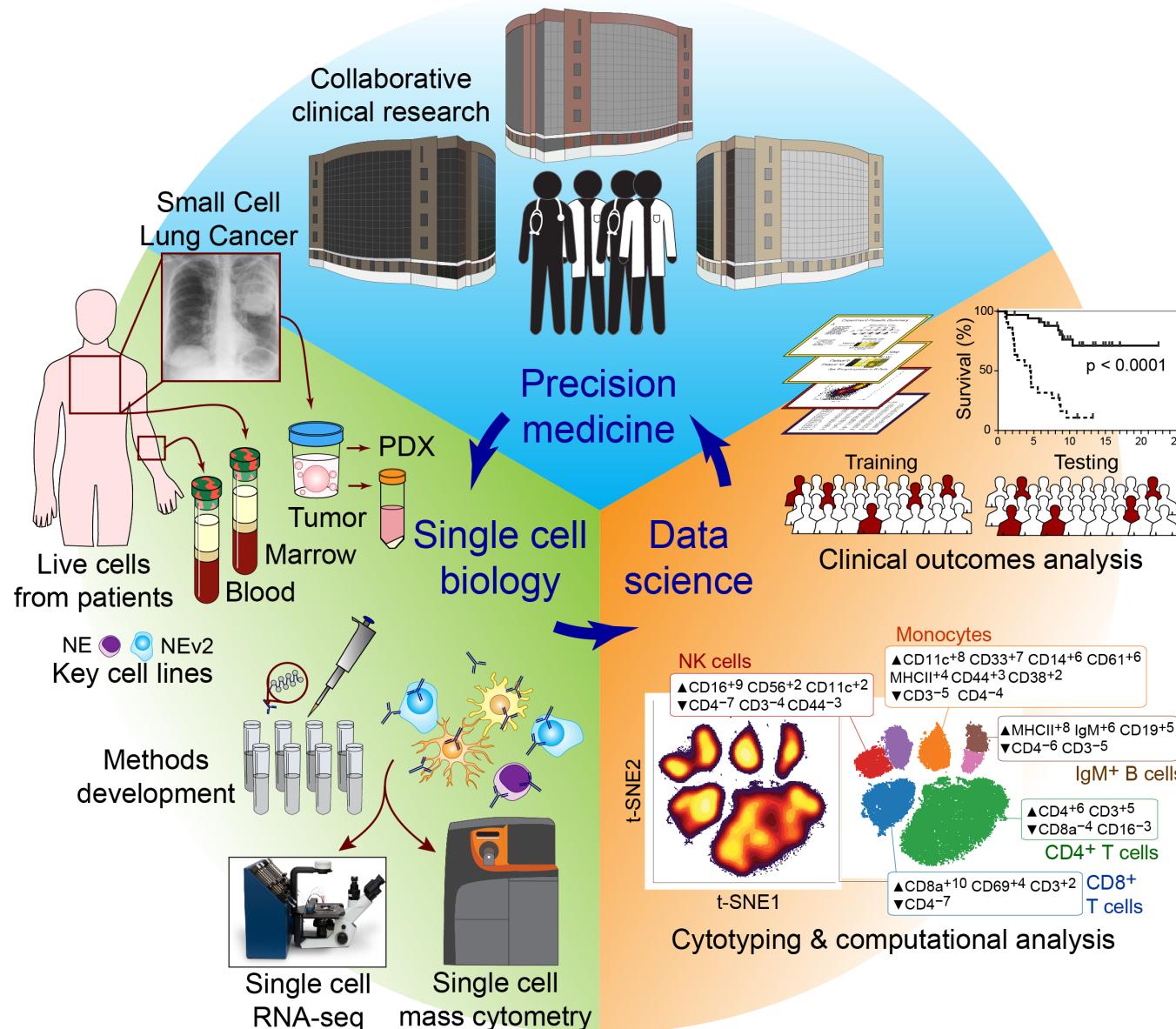
*Jonathan Irish, Ph.D.*  
Vanderbilt University, Nashville, TN  
Cell & Developmental Biology  
Pathology, Microbiology & Immunology

 @JonathanIrish

2022

# Introduction to Data Science and Computational Tools

# Goal: Systematically Dissect Cellular Mechanisms Across Time, Treatments, Tissues, & Tumor Types



# Imagine Finding Pieces of a Jigsaw Puzzle...

---

## Flow Cytometry

## Puzzle

Setup

**Manual review**  
(scaling, single cell gating,  
compensation, batch correction)

**Manual review**  
(make sure all the pieces  
are from the same puzzle)

Organization

**t-SNE, UMAP, PCA**  
(simplify the problem by  
organizing the data)

**Group pieces**  
(find corners, edges,  
pieces with distinct colors)

Grouping

**FlowSOM, SPADE, gating**  
(split cells into cell types  
like T cells or monocytes)

**Assemble parts**  
(connect similar pieces,  
create distinct shapes)

Interpretation

**Heatmaps, MEM, RMSD**  
(analyze group features,  
learn cell identities)

**Interpret picture**  
(see both the pieces  
and the whole picture)

Effective data analysis is critical in clinical research,  
& this now means working *with* computational tools  
that reveal and model patterns across data types

Tools from one area can be applied in others  
(economics, math, patients, cells, pixels, ...)

Data science workshop can be self-taught:

<https://github.com/cytolab/>

# Unsupervised Analysis: Not Using Prior Knowledge To Guide the Analysis

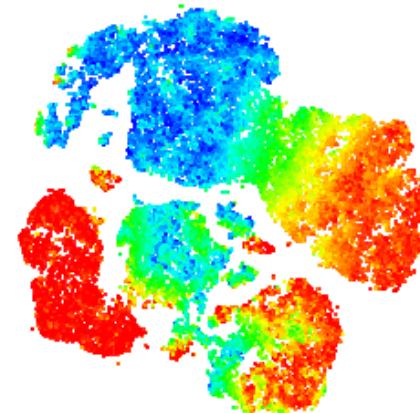
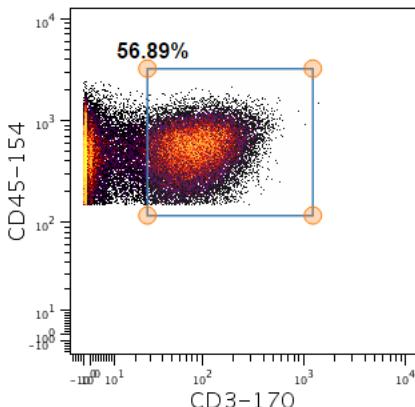
Prior knowledge examples: Stem cells express CD34, these samples were from patients that responded to drug

## Supervised Approaches

- Expert gating
- Citrus
- CellCNN (neural network)
- Wanderlust

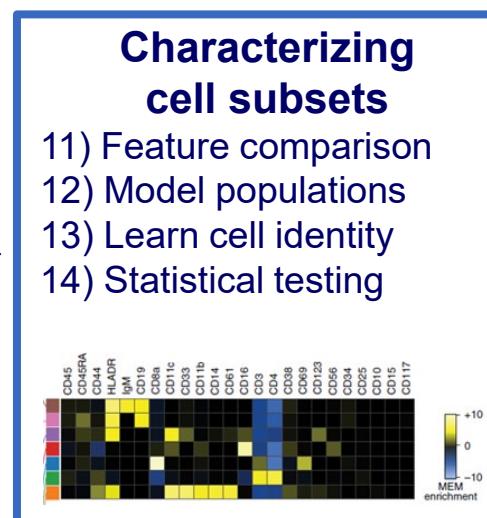
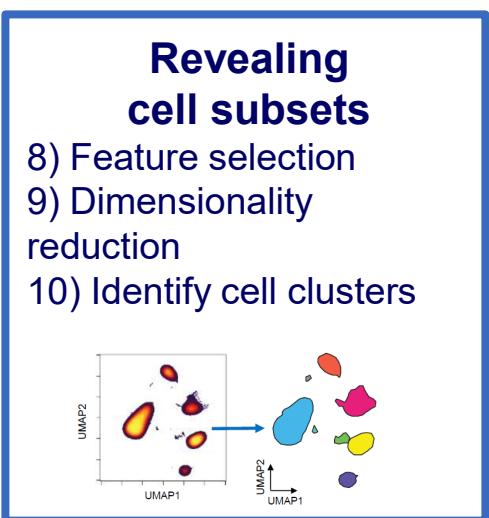
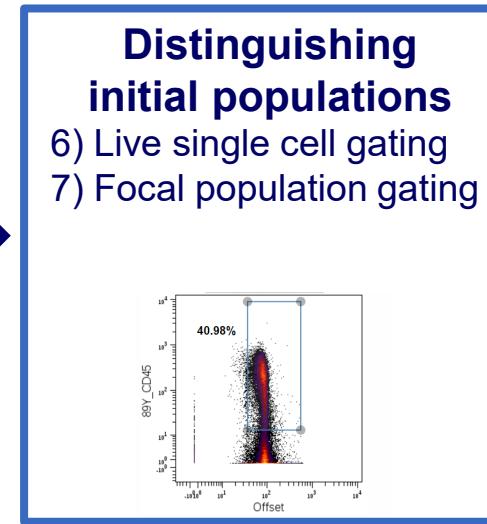
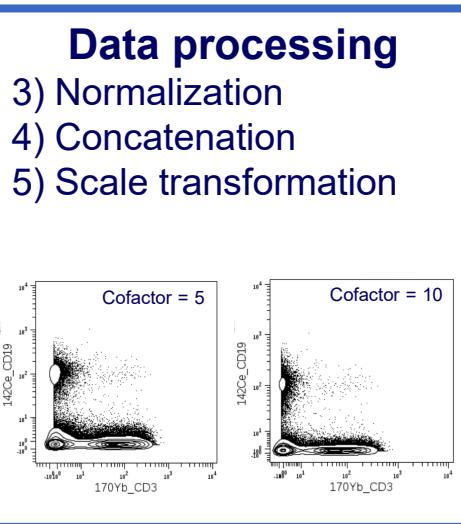
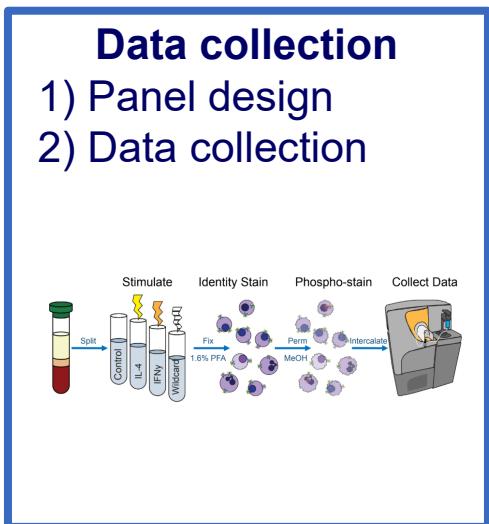
## Unsupervised Approaches

- Most heatmap clustering
- SPADE, FlowSOM
- t-SNE / viSNE, UMAP
- Phenograph



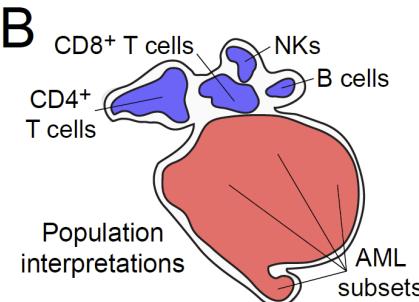
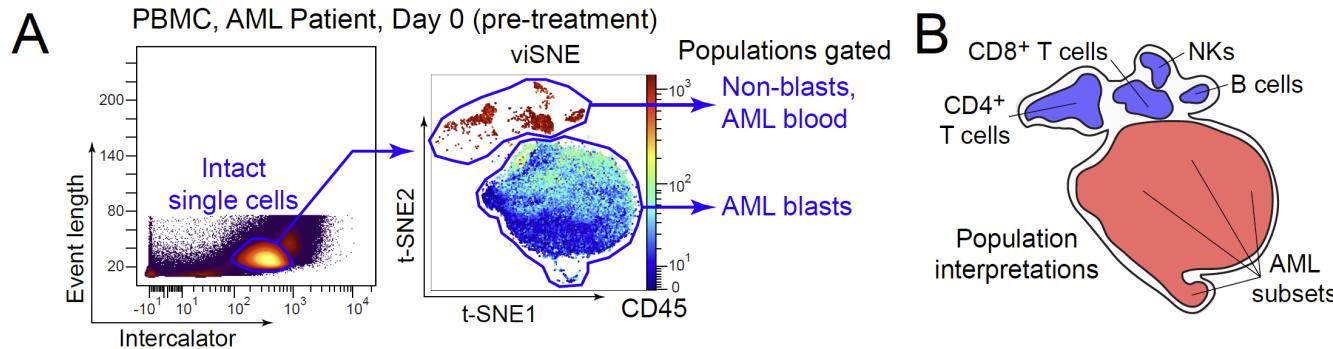
See Table 1 of Diggins et al., *Methods* 2015 for list of unsupervised tools

# Flow Cytometry Workflow from Data Collection to Deep Analysis



**How much can be automated?**  
**How do we select tools and use them well?**

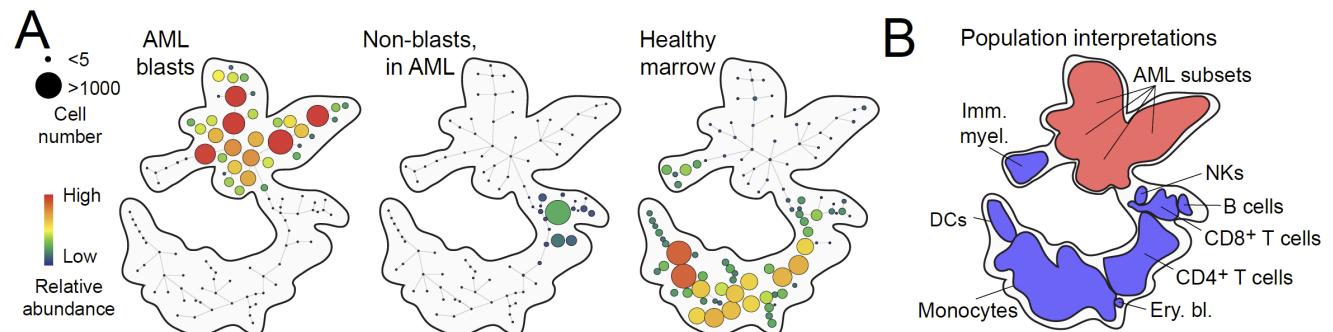
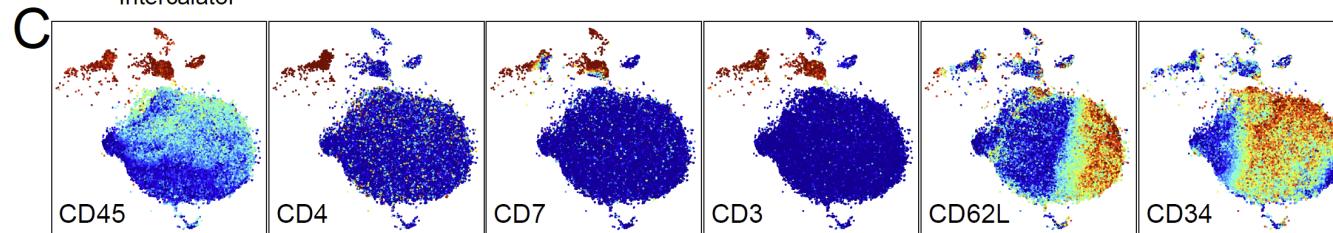
# Key Analysis Concepts: Dimensionality Reduction, Transformation, Clustering, Modeling, Visualization, & Integration



viSNE

Amir et al.

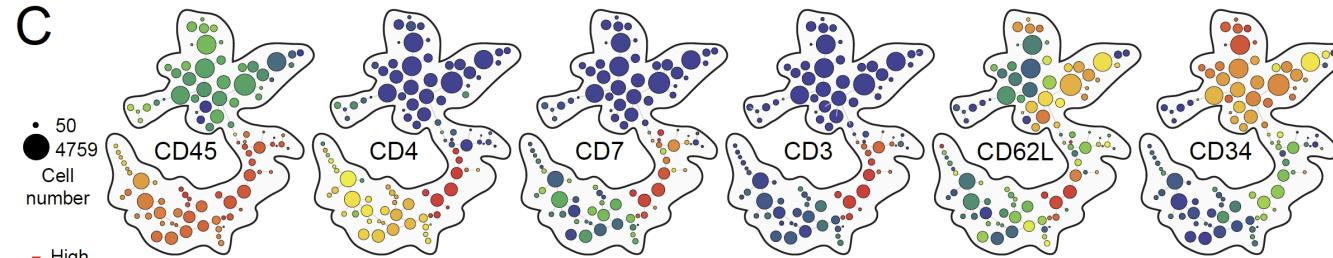
*Nature biotech* 2013



SPADE

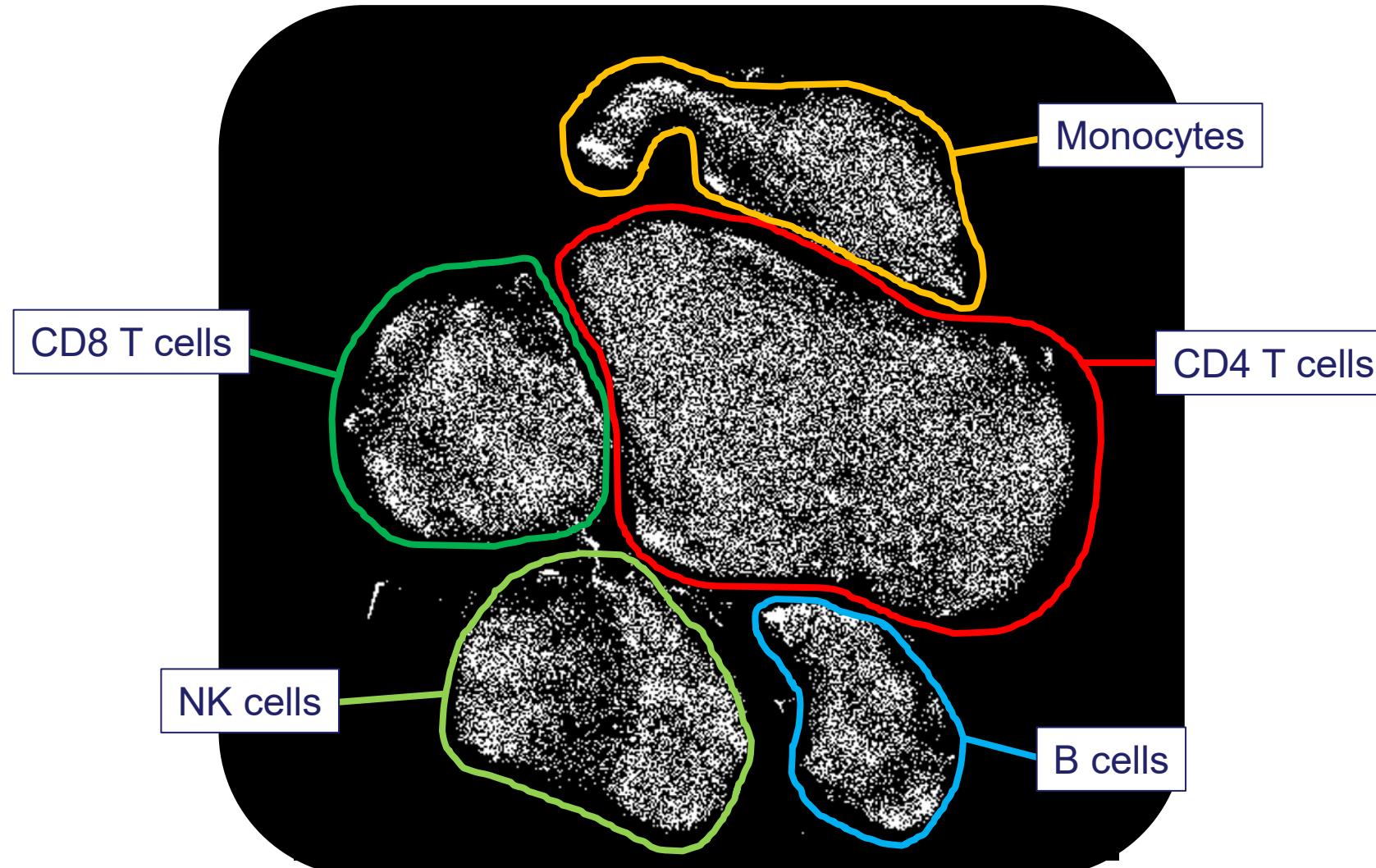
Qiu et al.

*Nature biotech* 2011



Diggins et al., *Methods* 2015

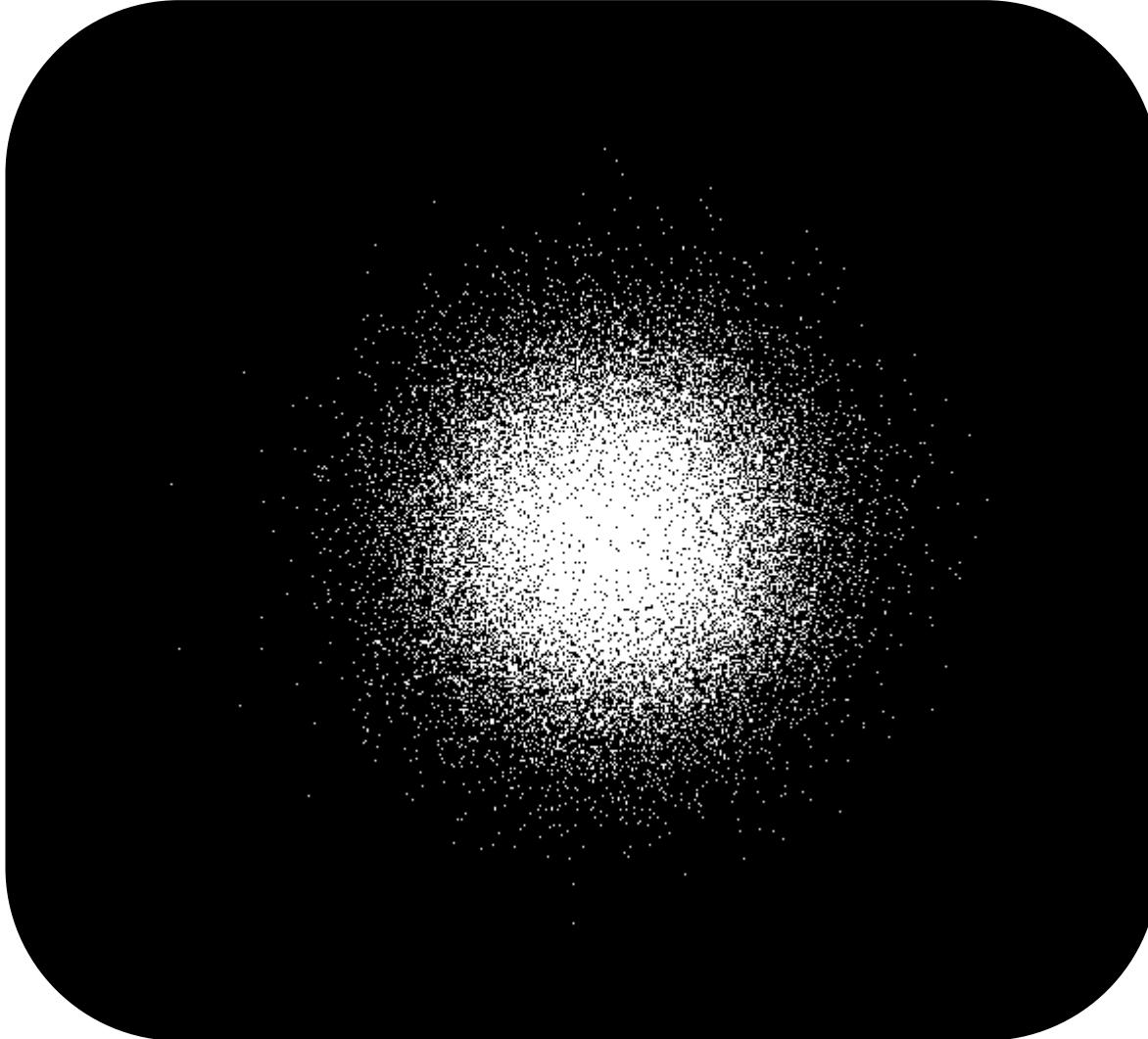
# viSNE / t-SNE Arranges Cells in 2D by Multi-D Similarity



Healthy human blood, mass cytometry,  
26 markers measured, viSNE analysis tool

Animation created by Cytobank team from iterations of viSNE / t-SNE using PBMC (26 features)

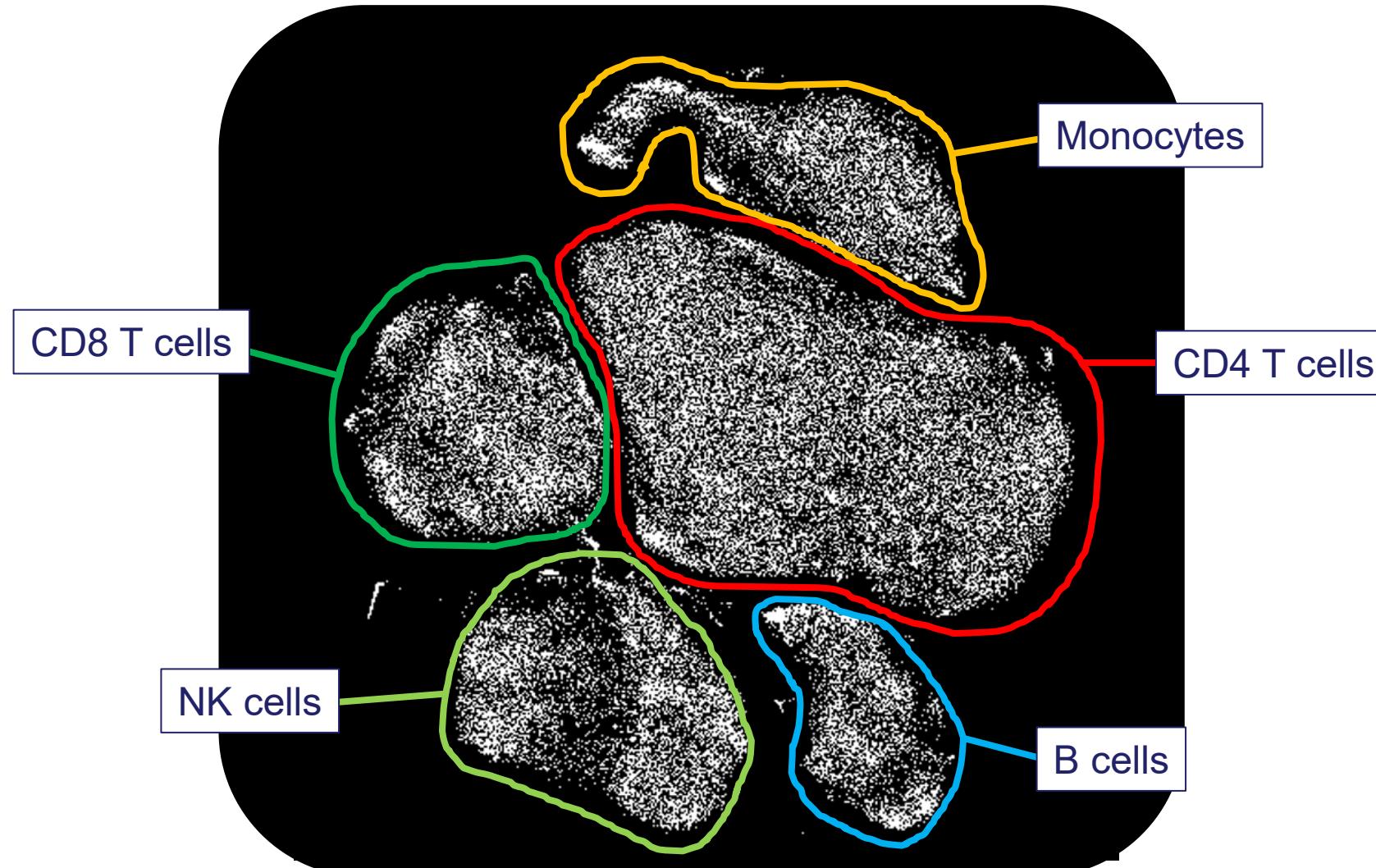
# viSNE / t-SNE Arranges Cells in 2D by Multi-D Similarity



Healthy human blood, mass cytometry,  
26 markers measured, viSNE analysis tool

Animation created by Cytobank team from iterations of viSNE / t-SNE using PBMC (26 features)

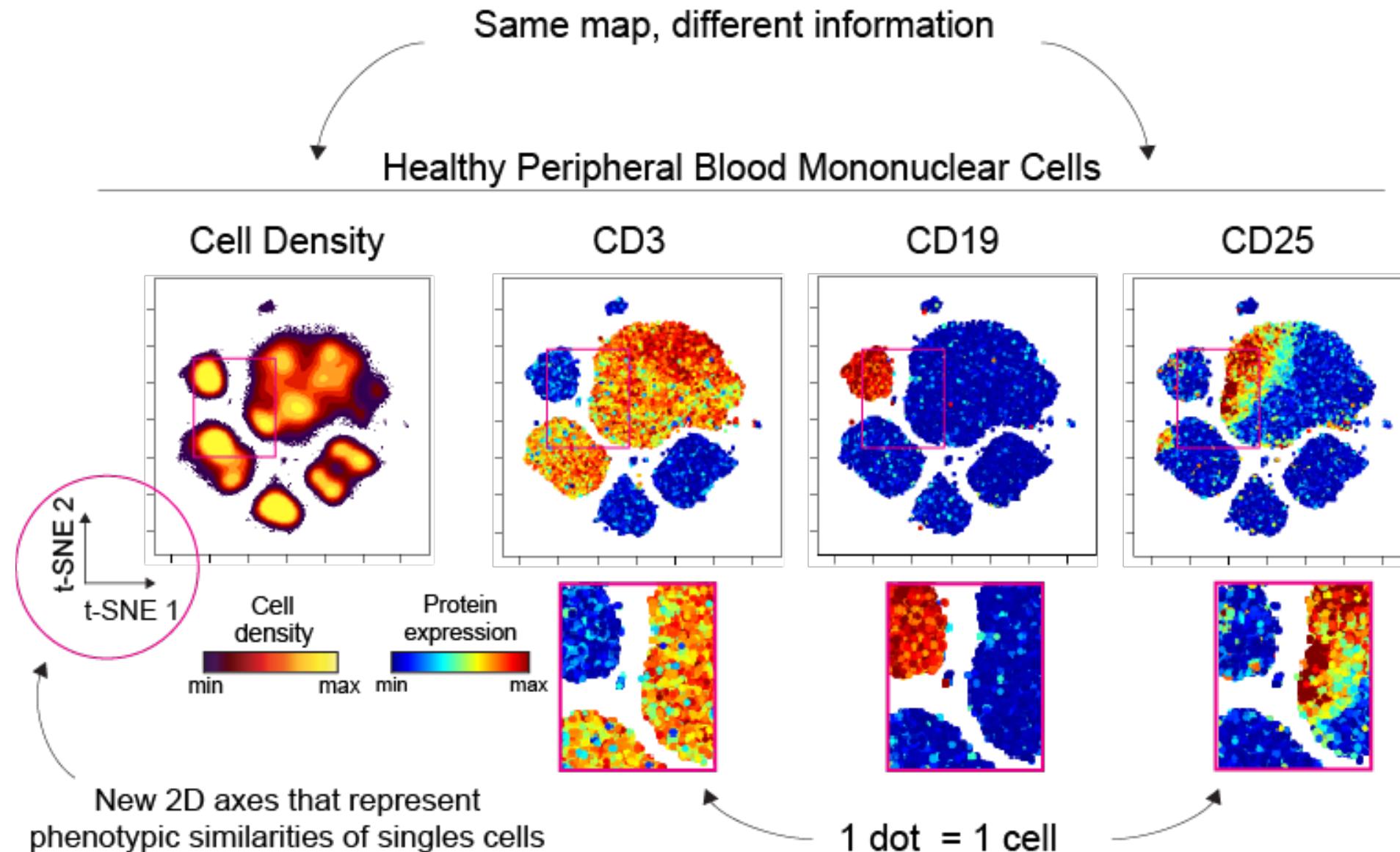
# viSNE / t-SNE Arranges Cells in 2D by Multi-D Similarity



Healthy human blood, mass cytometry,  
26 markers measured, viSNE analysis tool

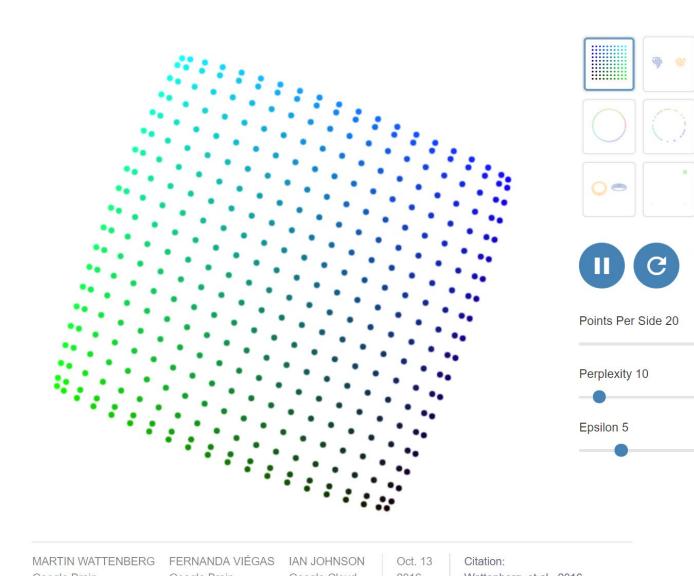
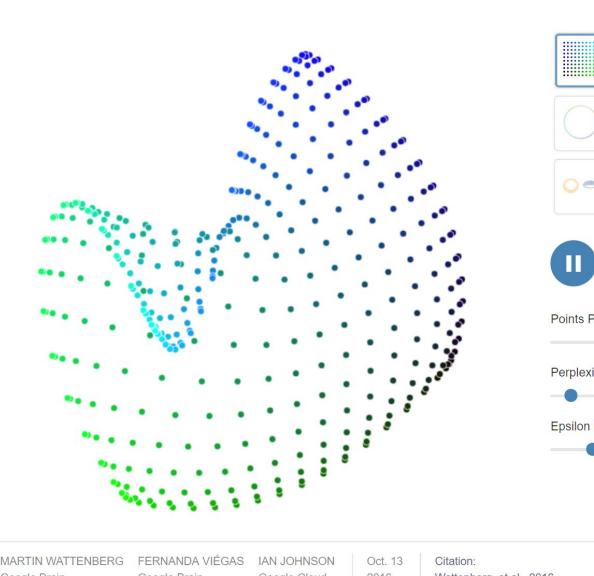
Animation created by Cytobank team from iterations of viSNE / t-SNE using PBMC (26 features)

# t-SNE Analysis Allows 2D Visualization of High Dimensional Single Cell Data

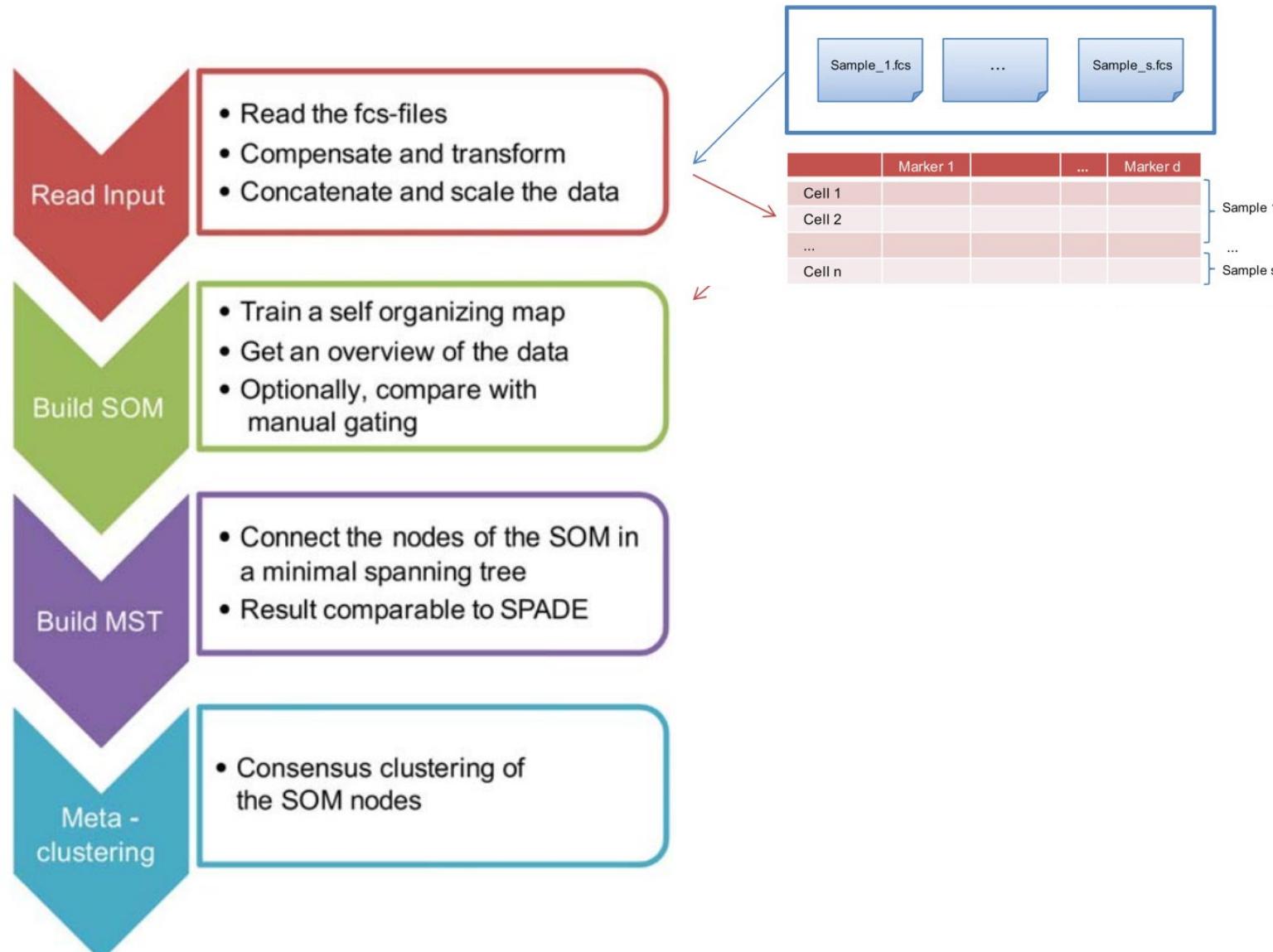


# t-SNE 2D Examples with Animations and Settings

<http://distill.pub/2016/misread-tsne/>

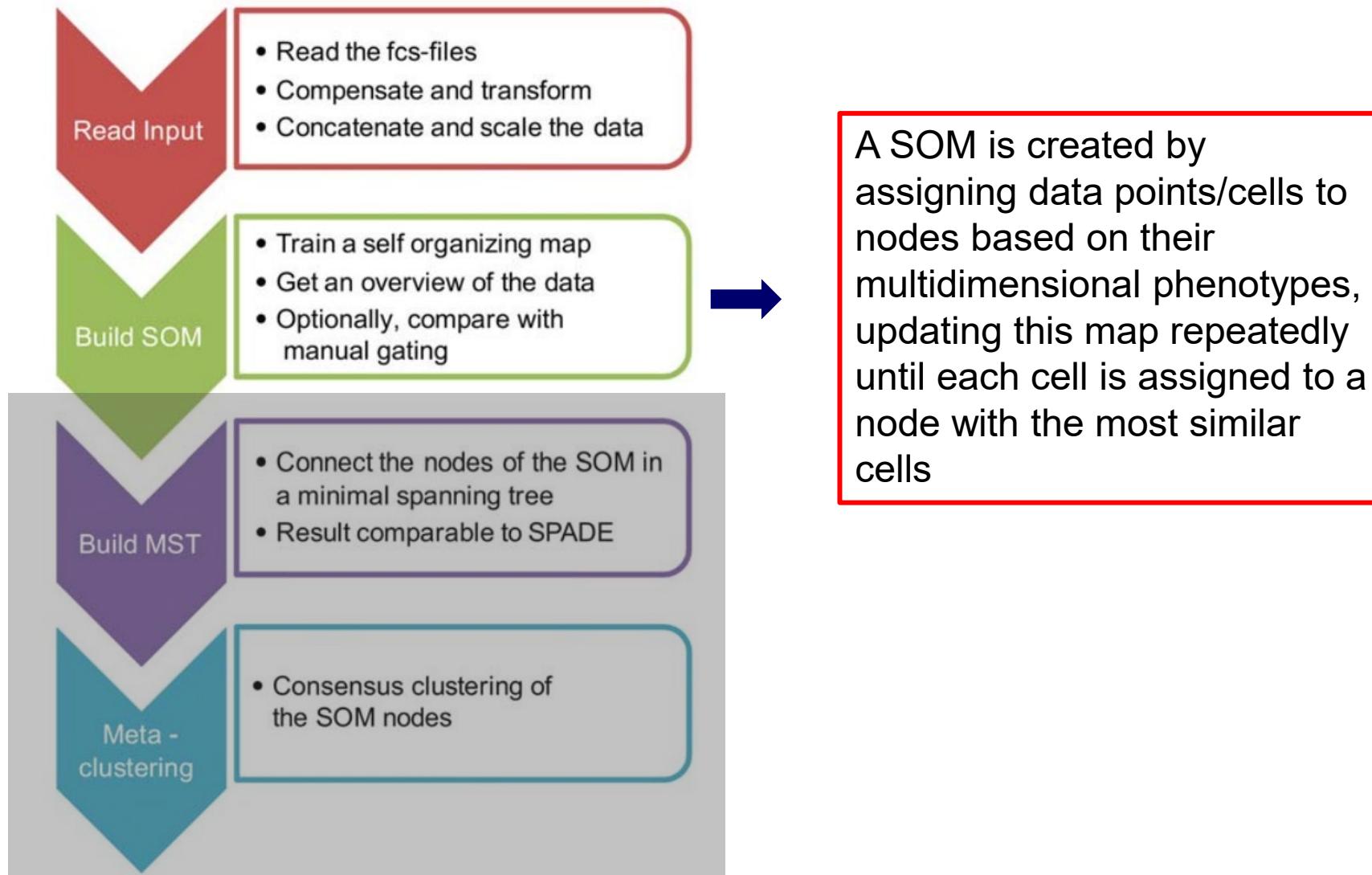


# Clustering with FlowSOM: Self-organizing Maps

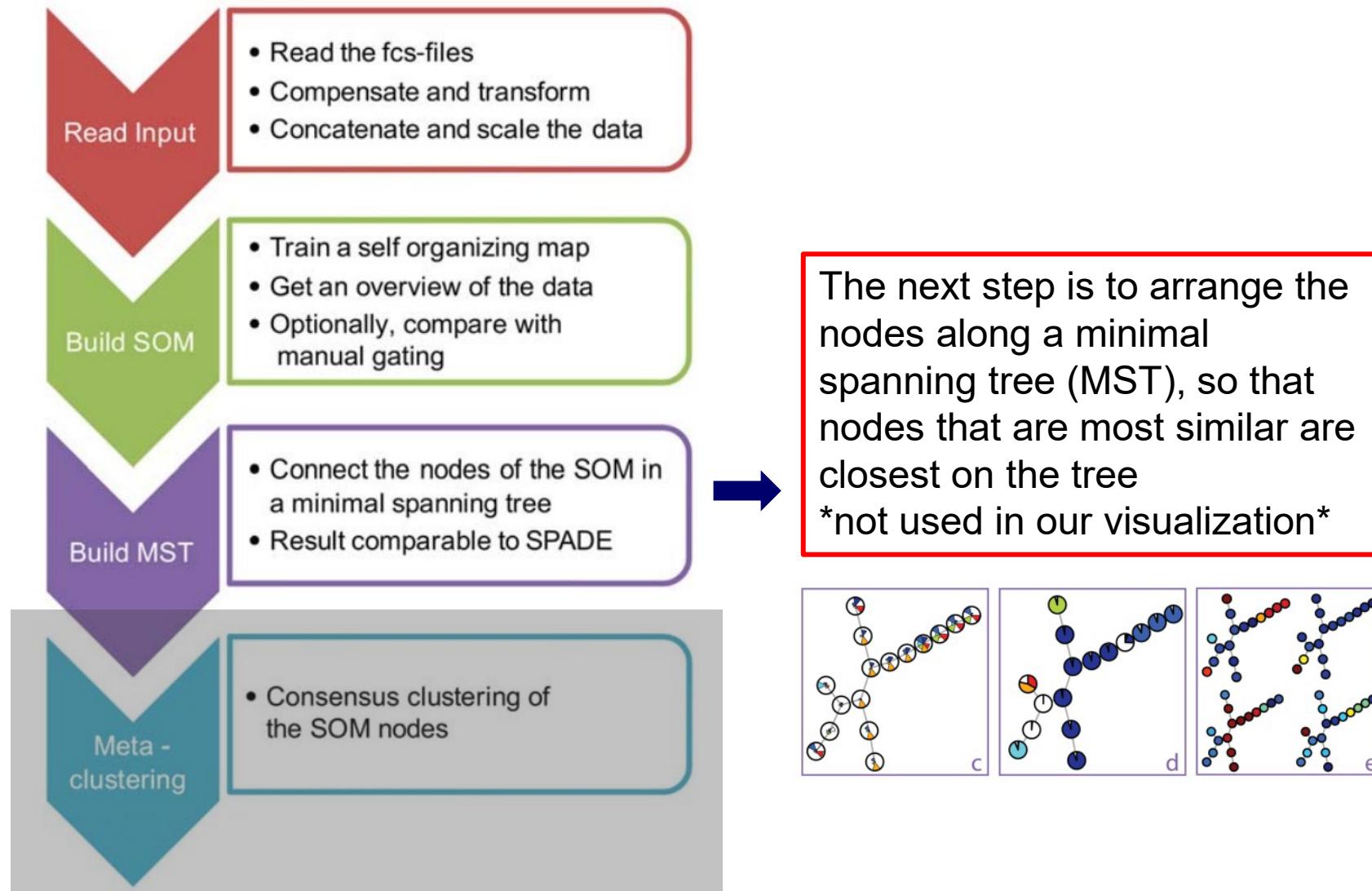


# Clustering with FlowSOM: Self-organizing Maps

---



# Clustering with FlowSOM: Self-organizing Maps

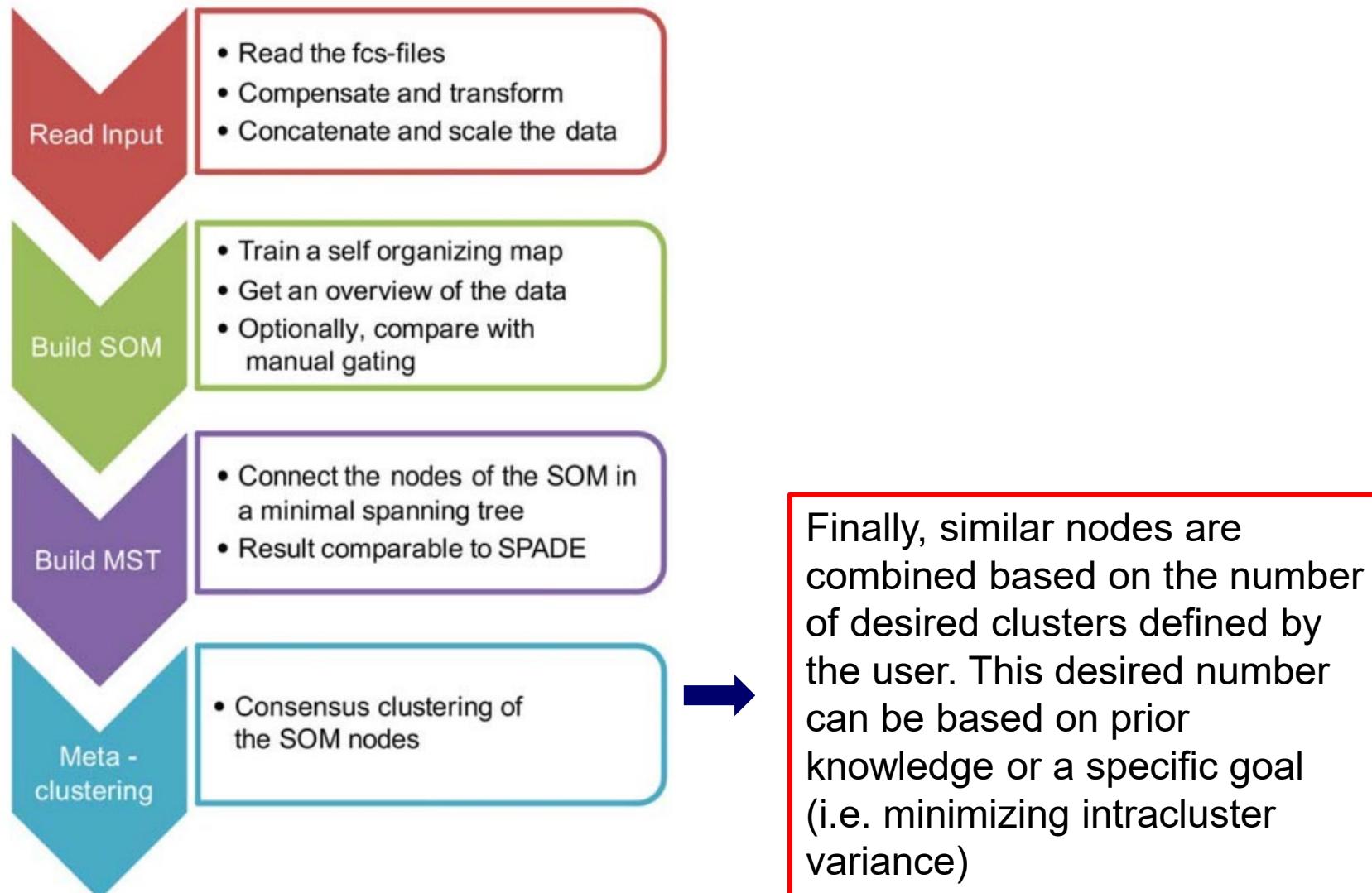


Van Gassen et al., *Cytometry A* 2015

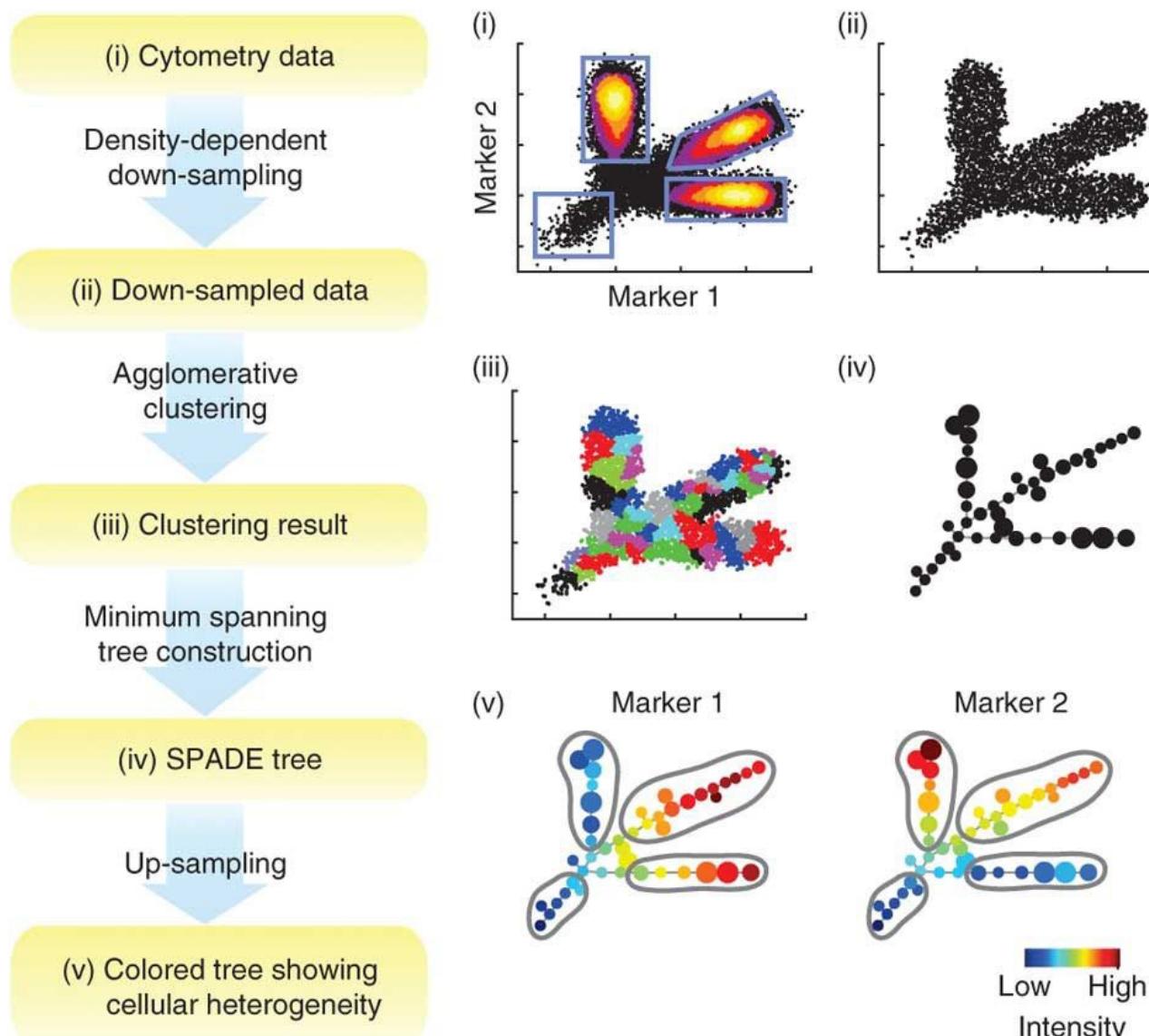
See also, FlowSOM does well in a comparison of clustering tools: Weber & Robinson, *Cytometry A* 2017

# Clustering with FlowSOM: Self-organizing Maps

---

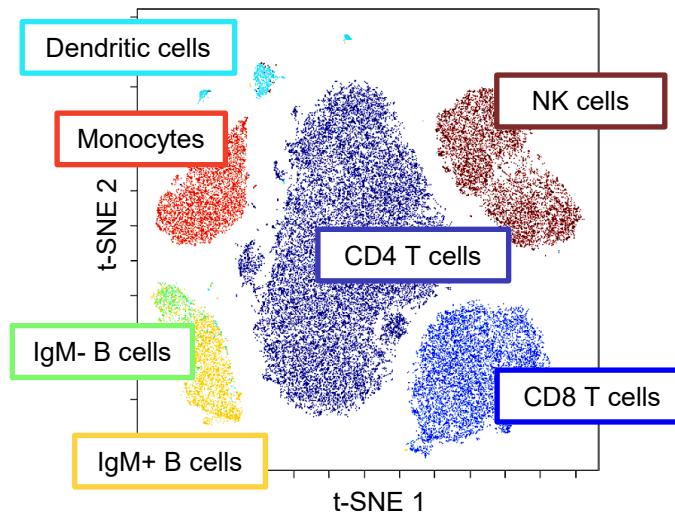


# Spanning-Tree Progression Analysis of Density-Normalized Events (SPADE) is an Alternative Clustering Tool

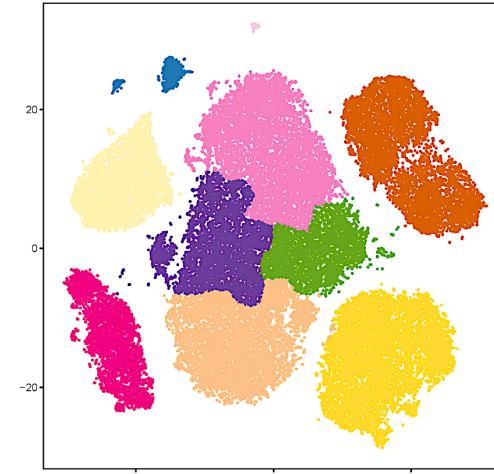


# FlowSOM Clusters are Dependent on Input Parameters

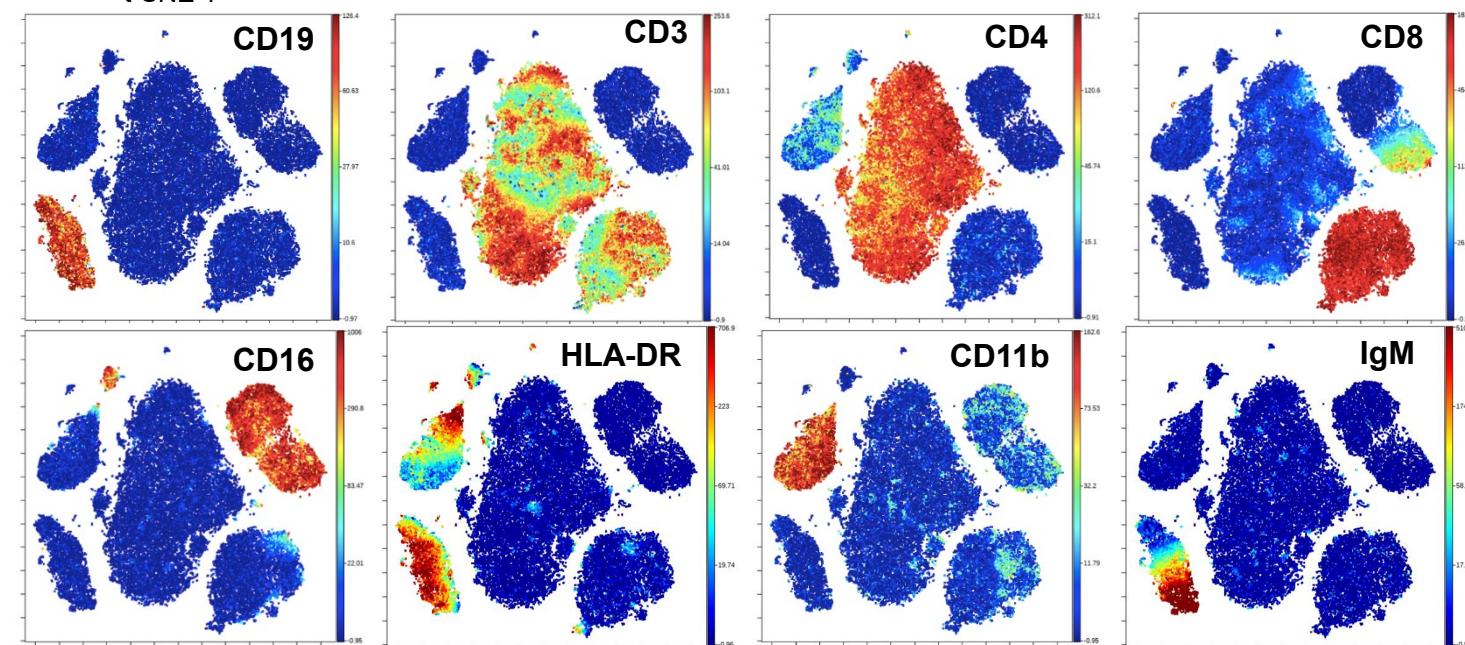
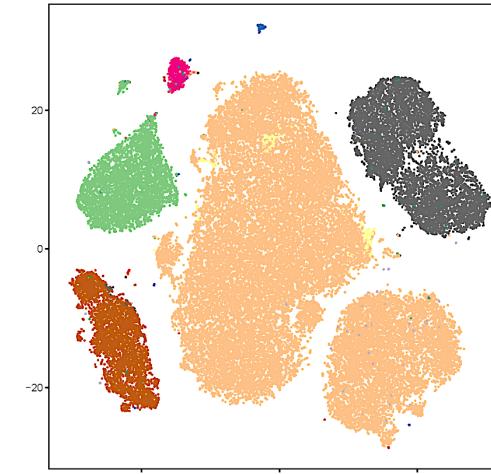
Major Populations Overlaid on t-SNE Axes



FlowSOM on t-SNE Axes (n = 10)

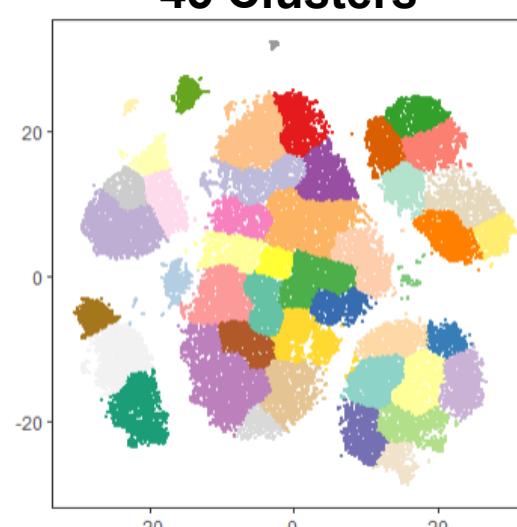
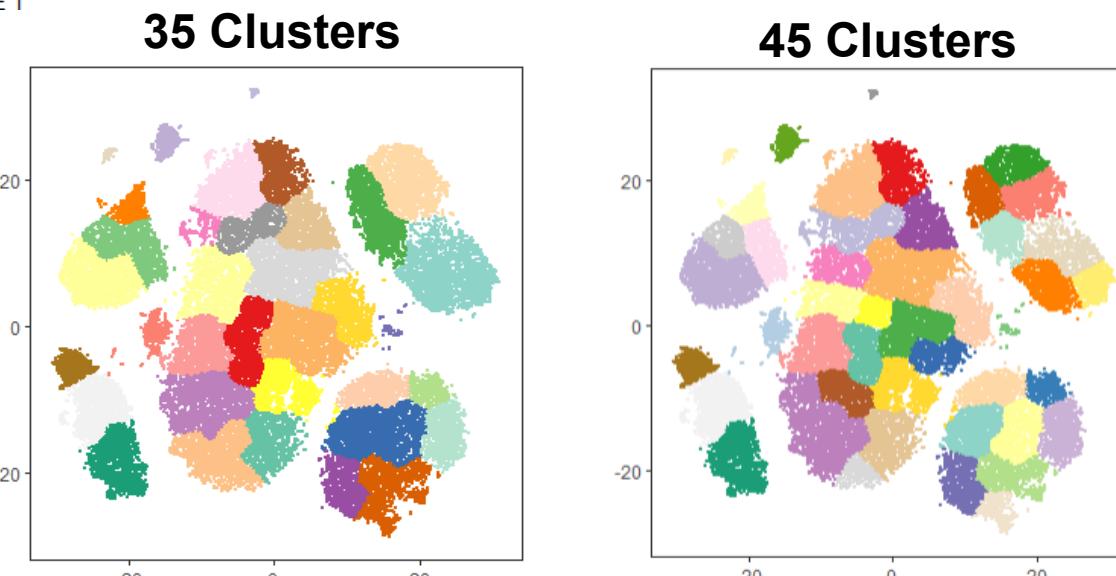
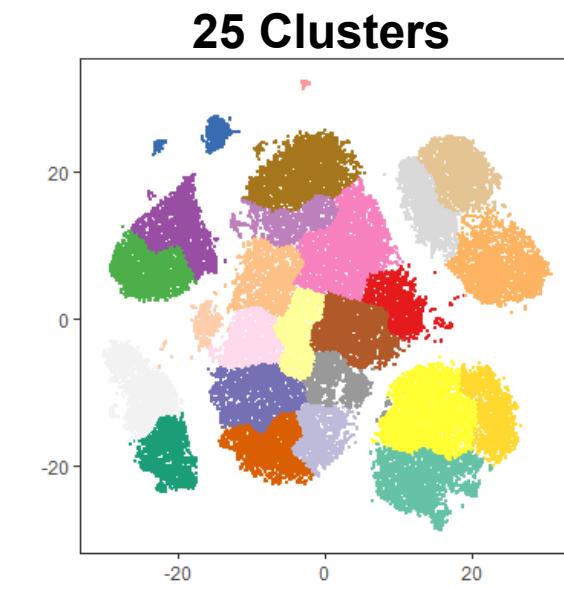
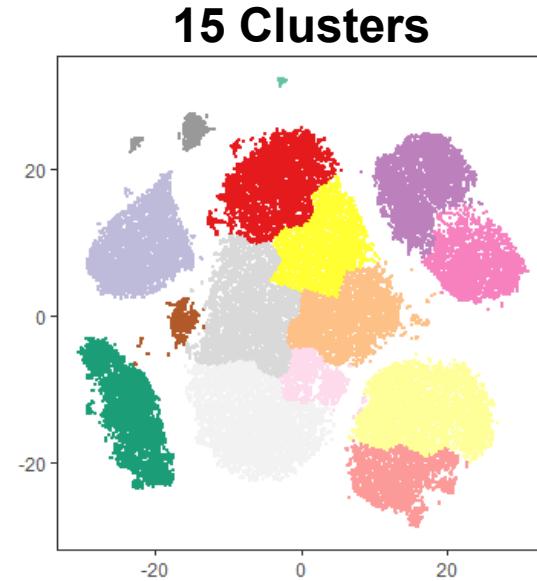
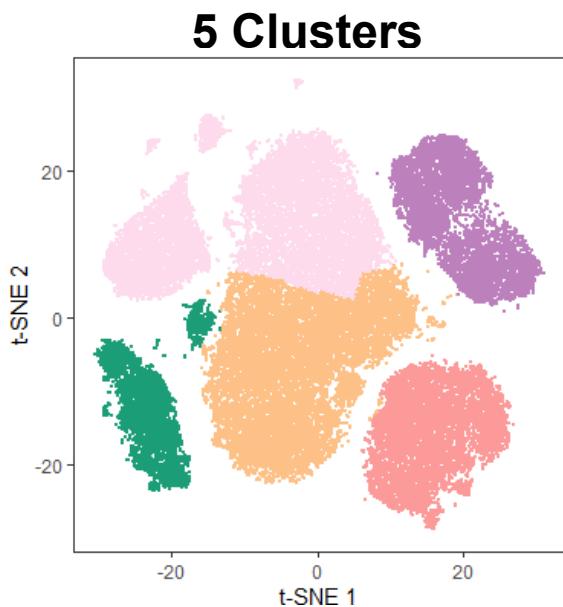


FlowSOM on Original Markers (n = 10)



# FlowSOM Requires that Users Choose a Number of Clusters

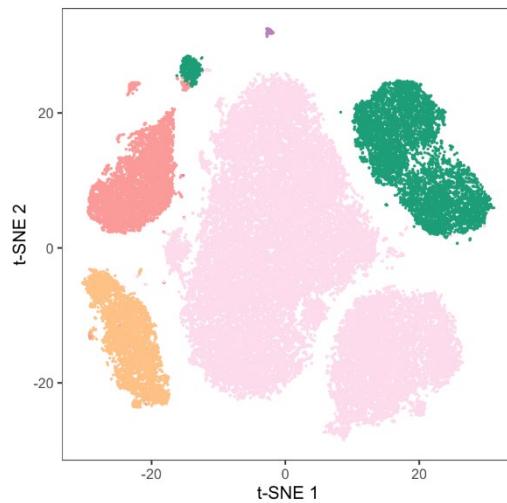
---



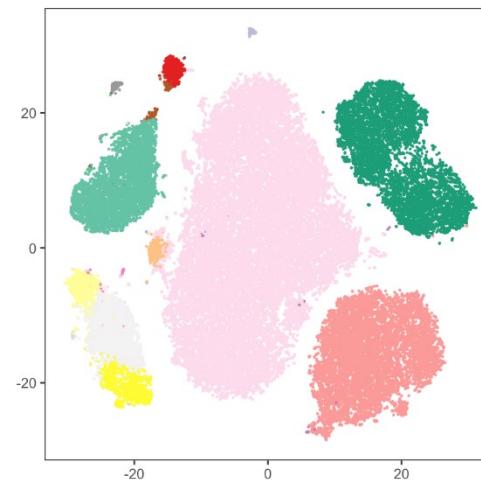
# FlowSOM Clusters are Dependent on Input Parameters

---

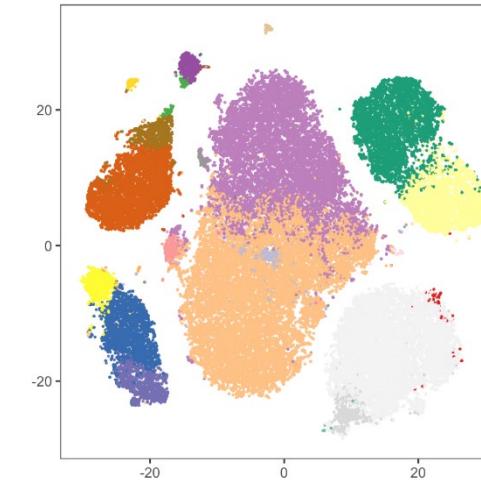
**5 Clusters**



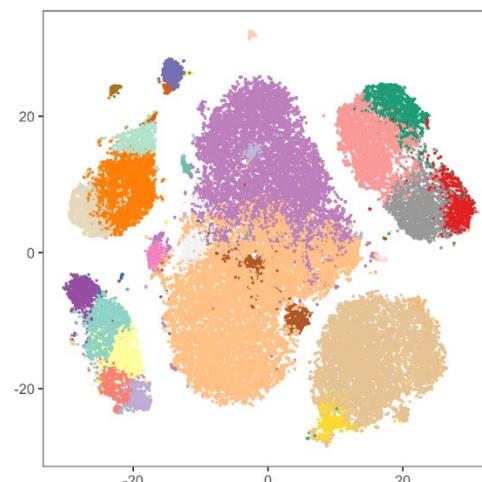
**15 Clusters**



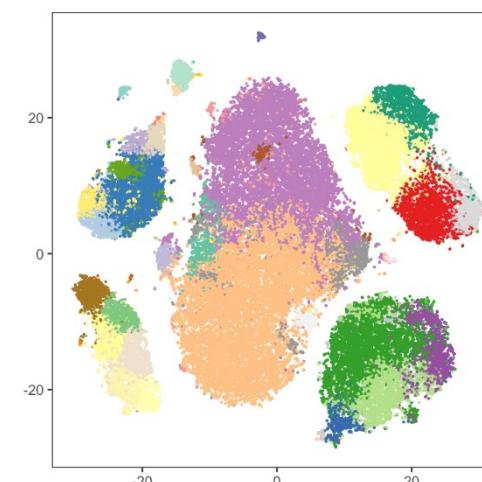
**25 Clusters**



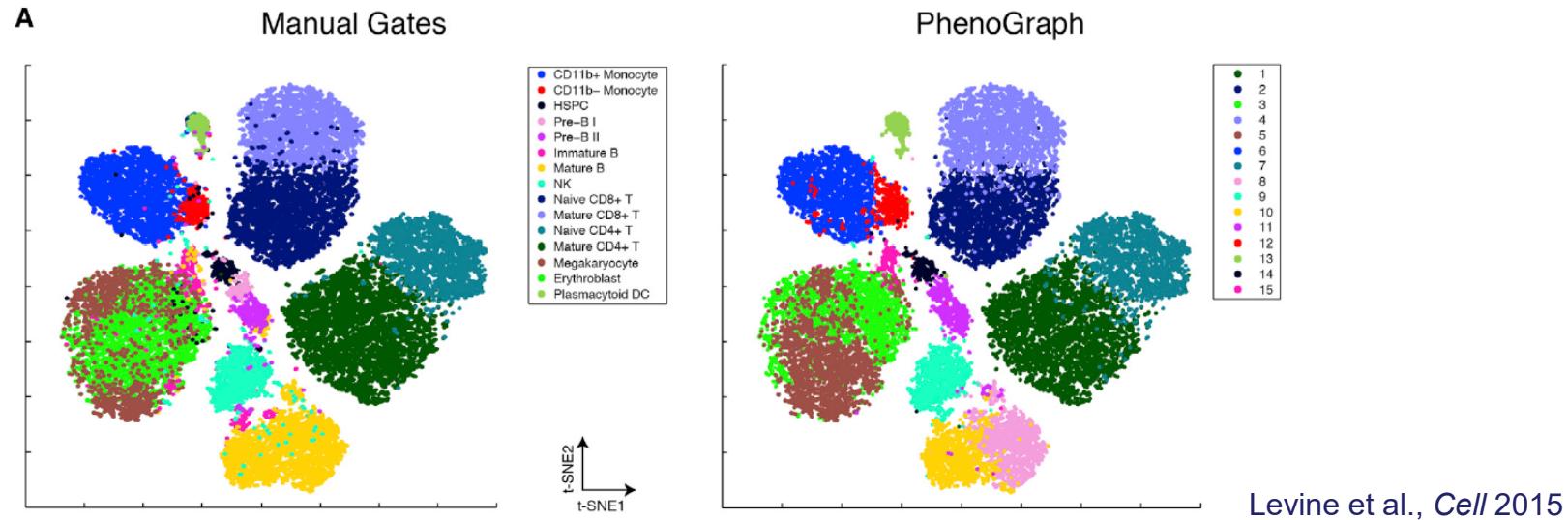
**35 Clusters**



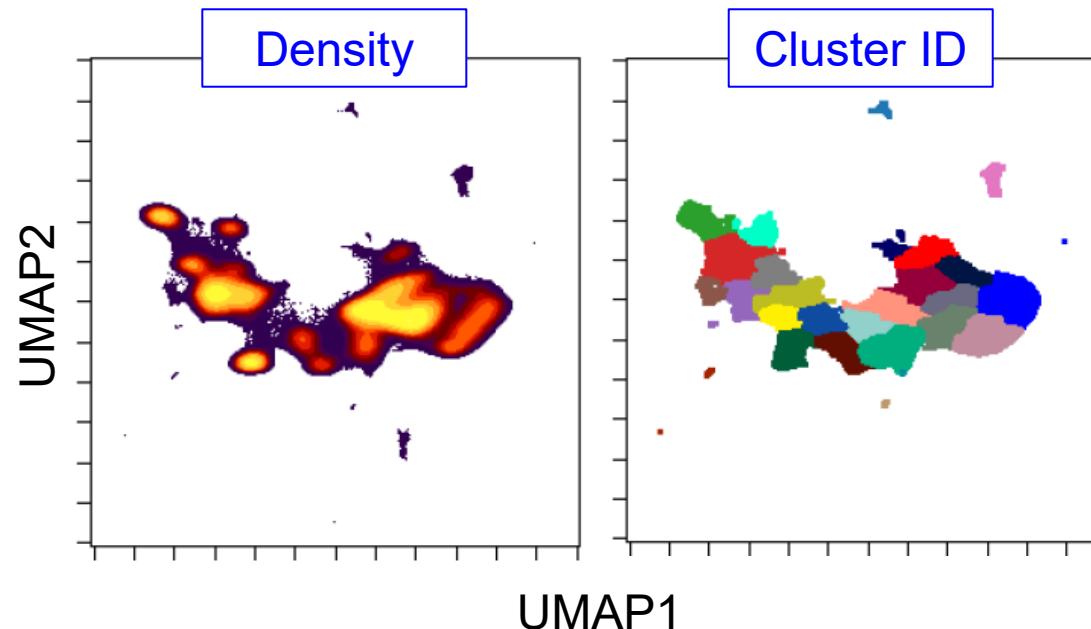
**45 Clusters**



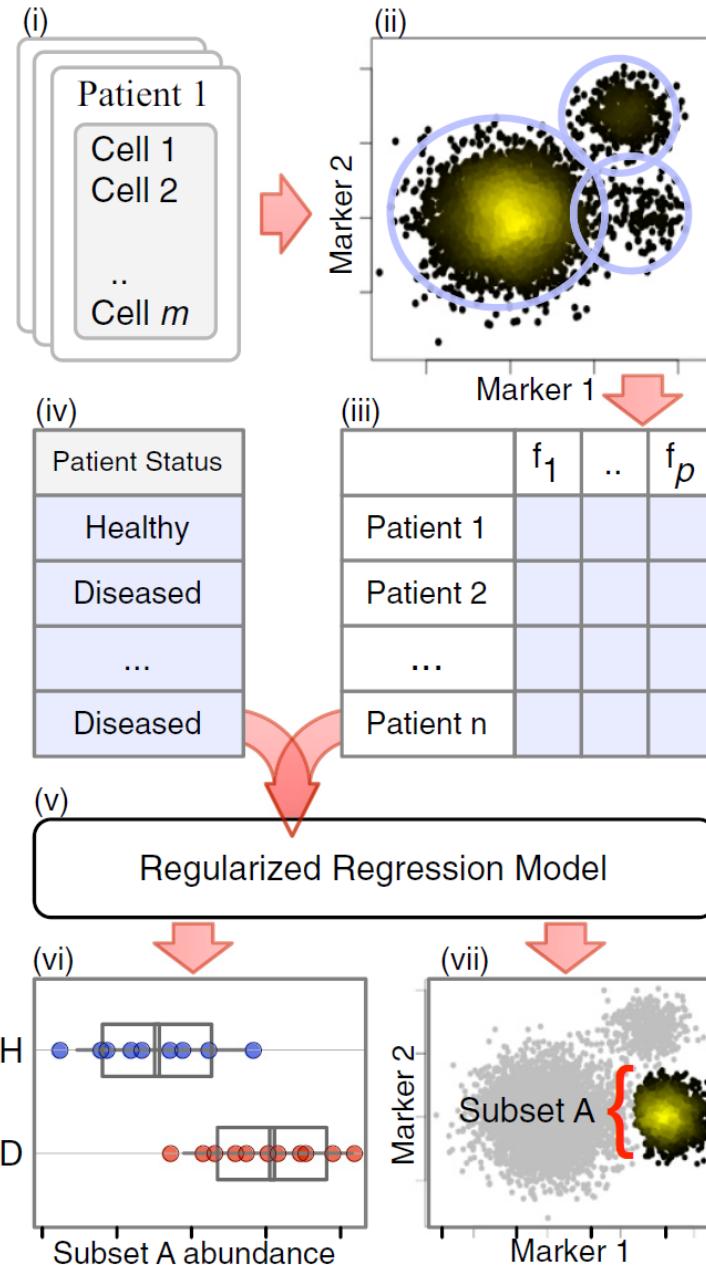
## Phenograph: Clustering 35 Features => t-SNE (Not the Reverse)



Diggins: t-SNE or UMAP on Features => Clustering on 2 axes



# Citrus: Supervised Population Finding

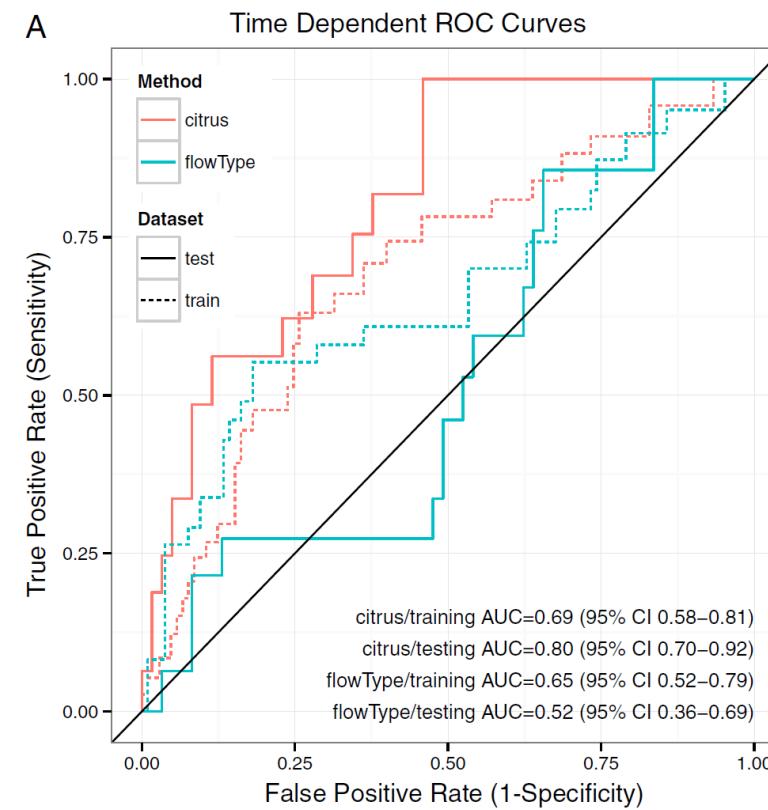


## Automated identification of stratifying signatures in cellular subpopulations

Robert V. Bruggner<sup>a,b</sup>, Bernd Bodenmiller<sup>c</sup>, David L. Dill<sup>d</sup>, Robert J. Tibshirani<sup>e,f,1</sup>, and Garry P. Nolan<sup>b,1</sup>

<sup>a</sup>Biomedical Informatics Training Program, Stanford University Medical School, Stanford, CA 94305; <sup>b</sup>Baxter Laboratory for Stem Cell Biology, Department of Microbiology and Immunology, and Departments of <sup>c</sup>Computer Science, <sup>e</sup>Health Research and Policy, and <sup>f</sup>Statistics, Stanford University, Stanford, CA 94305; and <sup>d</sup>Institute of Molecular Life Sciences, University of Zurich, CH-8057 Zurich, Switzerland

Contributed by Robert J. Tibshirani, May 14, 2014 (sent for review February 12, 2014)



# Citrus & RAPID Connect Cell Clusters to Clinical Outcomes, RAPID is Designed for Unsupervised Analysis of Survival

---

## Citrus

Bruggner, Tibshirani, et al., PNAS 2014

Finding  
cell clusters

Unsupervised  
(hierarchical clustering,  
cells may be in 2+ clusters)

Determining number  
of cell clusters to seek

Unsupervised  
(must be  $>5\%$  of sample)

Modeling  
cluster features

Supervised, multivariate  
(lasso regularized  
logistic regression,  
nearest shrunken centroid)

Splitting patients  
into groups

Supervised, happens at start  
(expert knows cut points,  
assigns patients to groups)

## RAPID

eLife 2020

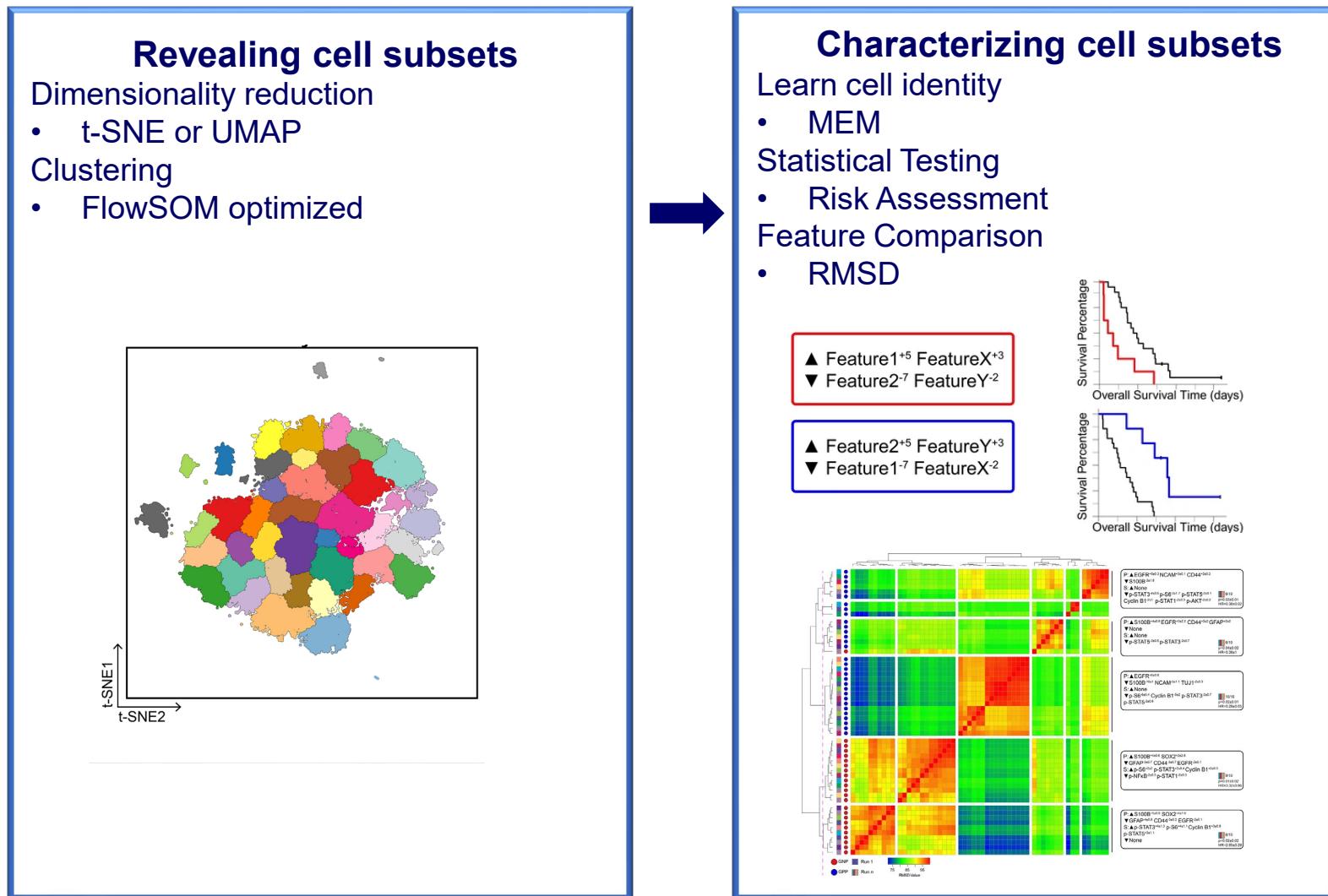
Unsupervised  
(various: FlowSOM, KNN,  
t-SNE + FlowSOM)

Unsupervised  
(seeks few clusters  
w/ low internal variation)

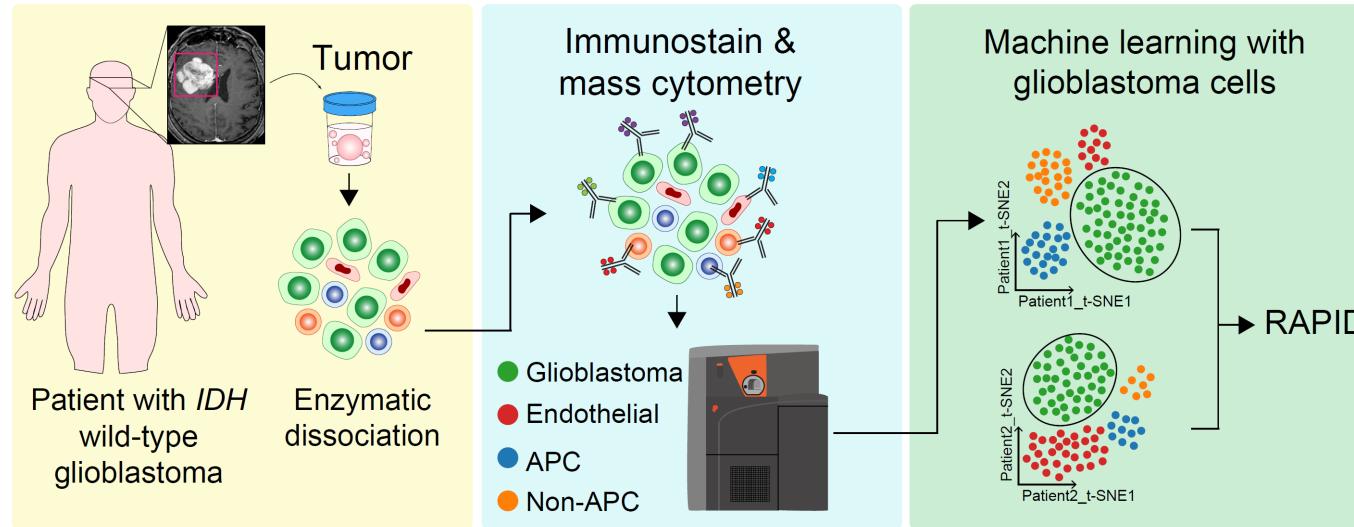
Unsupervised, univariate  
(median or MEM, simply a  
statistical description of cluster)

Unsupervised, happens at end  
(cluster abundance as cut point,  
Cox model of hazard)

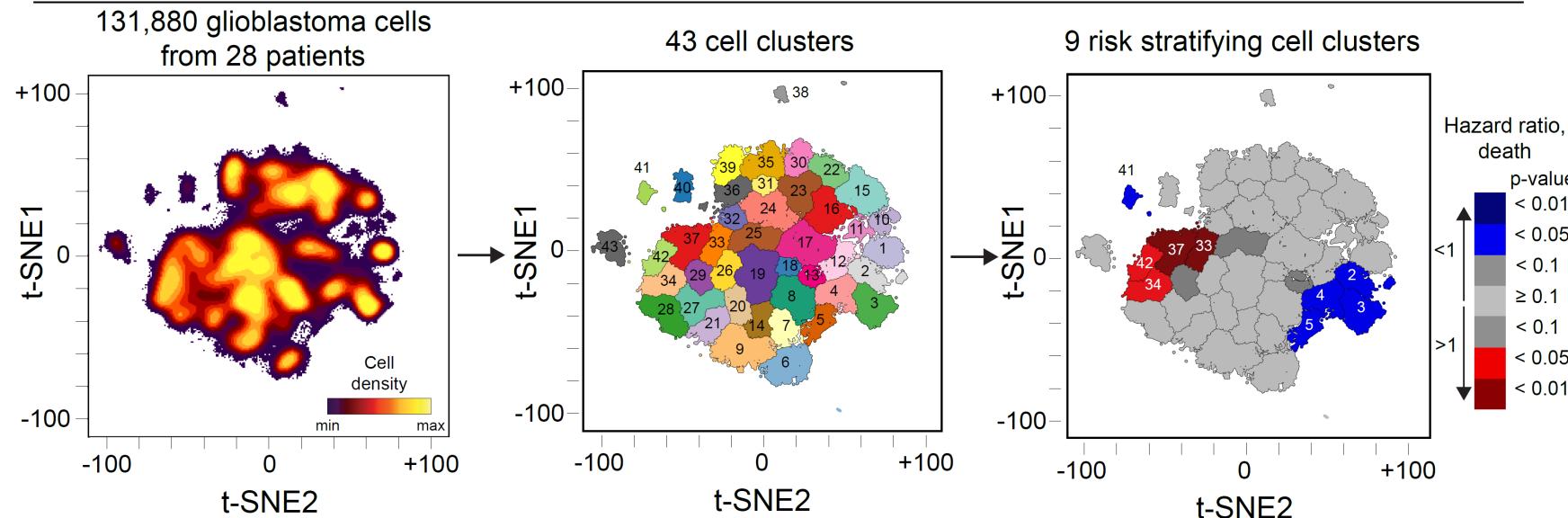
# Data Science Workflow using RAPID



# RAPID Maps Clinical Outcomes Onto Clusters (in t-SNE, UMAP, 2D image, original features, PCA, etc.)

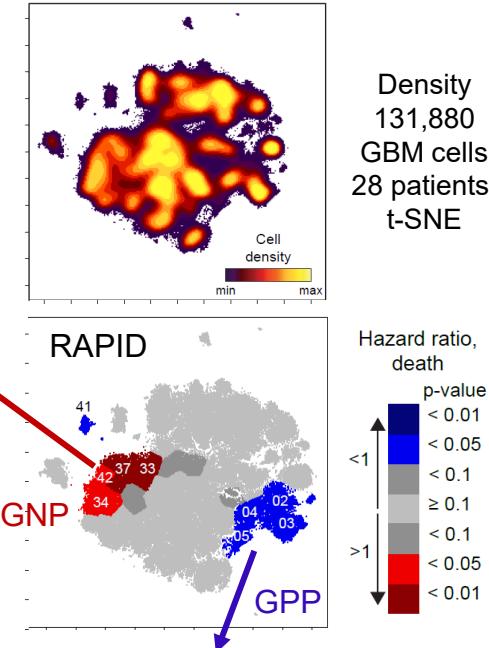
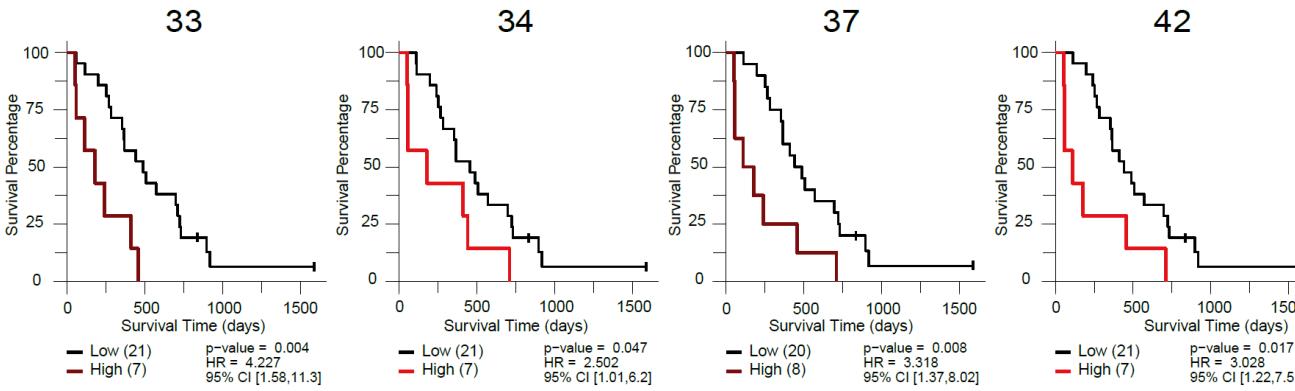


## Risk Assessment Population IDentification (RAPID) Maps Outcome onto t-SNE

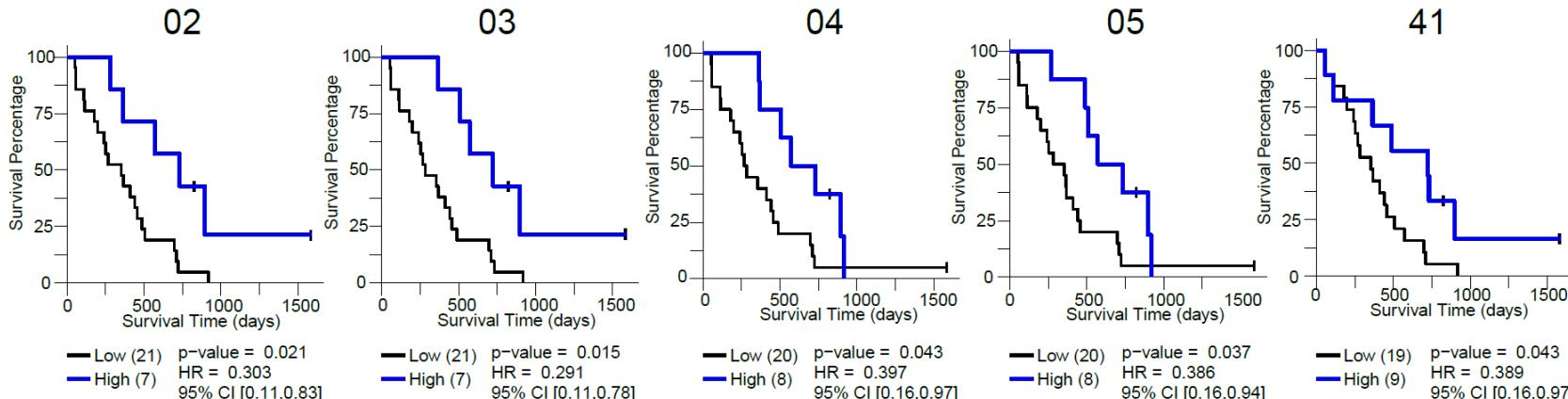


# RAPID Revealed Phenotypically Distinct Risk Stratifying Glioblastoma Cell Clusters

Poor survival of Glioma Negative Prognostic (GNP) high patients



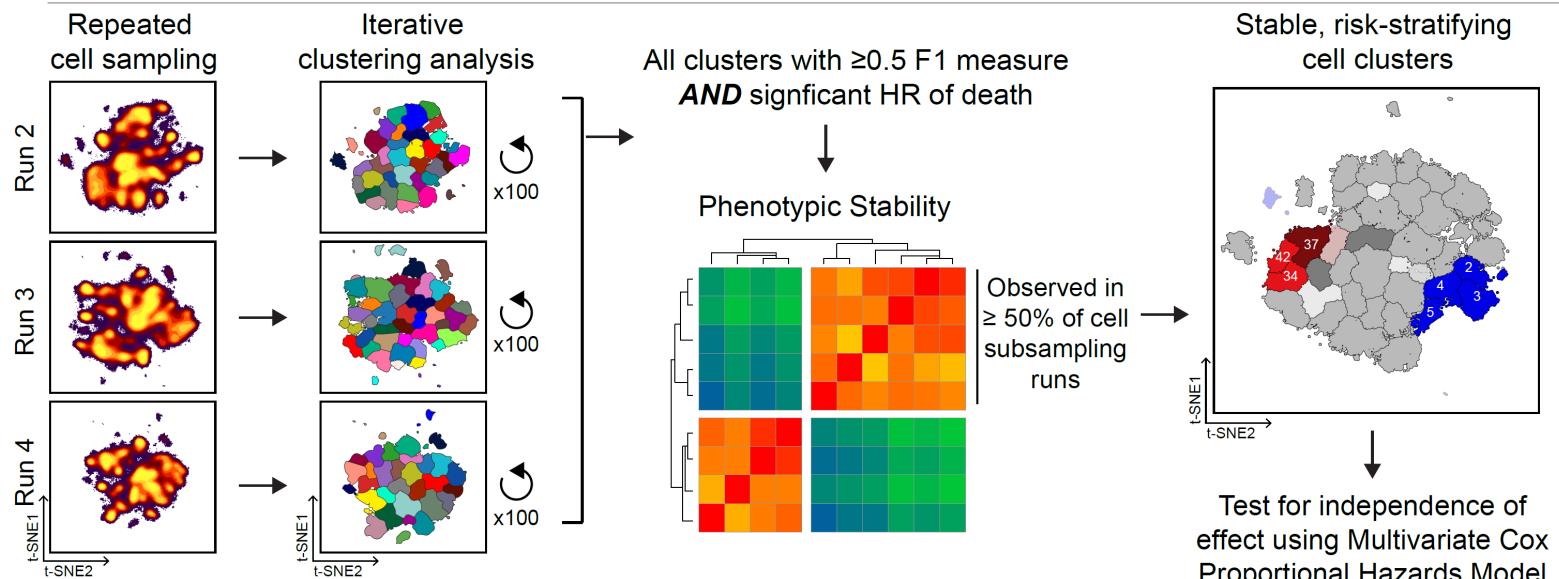
Better survival of Glioma Positive Prognostic (GPP) high patients



# Statistical & Biological Validation Are Essential Parts of Algorithm & Study Design

C

## Cluster and Phenotypic Stability Testing



d

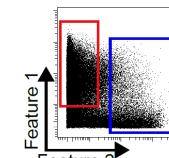
## Biological Validation

### Feature selection for validation

- ▲ Feature1<sup>+5</sup> FeatureX<sup>+3</sup>
- ▼ Feature2<sup>-7</sup> FeatureY<sup>-2</sup>

- ▲ Feature2<sup>+5</sup> FeatureY<sup>+3</sup>
- ▼ Feature1<sup>-7</sup> FeatureX<sup>-2</sup>

### Low dimensional gating strategy

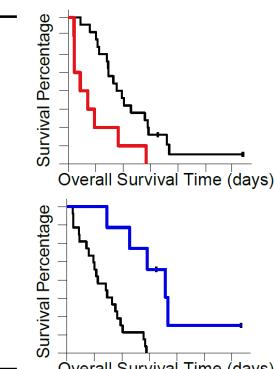


OR

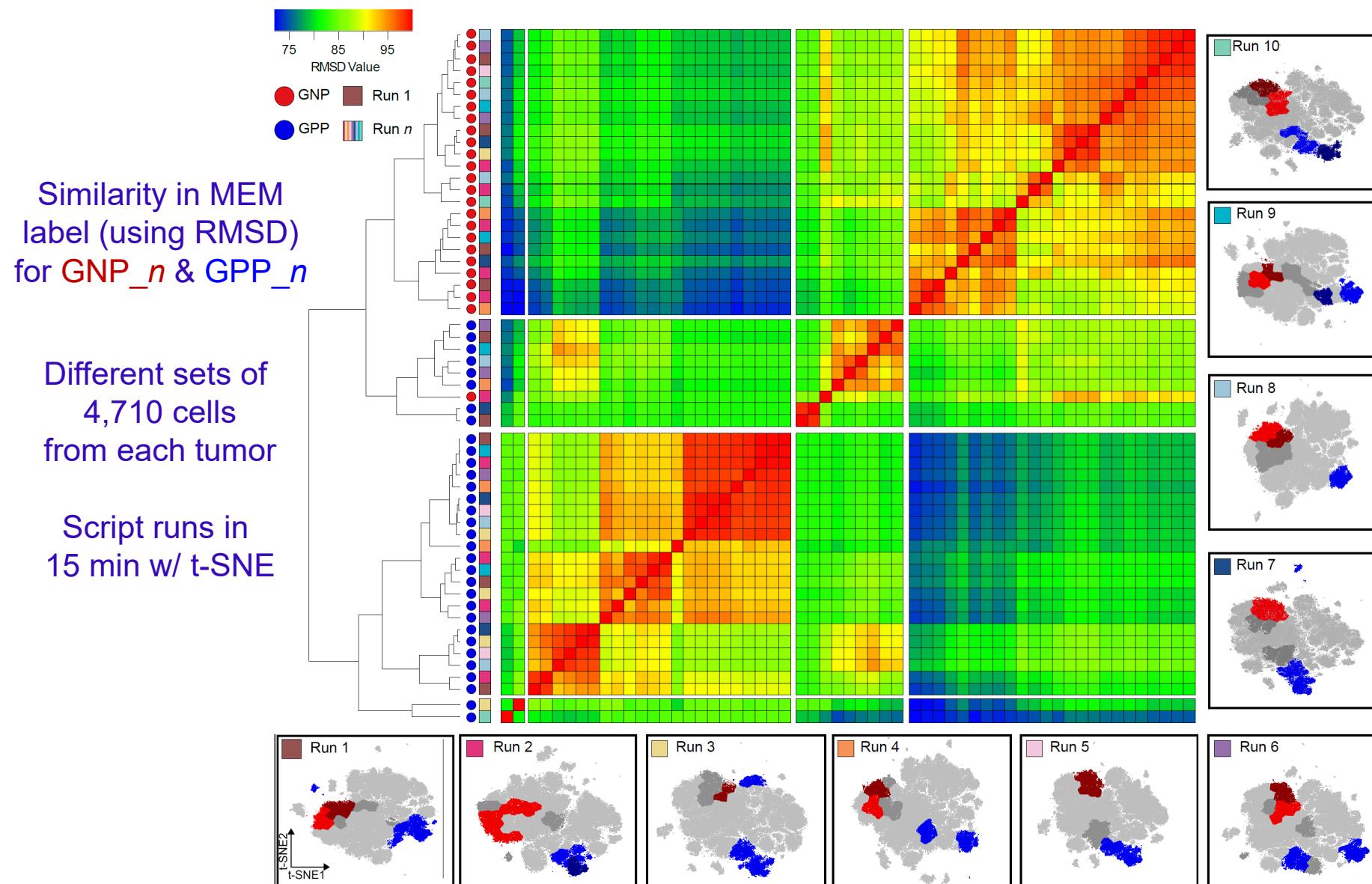
### Immunohistochemistry staining



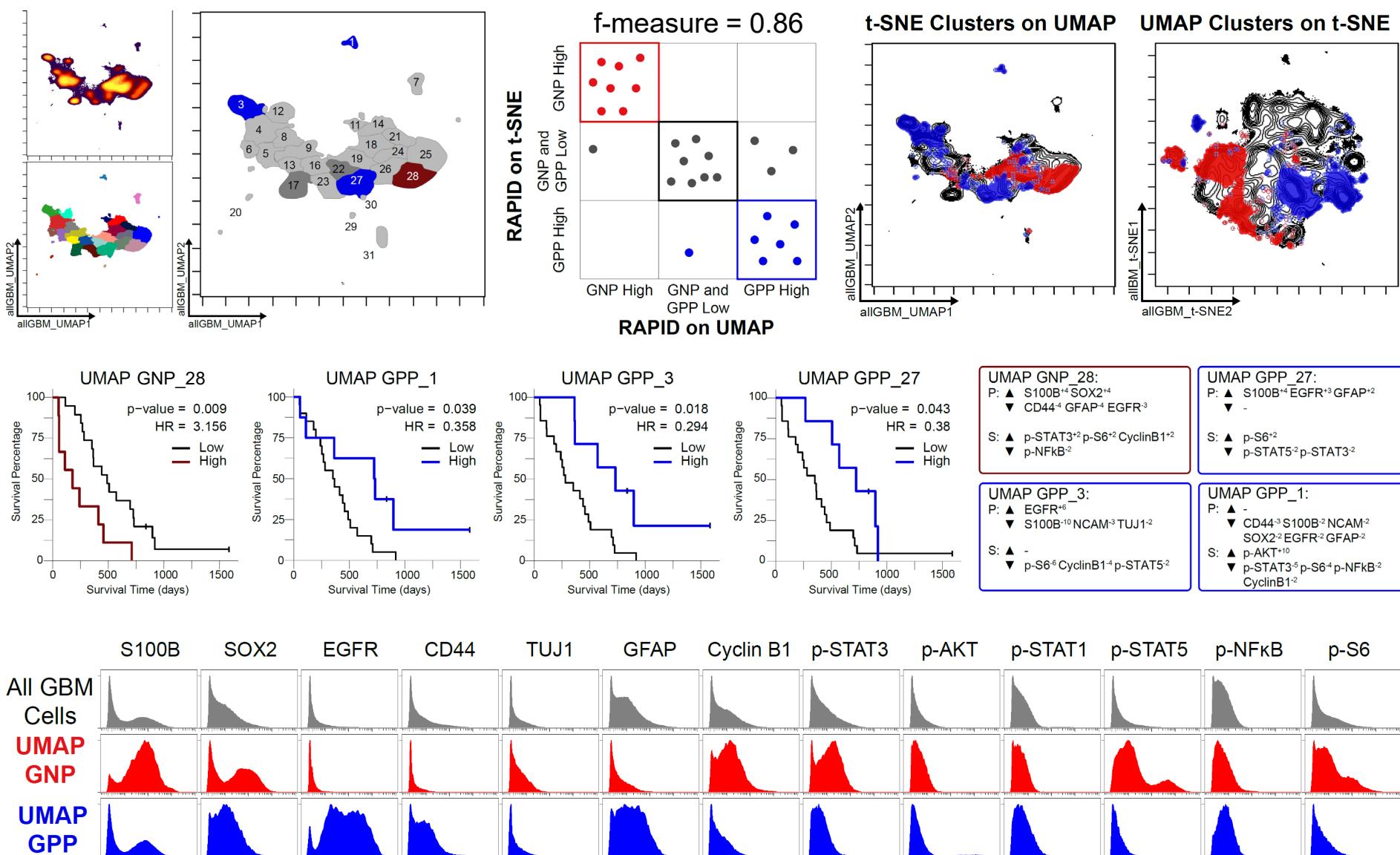
### Survival analysis



# Re-Running RAPID +9X with Different Cells from the Same Tumors Gave Similar GNP & GPP Phenotypes and Risk Stratification

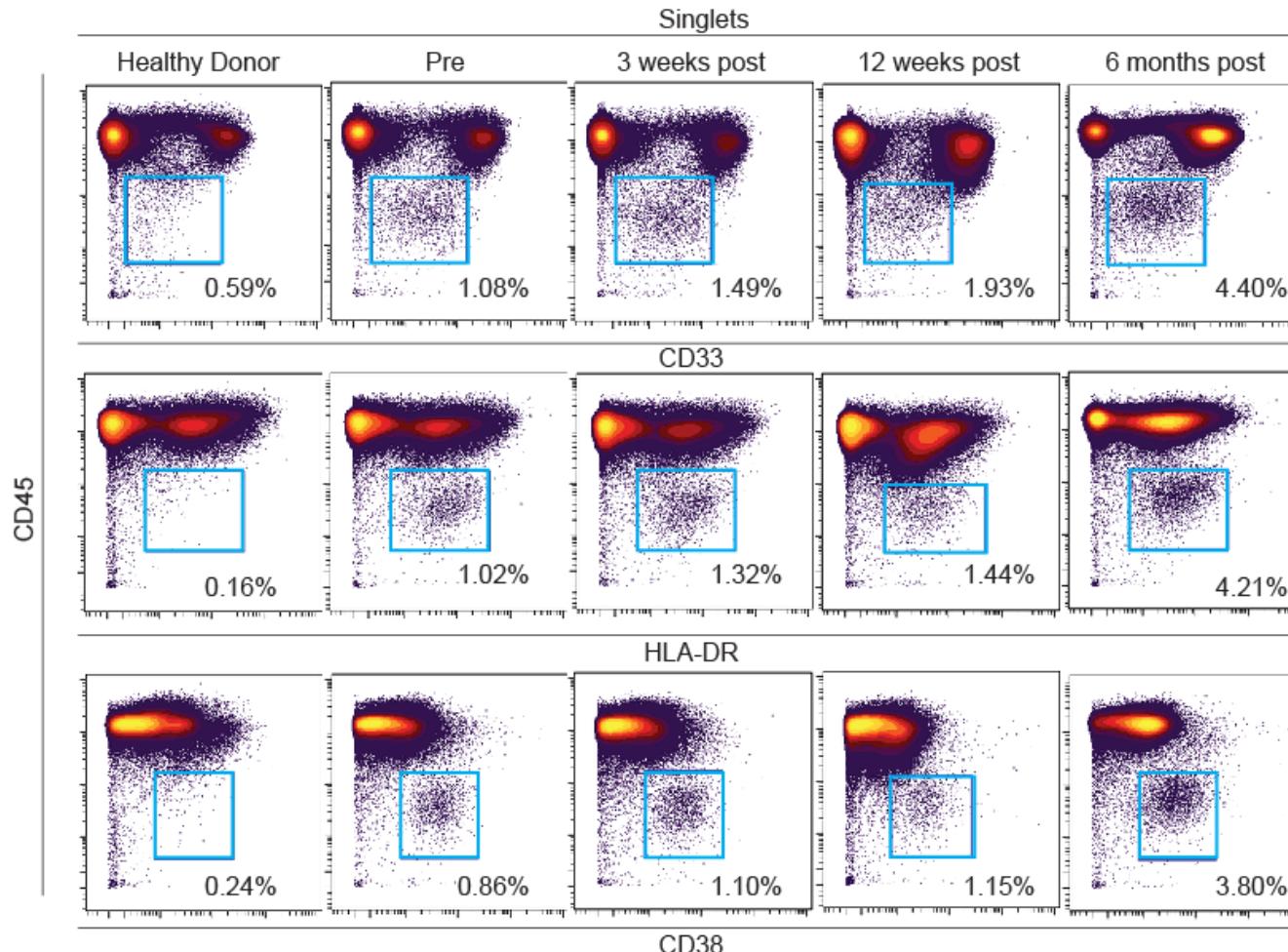


# Re-Running RAPID with UMAP Instead of t-SNE Gave Similar GNP & GPP Phenotypes and Risk Stratification



# A Case Study: Systems Immune Monitoring with Mass Cytometry Reveals A Clinically Significant Rare Cell Subset

## MDS in Melanoma Patient Revealed During $\alpha$ -PD-1 Therapy



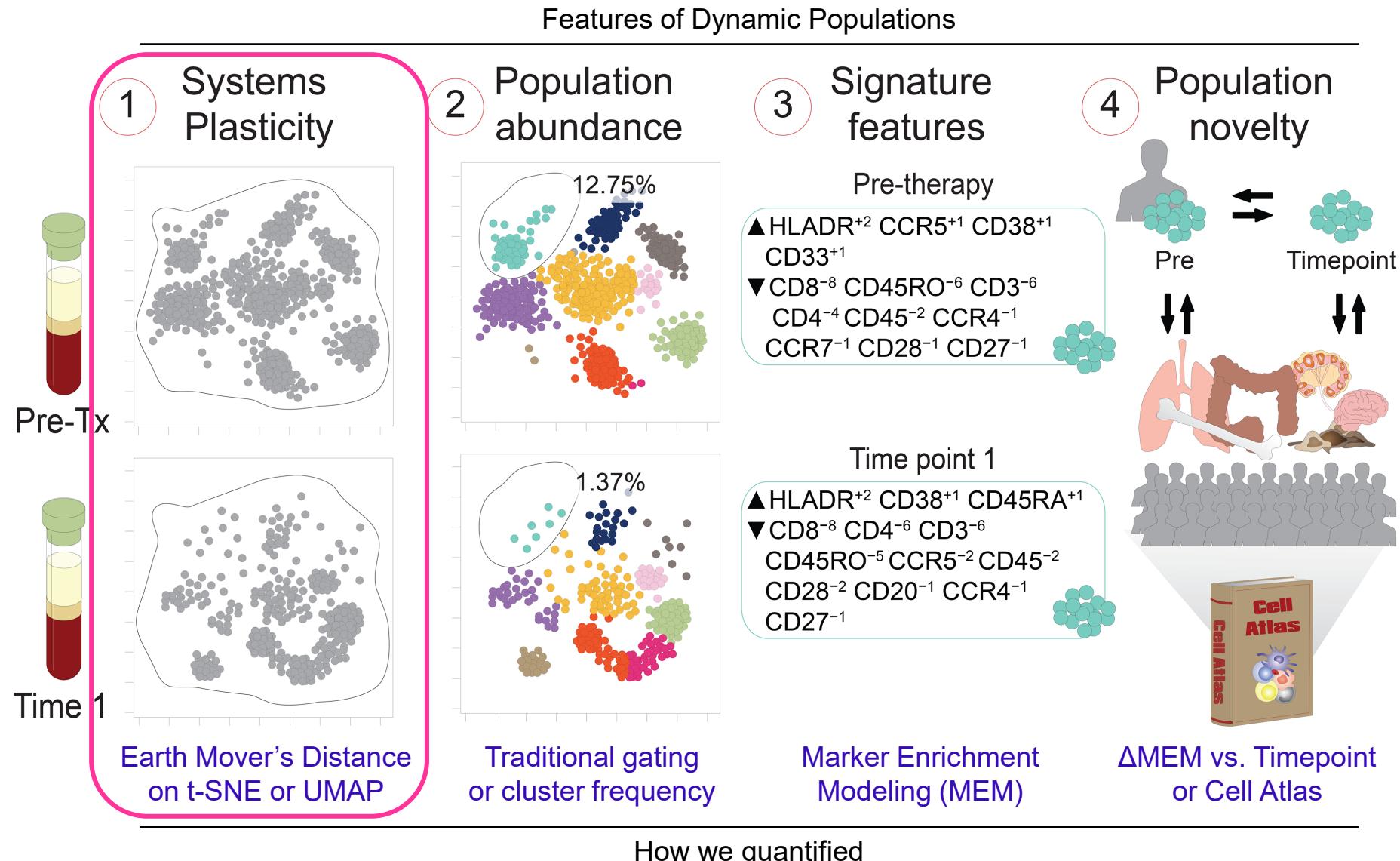
Mass cytometry data (CyTOF)

Healthy donor looks similar to melanoma in 2D views

At Pre-Tx, MDS blasts were not detected by standard CBC

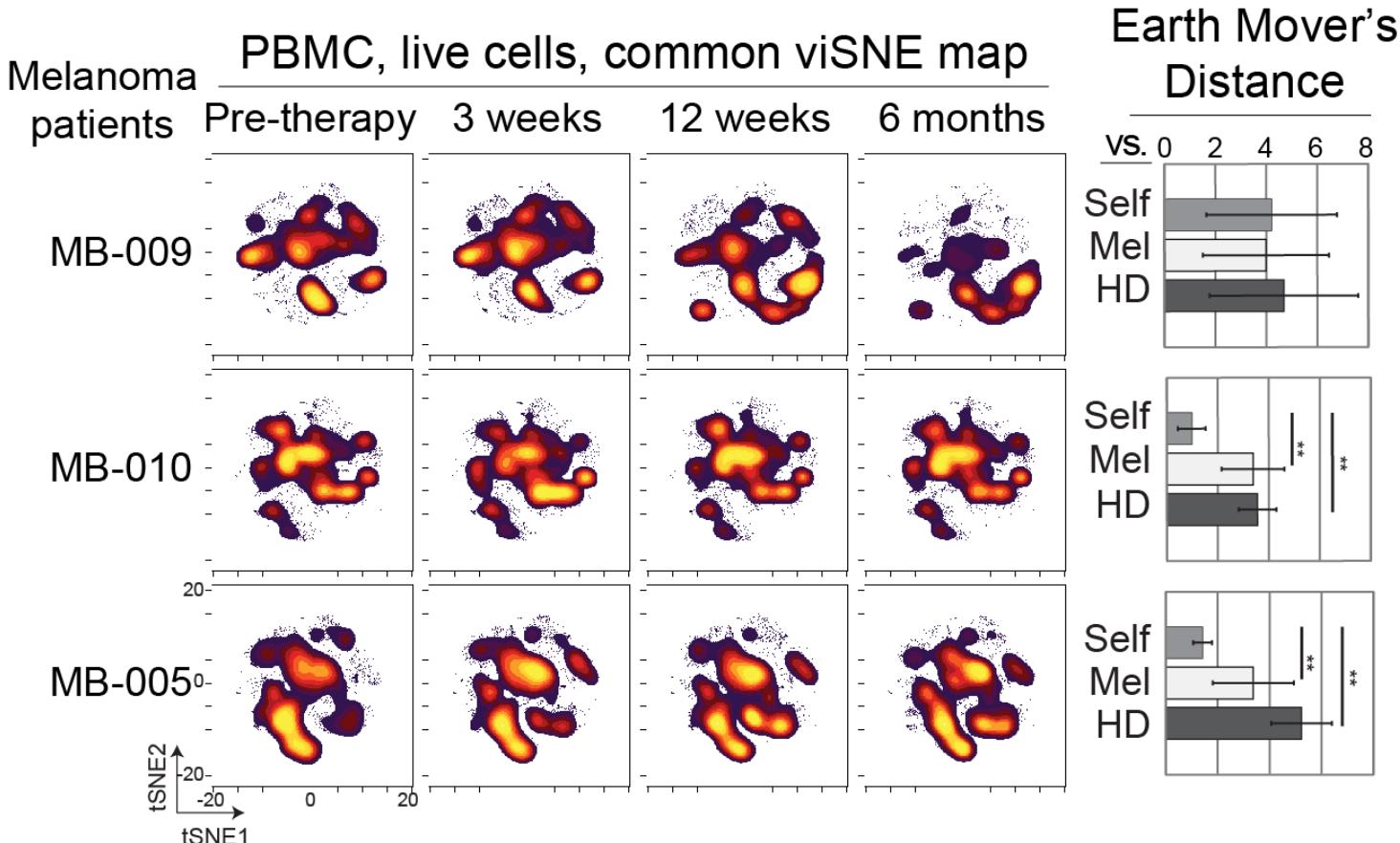
High dimensional panel allowed review of PD-1 on MDS blasts w/ existing data

# Clinical Trial Monitoring: What Do We Need to Know? Automate Four Key Readouts vs. Clinical Outcomes



# Plasticity / Stability: Earth Mover's Distance Quantifies Change Over Time Within a t-SNE Analysis

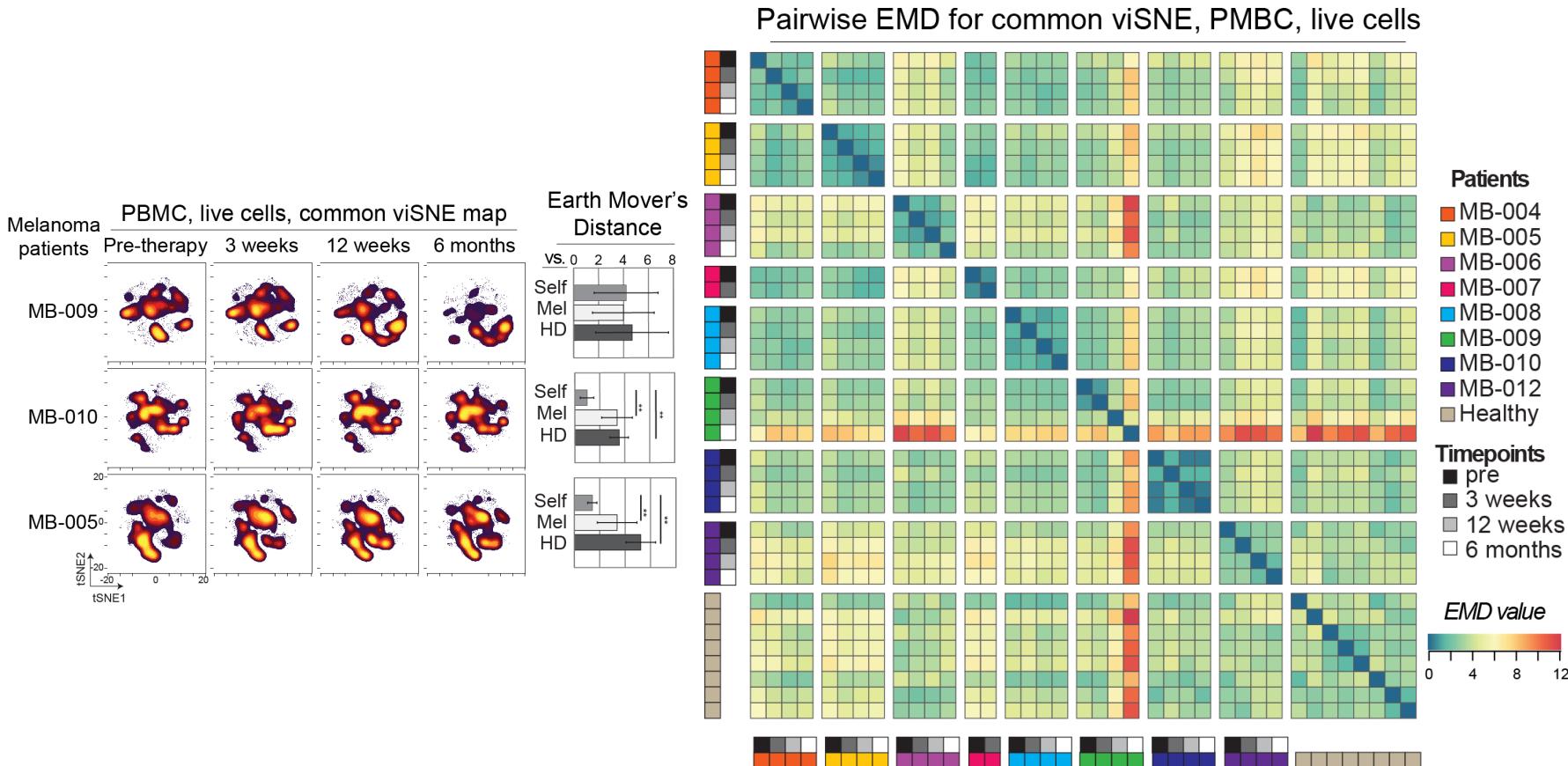
Melanoma Patients Treated with  $\alpha$ -PD-1 Therapy, Monitored by Mass Cytometry



Systems immune monitoring reveals an unexpected pattern in MB-009  
Individuals can be their own significantly stable baseline

# Plasticity / Stability: Earth Mover's Distance Quantifies Change Over Time Within a t-SNE Analysis

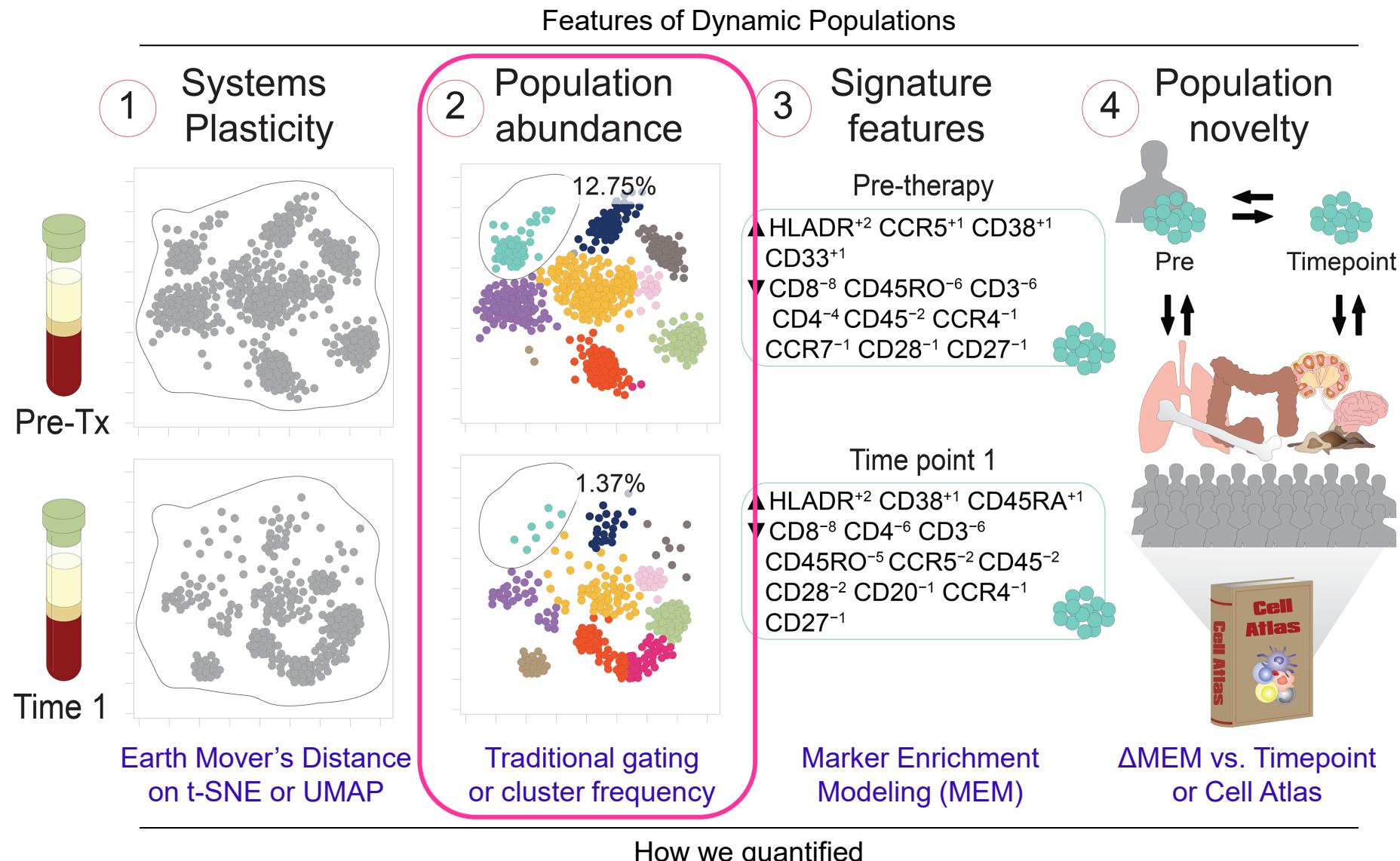
Melanoma Patients Treated with  $\alpha$ -PD-1 Therapy, Monitored by Mass Cytometry



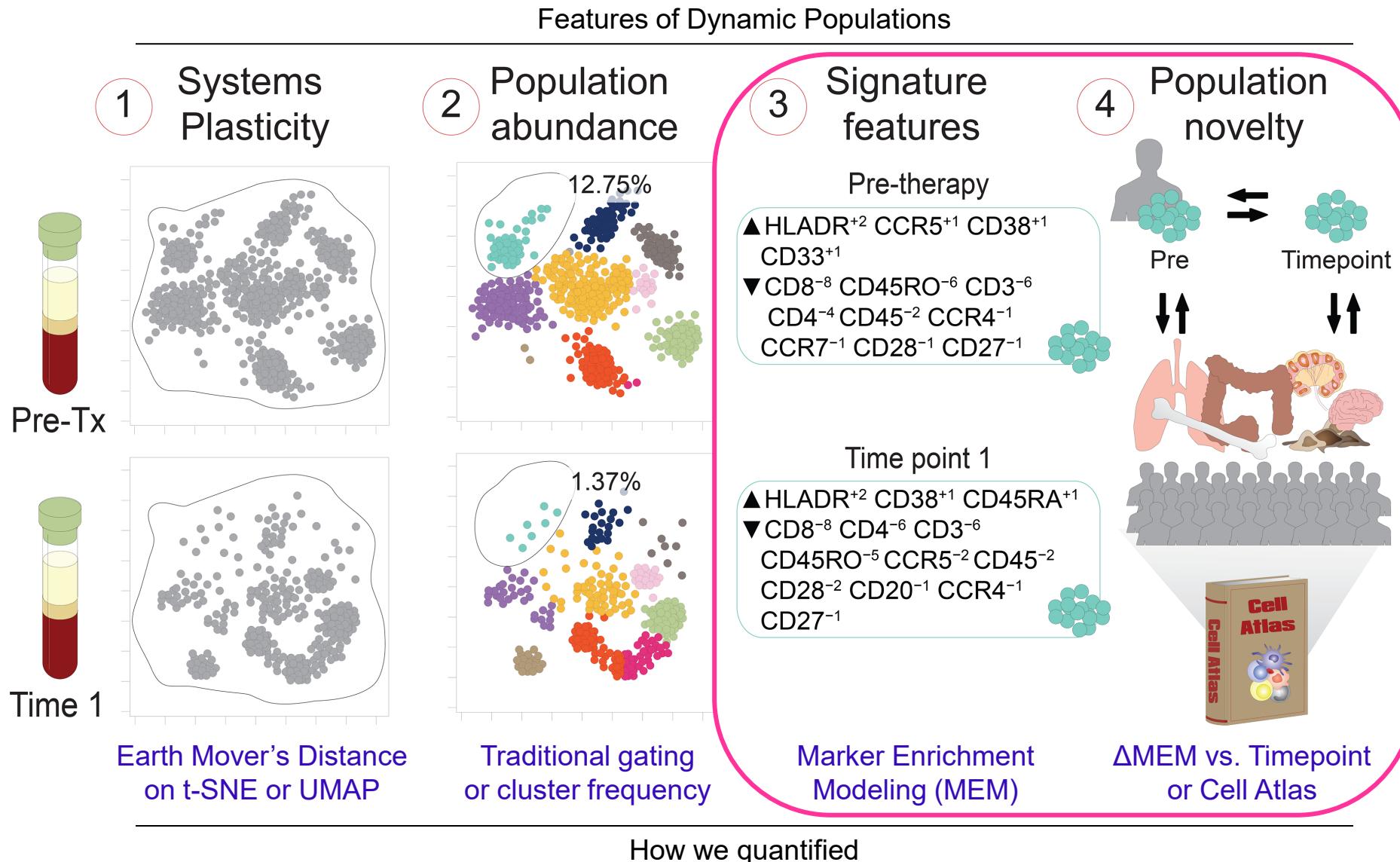
Systems immune monitoring reveals an unexpected pattern in MB-009

# Clinical Trial Monitoring: What Do We Need to Know?

## Automate Four Key Readouts vs. Clinical Outcomes

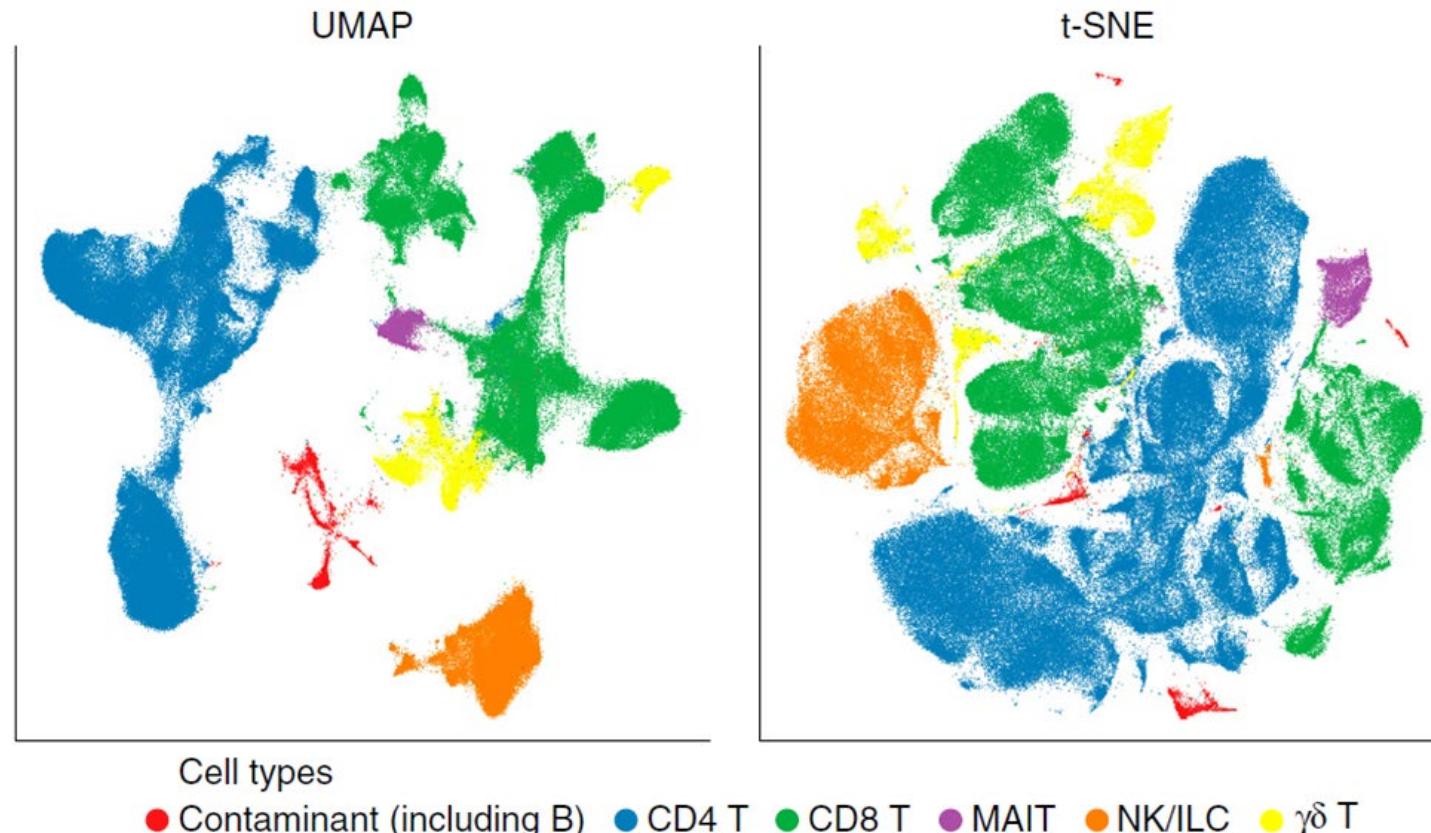


# Clinical Trial Monitoring: What Do We Need to Know? Automate Four Key Readouts vs. Clinical Outcomes



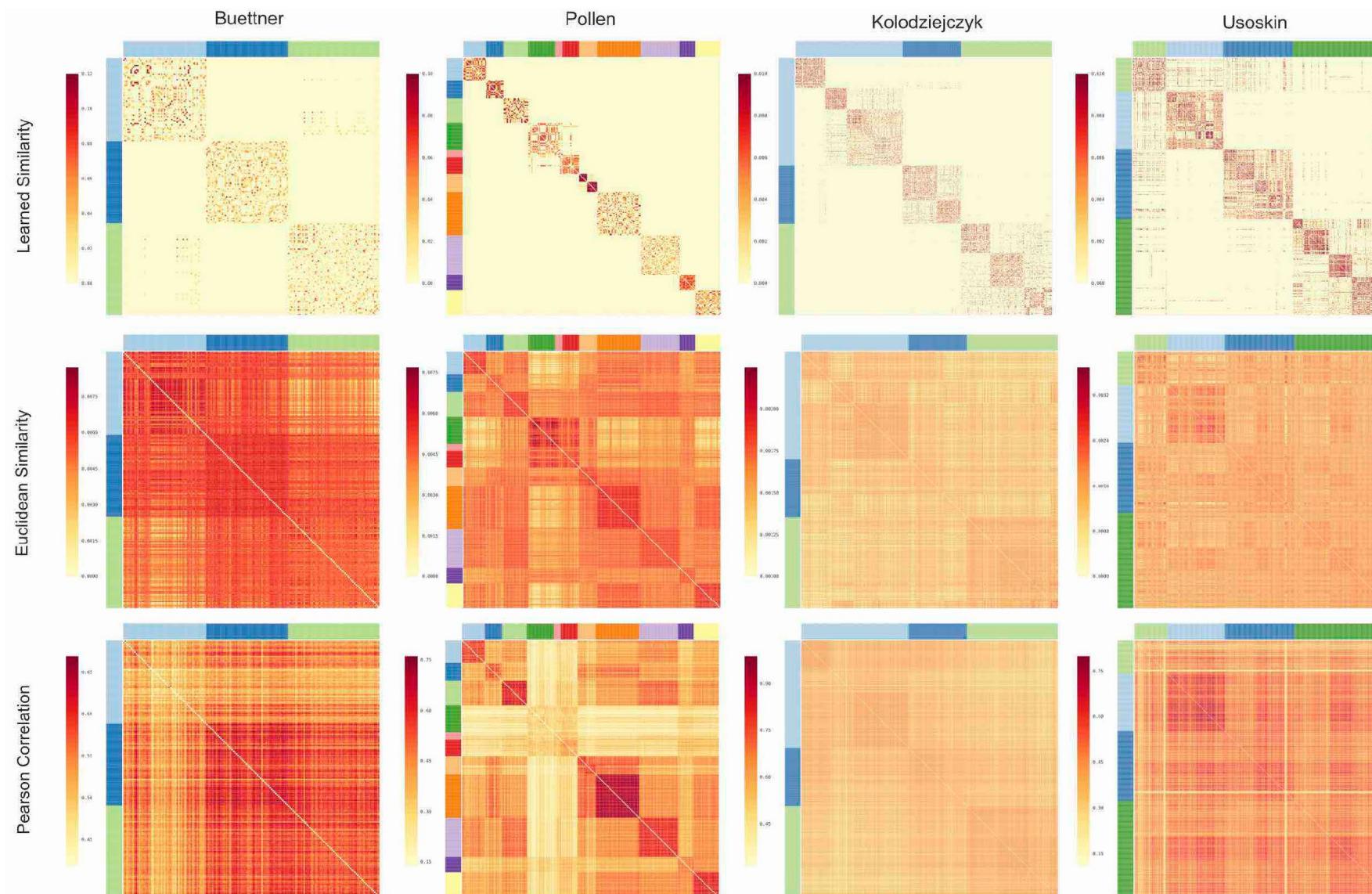
# Becht et al., UMAP Preserves Local and Global Structure (Analysis of Tissue T Cells; Color = Expert Knowledge / Source)

(a) UMAP better split CD8 T cells,  $\gamma\delta$  T cells, and contaminating cells



Dataset covering 35 samples originating from 8 distinct human tissues enriched for T and natural killer (NK) cells, of more than >300,000 cell events with 39 protein targets (Wong et al. dataset).

# Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning (SIMLR)



# Resources

---

## Normalization

<https://onlinelibrary.wiley.com/doi/full/10.1002/cyto.a.22271>

## Gaussian Gating

<http://cytoforum.stanford.edu/download/file.php?id=242&sid=37e5ec0a3dedb53865bbbcb6a023c316>

## t-SNE

<https://www.nature.com/articles/nbt.2594>

## Opt-SNE

<https://www.biorxiv.org/content/10.1101/451690v3.full>

## UMAP

<https://www.nature.com/articles/nbt.4314>

## FlowSOM

<https://www.ncbi.nlm.nih.gov/pubmed/25573116>

## SPADE

<https://www.nature.com/articles/nbt.1991>

## Phenograph

<https://www.sciencedirect.com/science/article/pii/S0092867415006376>

## MEM

<https://www.nature.com/articles/nmeth.4149>

## RAPID

<https://elifesciences.org/articles/56879>

## T-REX

<https://elifesciences.org/articles/64653>

## “A Beginner’s Guide to Analyzing and Visualizing Mass Cytometry Data”

<https://www.jimmunol.org/content/200/1/3>

## Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data

<https://www.ncbi.nlm.nih.gov/pubmed/27992111>

# Contact Info

---

## **Jonathan Irish**

Principal Investigator, Associate Professor

Cell & Developmental Biology Pathology, Microbiology & Immunology

Email: [jonathan.irish@vanderbilt.edu](mailto:jonathan.irish@vanderbilt.edu)

## **Cass Mayeda**

Web Apps Research Assistant

Email: [cass.mayeda@vanderbilt.edu](mailto:cass.mayeda@vanderbilt.edu)