

Part 2: 45 min

Systems Immune Monitoring & Tailoring Workflows

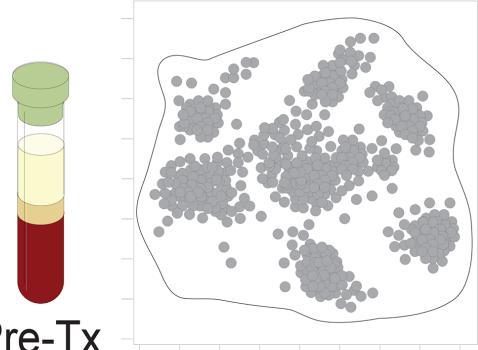
Samples Over Time Reveal Immune System Dynamics

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

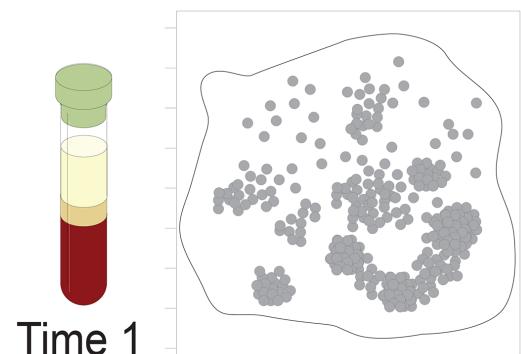
Clinical Trial Monitoring: What Do We Need to Know? Automate Four Key Readouts vs. Clinical Outcomes

Features of Dynamic Populations

1 Systems Plasticity



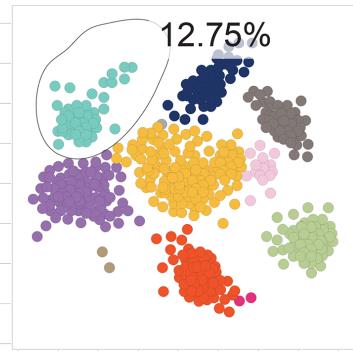
Pre-Tx



Time 1

Earth Mover's Distance
on t-SNE or UMAP

2 Population abundance



Traditional gating
or cluster frequency

3 Signature features

Pre-therapy

- ▲ HLA^{DR}⁺² CCR5⁺¹ CD38⁺¹
CD33⁺¹
- ▼ CD8⁻⁸ CD45RO⁻⁶ CD3⁻⁶
CD4⁻⁴ CD45⁻² CCR4⁻¹
CCR7⁻¹ CD28⁻¹ CD27⁻¹

Time point 1

- ▲ HLA^{DR}⁺² CD38⁺¹ CD45RA⁺¹
- ▼ CD8⁻⁸ CD4⁻⁶ CD3⁻⁶
CD45RO⁻⁵ CCR5⁻² CD45⁻²
CD28⁻² CD20⁻¹ CCR4⁻¹
CD27⁻¹

Marker Enrichment
Modeling (MEM)

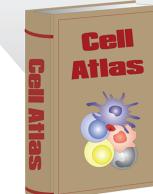
4 Population novelty



Pre



Timepoint

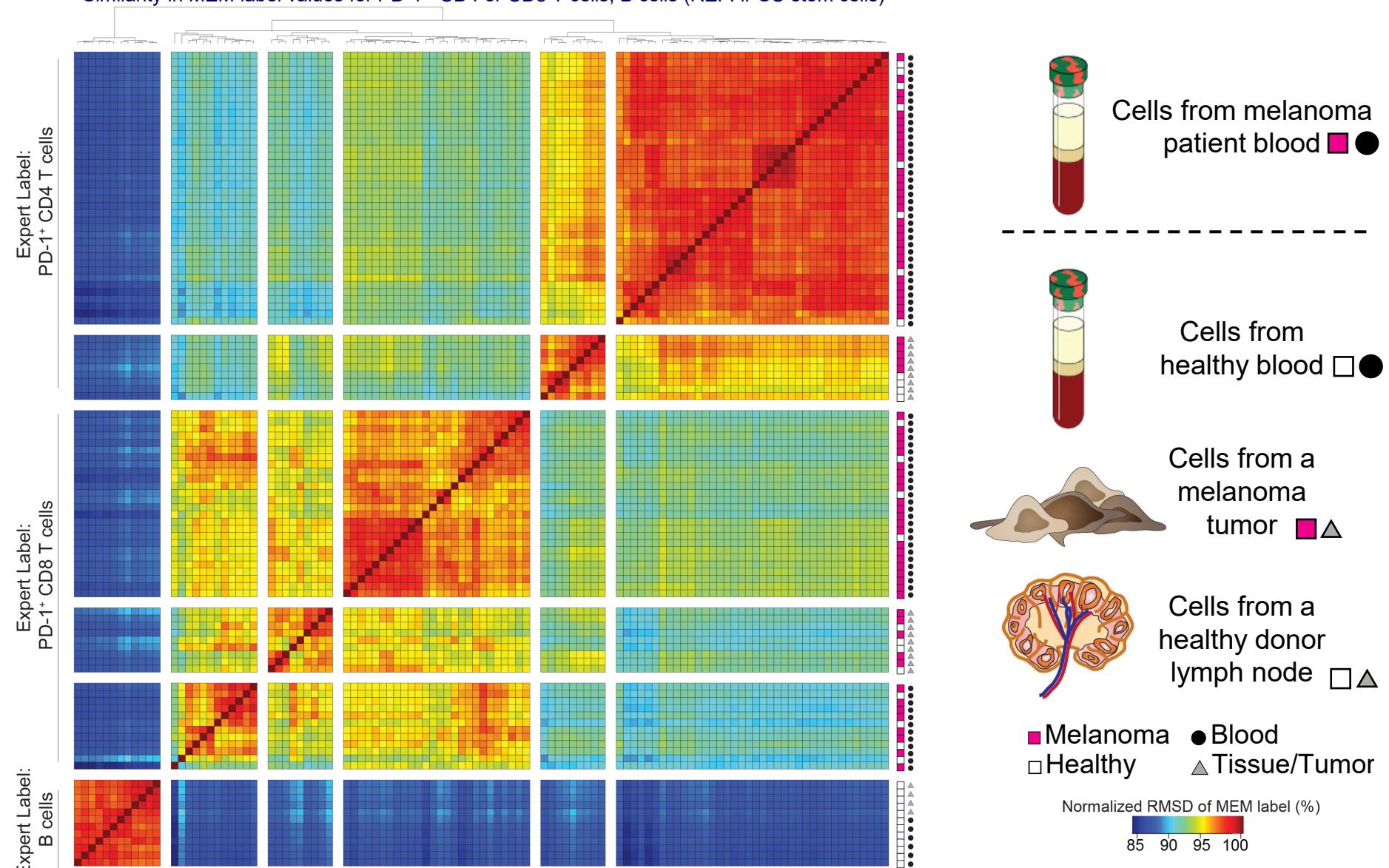


ΔMEM vs. Timepoint
or Cell Atlas

How we quantified

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

Similarity in MEM label values for PD-1⁺ CD4 or CD8 T cells, B cells (REF: iPSC stem cells)



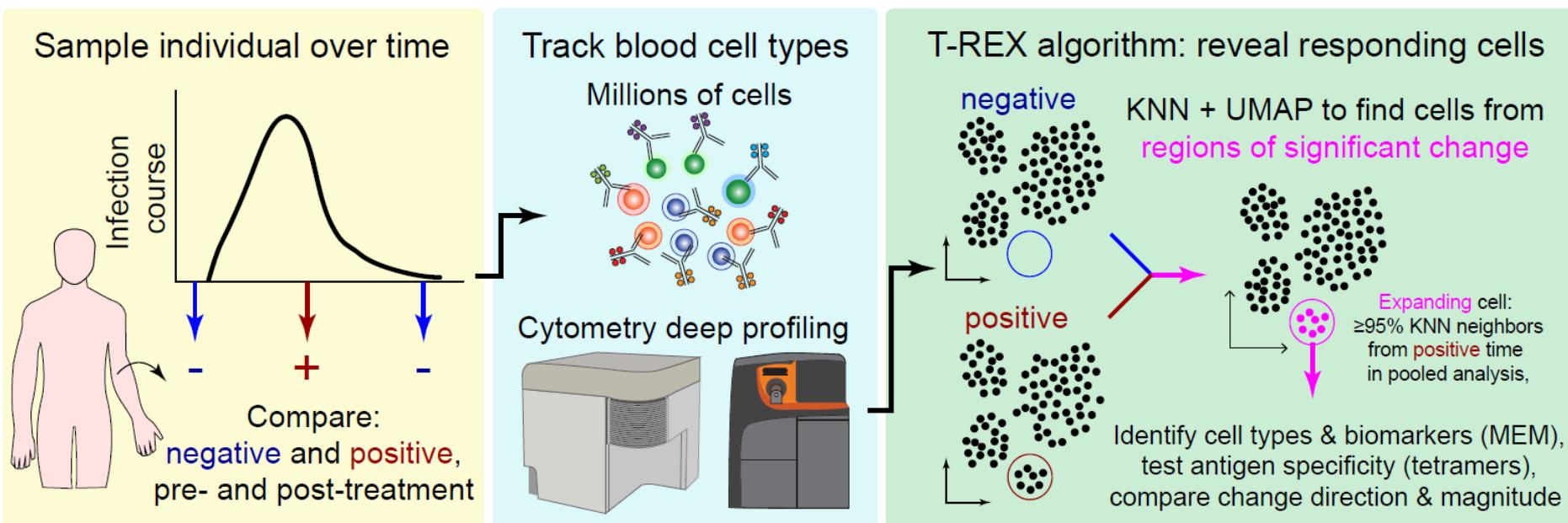
Greenplate et al., *Cancer Immunology Research* 2019

Methods: Diggins et al., *Nature Methods* 2017; *Curr Prot Cyt* 2018

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

Comparison of Tools

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



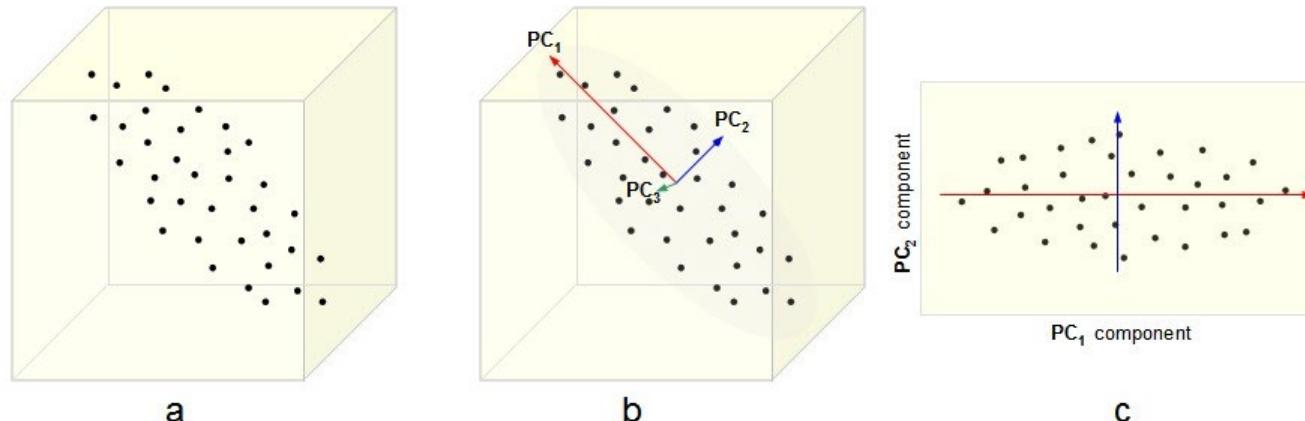
T-REX (Tracking Responders EXpanding) identifies phenotypic hotspots undergoing great change between conditions (e.g., +/- infection)

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

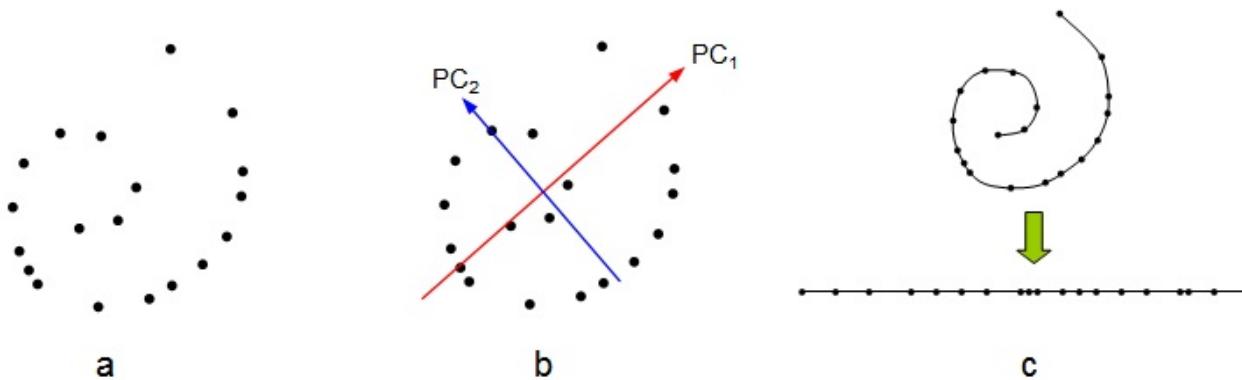
Manuscript: <https://elifesciences.org/articles/64653>



PCA is a Linear Dimensionality Reduction Tool



An illustration of PCA. **a)** A data set given as 3-dimensional points. **b)** The three orthogonal Principal Components (PCs) for the data, ordered by variance. **c)** The projection of the data set into the first two PCs, discarding the third one.



Effects of dimensionality reduction on an inherently non-linear data set. **a)** The original data given as a two-dimensional set. **b)** PCA identifies two PCs as contributing significantly to explain the data variance. **c)** However, the inherent topology (connectivity) of the data helps identify the set as being one-dimensional, but non-linear.

Is t-SNE Like PCA?

| | PCA | t-SNE |
|--------------------------------|--|---|
| Finding cell clusters | Principal Component Analysis N/A (algorithm does not assign items to clusters) | t-distributed Stochastic Neighbor Embedding N/A (algorithm does not assign items to clusters) |
| Dimensionality reduction | Linear, user determined (user chooses eigenvectors or variance amount to keep) | Non-linear, user determined (user chooses number of dimensions, usually 2) |
| Modeling dataset features | Unsupervised, linear, deterministic statistical model | N/A (algorithm does not model data) |
| Splitting patients into groups | N/A (algorithm does not analyze groups, outcomes) | N/A (algorithm does not analyze groups, outcomes) |

Is UMAP Like t-SNE?

UMAP

Uniform Manifold Approximation

Finding
cell clusters

N/A

(algorithm does not assign
items to clusters)

Dimensionality
reduction

Non-linear, user determined
(user chooses number
of dimensions, usually 2)

Modeling
dataset features

Unsupervised, non-linear,
statistical model

Splitting patients
into groups

N/A

(algorithm does not
analyze groups, outcomes)

t-SNE

t-distributed Stochastic
Neighbor Embedding

N/A

(algorithm does not assign
items to clusters)

Non-linear, user determined
(user chooses number
of dimensions, usually 2)

N/A

(algorithm does not
model data)

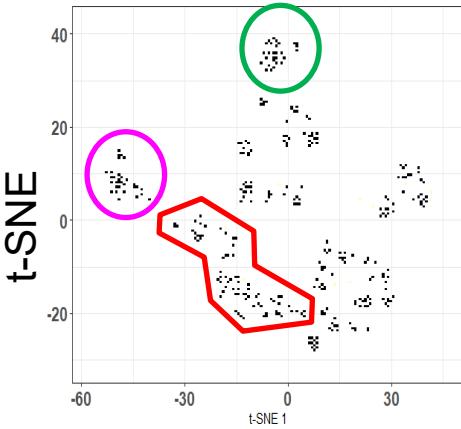
N/A

(algorithm does not
analyze groups, outcomes)

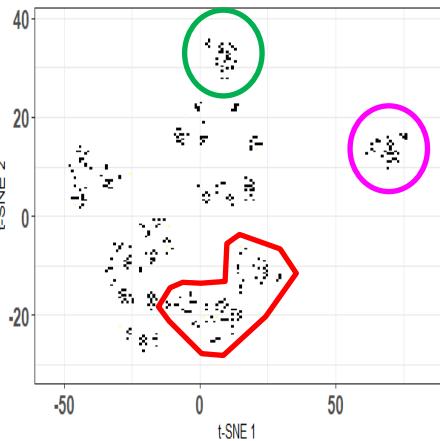
Multiple Runs of t-SNE vs. UMAP on a Patient Dataset (n = 339)

Gandelman et al., cGVHD Patient Dataset

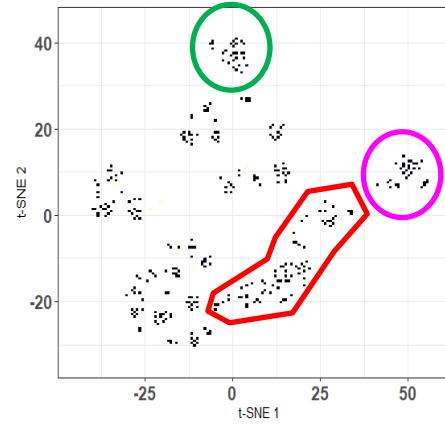
Run 1



Run 2

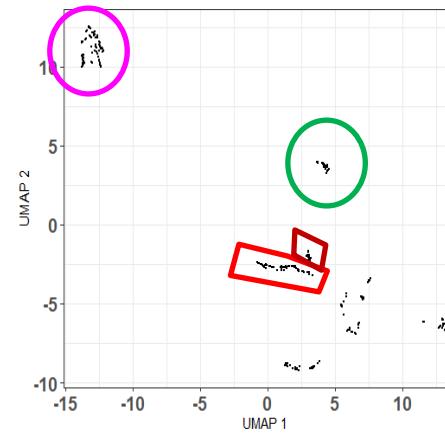
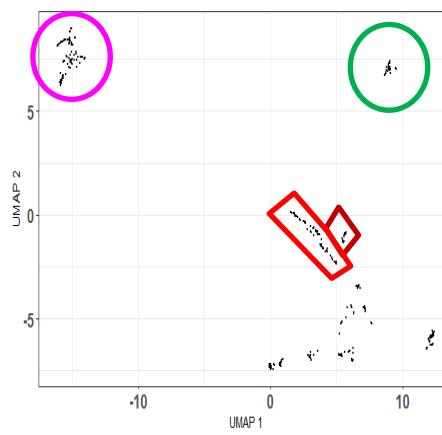
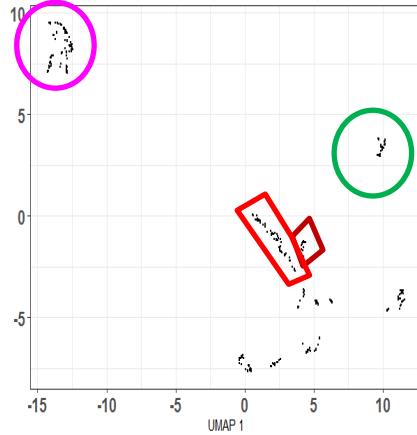


Run 3



In the t-SNE plots, the relative relationship of the major islands (“global structure”) alters between runs; t-SNE focuses on local structure

UMAP



Relative island position (“global structure”) is more stable & reflects original measurements in UMAP

Principle Component Analysis (PCA) is linear and deterministic, meaning that it strictly preserves global structure (and can overlook significant local structures / paths / trajectories)

Now, McInnes et al., UMAP Preserves Local and Global Structure

Datasets

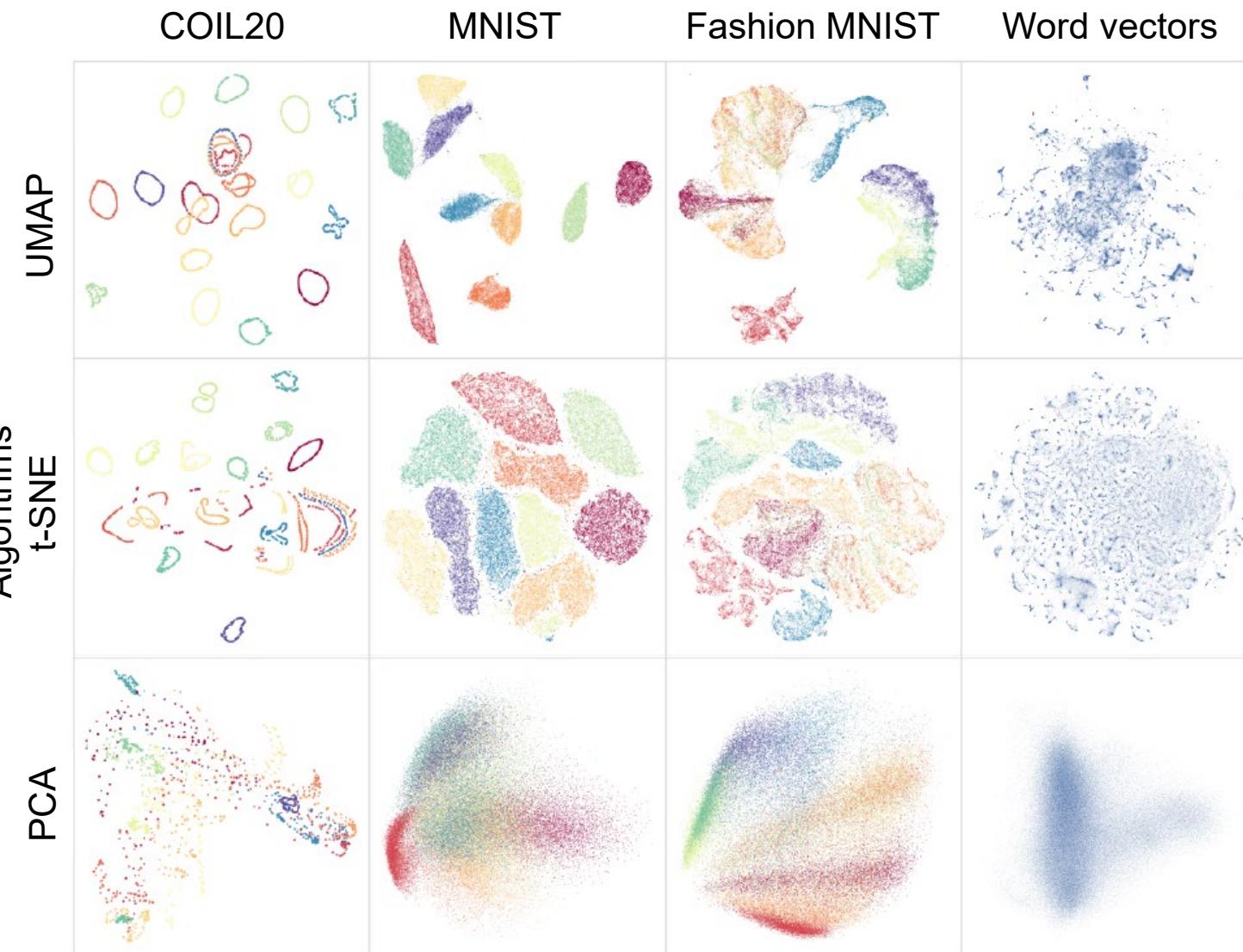


Figure 2: A comparison of several dimension reduction algorithms. We note that UMAP successfully reflects much of the large scale global structure that is well represented by Laplacian Eigenmaps and PCA (particularly for MNIST and Fashion-MNIST), while also preserving the local fine structure similar to t-SNE and LargeVis.

UMAP & Fit-SNE are Much Faster than Traditional t-SNE

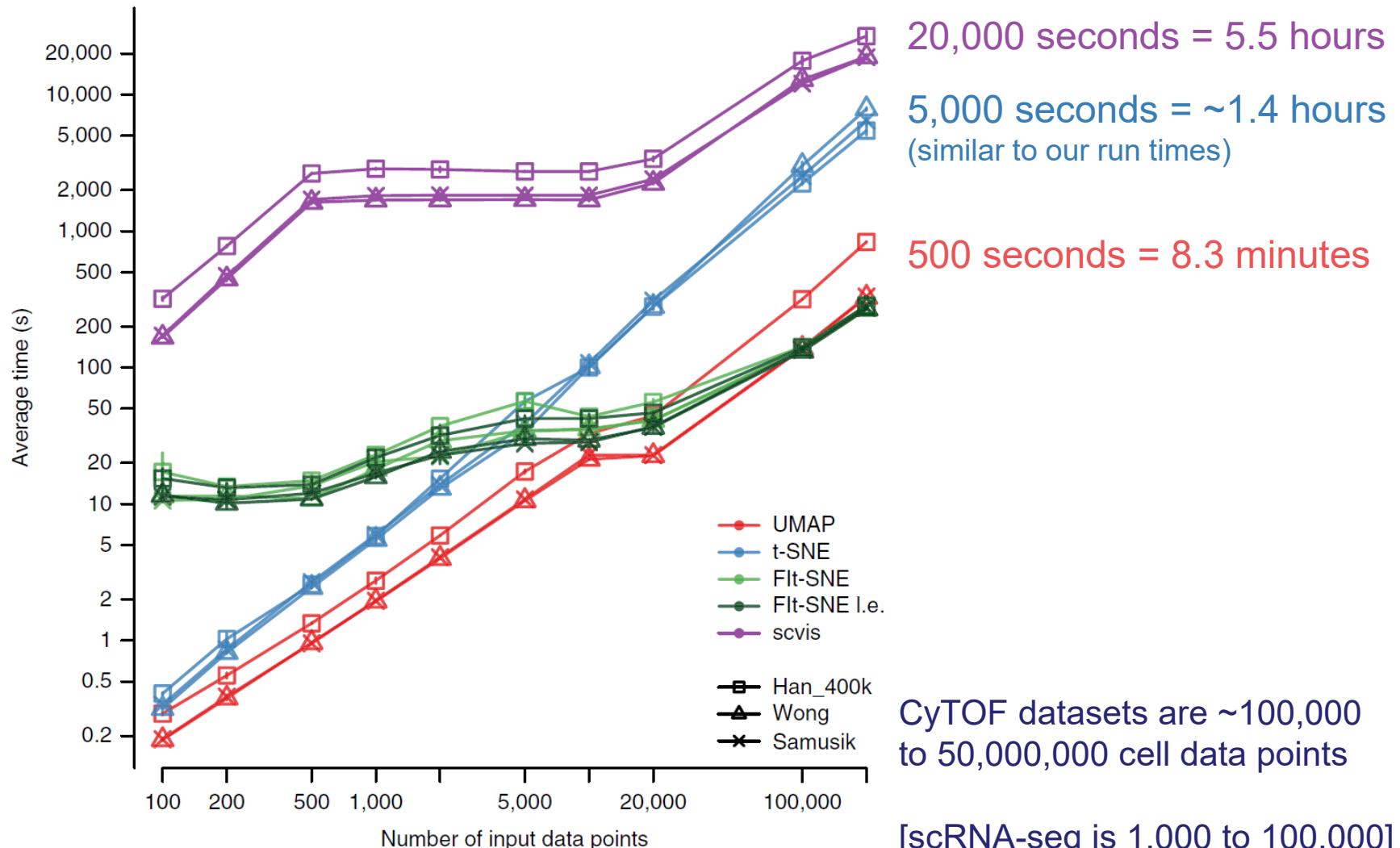
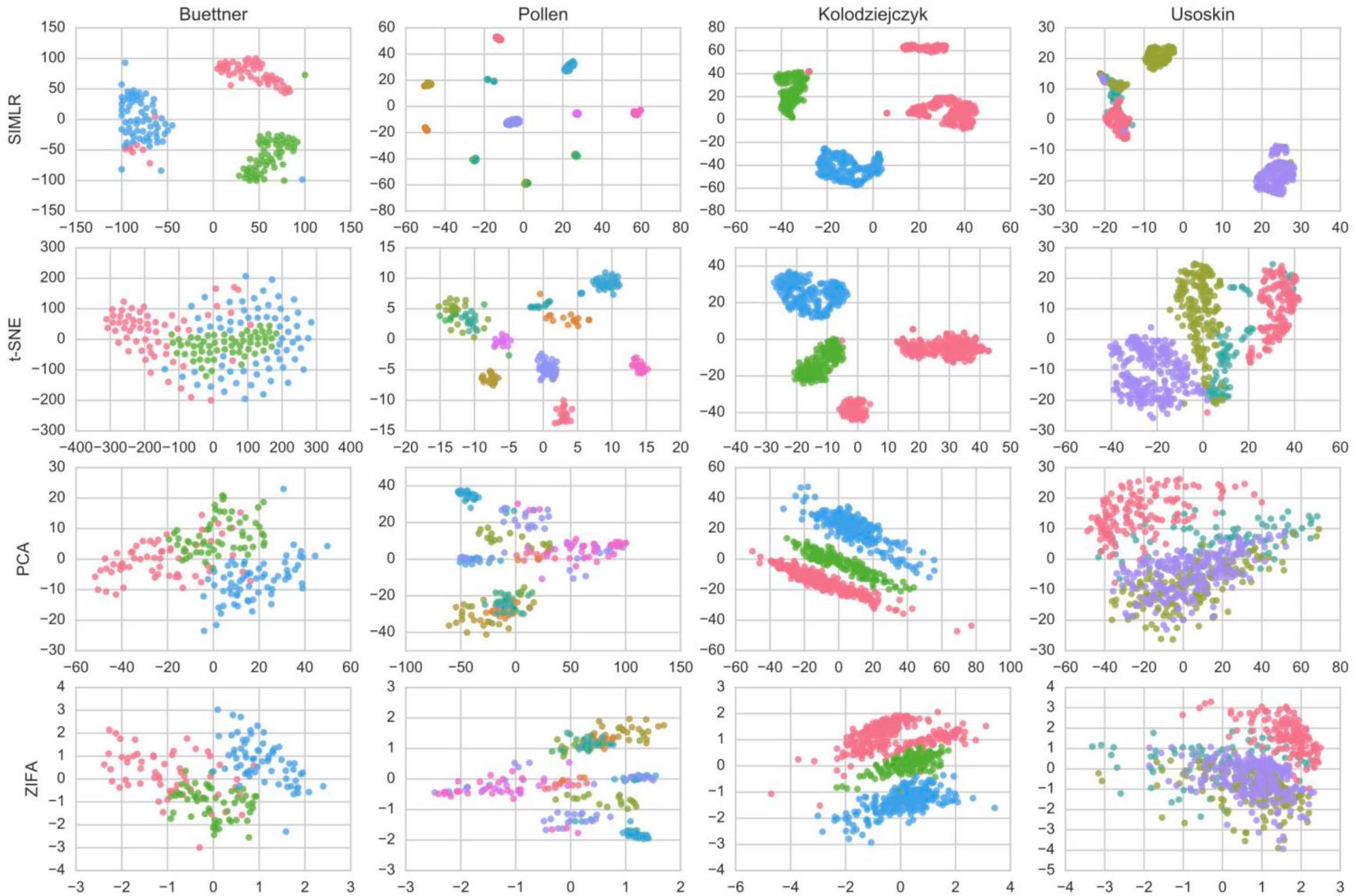


Figure 3 Run times of five dimensionality reduction methods for inputs of varying sizes. The average run time of three random subsamples is represented, with vertical bars representing s.d. after log-transforming the run times.

SIMLR vs. t-SNE vs. PCA on Four scRNA-seq Datasets



opt-SNE Provides Automated Optimization of t-SNE Parameters and PCA Initialization (Fast, Reliable)

nature communications

[View all journals](#)

Search  [Login !\[\]\(5c47ab4c3cccdc3719e6f40c67bea40b_img.jpg\)](#)

[Explore content](#)  [About the journal](#)  [Publish with us](#) 

[nature](#) > [nature communications](#) > [articles](#) > [article](#)

Article | [Open Access](#) | Published: 28 November 2019

Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets

Anna C. Belkina , Christopher O. Ciccolella, Rina Anno, Richard Halpert, Josef Spidlen & Jennifer E. Snyder-Cappione

[Nature Communications](#) **10**, Article number: 5415 (2019) | [Cite this article](#)

12k Accesses | **53** Citations | **67** Altmetric | [Metrics](#)

Abstract

Accurate and comprehensive extraction of information from high-dimensional single cell datasets necessitates faithful visualizations to assess biological populations. A state-of-the-art algorithm for non-linear dimension reduction, t-SNE, requires multiple heuristics and fails to produce clear representations of datasets when millions of cells are projected. We develop opt-SNE, an automated toolkit for t-SNE parameter selection that utilizes Kullback-Leibler divergence evaluation in real time to tailor the early exaggeration and overall number of gradient descent iterations in a dataset-specific manner. The precise calibration of early exaggeration together with opt-SNE adjustment of gradient descent learning rate dramatically improves computation time and enables high-quality visualization of large cytometry and transcriptomics datasets, overcoming limitations of analysis tools with hard-coded parameters that often produce poorly resolved or misleading maps of fluorescent and mass cytometry data. In summary, opt-SNE enables superior data resolution in t-SNE space and thereby more accurate data interpretation.

Check out Dr. Anna Belkina's opt-SNE Webinar on ISAC's CYTO U Learning Portal



[Donate](#) [Contact Us](#) [Sign In](#) [Join](#)

[About](#) [Membership](#) [CYTO U](#) [Publications](#) [Events](#) [Programs](#) [Resources](#) [News](#)

[« Go to Upcoming Event List](#)

The promotional graphic for the CYTO U webinar. It features a purple header with the ISAC logo. Below it, a grey bar contains the text "NEW WEBINAR". The main content area is purple and contains the title "Don't Leave Home Without a Map: Powerful Dimensionality Reduction Methods for Cytometry Data Visualization" and the presenter information "Presented by: Anna Belkina, MD, PhD Boston University School of Medicine". At the bottom left is the CYTO UNIVERSITY logo with the tagline "Your Cytometry Learning Portal". A white call-to-action button on the right says "Register today at <https://learning.isac-net.org>". On the right side, there is a portrait photo of Dr. Anna Belkina, a woman with long, curly brown hair, looking directly at the camera. A vertical text box on the far right lists the webinar details: "Wednesday August 25th 12pm EDT".

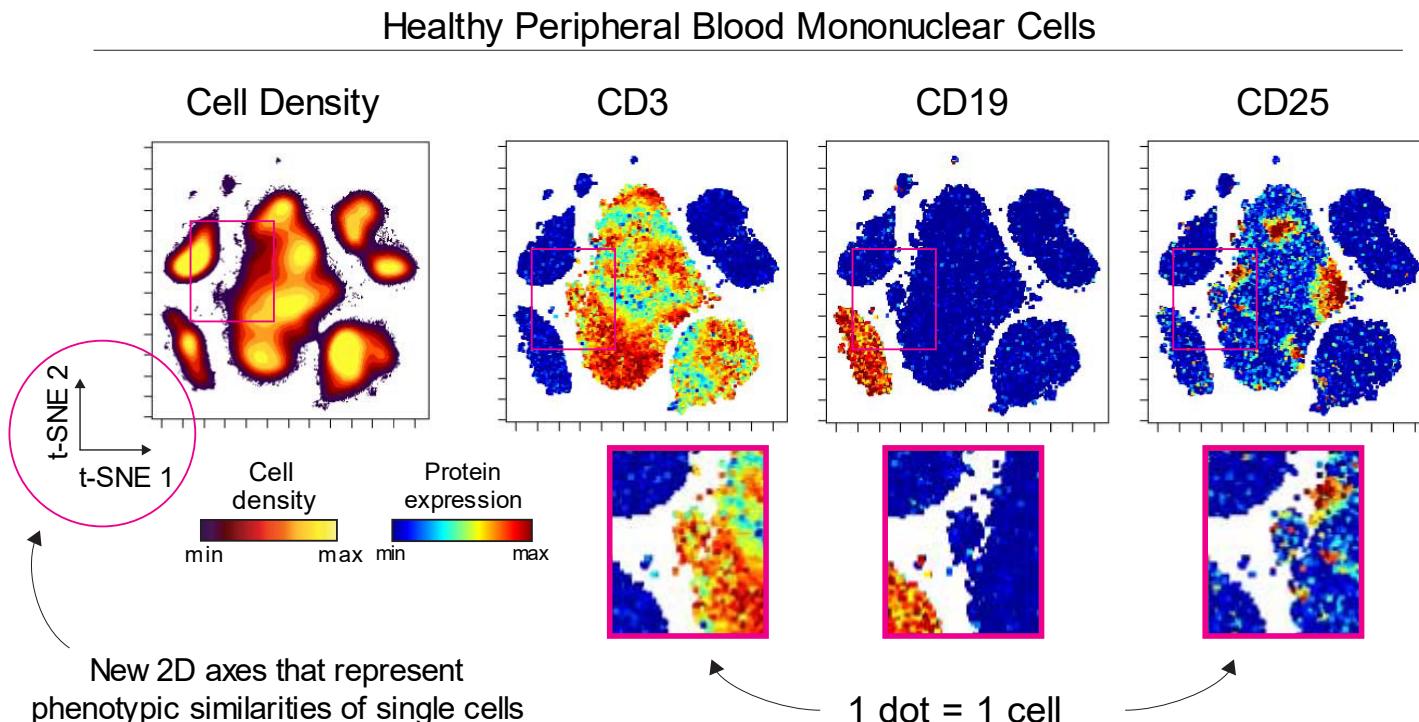
CYTO U Webinar: <https://learning.isac-net.org/products/dont-leave-home-without-a-map-powerful-dimensionality-reduction-methods-for-cytometry-data-visualization>

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:

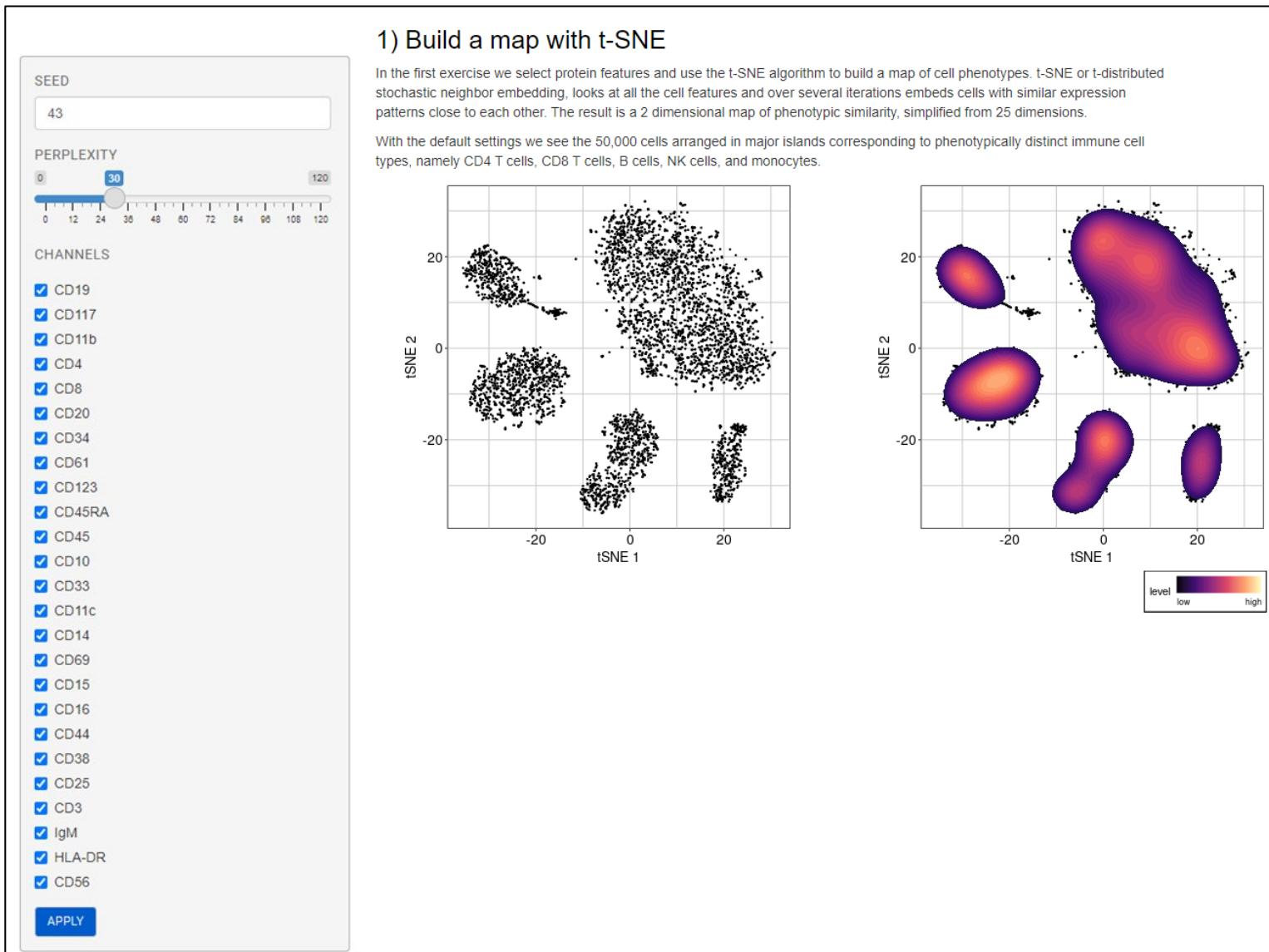
| | CD19 | CD117 | CD11b | CD4 | CD8 | CD20 | CD34 | CD61 | CD123 | CD45RA | CD45 | CD10 | CD33 | CD11c | CD14 | CD69 | CD15 | CD16 | CD44 | CD38 | CD25 | CD3 | IgM | HLA-DR | CD56 |
|--------|------|-------|-------|-----|-----|------|------|------|-------|--------|------|------|------|-------|------|------|------|------|------|------|------|-----|-----|--------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

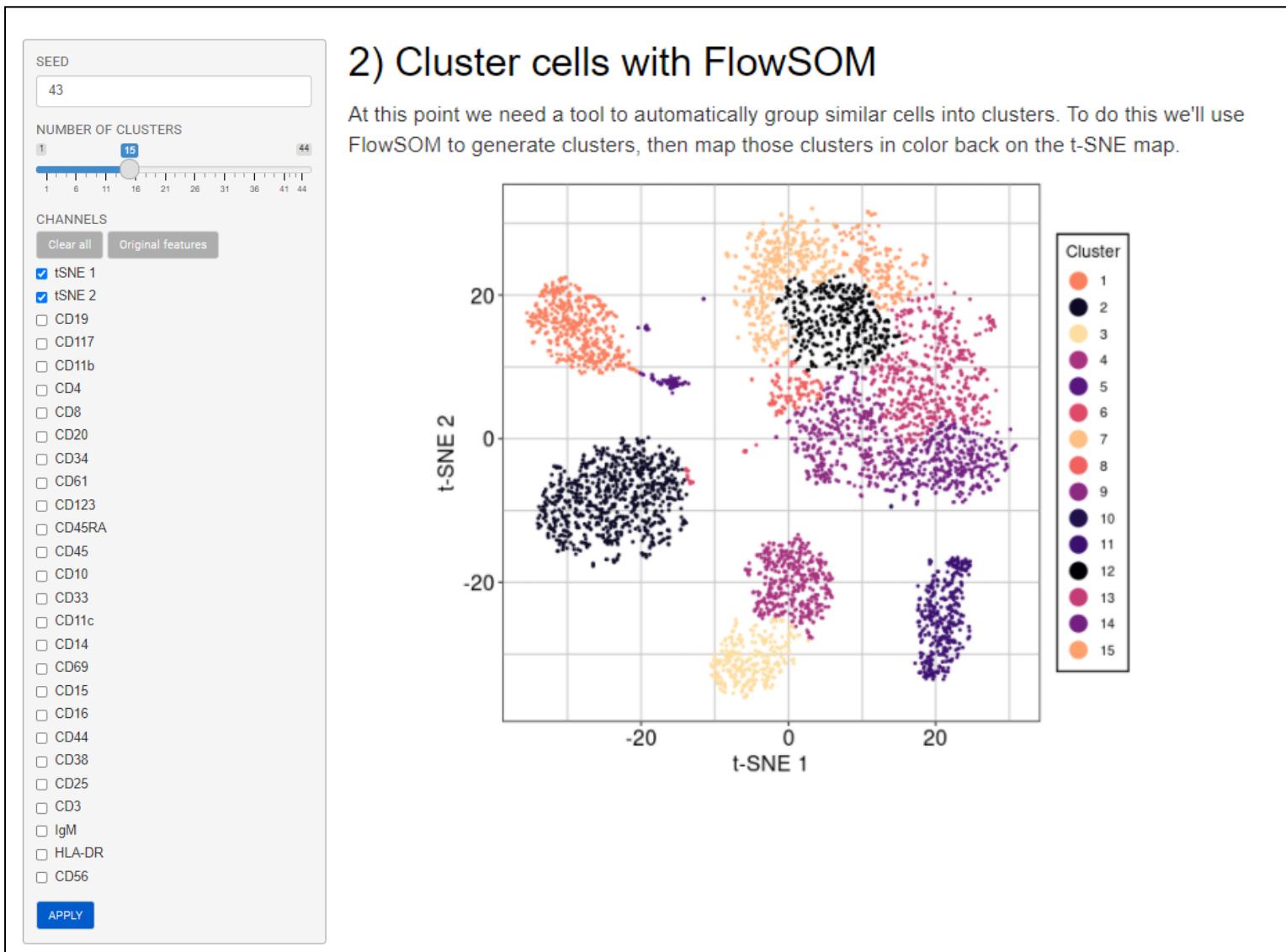
For this tutorial, we've taken a random sample of 5,000 cells from the 50,000 to run analyses on. If you'd like a larger or smaller sample size, you have the option to change that in the following menu. Alternatively if you'd like to reset your session, you can use the clear session button.

SAMPLE SIZE

APPLY

CLEAR SESSION





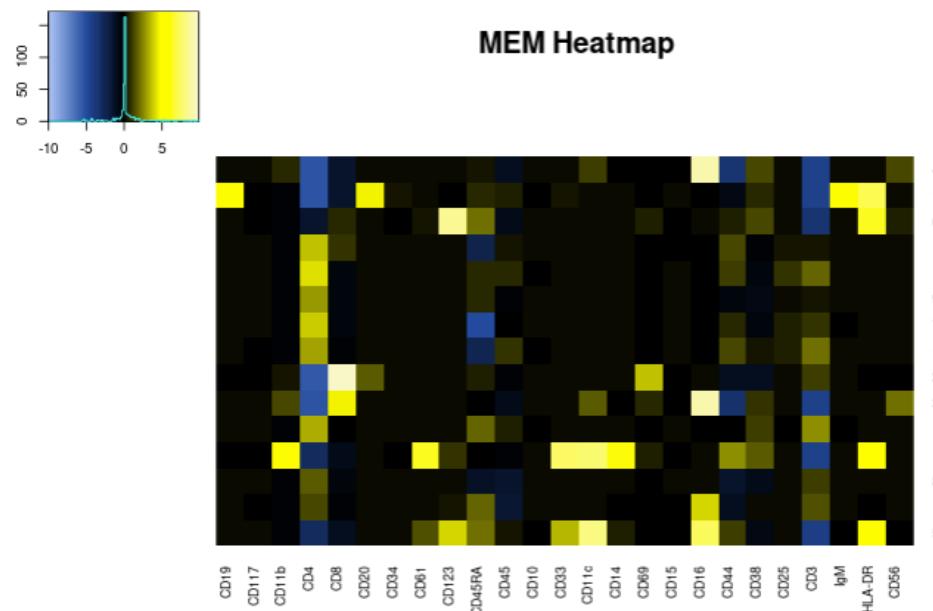
Referenceless?

APPLY

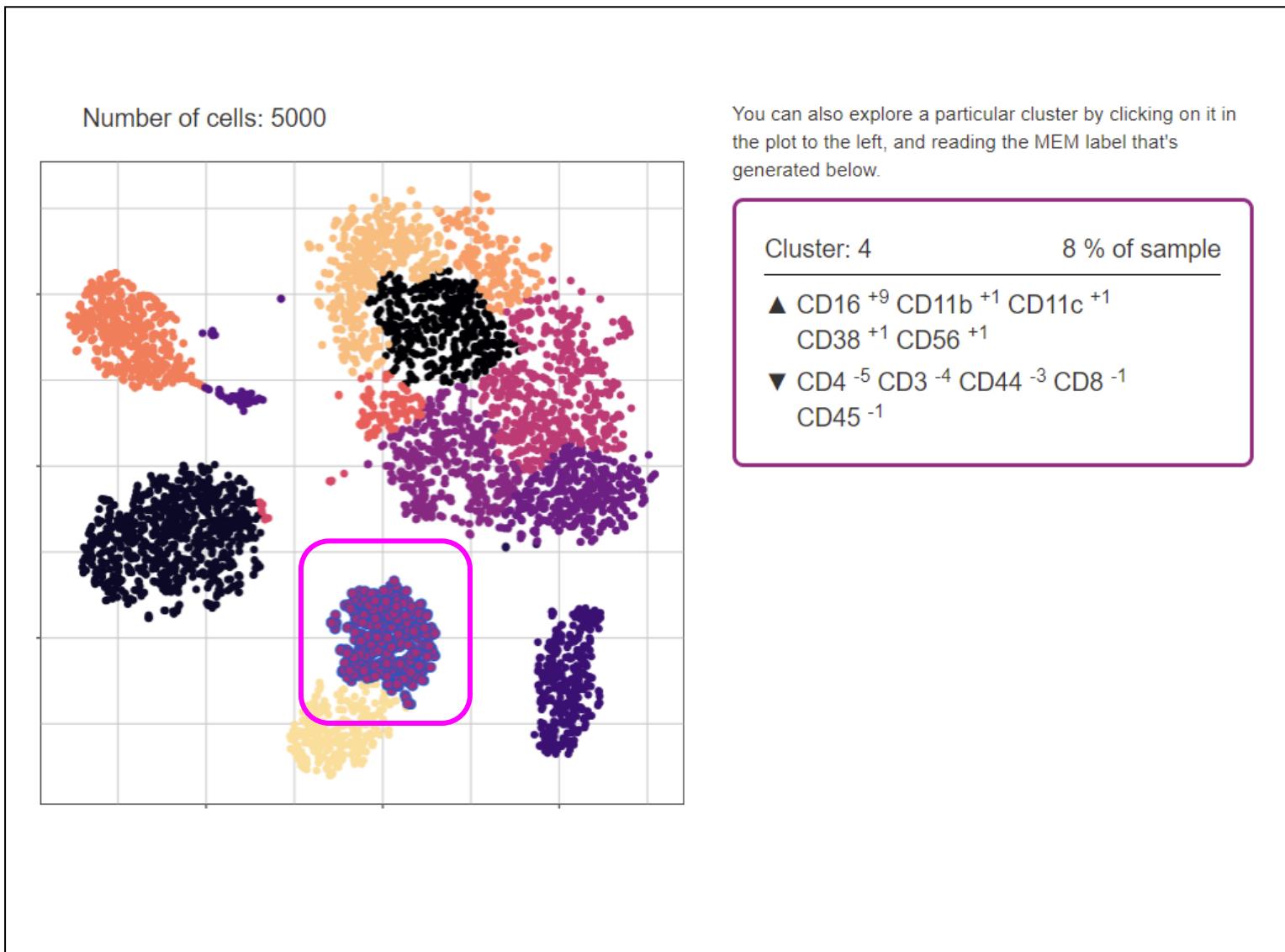
3) Describe clusters with MEM

Once we've grouped cells into clusters, how can we identify what kind of cells are in each cluster? You can look up marker expression values in the "spreadsheet" view of the data, or run an algorithm like MEM which calculates features that are enriched within the various groupings. A MEM label often provides enough information to infer identity if it is a known cell type, or guess at its biological significance. A methods paper explaining the MEM algorithm and going through examples is available in [Current Protocols in Cytometry](#).

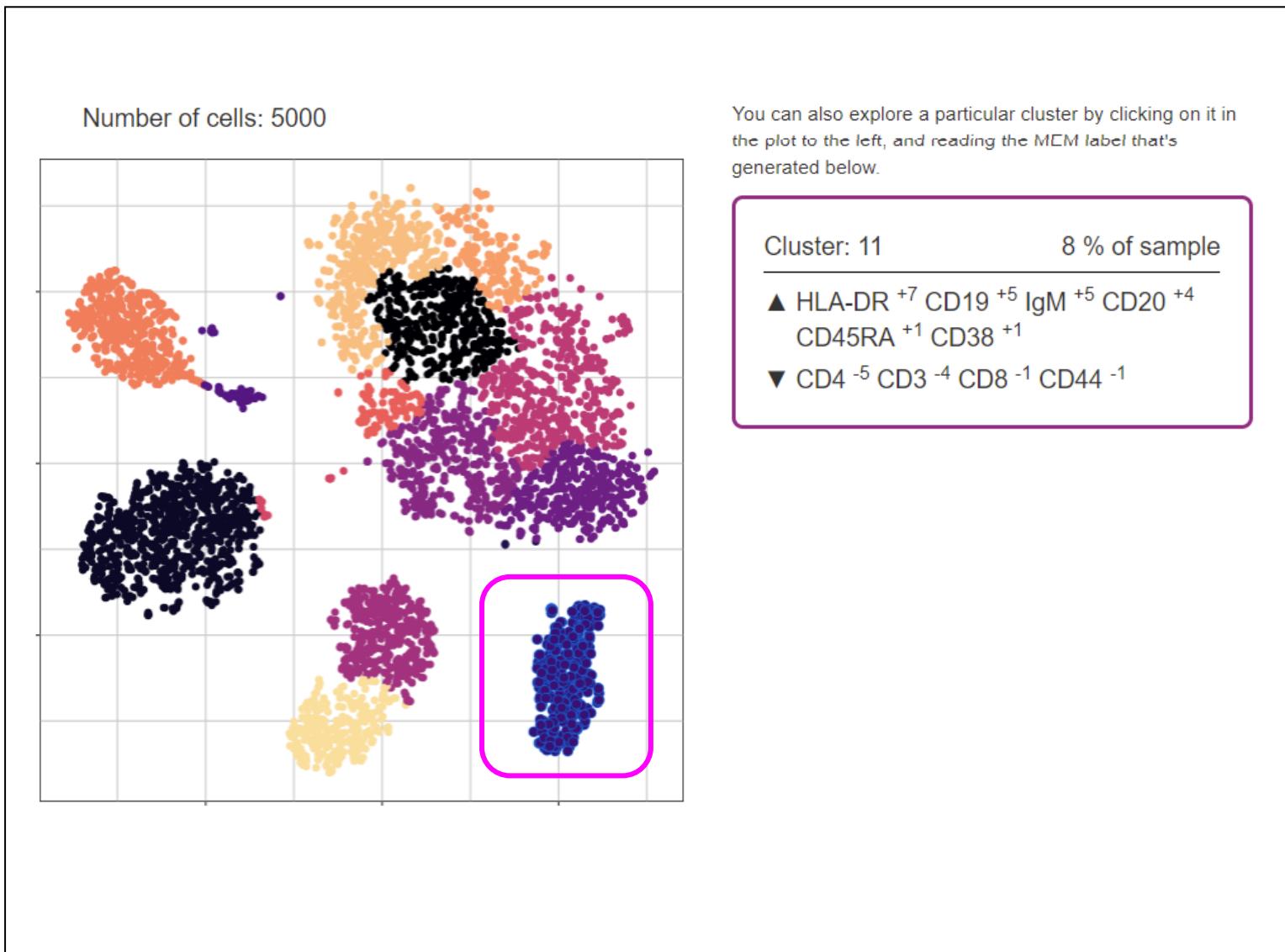
Here MEM outputs a heatmap of the relative expression of each protein organized by cluster.

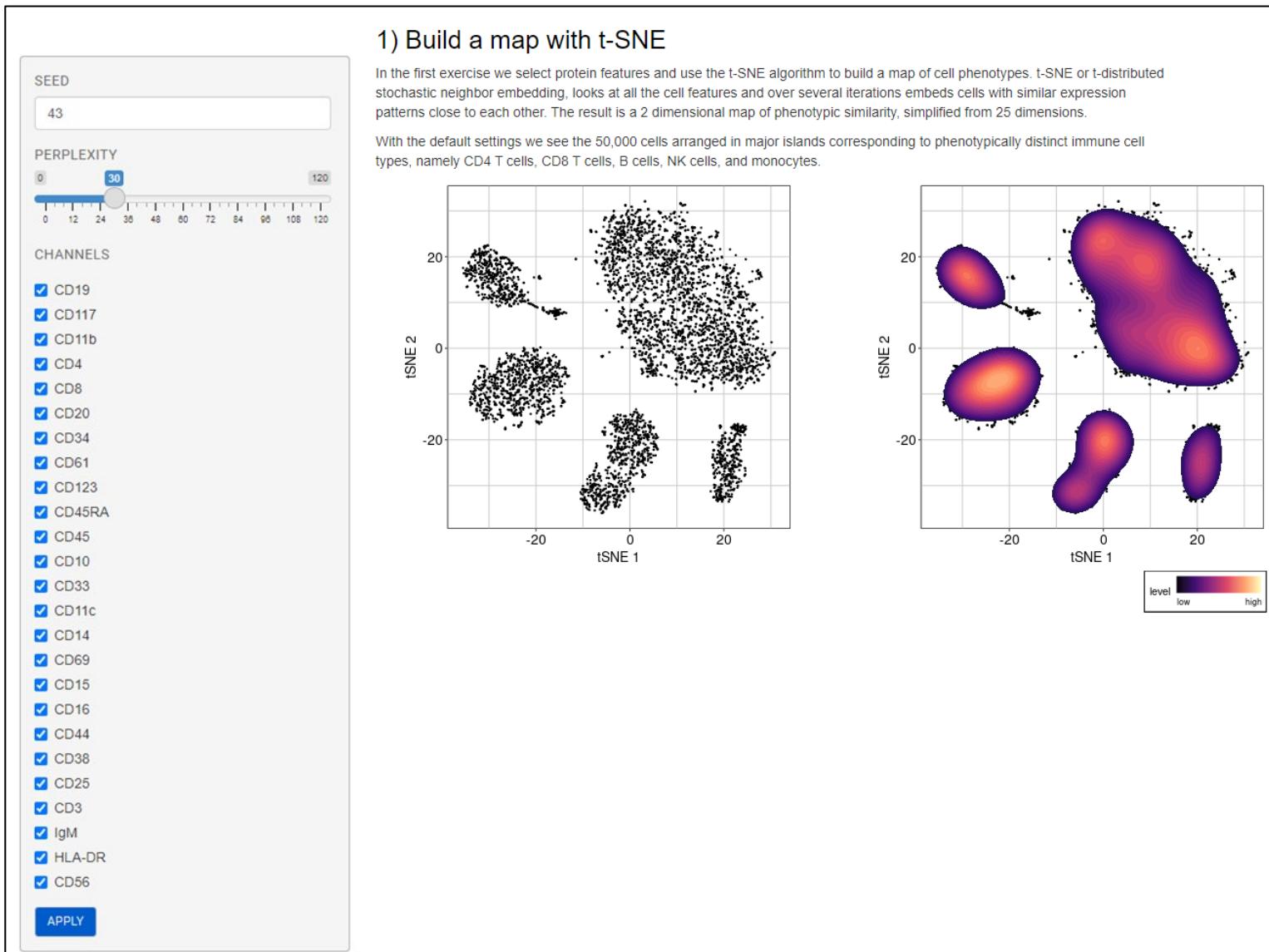


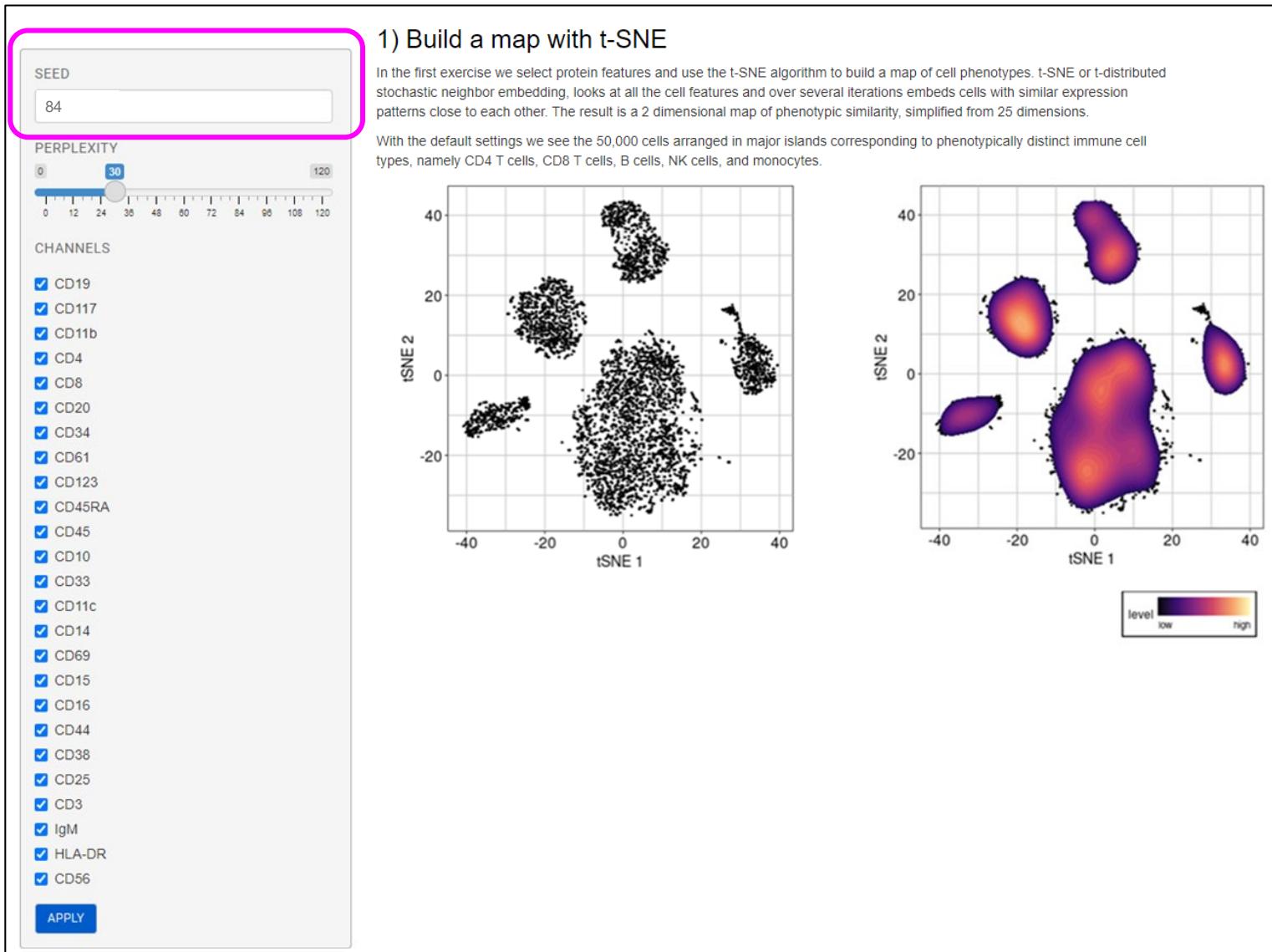
<https://cytolab.shinyapps.io/PBMC/>



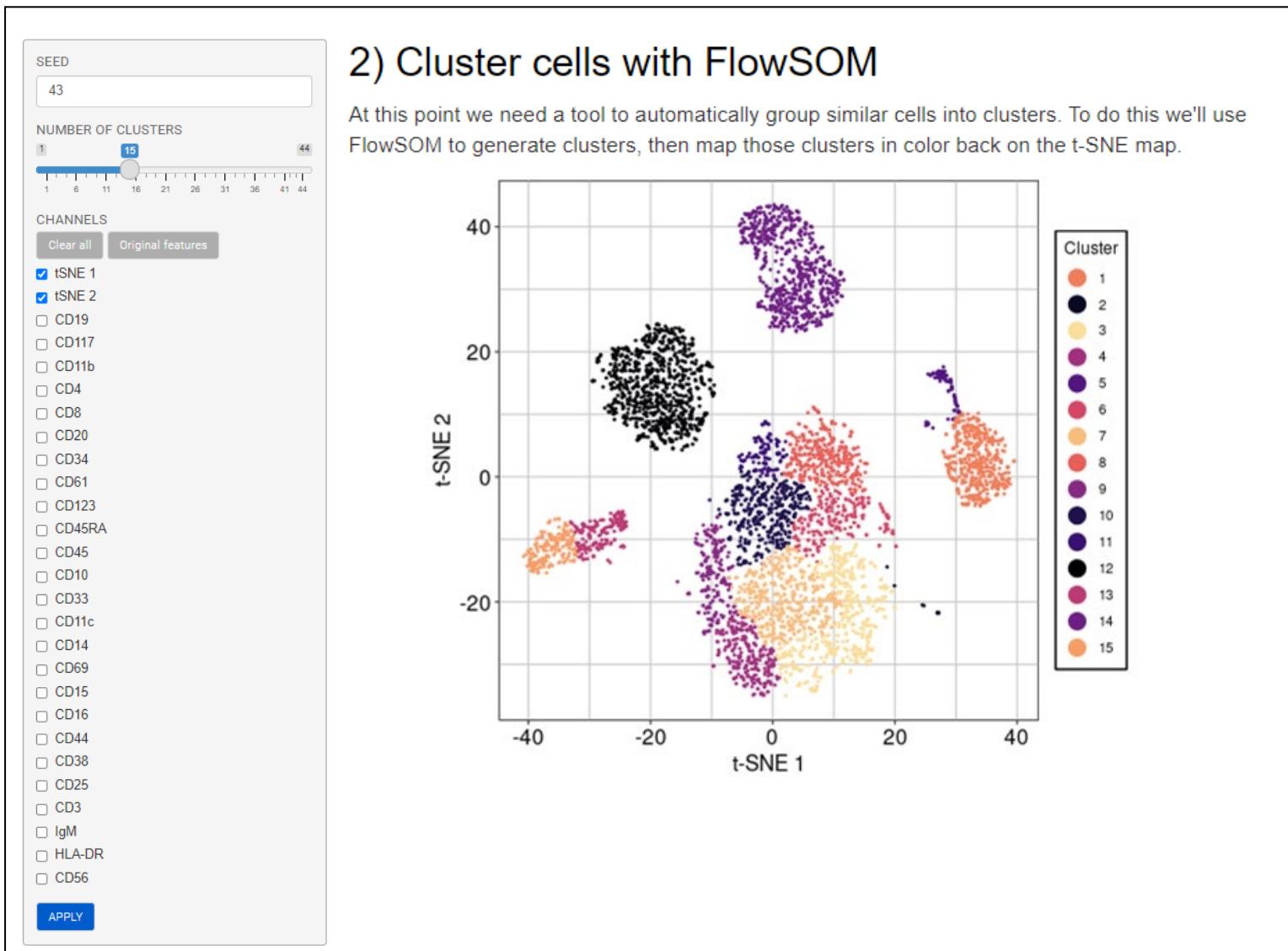
<https://cytolab.shinyapps.io/PBMC/>

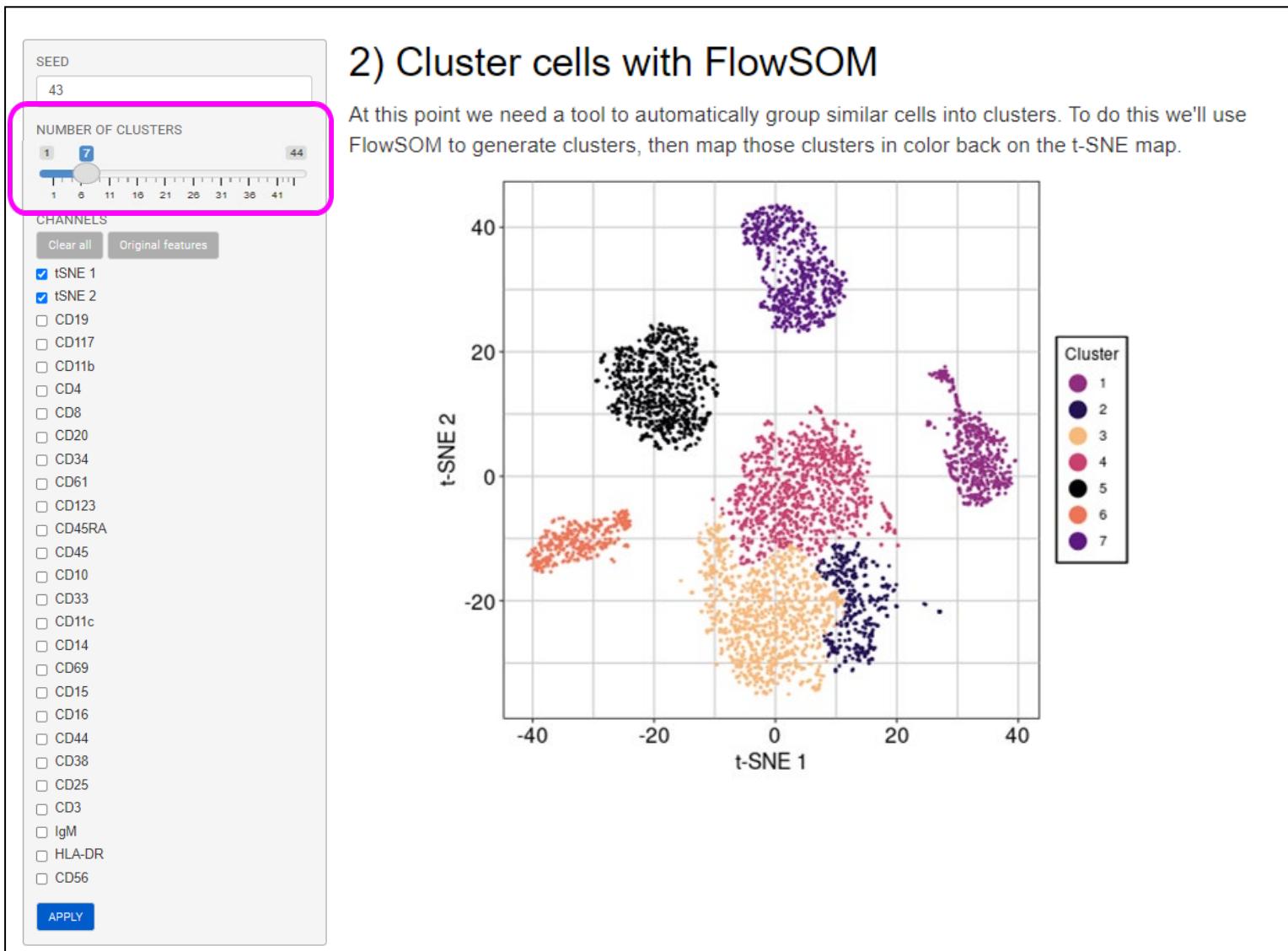






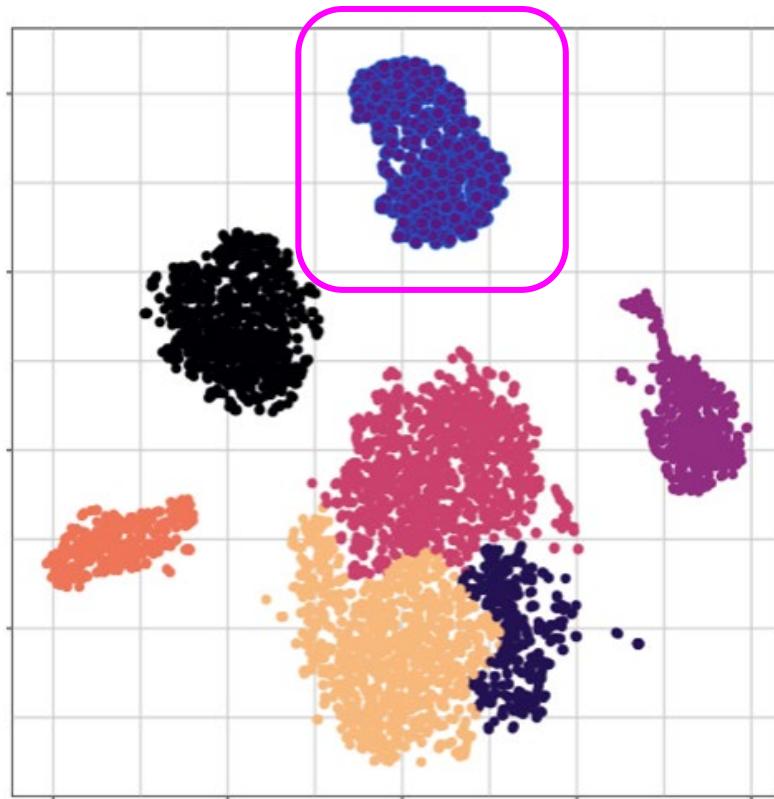
<https://cytolab.shinyapps.io/PBMC/>





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

Number of cells: 5000



You can also explore a particular cluster by clicking on it in the plot to the left, and reading the MEM label that's generated below.

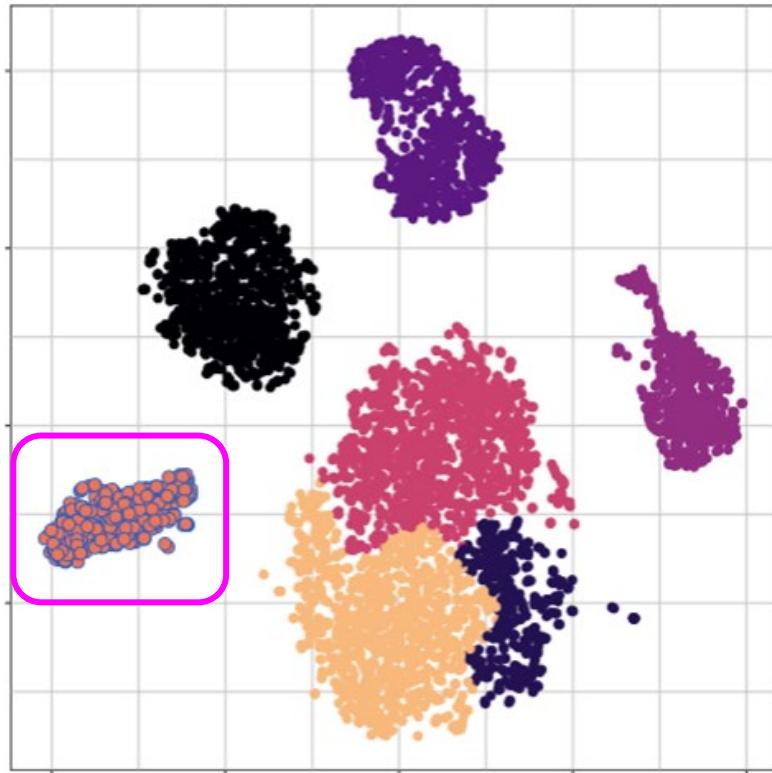
Cluster: 7

14 % of sample

- ▲ CD16⁺⁹ CD56⁺² CD11b⁺¹
CD11c⁺¹ CD38⁺¹
- ▼ CD4⁻⁶ CD44⁻⁴ CD3⁻⁴ CD45⁻¹

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

Number of cells: 5000



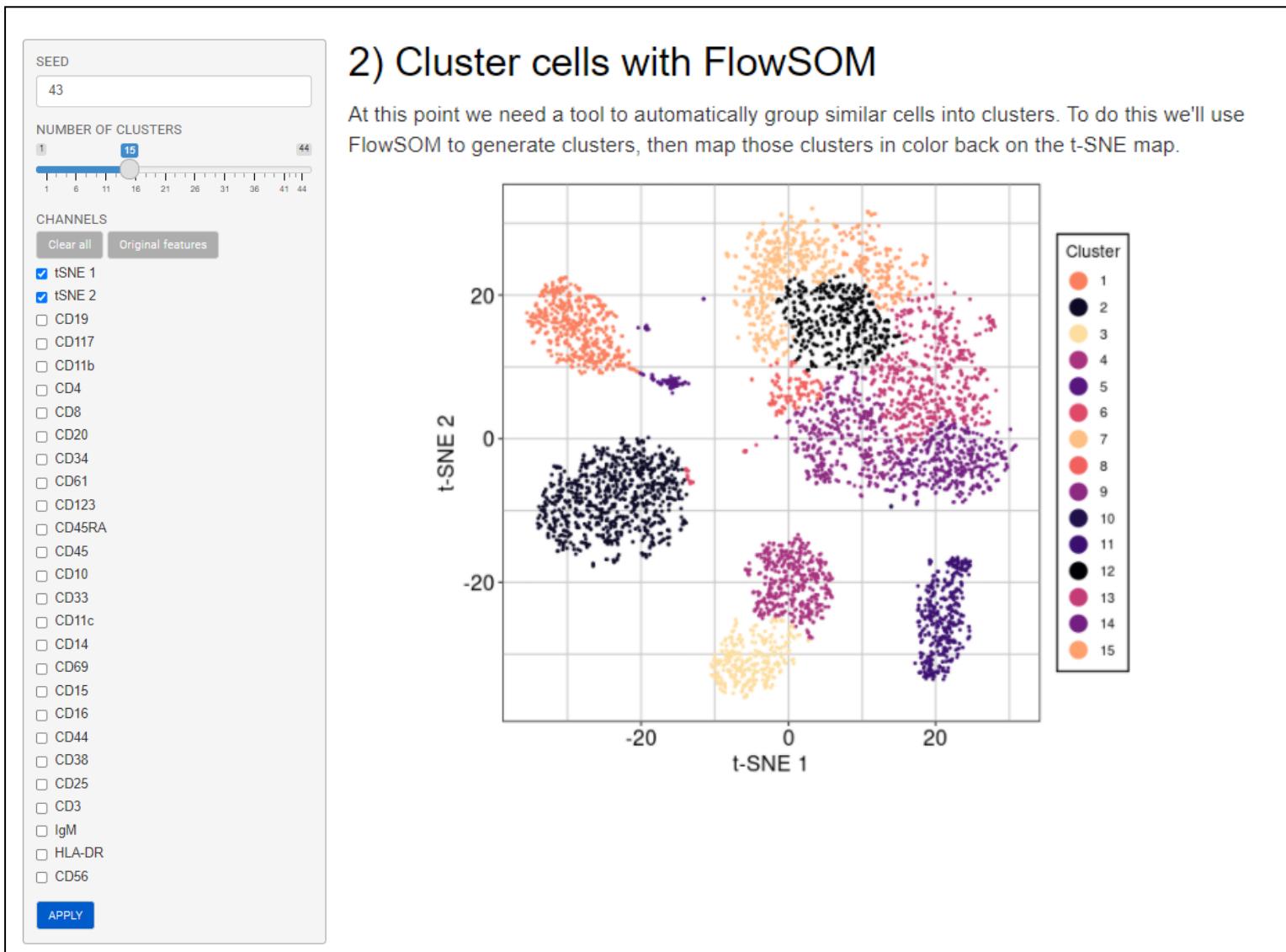
You can also explore a particular cluster by clicking on it in the plot to the left, and reading the MEM label that's generated below.

Cluster: 6

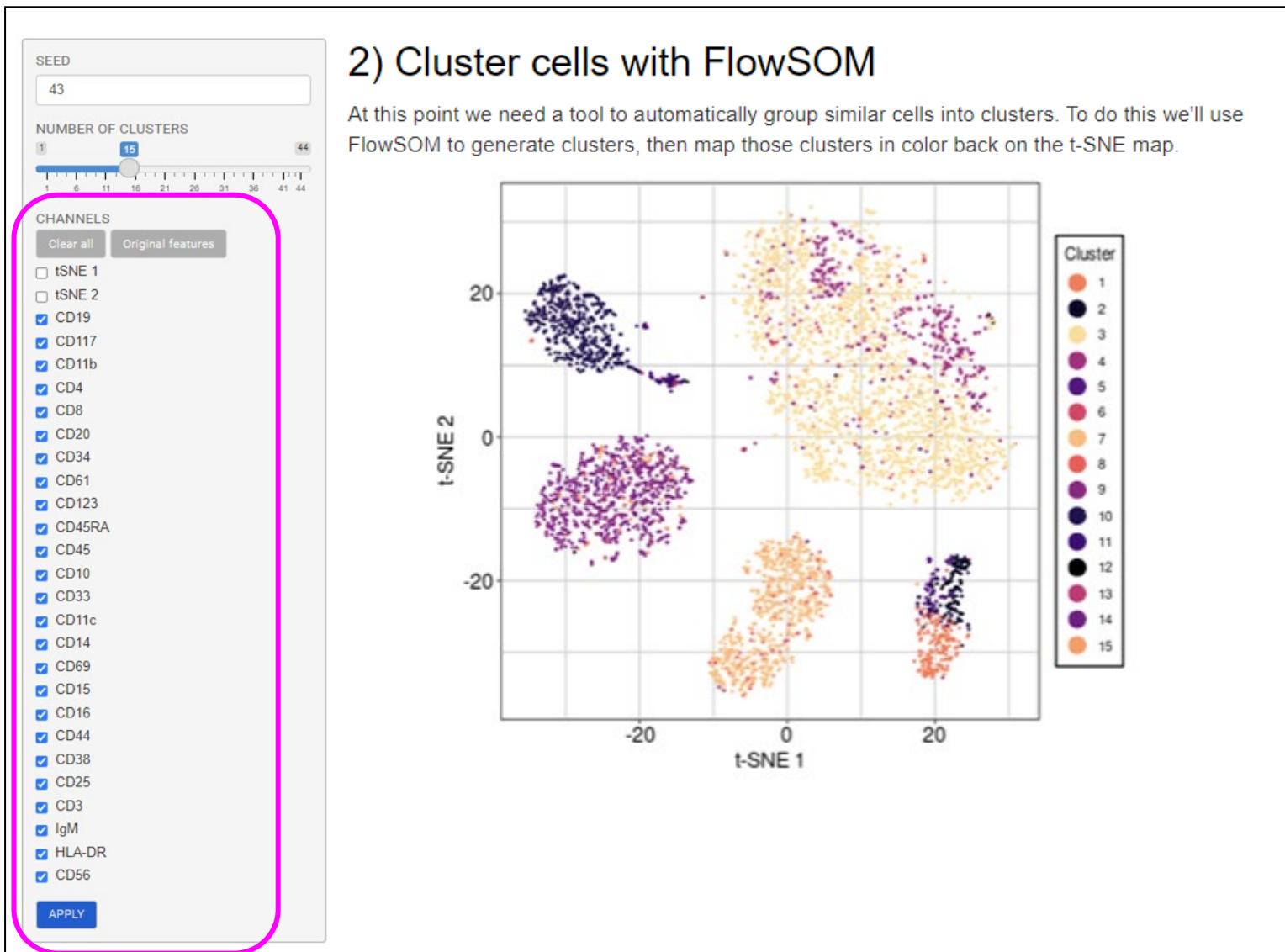
8 % of sample

- ▲ HLA-DR ⁺⁷ CD19 ⁺⁵ IgM ⁺⁵
CD20 ⁺⁴ CD45RA ⁺¹ CD38 ⁺¹
- ▼ CD4 ⁻⁵ CD3 ⁻⁴ CD8 ⁻¹ CD44 ⁻¹

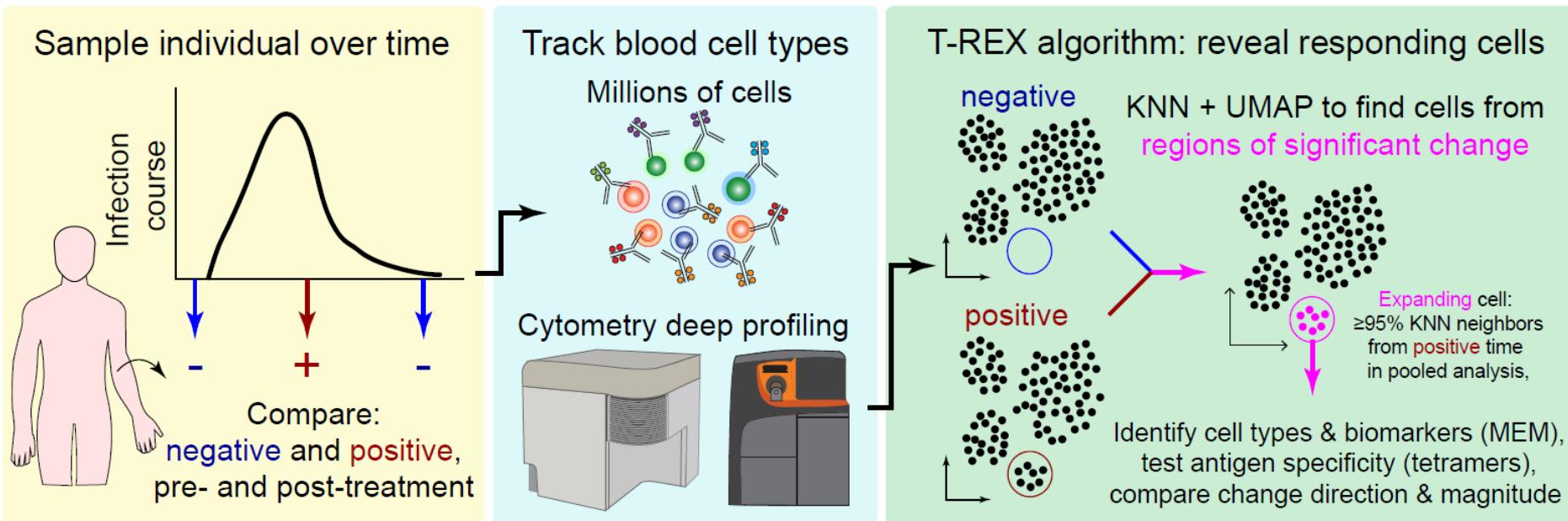
<https://cytolab.shinyapps.io/PBMC/>



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



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

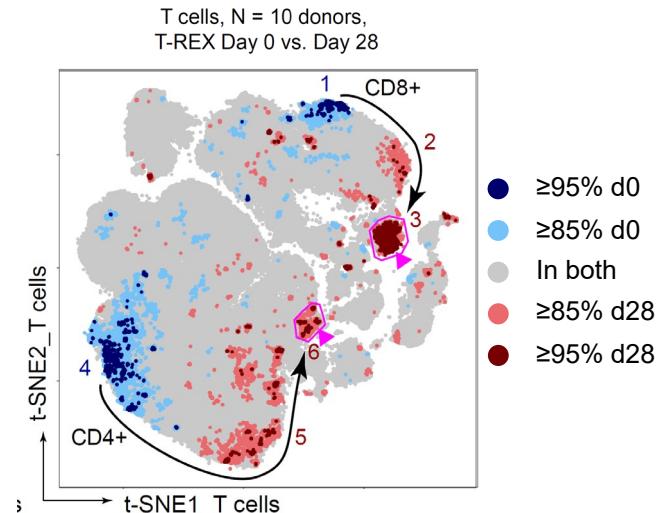
Revealing very rare cells or cells changing significantly

Dimensionality reduction

- t-SNE or UMAP

Clustering

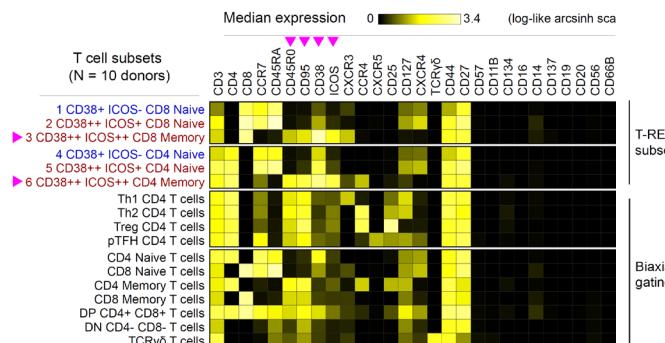
- KNN on all cells
- DBSCAN on cells in regions changing significantly



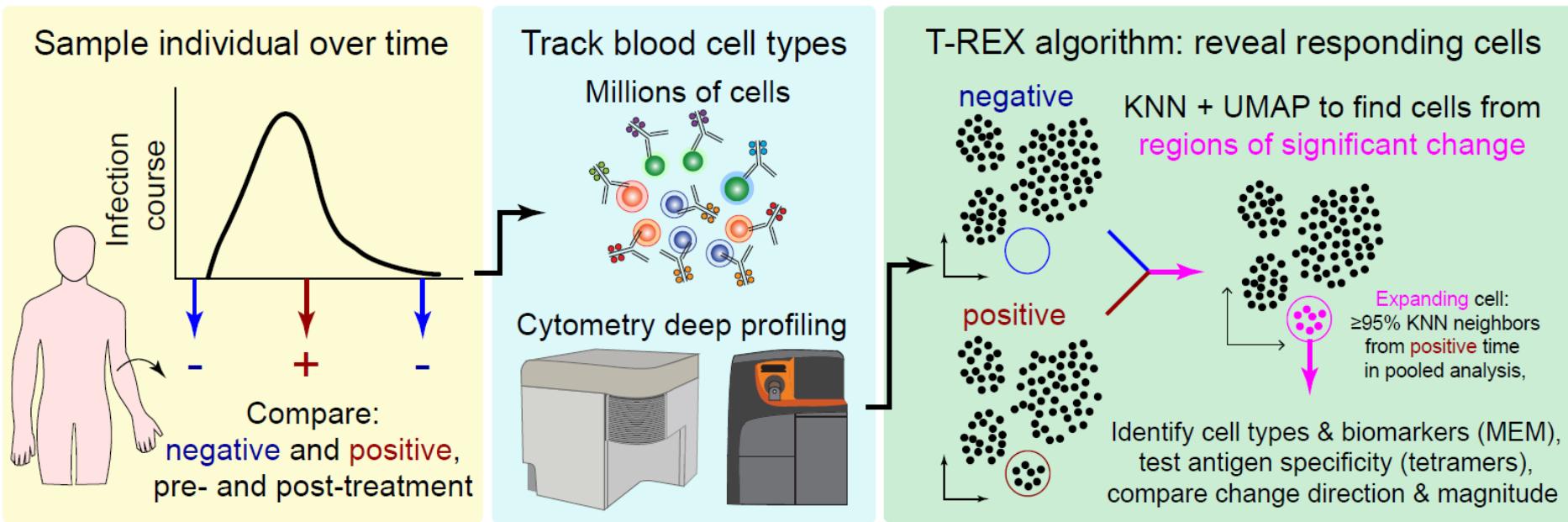
Characterizing cells expanding or contracting

Learn cell identity

- MEM

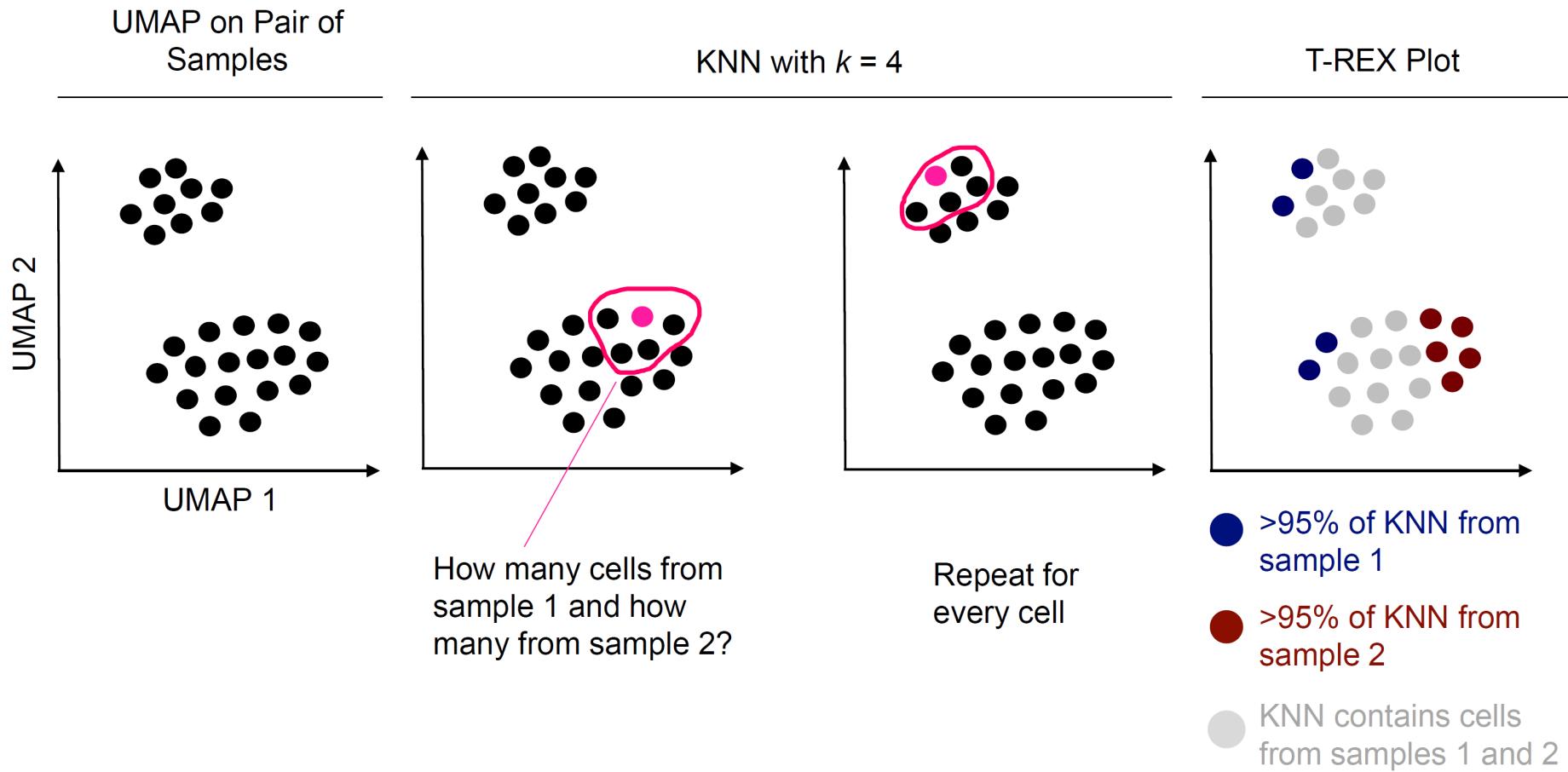


Key Ideas & Findings in Today's Talk



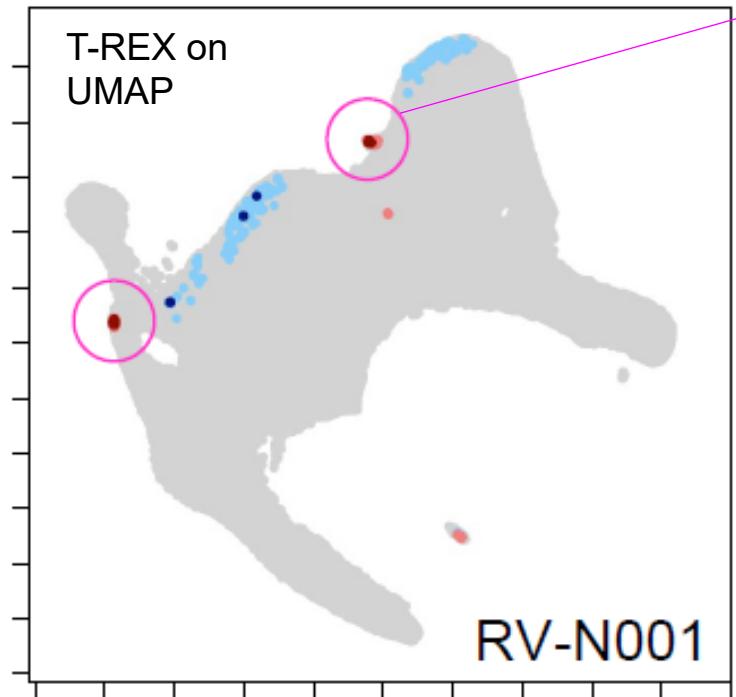
- Idea 1: T-REX automatically reveals virus-specific T cells in rhinovirus & SARS-CoV-2 vaccine response (without the need for tetramers, sorting, or sequencing)
- Idea 2: Approach focuses on extreme change & can summarize disease, therapy, or perturbation response (direction & magnitude of change; rhinovirus, 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

Live CD4+ T cells



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)

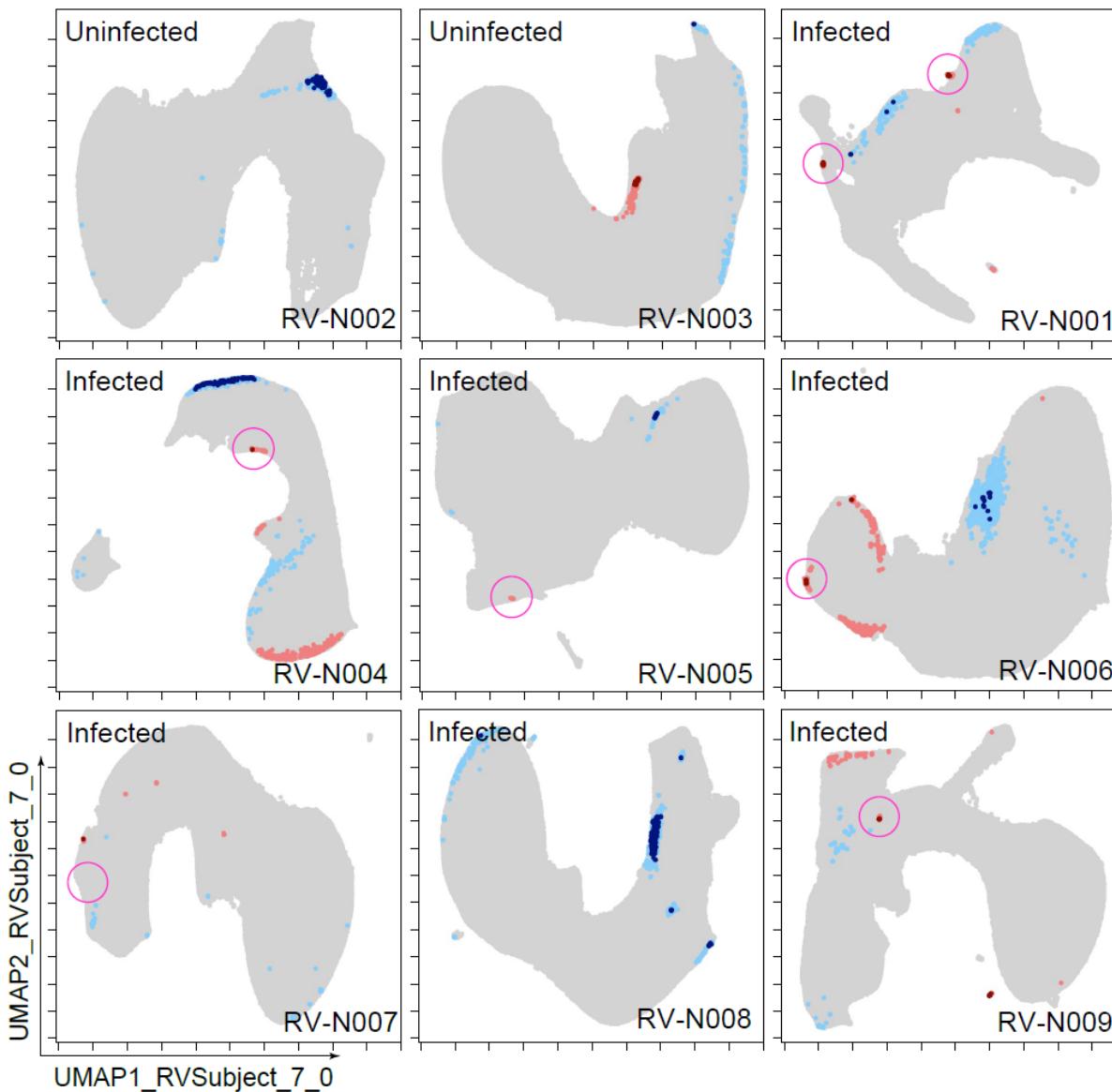


- >95% from d0 ● >95% from d7
- >75% from d0 ● >75% from d7
- >5% of neighbors tetramer+

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

● $\geq 95\%$ from day 0 ● $> 85\%$ from day 0 ● $> 85\%$ from day 7 ● $\geq 95\%$ from day 7 ○ Tetramer+ hotspot



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

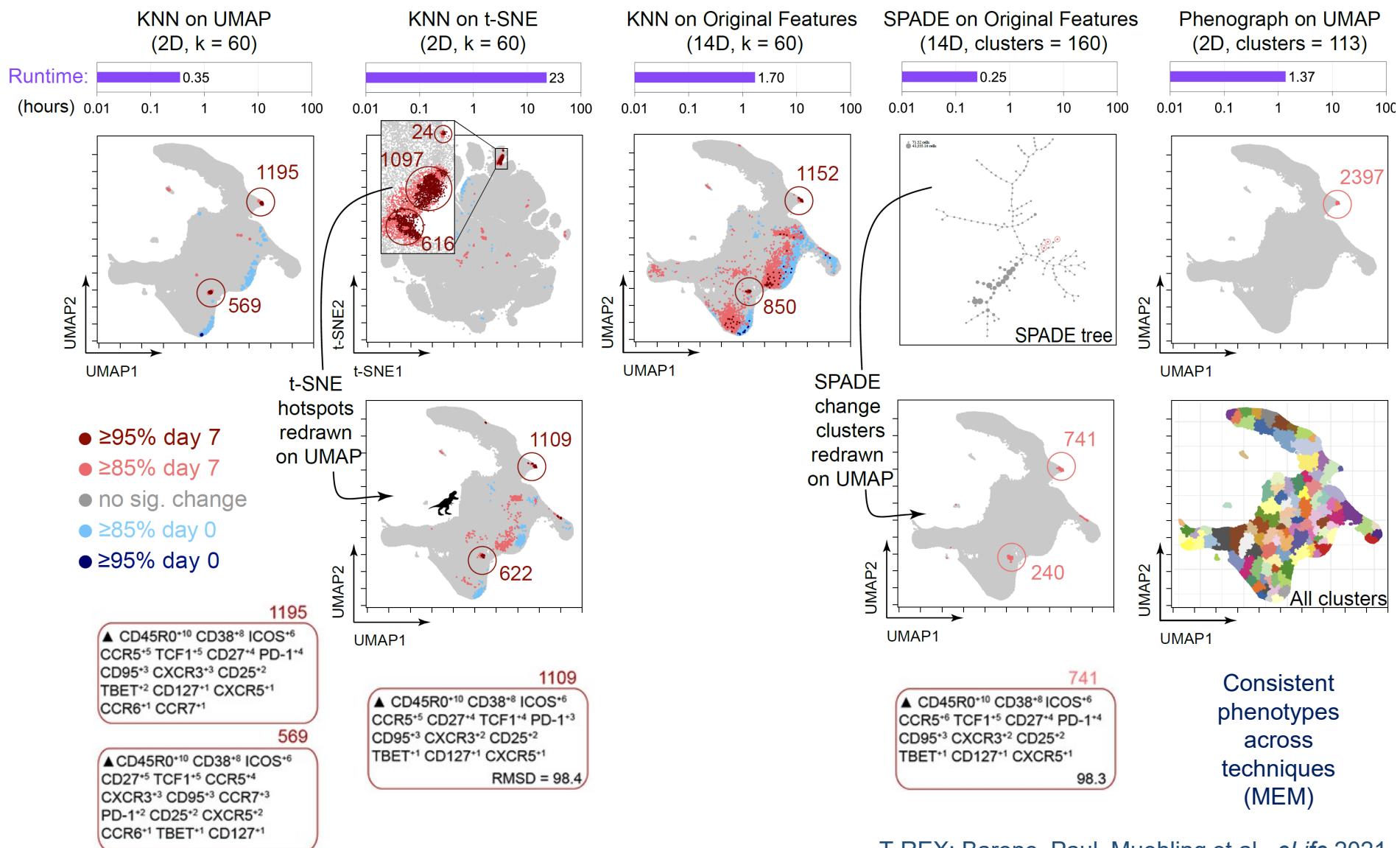
T-REX revealed virus-specific T cells without tetramers



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

Methods that identified at least one >85% change cluster (T-REX hotspot of change)



T-REX revealed virus-specific T cells without tetramers

Also found to work for:

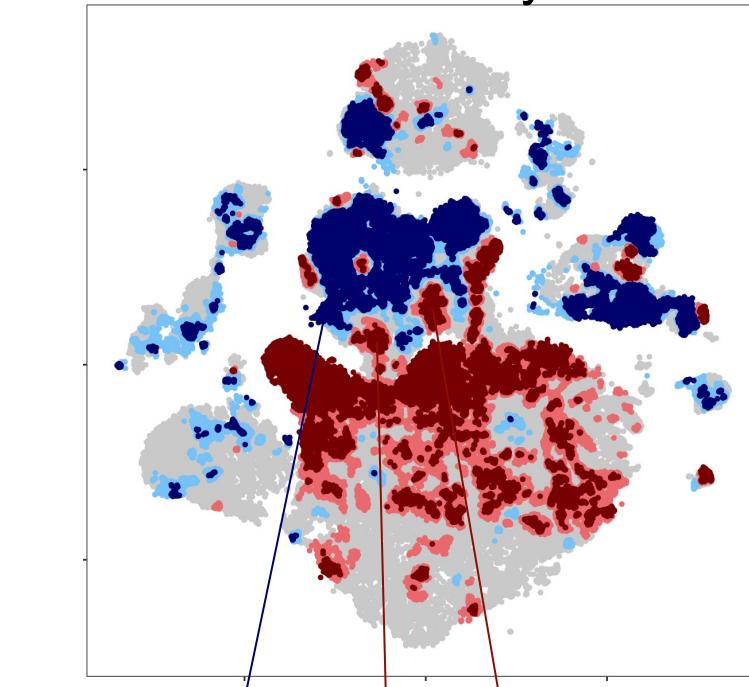
- a range of k-values ($k = 60$ was optimal)
- post-infection as the comparison point to day 7
- data from a range of cytometers, studies, and labs
- COVID-19, melanoma immunotherapy response, AML

(see the manuscript for this & more!)

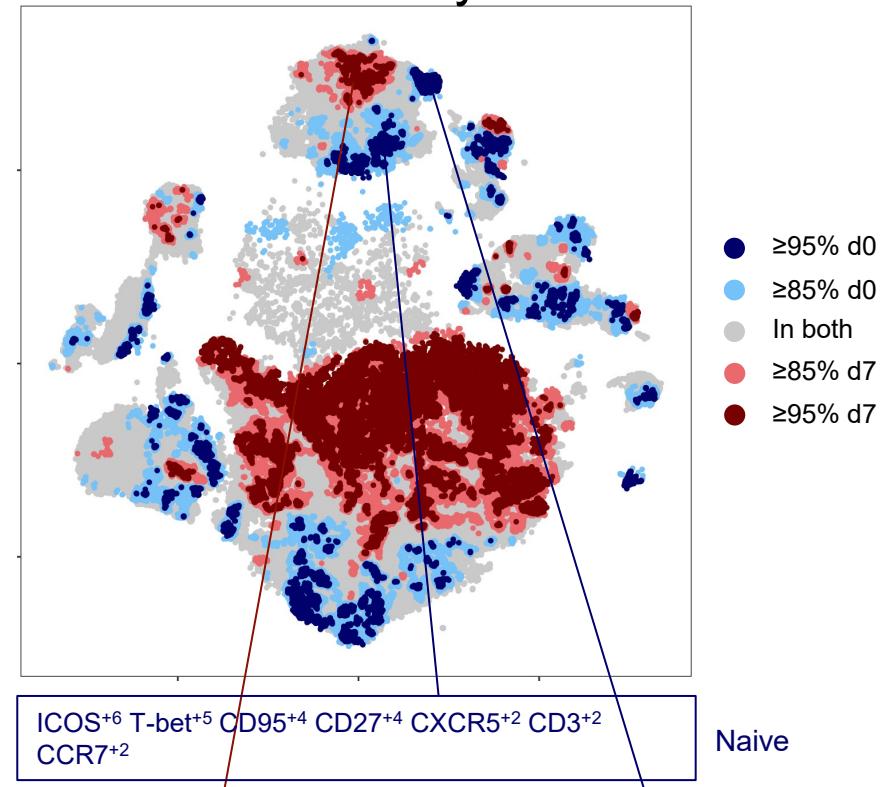


Massive Immune Change, Common Shifts in Expanding Cell Subsets Observed Between Day 0 and Day 7 in COVID-19

COV-994535 Day 0 vs. 7



COV-994536 Day 0 vs. 7



CM

ICOS⁺⁵ CD27⁺⁴ CD95⁺³ CCR7⁺³ T-bet⁺³ CXCR5⁺²
CD3⁺²

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

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

Naive

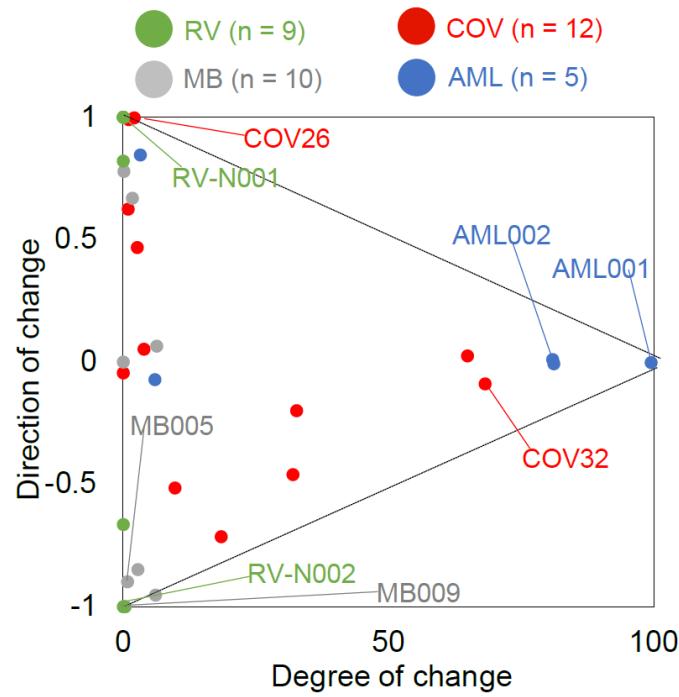
ICOS⁺⁶ T-bet⁺⁵ CD95⁺⁴ CD27⁺⁴ CXCR5⁺² CD3⁺²
CCR7⁺²

CD45RA⁺⁵ CD27⁺⁵ T-bet⁺⁵ ICOS⁺⁴ CD3⁺³ CCR7⁺³
CXCR5⁺²

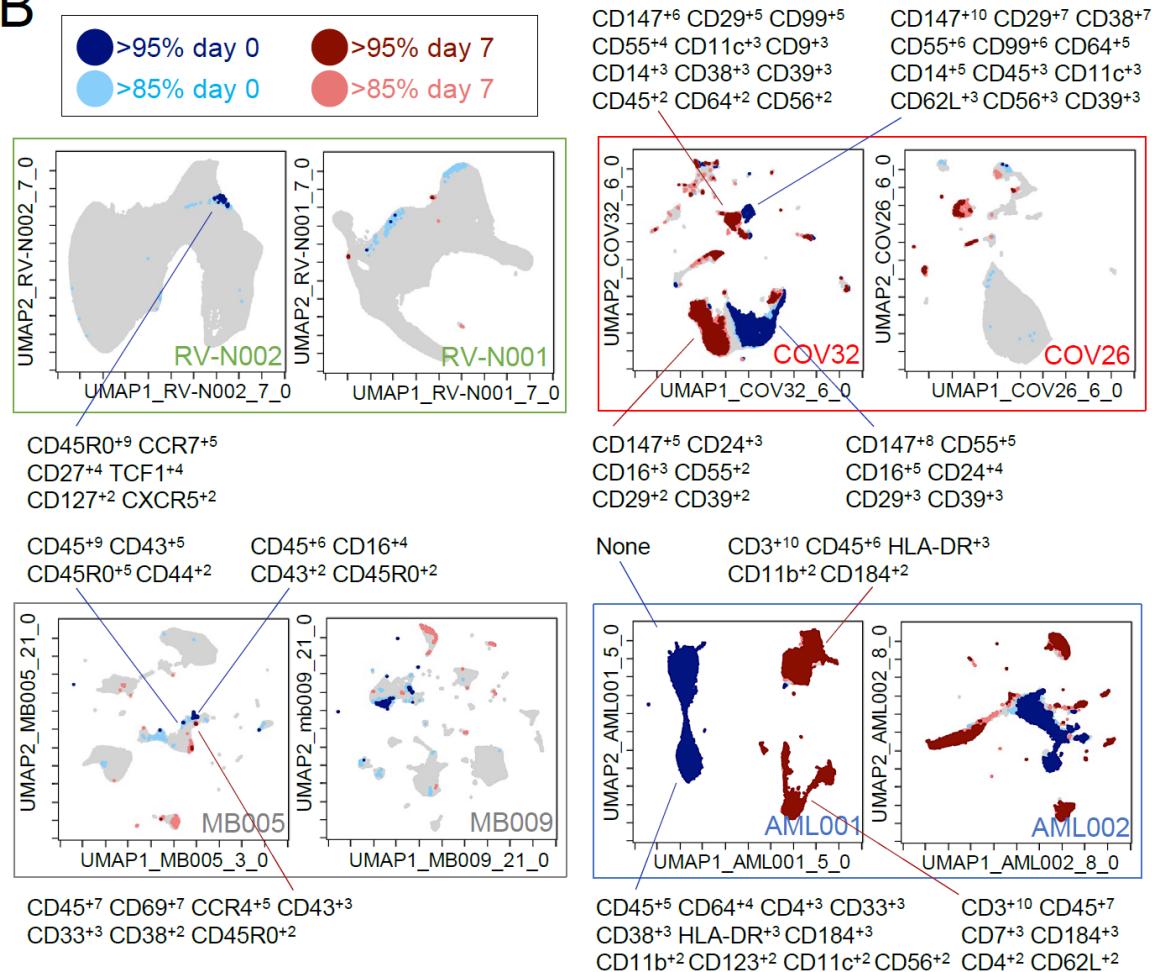
CX3CR1⁺⁶ T-bet⁺⁶ HLA-DR⁺⁶ CD95⁺⁴ CD39⁺⁴
CXCR5⁺³ CD38⁺³

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

A



B

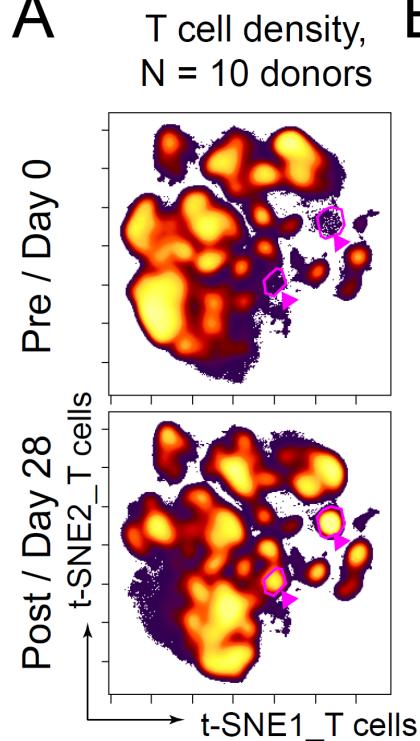


T-REX revealed virus-specific T cells without tetramers
& characterized massive immune changes in COVID-19

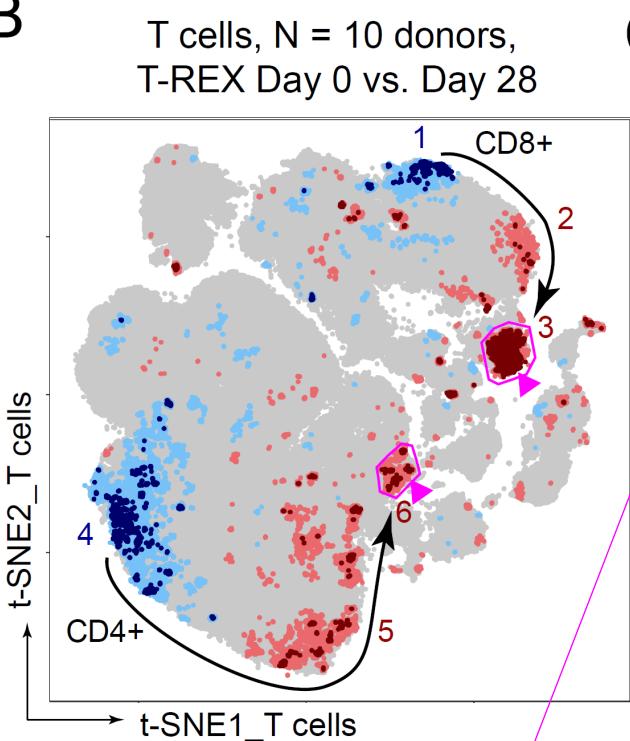
Would it also work to characterize SARS-CoV-2 vaccine response? 

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

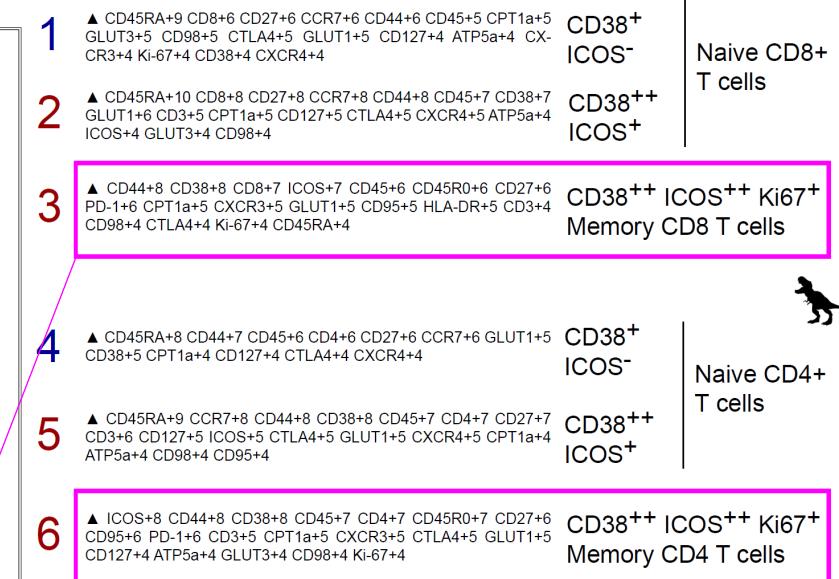
A



B



C



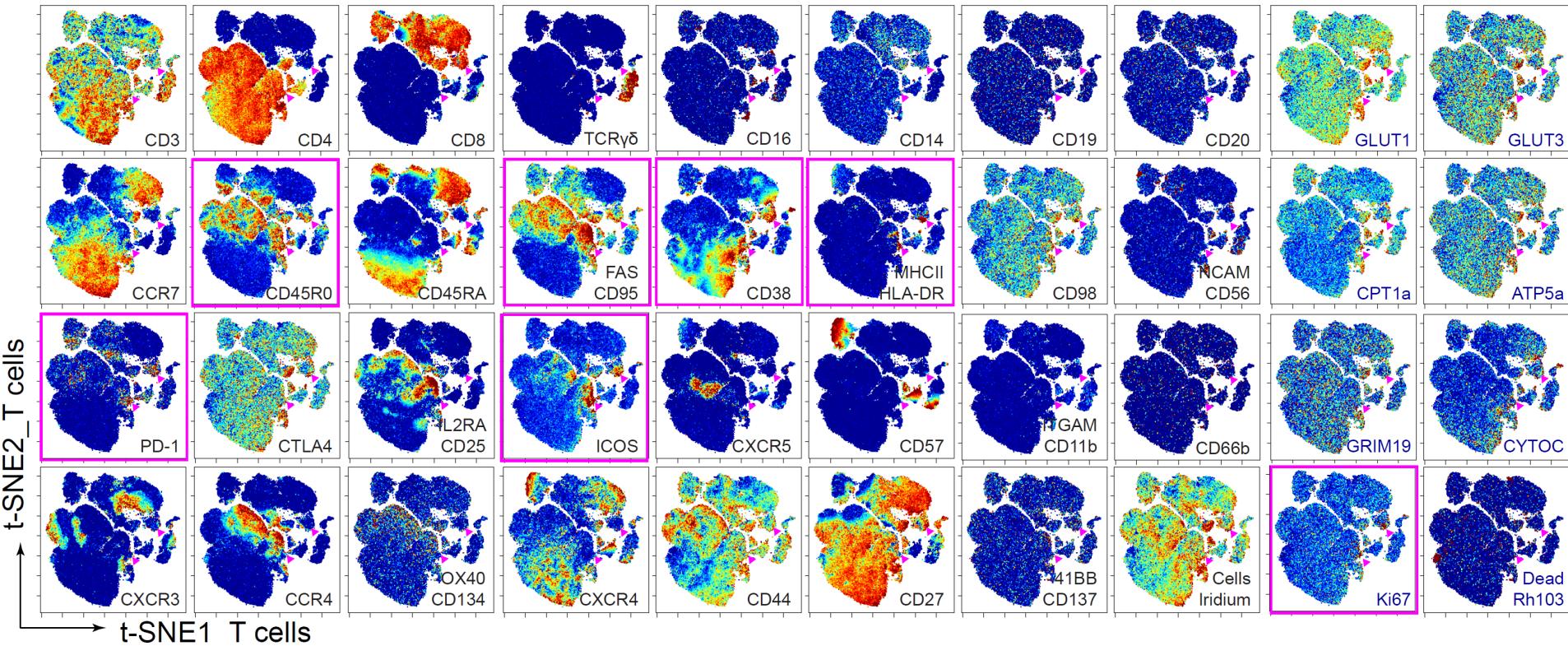
3

▲ CD44+8 CD38+8 CD8+7 ICOS+7 CD45+6 CD45R0+6 CD27+6 PD-1+6 CPT1a+5 CXCR3+5 GLUT1+5 CD95+5 HLA-DR+5 CD3+4 CD98+4 CTLA4+4 Ki-67+4 CD45RA+4

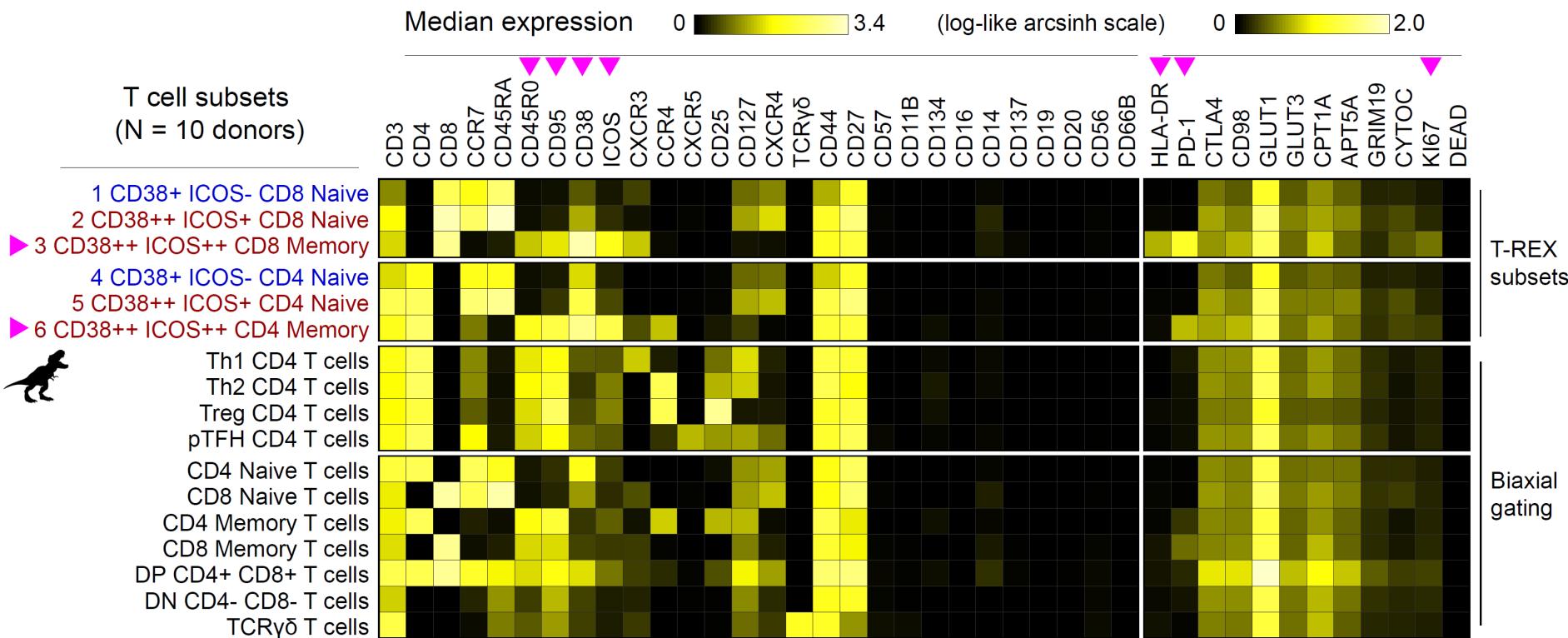
**CD38⁺⁺ ICOS⁺⁺ Ki67⁺
Memory CD8 T cells**

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

T cell mass cytometry panel on merged post-vaccine data (Day 28, N = 10)



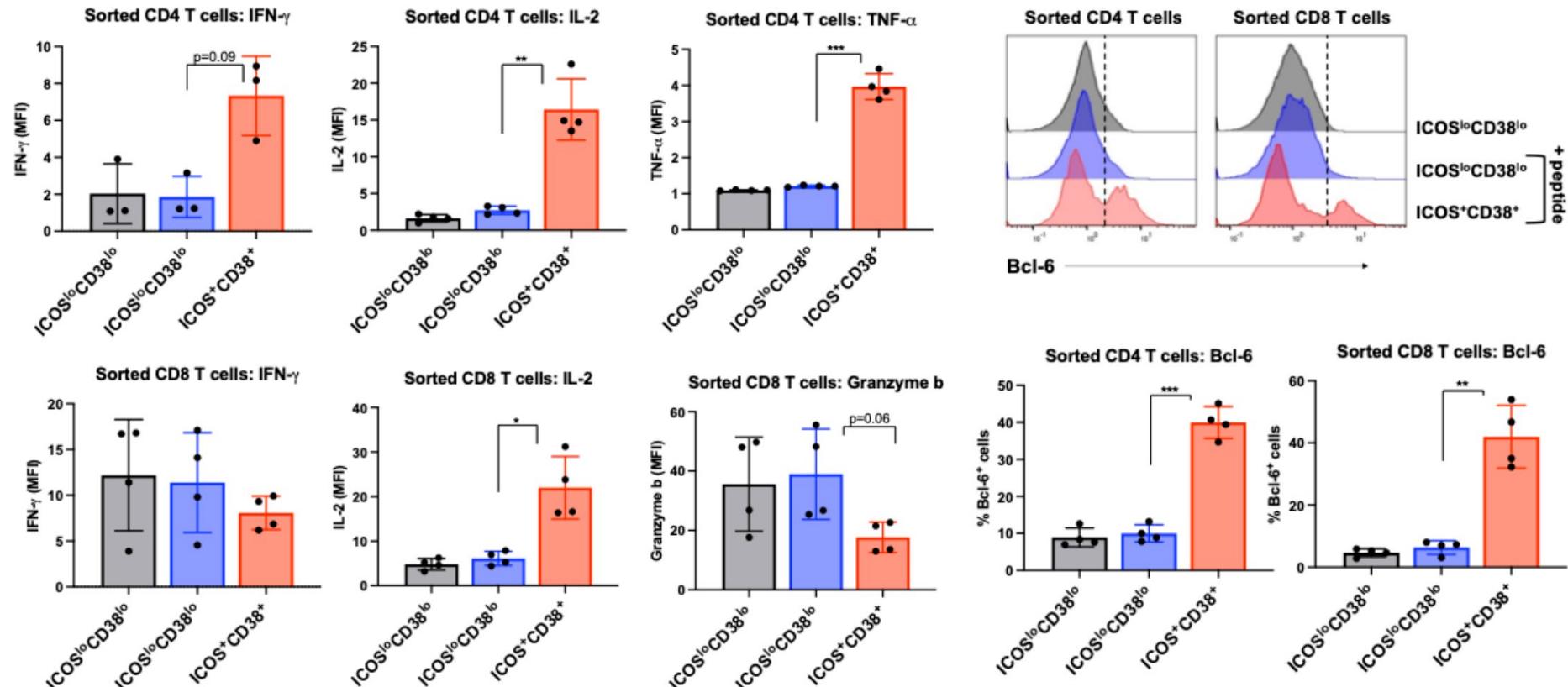
Mass Cytometry Phenotyping of ICOS+ CD38+ PD-1+ CXCR5- Memory CD4 & CD8 T Cells following SARS-CoV-2 Vaccination



Vaccine response: Kramer, Wilfong, Voss et al., *bioRxiv* 2021

T-REX: Barone, Paul, Muehling et al., *eLife* 2021

Sorting T cells on T-REX MEM Phenotype (ICOS⁺⁺ CD38⁺⁺) 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.