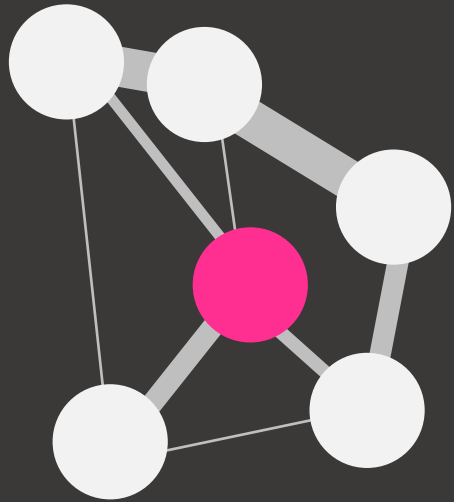




2

TRANSFORMATIVE CONNECTIONS

ONE
CELL
SOCIAL
NETWORK

TWO
REDUCTION
&
VISUALIZATION

100

milliseconds

...is (literally) a blink of an eye

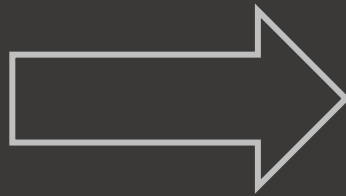


Created by Sewon Park
from Noun Project

...is plenty of time for a decision



Created by W. X. Chee
from Noun Project



Olivola & Todorov J Nonverbal Behavior, 2010

...is plenty of time for advertisement targeting

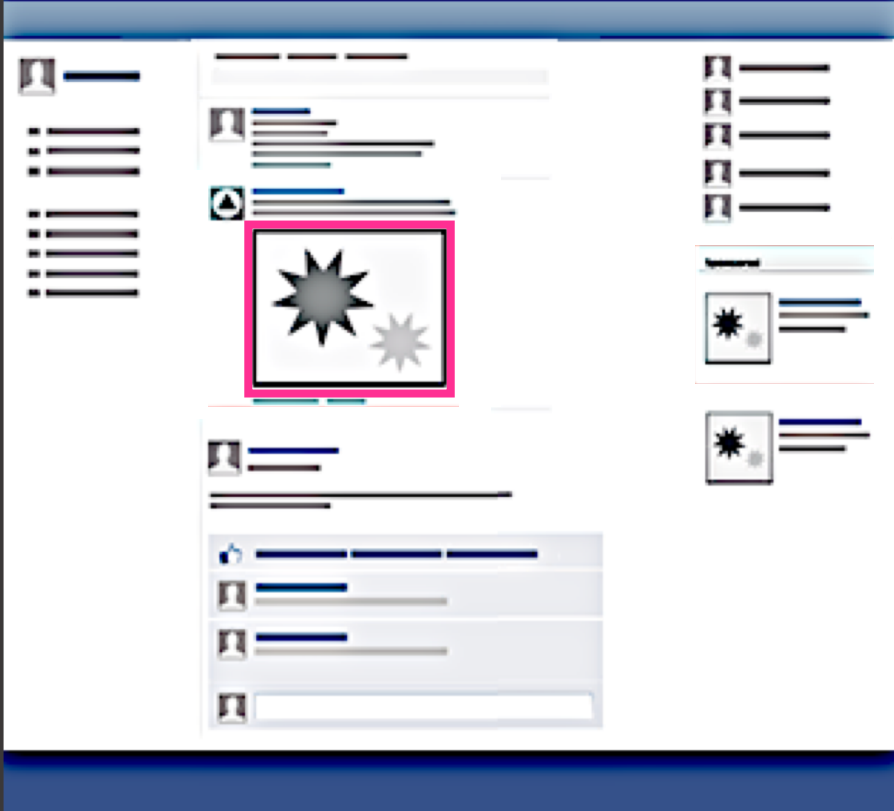


image credit: Plinio Filho

demographic information

+

interests

+

connections

+

friends of connections

"About the delivery system: Ad auctions" Facebook.com
"Most popular Facebook Ad Criteria." socialfresh.com

What are you most likely to buy/click-on right now?

Who are you?

What do you like?

What do you post?

With whom are you connected?

With whom are they connected?

identity

internal
sentiment

network

Output/Goal:

your ad here!

What is a healthy immune system?

What kind of cell are you?

What genes are you expressing?

What proteins are you expressing?

With what cells do you interact?

With what cells do they interact?

**test for a healthy immune system
+ model for perturbations**

WHAT IS A HEALTHY IMMUNE SYSTEM?

Can we use a “near-complete” understanding of immune cell protein and RNA expression to reveal the full interactome and immune cell network in healthy patients?

If we perturb this network in the context of disease, can we determine the effects on the interactions of cells and predictors of disease?

Immune Cell Social Network

**experimental
setup**

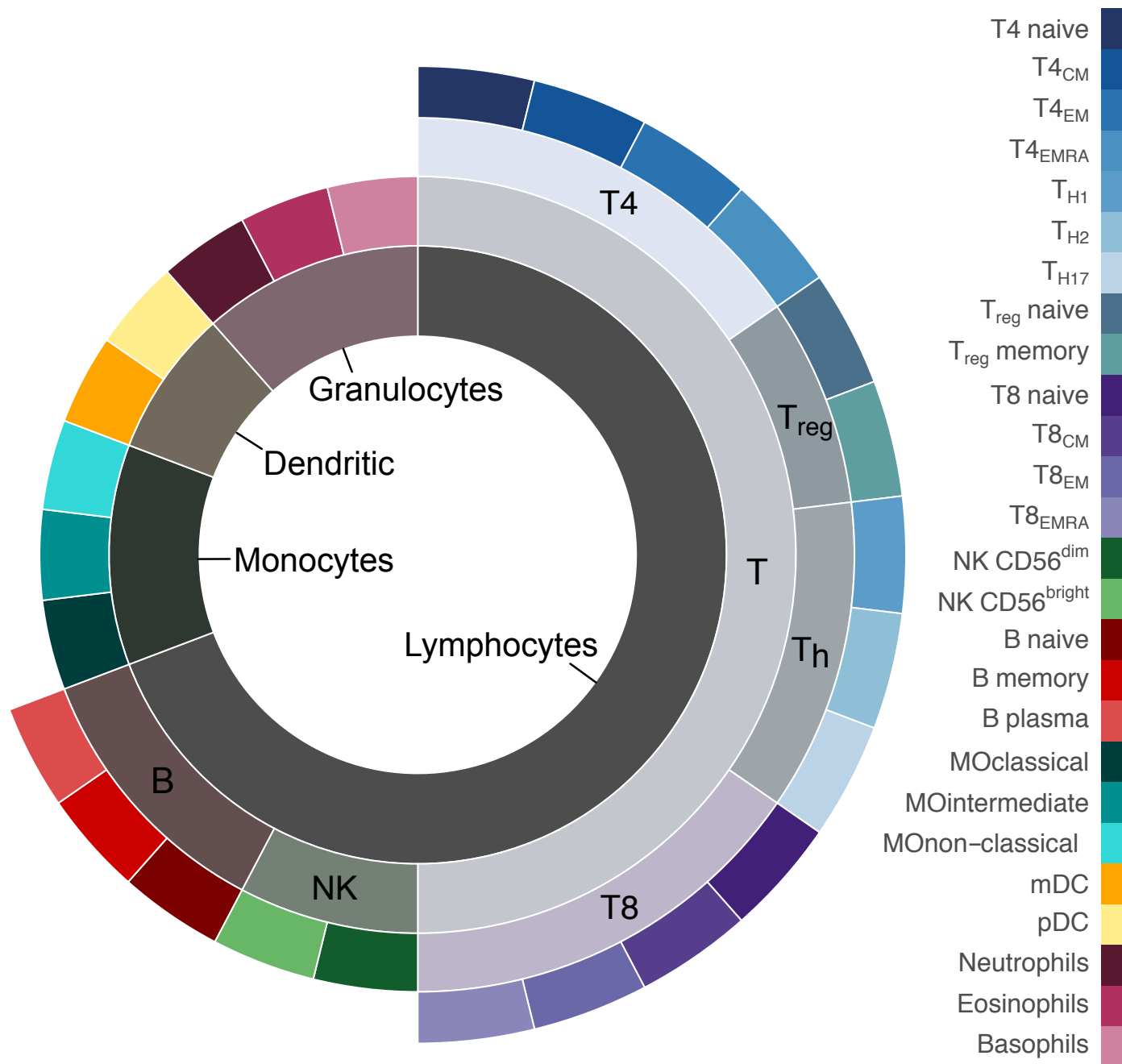
**surface & secreted
interactome
analysis**

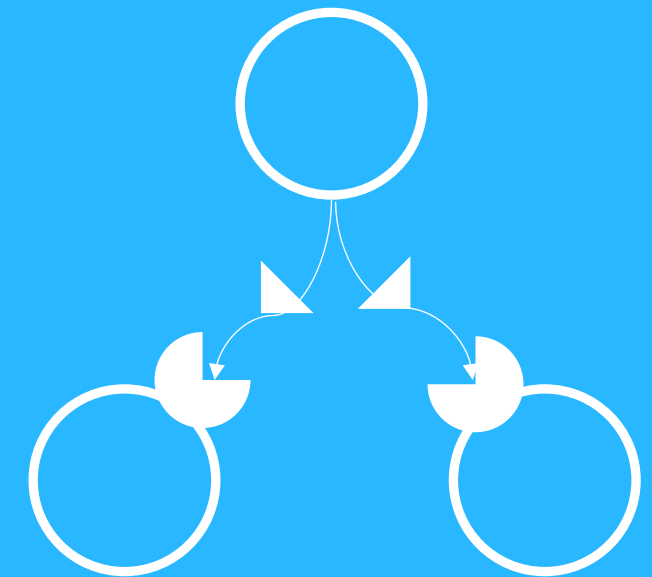
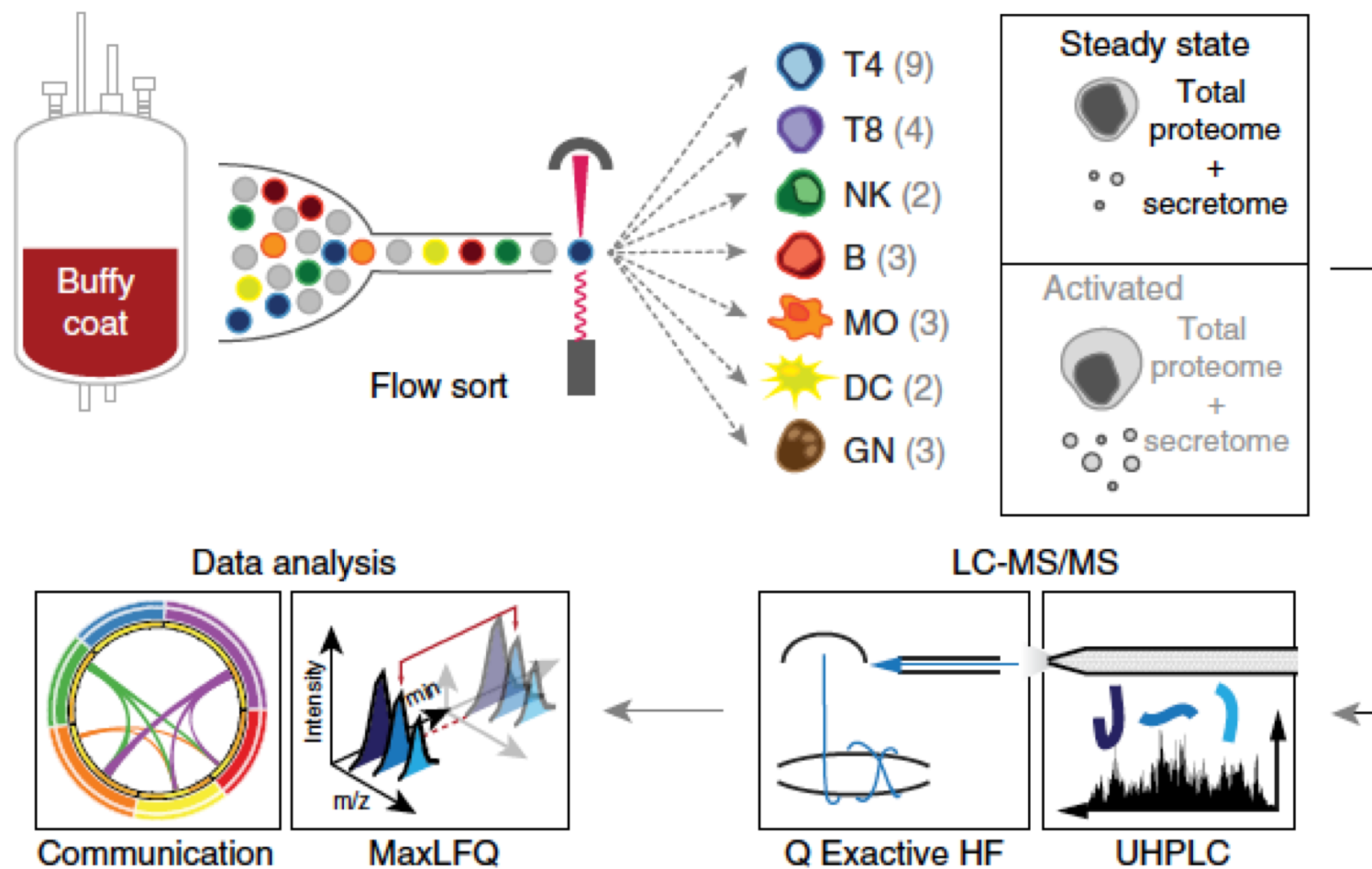
**26 steady state
immune cell models**

**predicting network
changes in disease**



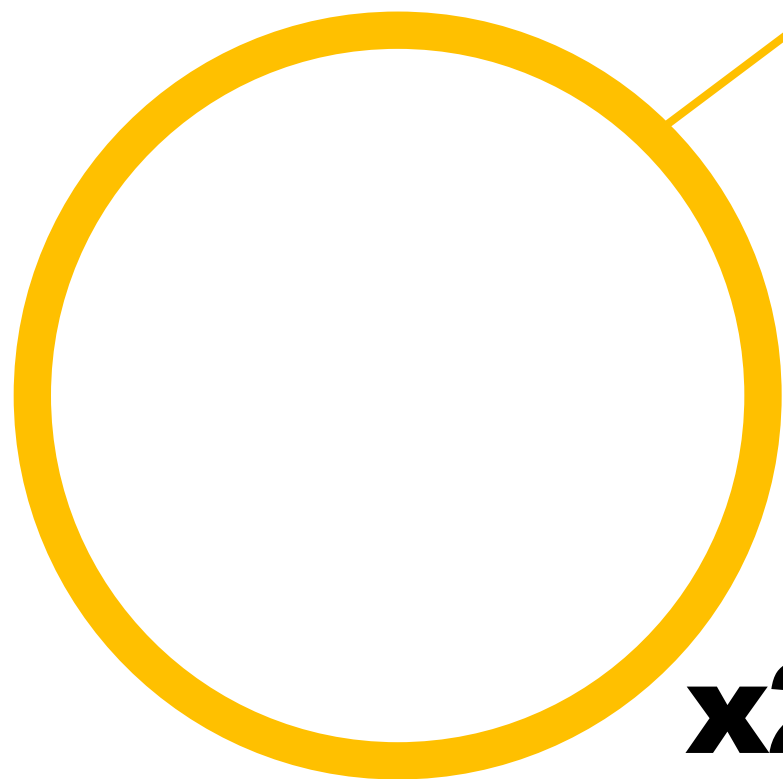
26 IMMUNE CELL SUBSETS





secretome focus:
cytokine &
chemokine-
receptor pairs

(Rieckmann et. al Nature Immunology 2017)



10,520 proteins
n=3-4 healthy donors
bulk sorted (FACSAria)

Focus: secretome

LC-MS/MS (Rieckmann et. al Nature Immunology 2017)

x26

OUR WORK

Focus: entire interactome

=

immune synapse-interactome

+

merge with single cell sequencing

for all 26 subsets

→ examine translation regulation

→ use as input for predictive modeling

cytokine, chemokine-
receptor pairs
(SwissProt & Trembl)

all putative surface
receptors and ligands
(Cell Surface Protein Atlas)

```
graph TD; A[cytokine, chemokine-receptor pairs (SwissProt & Trembl)] --> C[5982 Proteins]; B[all putative surface receptors and ligands (Cell Surface Protein Atlas)] --> C; C --> D[5198 Proteins in MS/MS data]; D --> E[1767 Genes in scRNA-Seq data];
```

5982 Proteins

5198 Proteins
in MS/MS data

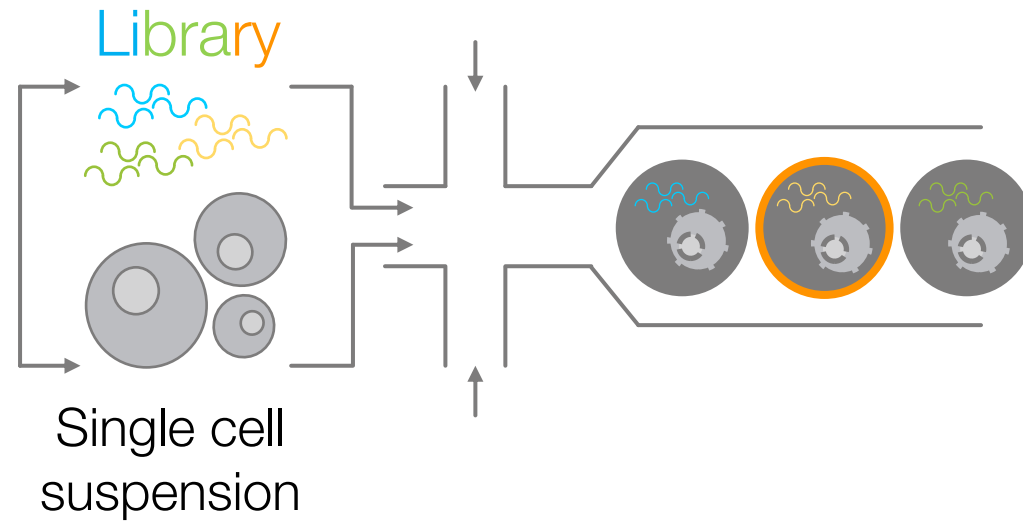
+

1767 Genes in
scRNA-Seq data



How do we get to single cells?

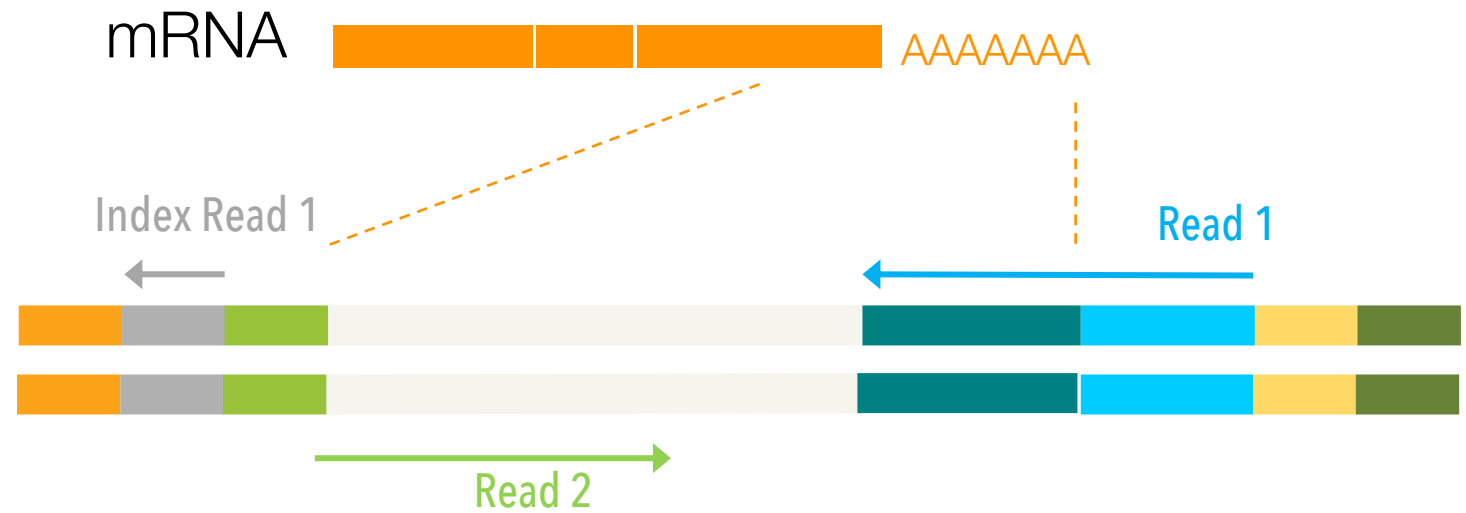
barcoding





Index Read 1: identify sample

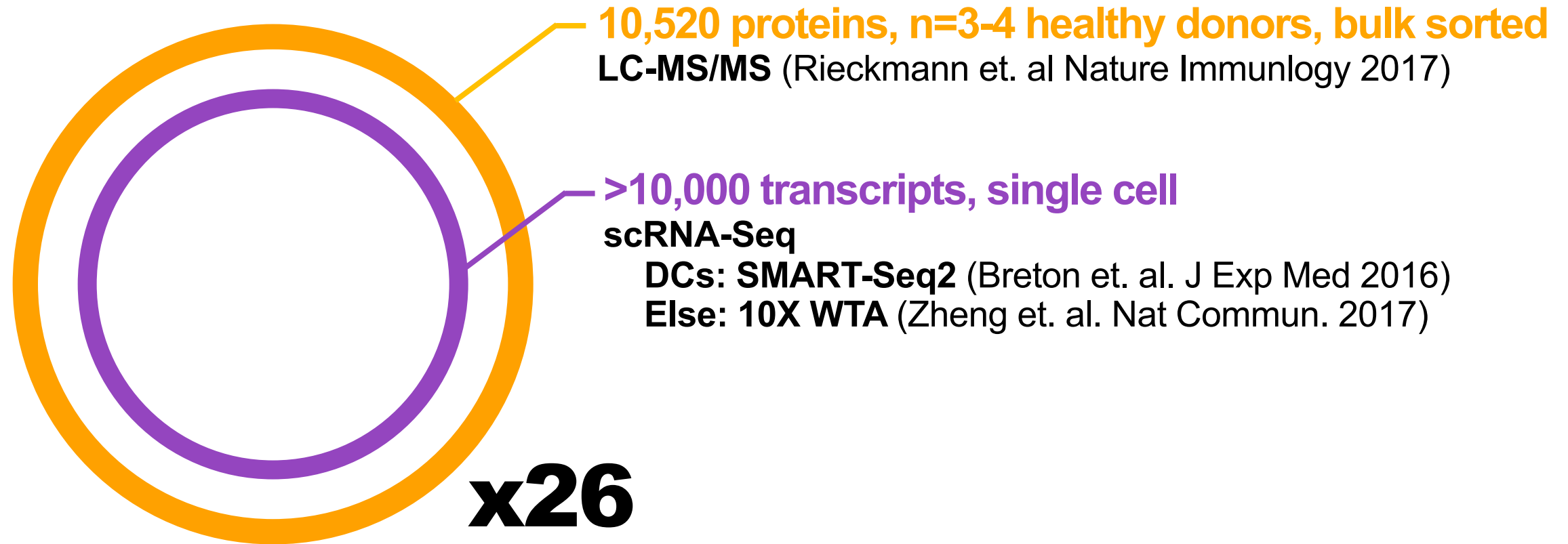
READ 1: identify cell barcode & UMI for each read



READ 2: Alignment

what sample + what cell + what molecule + what gene = expression matrix

OUR WORK



Immune Cell Social Network

**experimental
setup**

**surface & secreted
interactome
analysis**

**26 steady state
immune cell models**

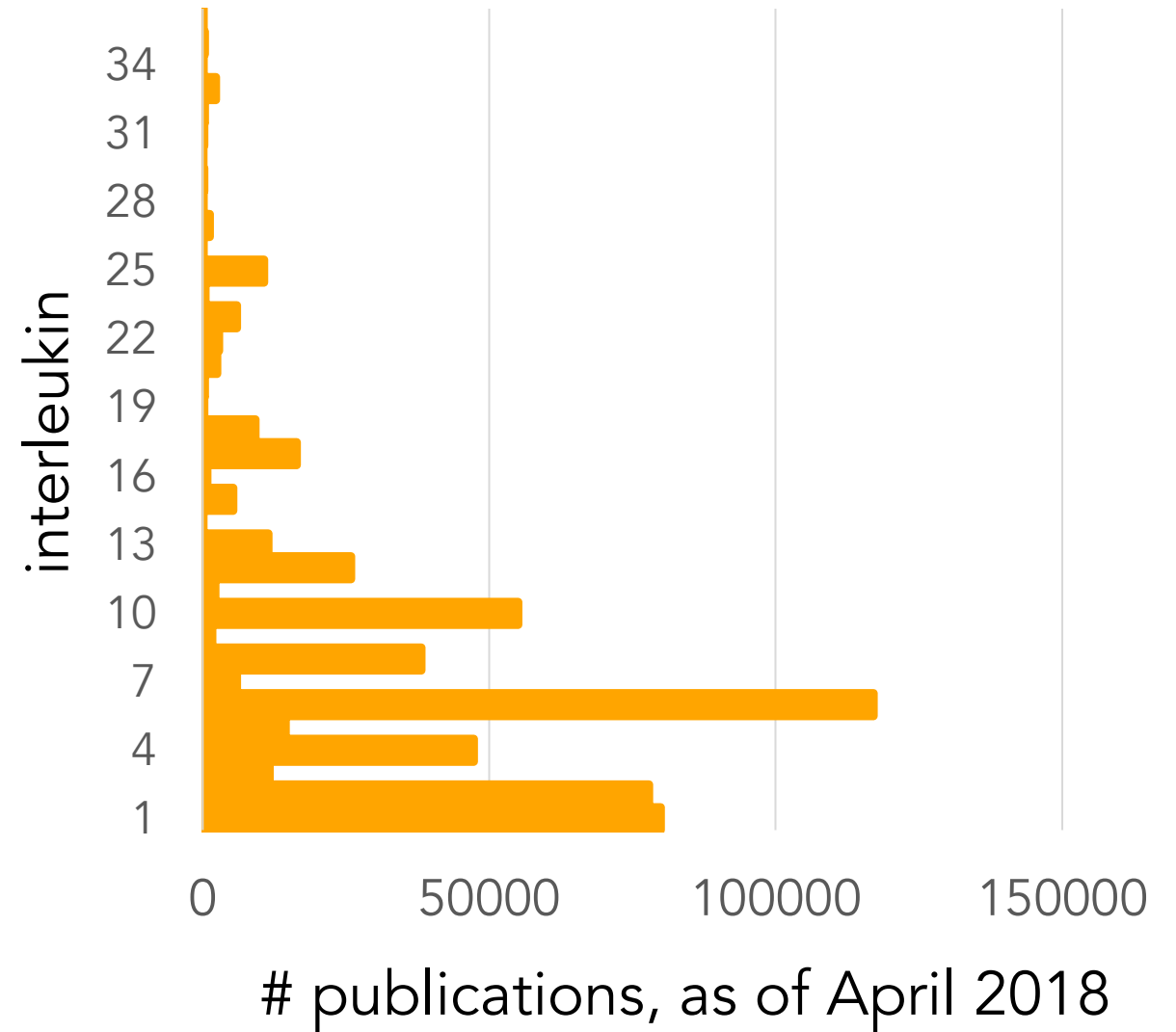
**predicting network
changes in disease**



DEFINING THE INTERACTOME

Database	Collected interactions	Reference
IntAct	1681	Hermjakob et. al.. Nucl. Acids. Res. 2004
DIP	286	Xenarios et. al Nucl. Acids. Res. 2000
String	14,467	Szklarczyk et. al. Nucleic Acids Res 2015.
15,938 interactions		*all with experimental evidence interact with at least receptor score >0.4 or >0.9 combined score

MUCH TO LEARN



Immune Cell Social Network

**experimental
setup**

**surface & secreted
interactome
analysis**

**26 steady state
immune cell models**

**predicting network
changes in disease**



CHALLENGES

Immune cells have writer's block:

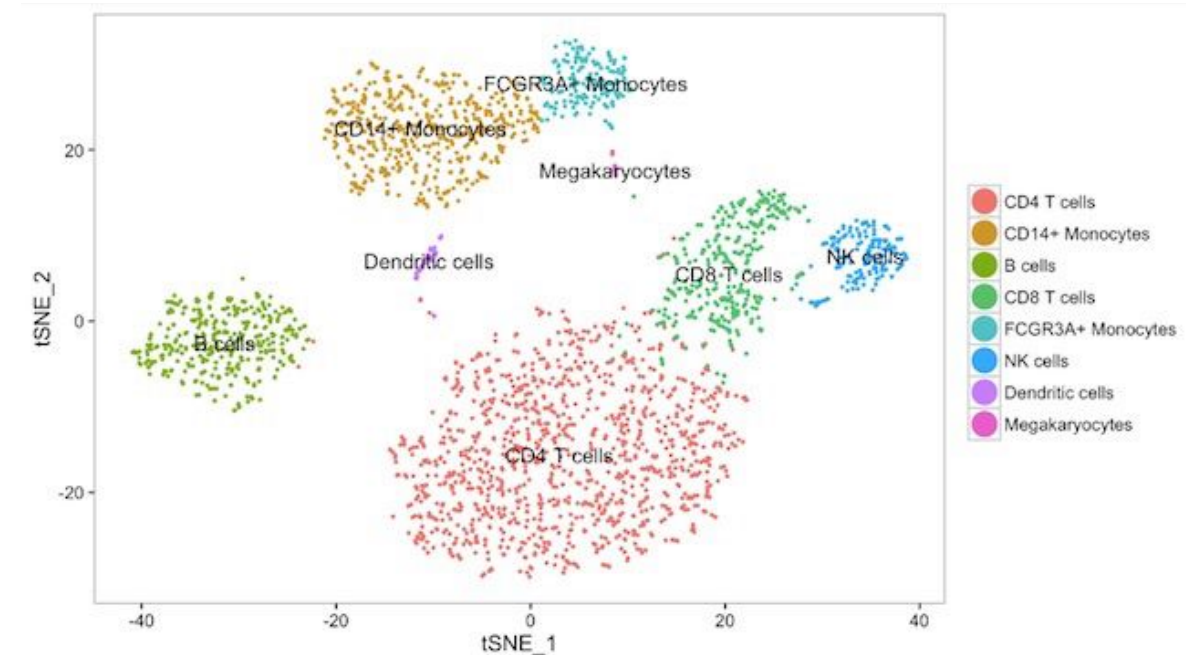
- immune cells express very low copies of RNA
- mRNA and protein expression levels not correlated (Tian et. al. 2004, de Sousa Abreu et. al. 2009; Maier et. al. FEBS Letters 2009 Vogel et. al. 2010, Lundberg et. al. 2010, Schwänhauser et. al. 2011)
- mRNA levels explain only around 40% of the variability in protein levels (Schwänhauser et. al. Nature 2011)

State of field:

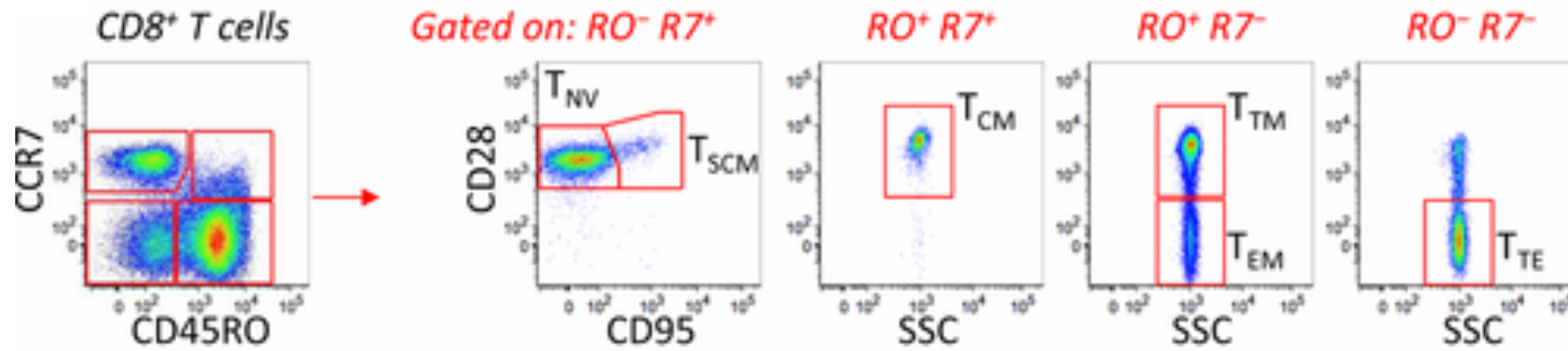
- only 2-3 differentially expressed genes used per immune cell subset for identification (Seurat)
- No link to “gold standard” immunophenotyping

Result:

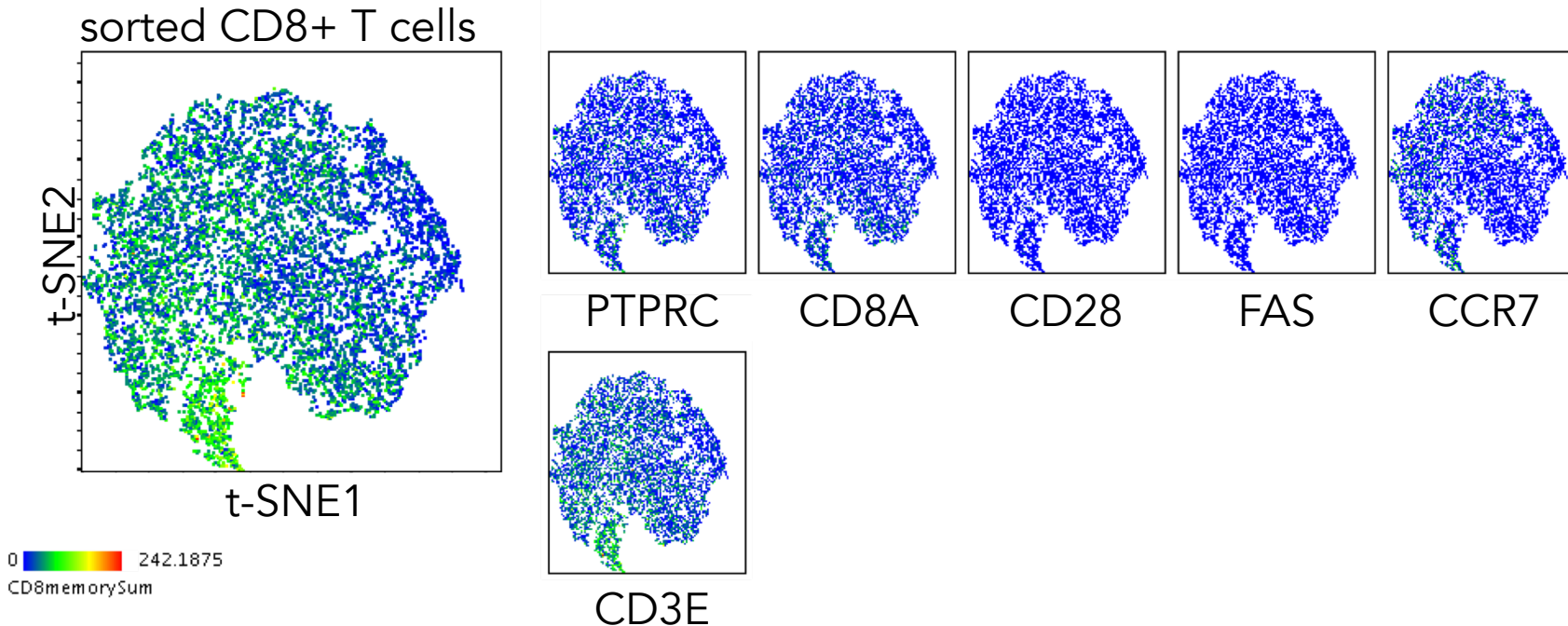
no canonical, validated gene sets for immune cell populations



SUBSETTING

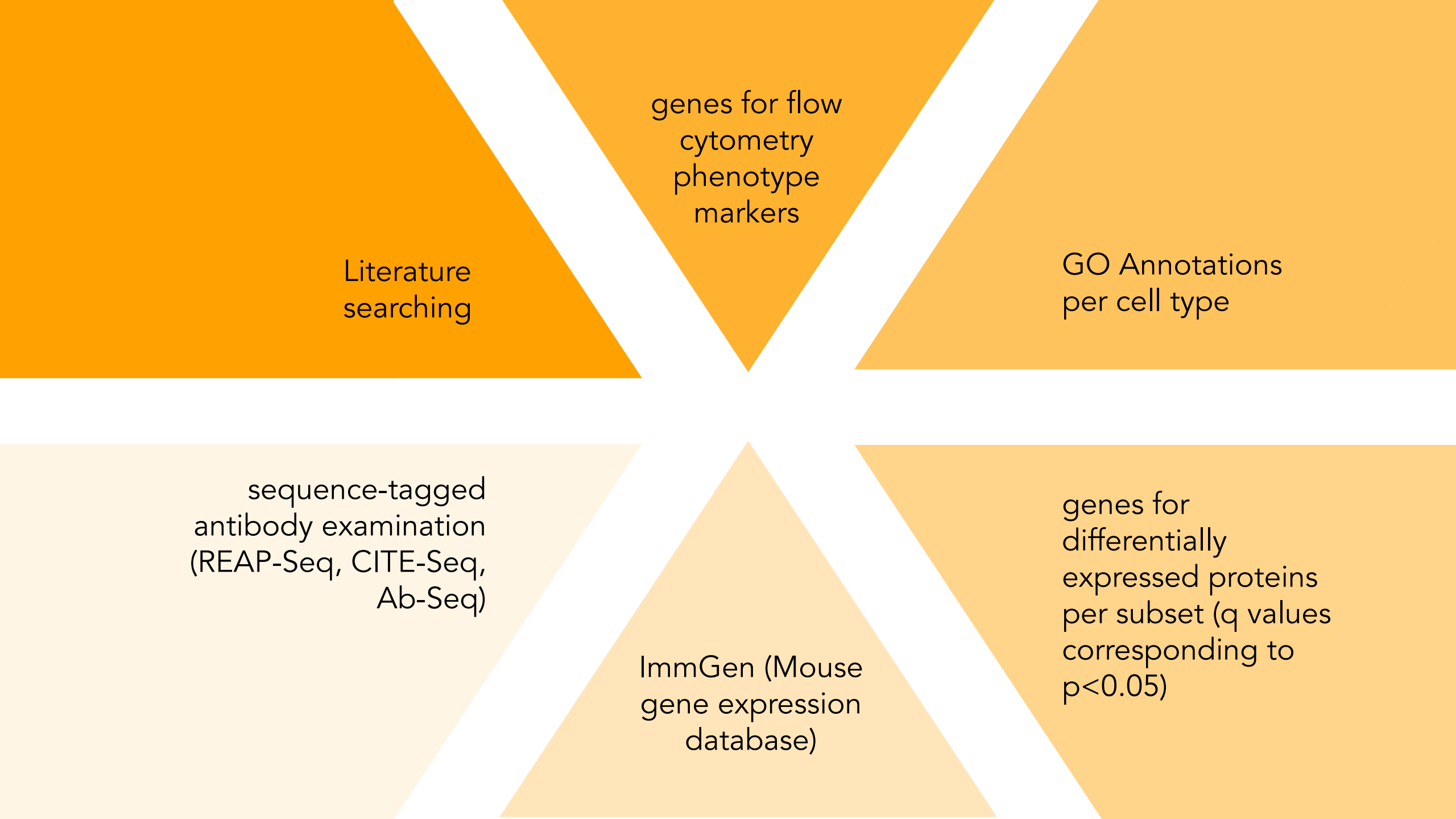


Mahnke et. al. EJI 2013.



challenge: cannot use canonical surface markers (also genes for cytokines, TFs and their network) on WTA

t-SNE ...not useful



genes for flow
cytometry
phenotype
markers

GO Annotations
per cell type

genes for
differentially
expressed proteins
per subset (q values
corresponding to
 $p < 0.05$)

ImmGen (Mouse
gene expression
database)

sequence-tagged
antibody examination
(REAP-Seq, CITE-Seq,
Ab-Seq)

Literature
searching

genes for flow
cytometry
phenotype
markers

Literature
searching

GO Annotations
per cell type

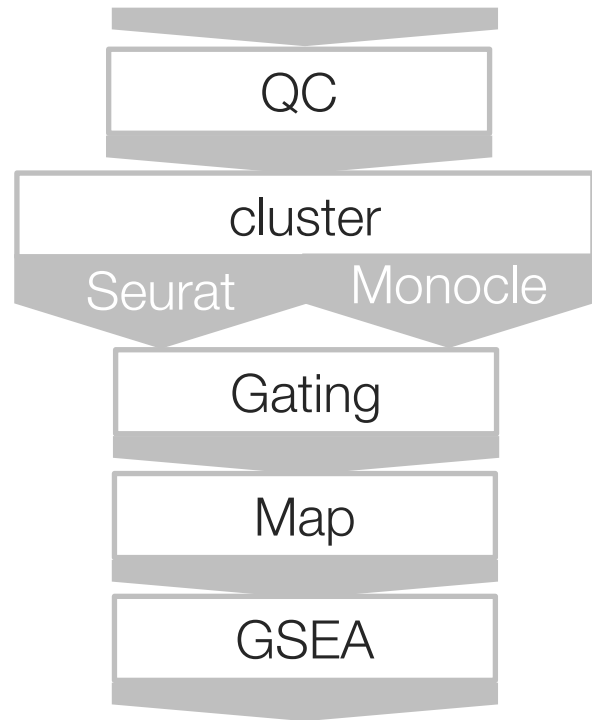
sequence-tagged
antibody
examination

ImmGen (Mouse
gene expression
database)

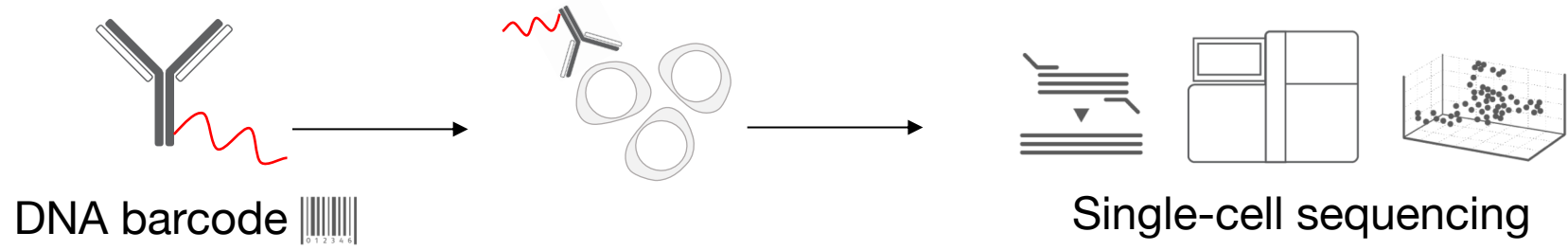
genes for
differentially
expressed proteins
per subset (q values
corresponding to
 $p < 0.05$)

SUBSETTING

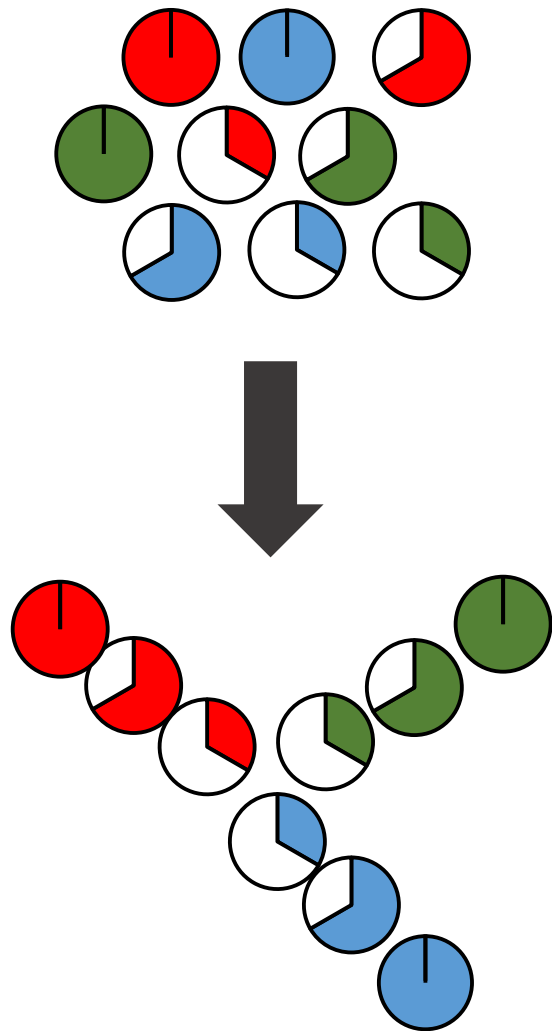
Ab-Seq & RNA-Seq



GSEA & surface “validated” gene sets
+
subsets for downstream analysis



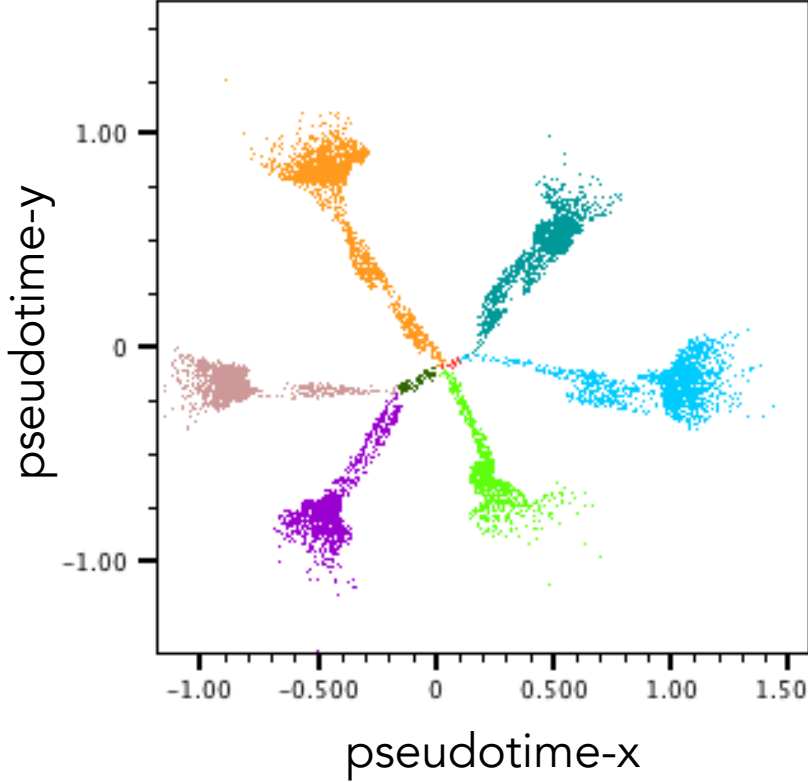
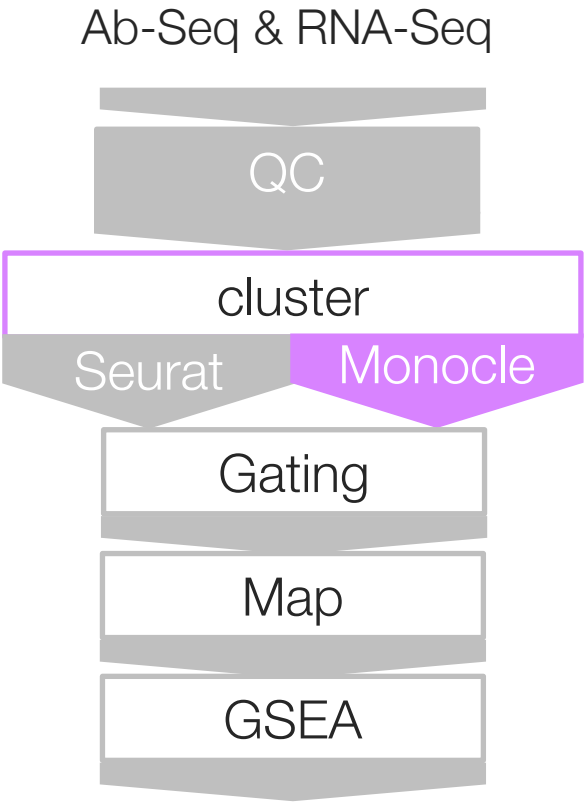
adapted from Chang et. al. ASHG 2017 & Science 2017.



Monocle

- Input: population & gene set
- builds single cell trajectory by comparing cells and their asynchronous progression through biological process
- Produces:
 - ordered cells by learning explicit principal graph (Reversed Graph Embedding)
 - clusters cells into predicted “States” of progression.

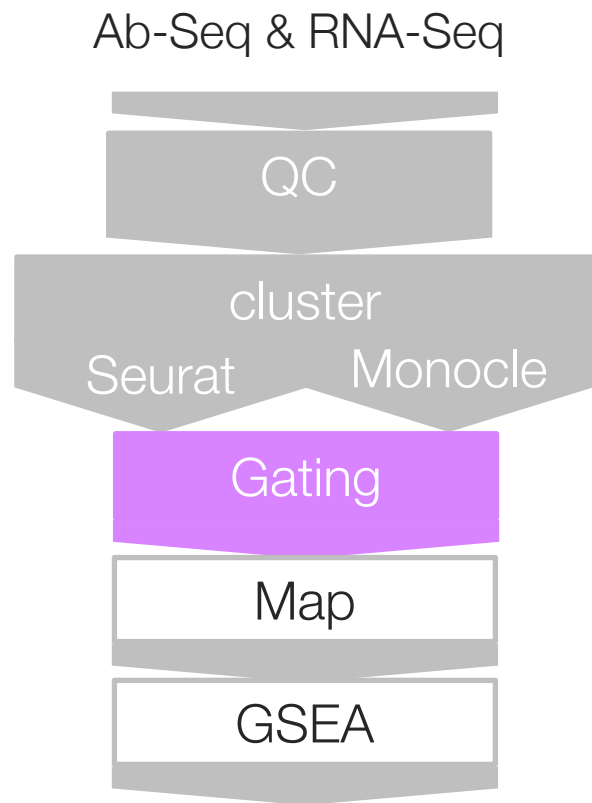
MONOCLE CLUSTERING



Cluster	Count	Freq. of Parent
1	1718	18.4
2	18	0.2
3	2097	22.5
4	1040	11.2
5	77	0.8
6	1356	14.5
7	1377	14.8
8	1644	17.6

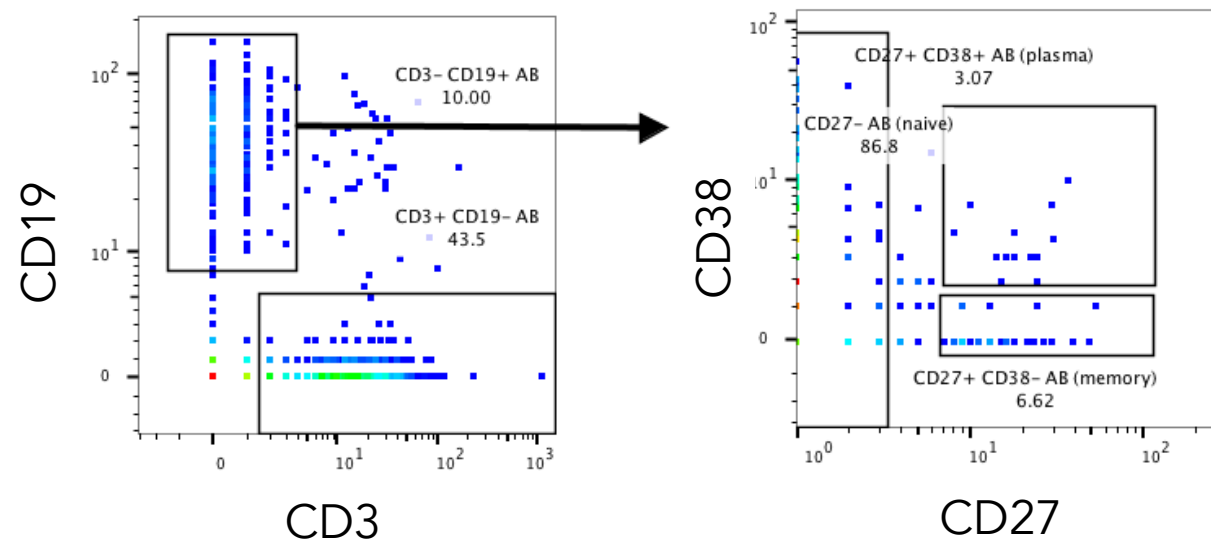
GSEA & surface “validated” gene sets
+
subsets for downstream analysis

GATING

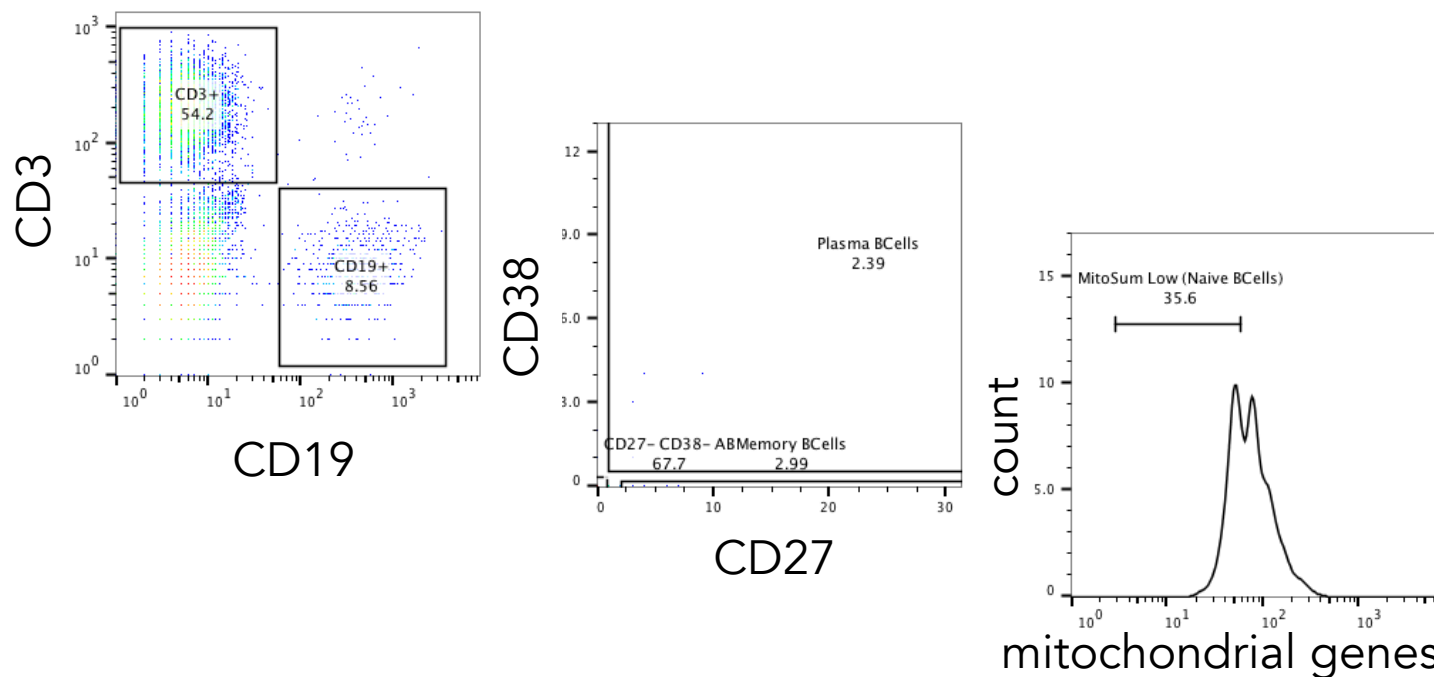


GSEA & surface “validated” gene sets
+
subsets for downstream analysis

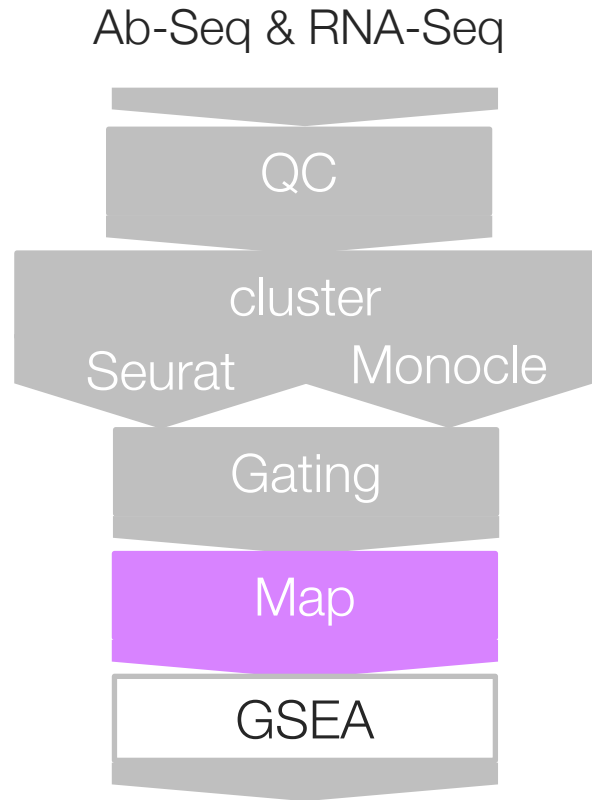
AbSeq



CITE-Seq

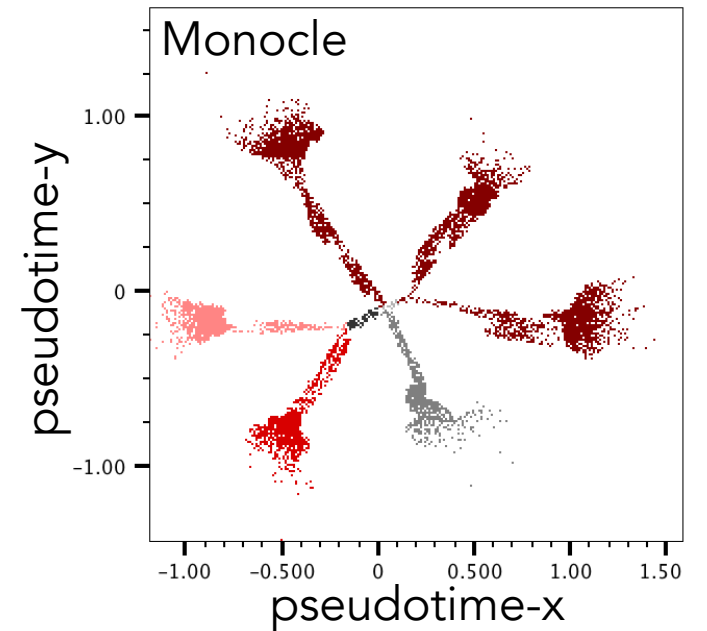
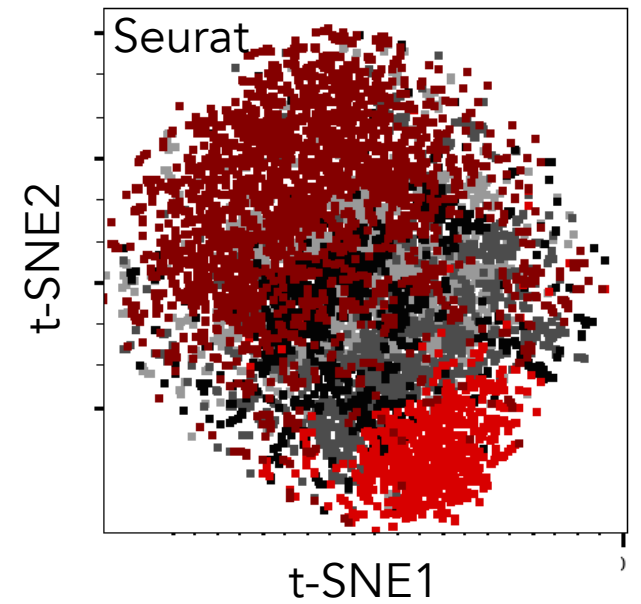
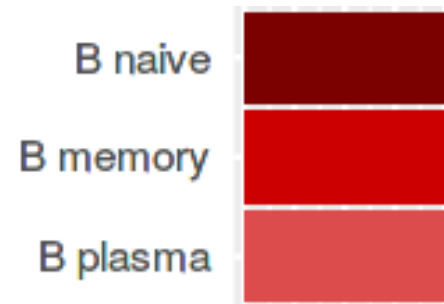


MAP

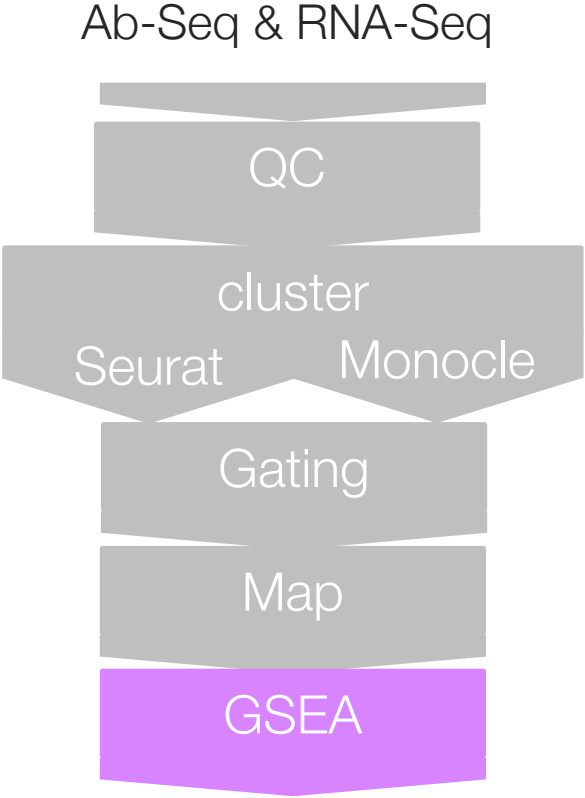


GSEA & surface “validated” gene sets
+
subsets for downstream analysis

from AbSeq
methods:



GSEA



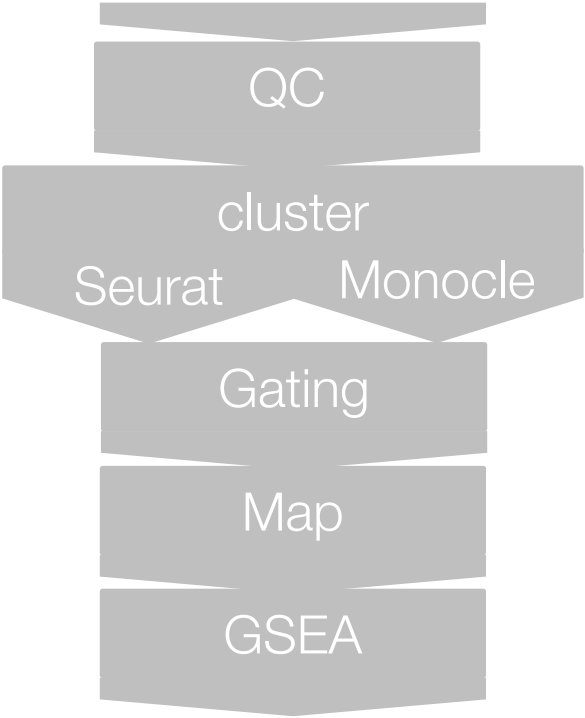
Method	Cluster(s)	Gene Set	Adjusted p-value	Common Genes	Out of
Seurat	0, 1, 4	Naïve B	<0.05	3	17
Monocle	3	Naïve B	7×10^{-6}	3	17
Monocle	6	Plasma B	2×10^{-7}	14	92
Monocle	7	Memory B	5×10^{-6}	4	8

Fisher’s Exact Test

GSEA & surface “validated” gene sets
+
subsets for downstream analysis

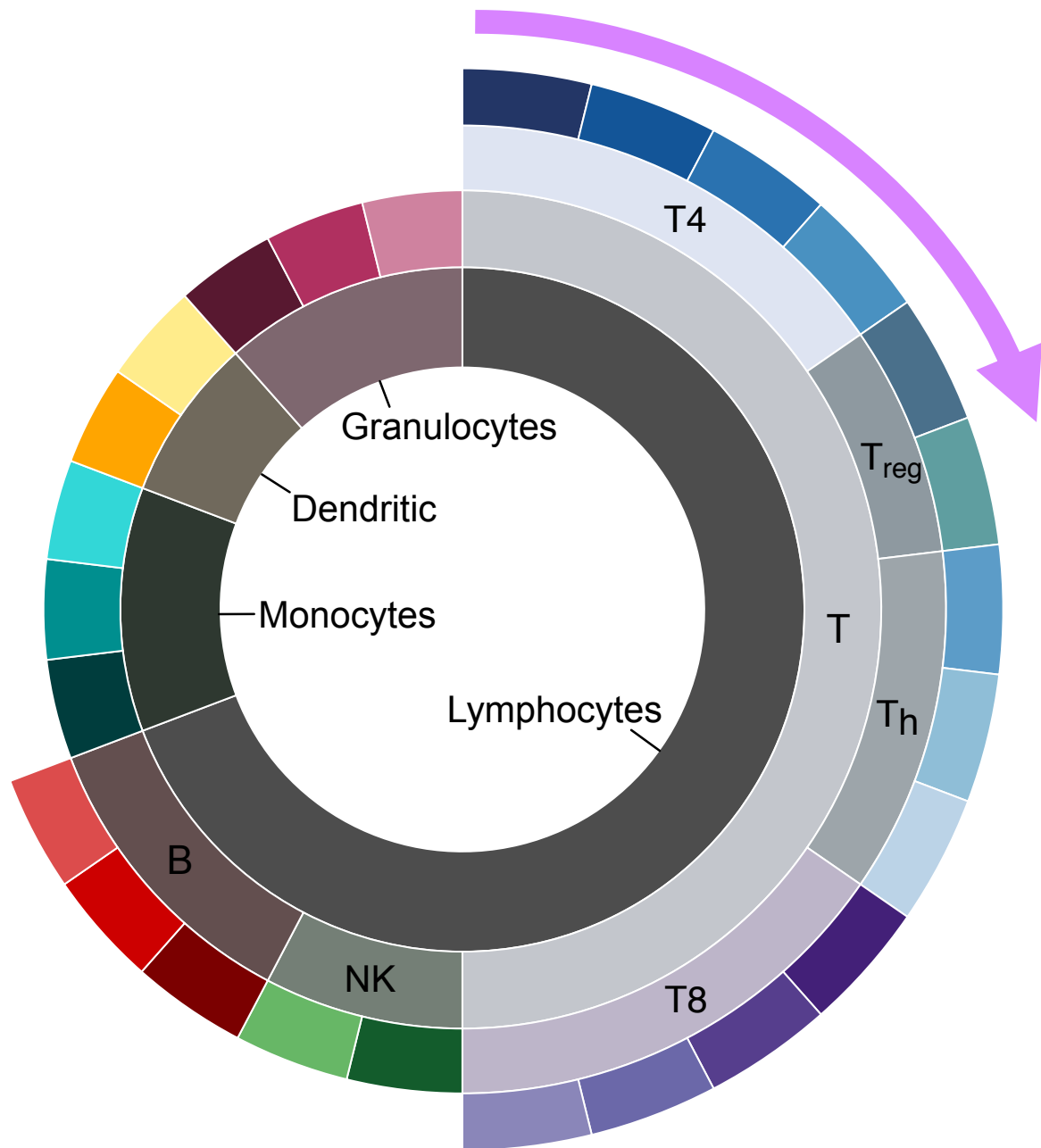
AbSeq & GSEA VALIDATED

Ab-Seq & RNA-Seq

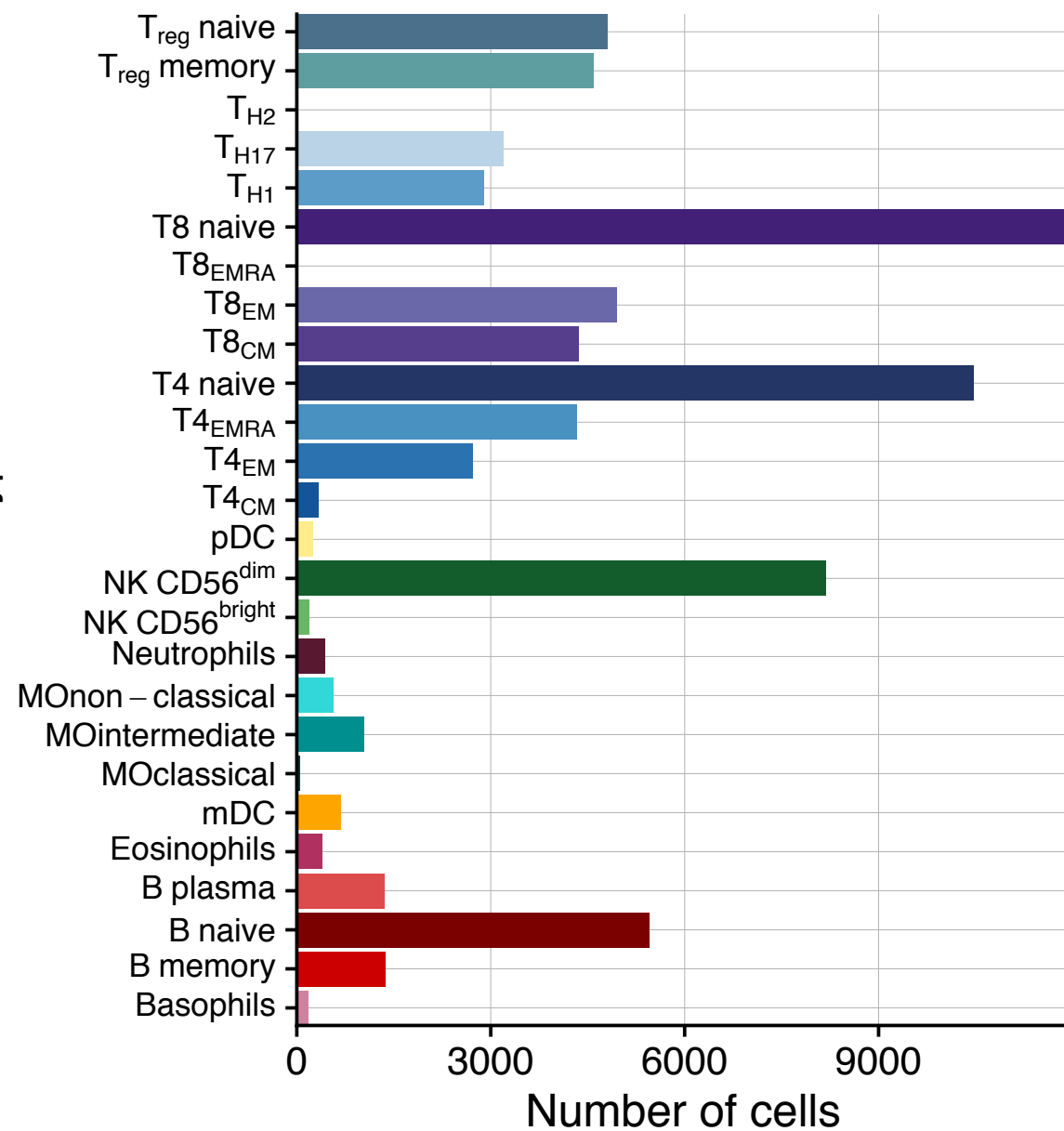


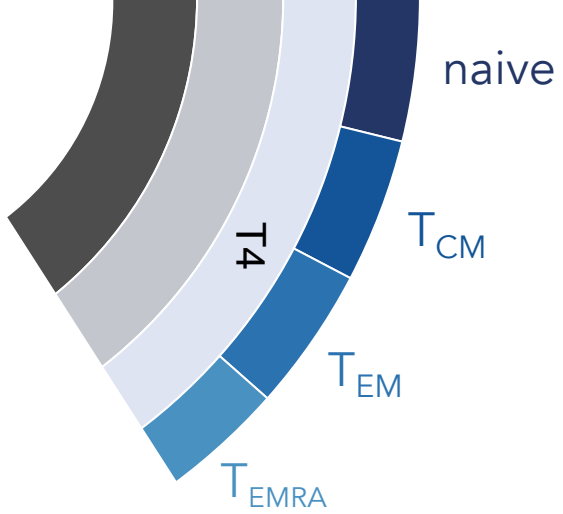
GSEA & surface “validated” gene sets
+
subsets for downstream analysis

Naïve B	Plasma B				Memory B
BTG1	FKBP11	TMED9	PCNA	XBP1	TMEM55B
CD72	UFD1L	VBP1	RGS14	SDF2L1	WARS
IL4R	COX7B	CD300A	RPN2	APOBEC3G	ITGAM
TNFRSF13B	GPAA1	SSR3	DCTN3	STAT1	NME3
CXCR4	TRADD	CNPY2	NEDD8	MCRS1	
YBX3	MRPS12	COPZ1	ERLEC1	PRDX6	
IGHM_membrane	IGHA1_secreted	DDOST	OS9	MANF	
FCER2	ITGB2	PDIA4	CHST12	CD2	
HSH2D	NUDT1	WDR33	ASNA1	COPS6	
MARCKS	AQP3	TPST2	LMAN2	PSMC3	
TCL1A	JCHAIN	NDUFS7	EMC6	NDUFAB1	
IGHD_membrane	LGALS3	CRELD2	ZC3H13	EIF3G	
PIK3IP1	CCND2	ERGIC2	IGHA2	RNF181	
SSPN	BSG	VAMP5	HMOX2	BAX	
DUS1L	SSSCA1	CD44	S100A10	ITM2A	
CD69	CYTIP	NAPRT1	TIMM8B	NUCB2	
CD200	CTSD	SLBP	DARS	ADA	



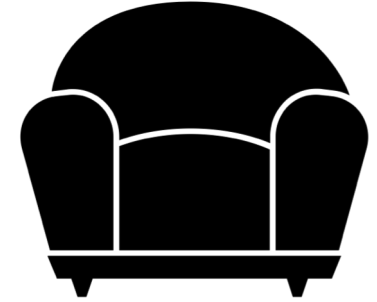
Cell type



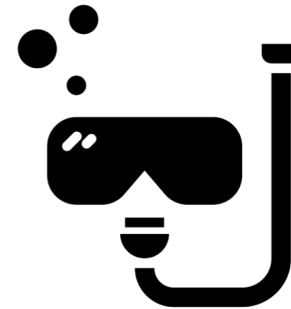


Subset

sort gating for
LC-MS/MS



Created by Les vieux garçons
from Noun Project



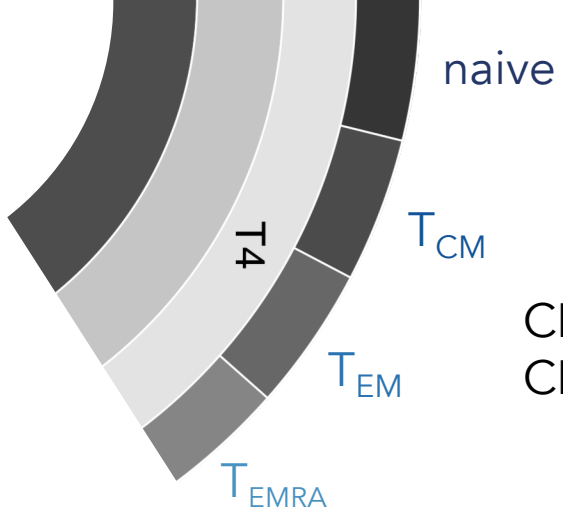
Created by anbilero adaleru
from Noun Project

identifying genes
(subsetting as
described)



Created by Nathan Newell Griffith
from Noun Project

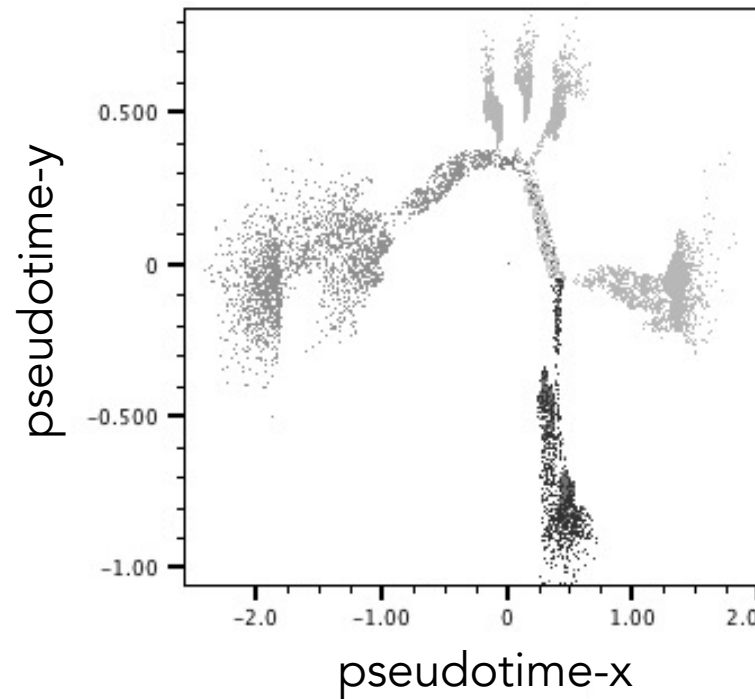
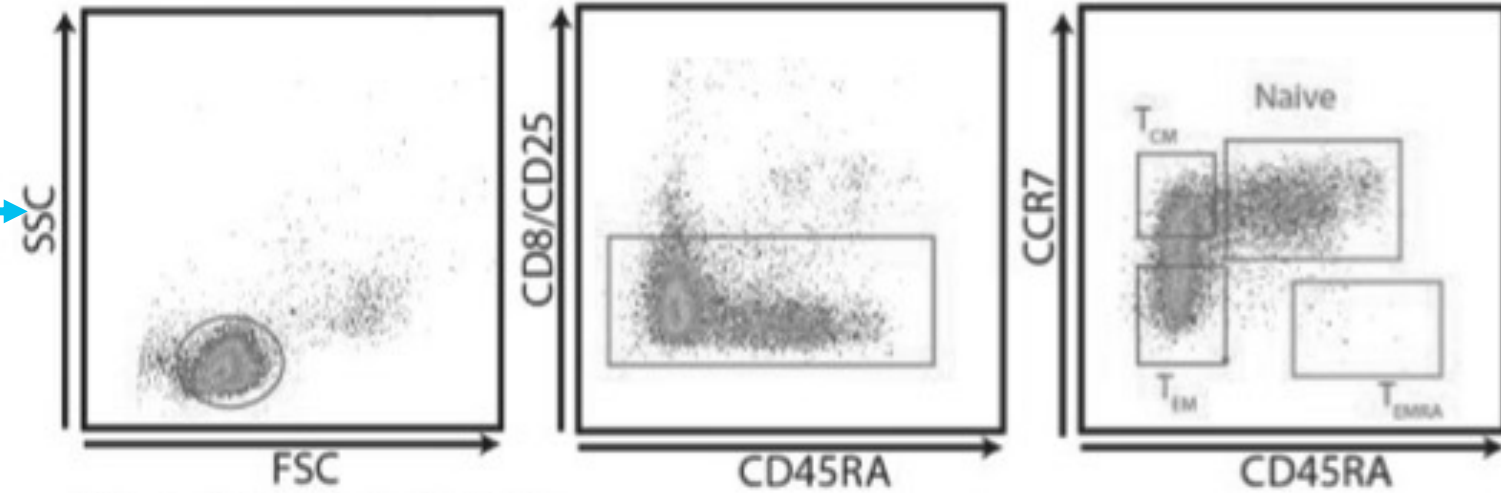
differential gene
expression for each
subset v. all cells



CD4+ enriched,
CD8/25/19-

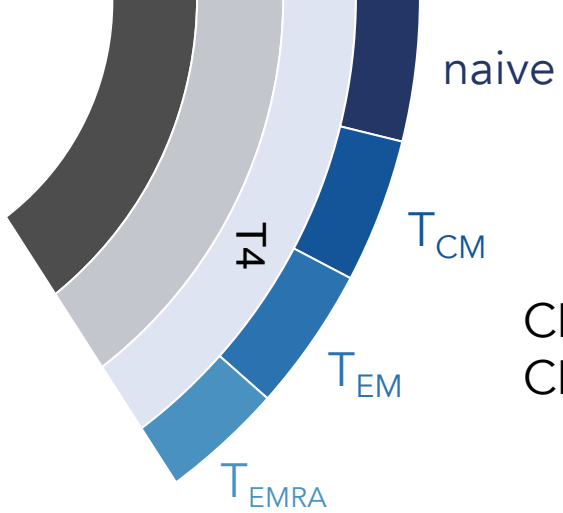


CD4+ T Memory

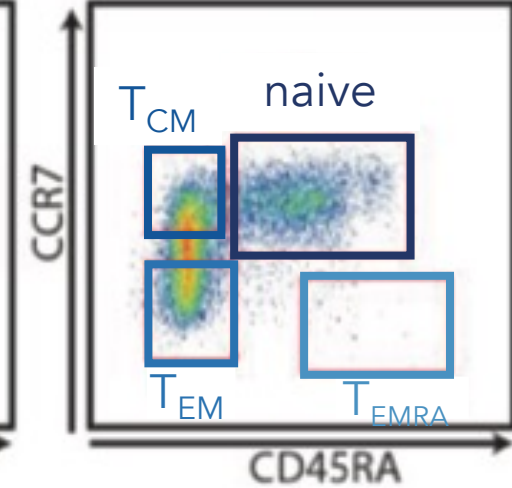
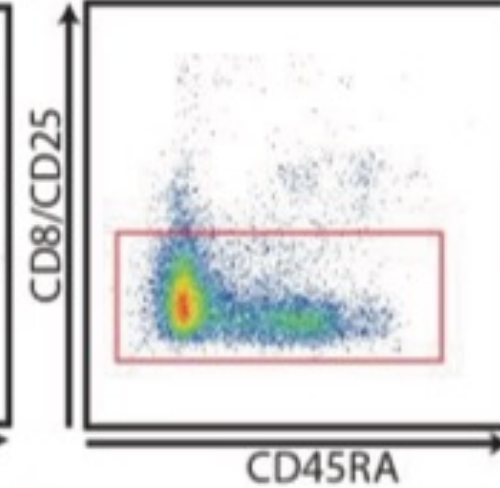
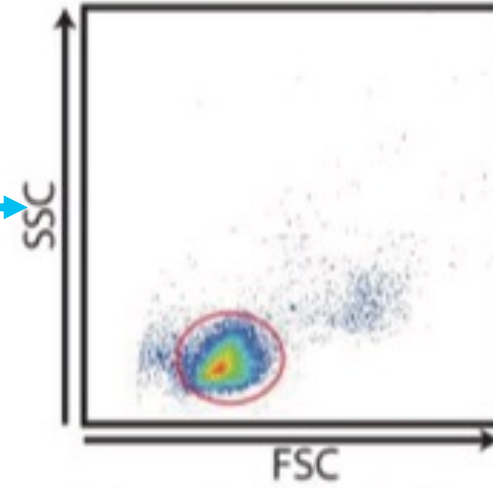


for research purposes only, not for use in diagnostic procedures

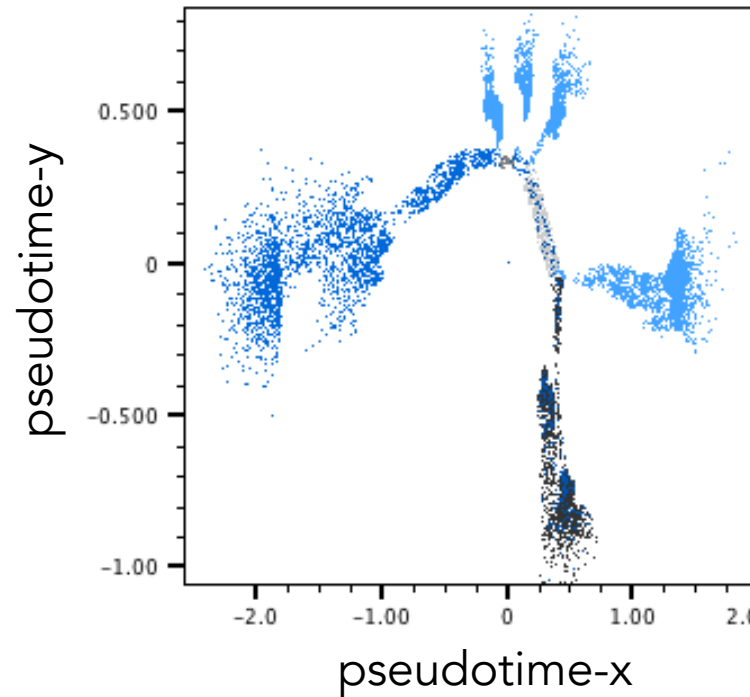
flow: Rieckmann et. al Nature Immunology 2017
single cell RNA-Seq: Team FlowJo



CD4+ enriched,
CD8/25/19-



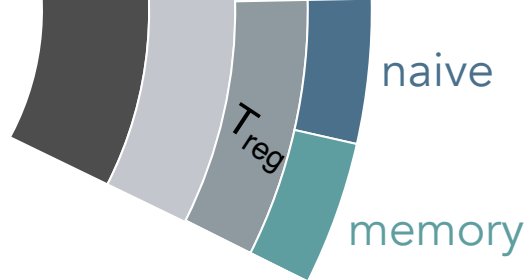
naïve sorted
separately



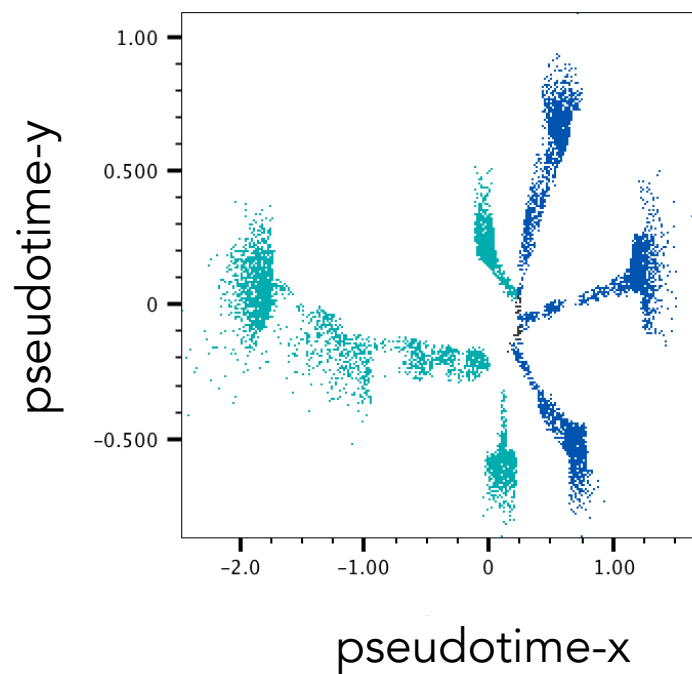
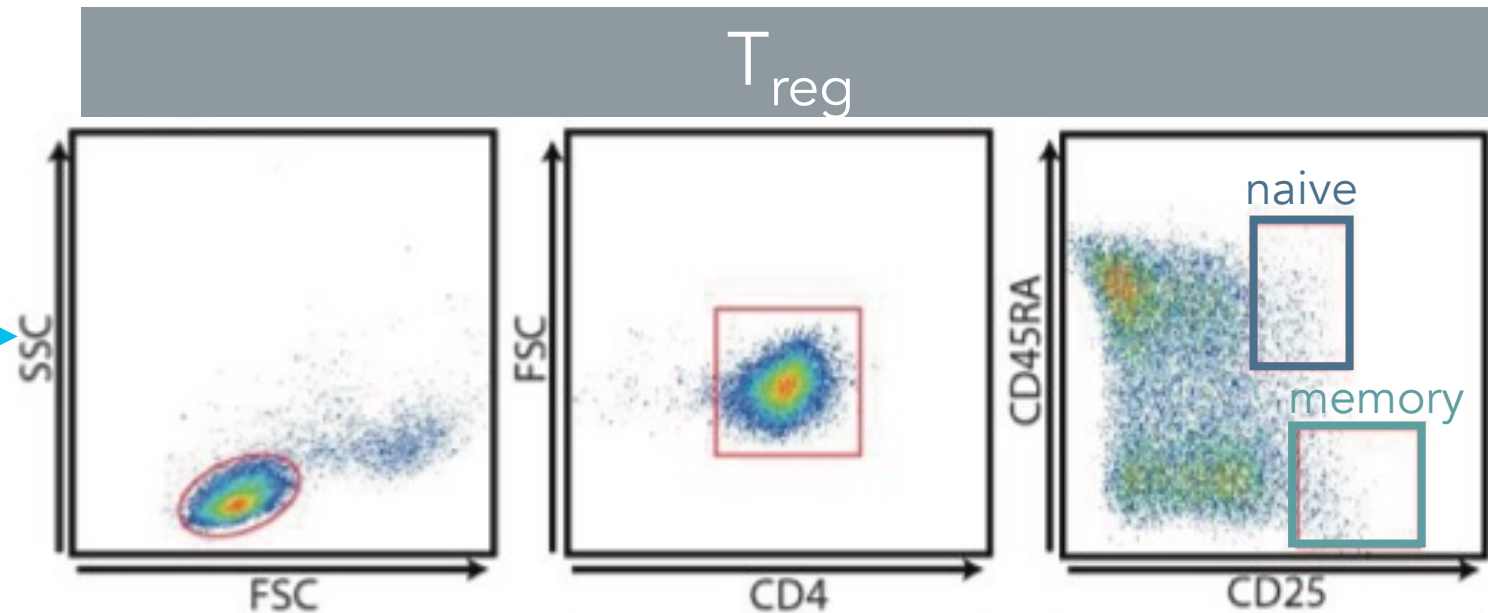
CD4+ T Memory

for research purposes only, not for use in diagnostic procedures

flow: Rieckmann et. al Nature Immunology 2017
single cell RNA-Seq: Team FlowJo

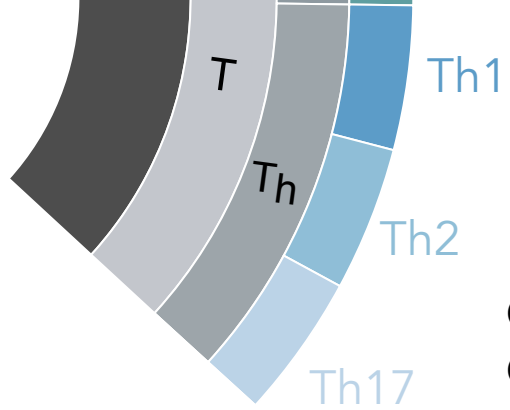


CD4+ enriched,
CD8/25/19-

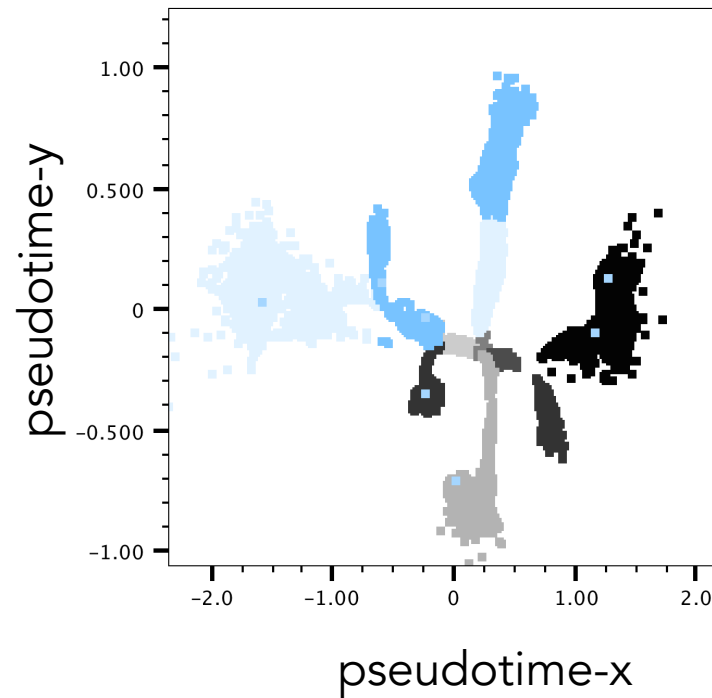
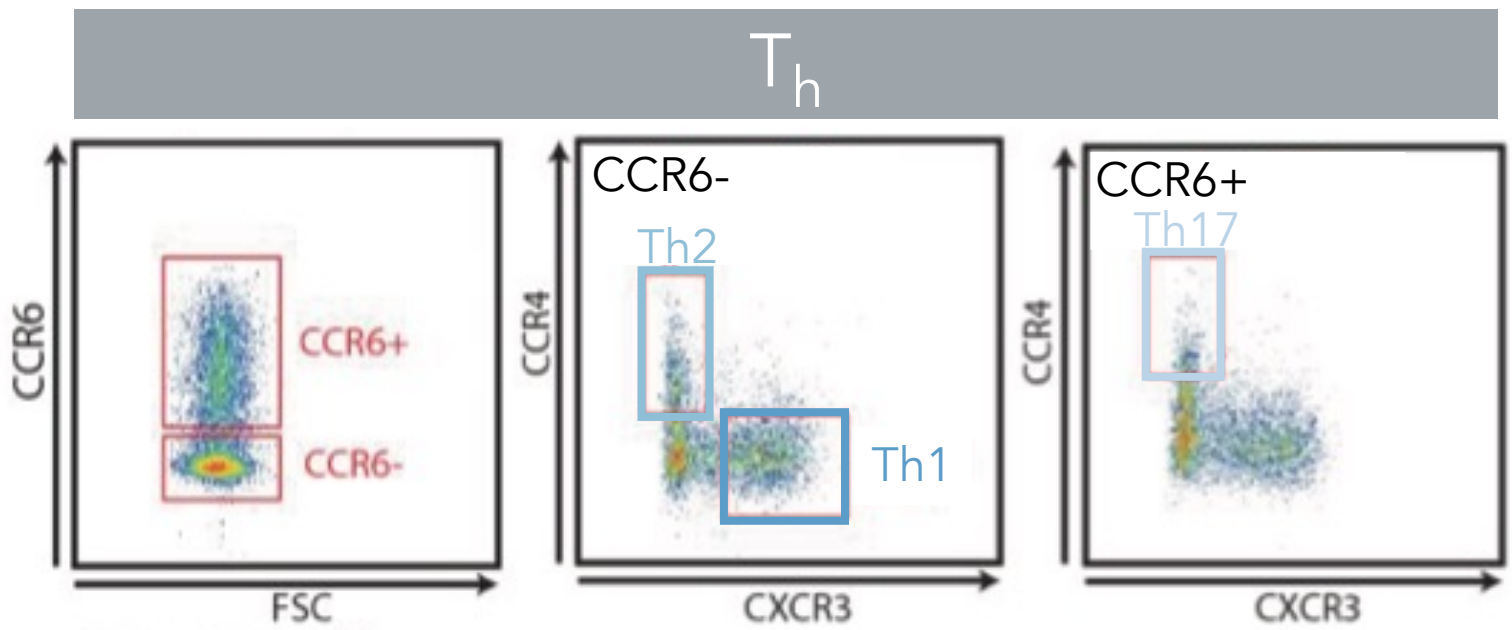


for research purposes only, not for use in diagnostic procedures

flow: Rieckmann et. al Nature Immunology 2017
single cell RNA-Seq: Team FlowJo

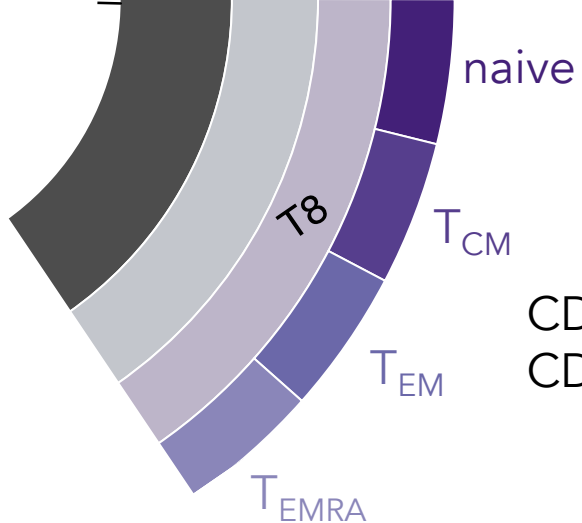


CD4+ enriched,
CCR7+CD45RA-
CD8/25-

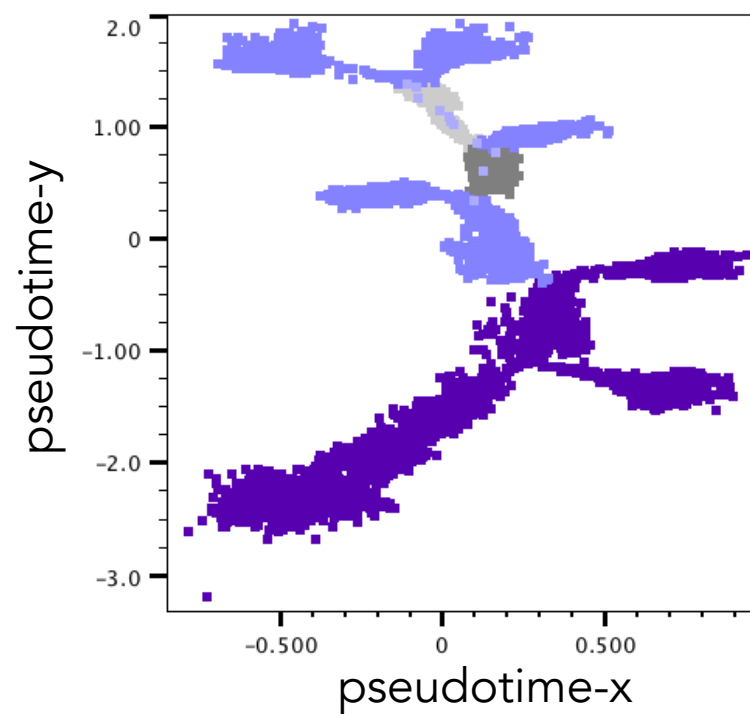
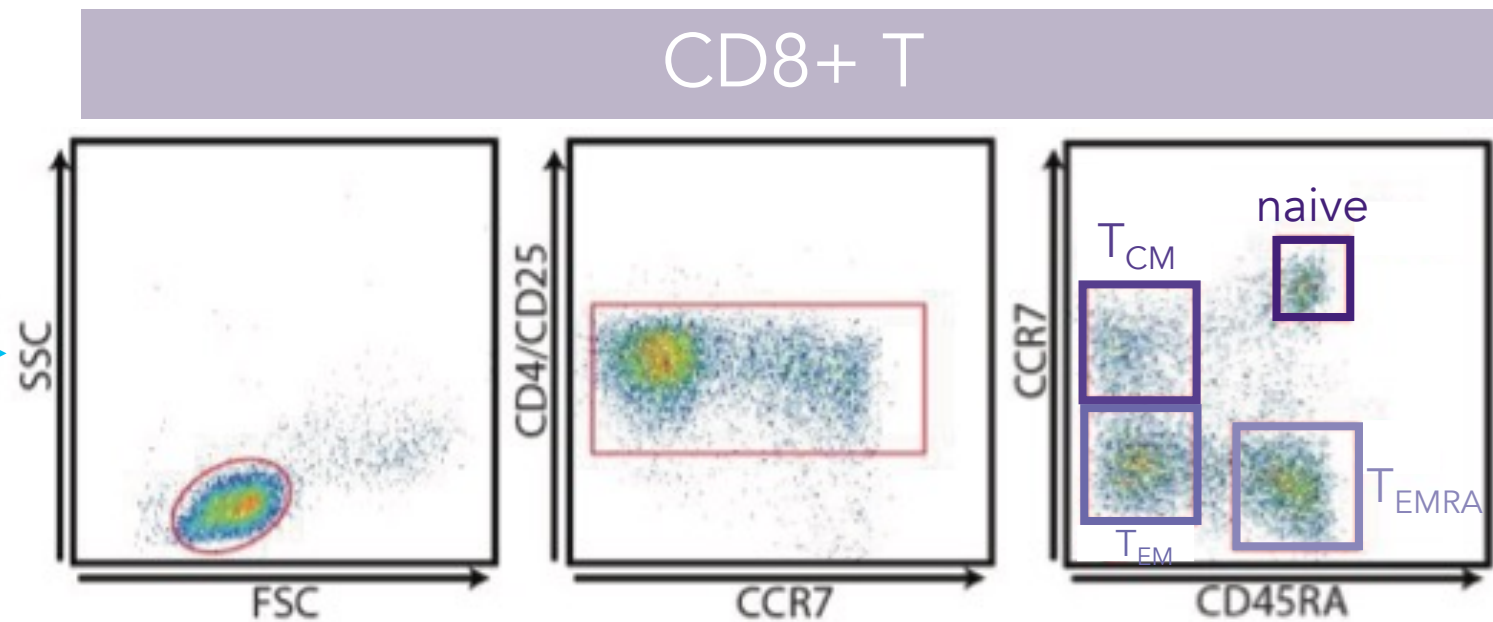


for research purposes only, not for use in diagnostic procedures

flow: Rieckmann et. al Nature Immunology 2017
single cell RNA-Seq: Team FlowJo



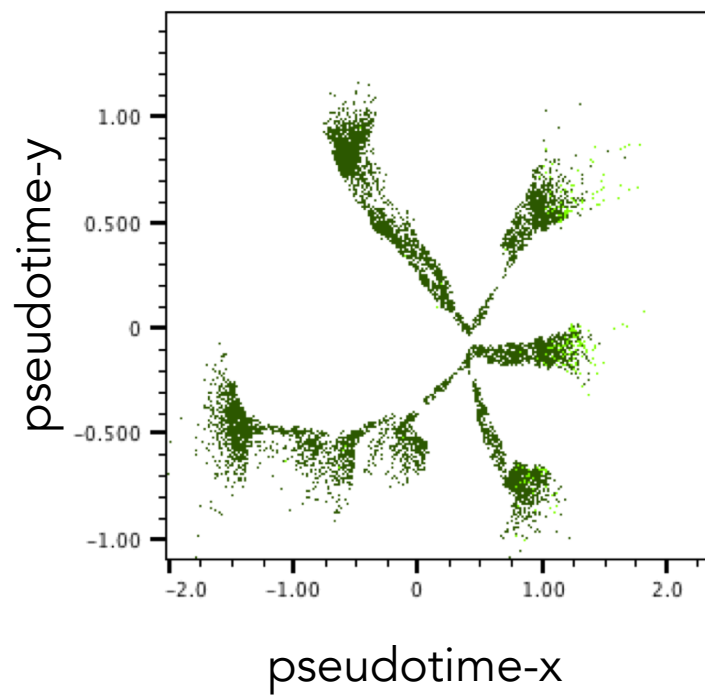
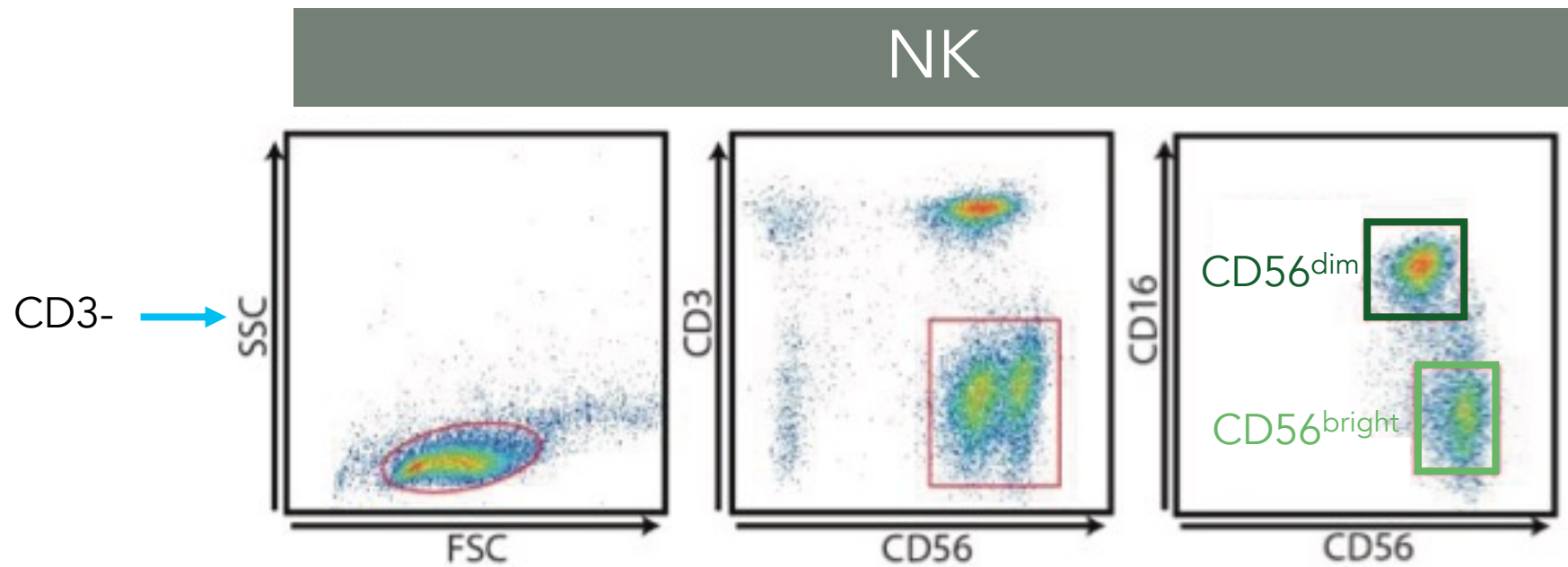
CD8+ enriched,
CD4/25-



for research purposes only, not for use in diagnostic procedures

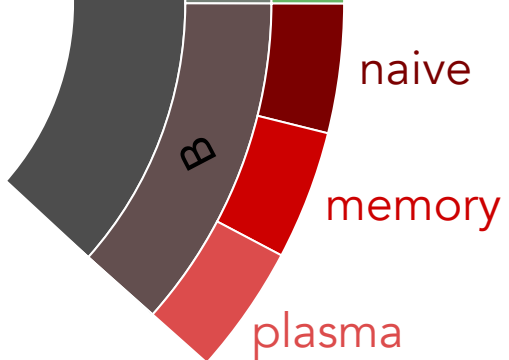
flow: Rieckmann et. al Nature Immunology 2017

single cell RNA-Seq: Team FlowJo

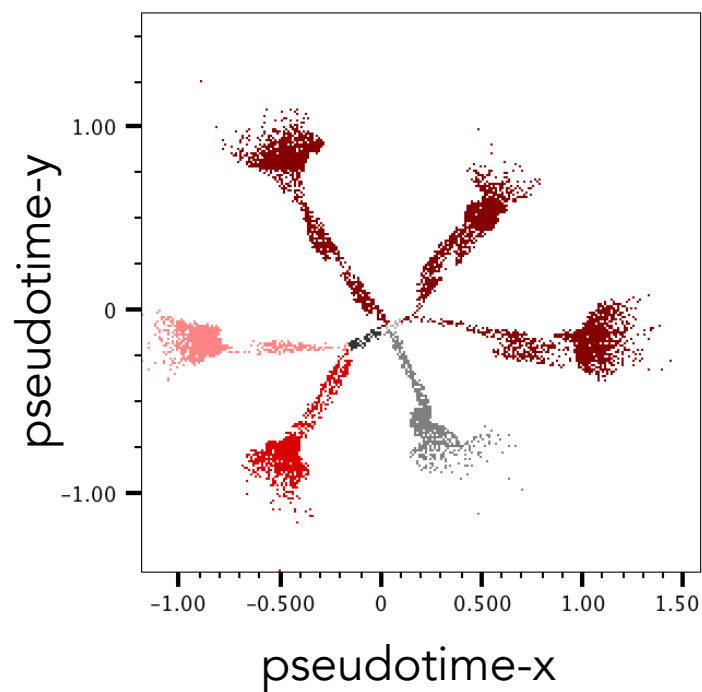
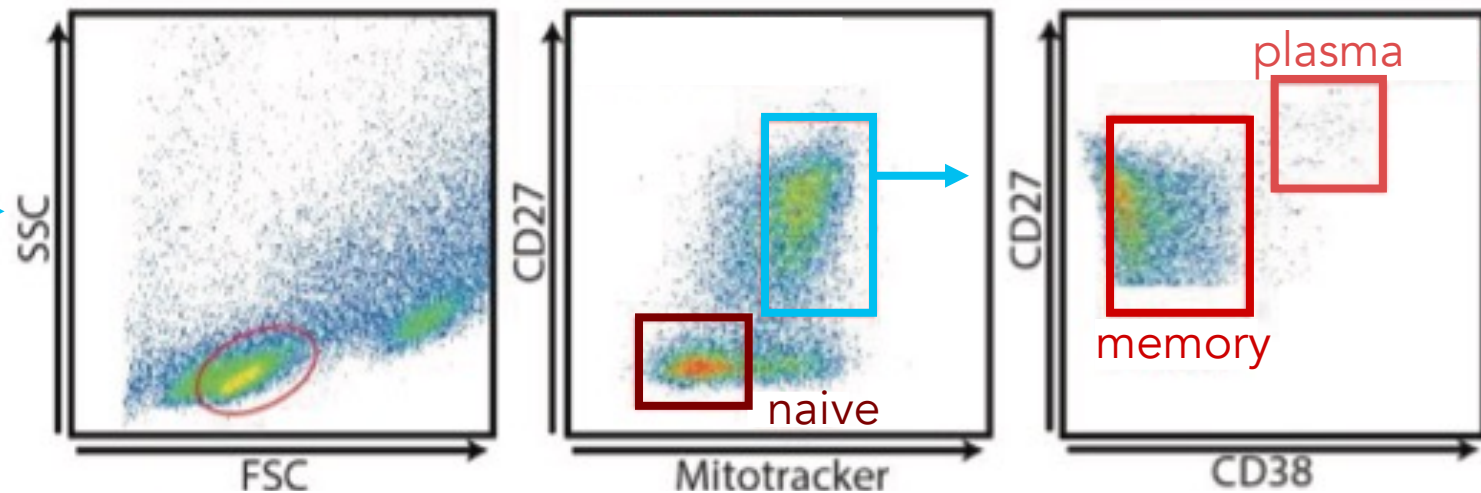


for research purposes only, not for use in diagnostic procedures

flow: Rieckmann et. al Nature Immunology 2017
single cell RNA-Seq: Team FlowJo



CD19
enriched



for research purposes only, not for use in diagnostic procedures

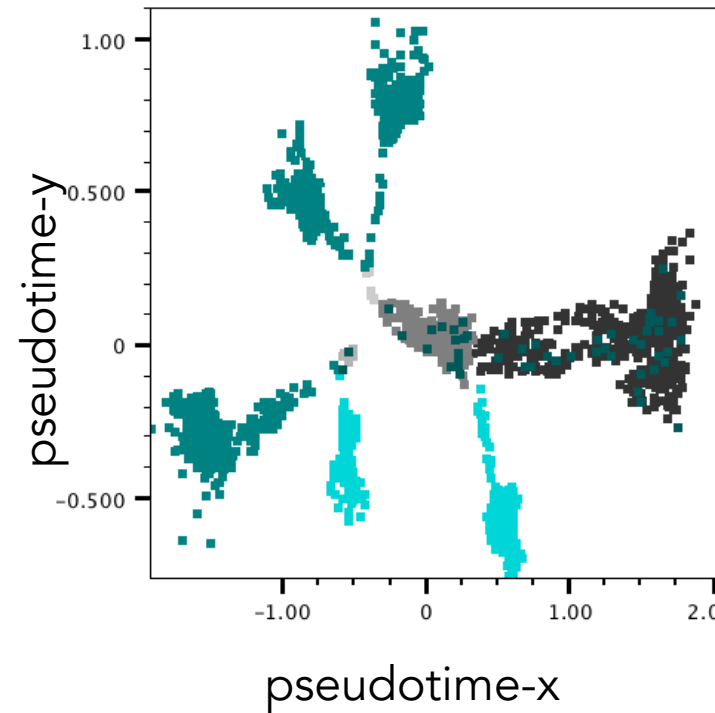
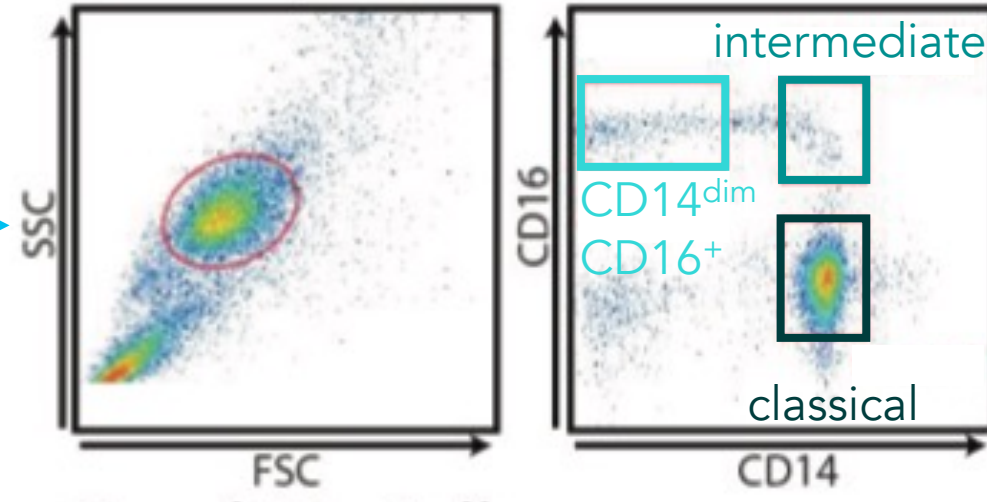
flow: Rieckmann et. al Nature Immunology 2017

single cell RNA-Seq: Team FlowJo

Monocytes



pan neg
sel kit

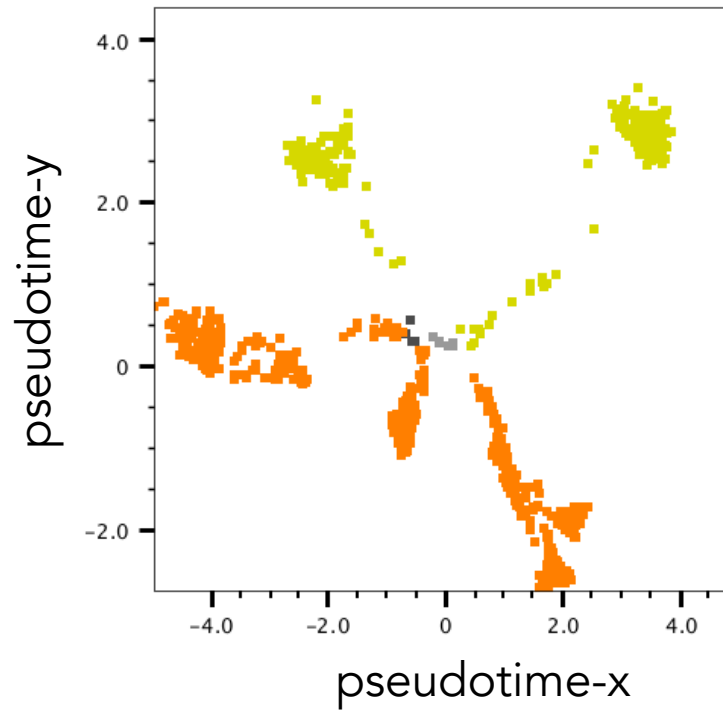
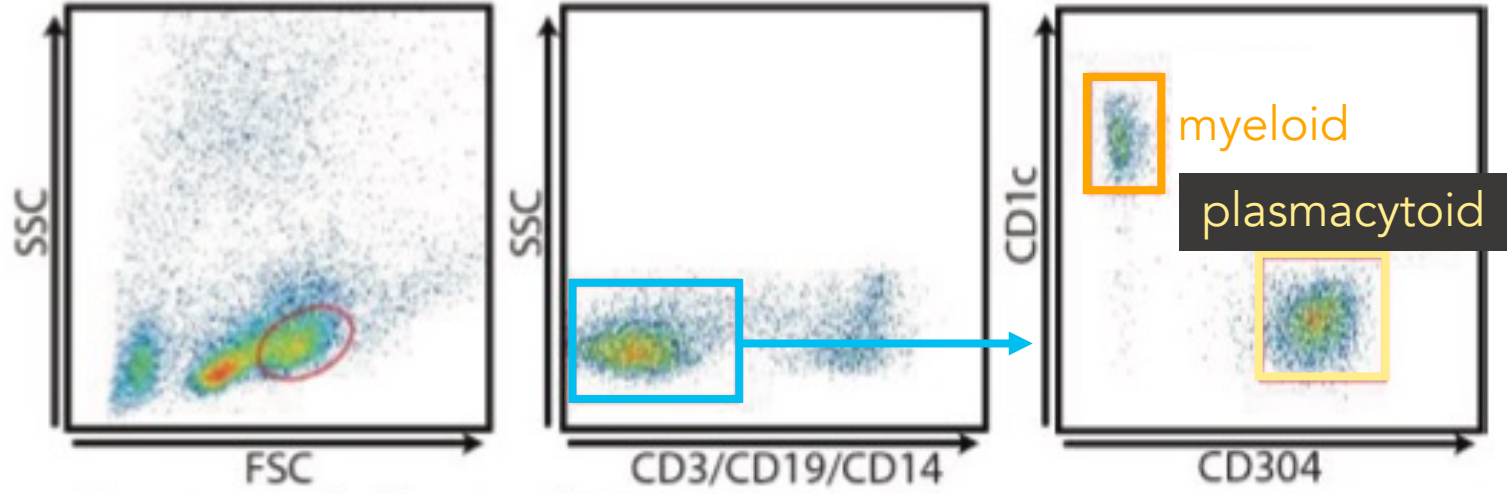


for research purposes only, not for use in diagnostic procedures

flow: Rieckmann et. al Nature Immunology 2017
single cell RNA-Seq: Team FlowJo



CD304
enriched,
CD14/19/3-



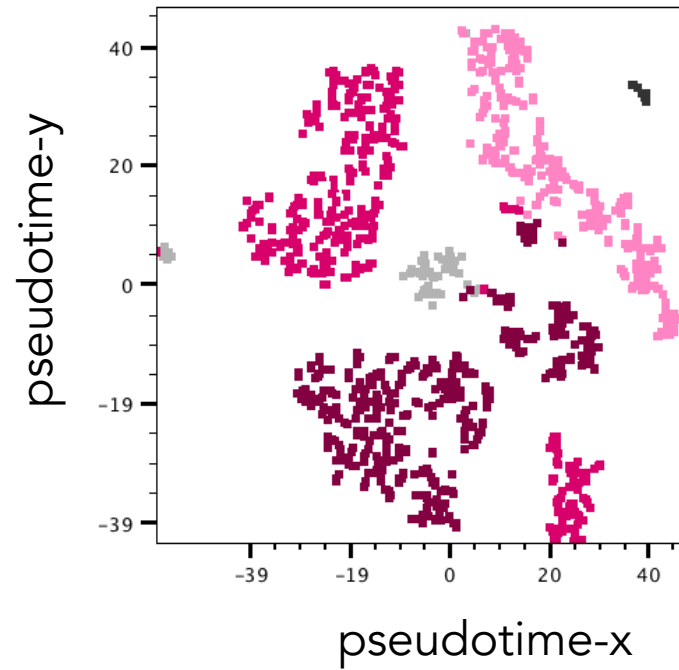
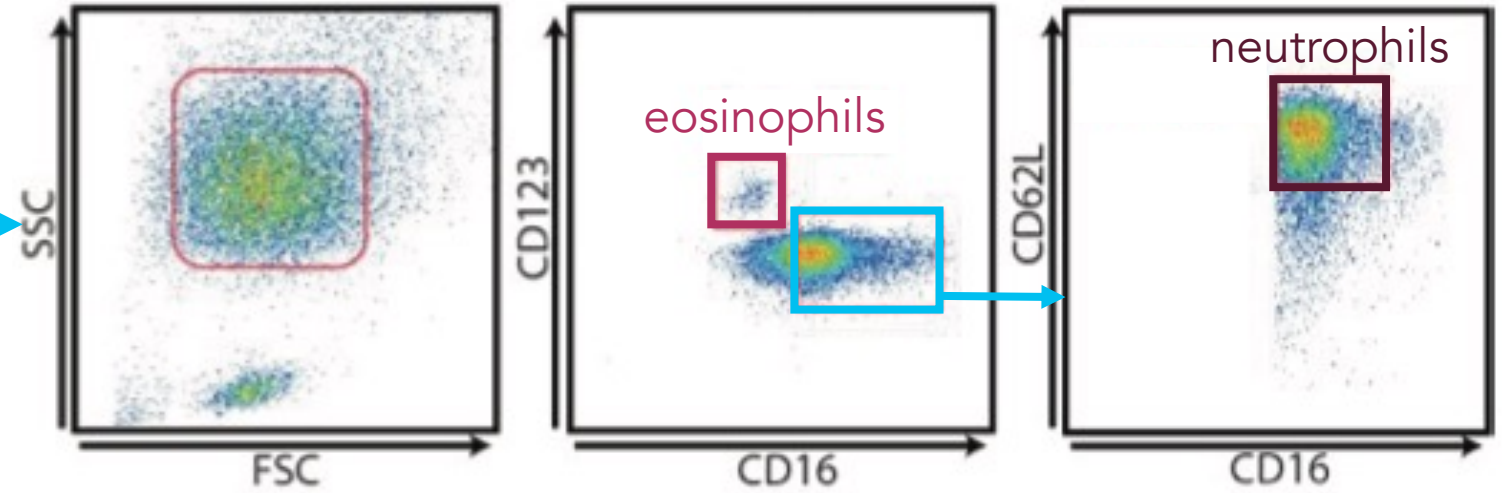
for research purposes only, not for use in diagnostic procedures

flow: Rieckmann et. al Nature Immunology 2017
single cell RNA-Seq: Team FlowJo



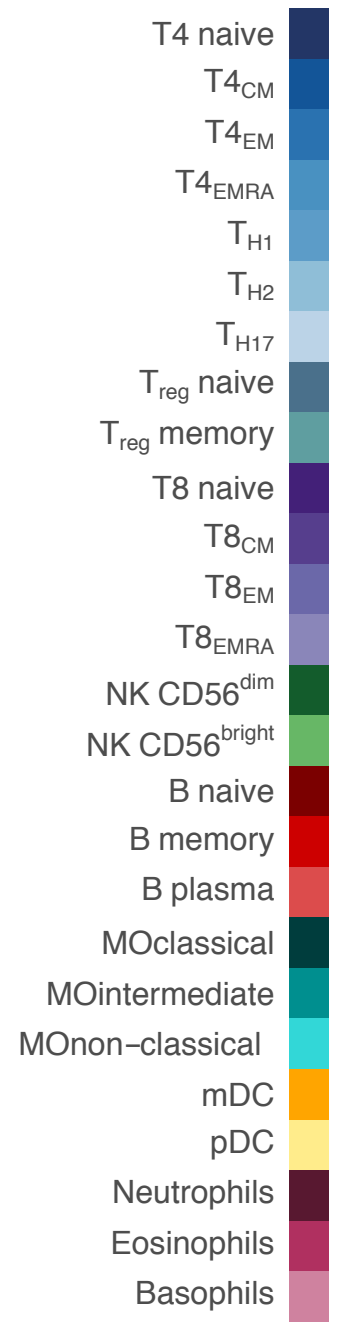
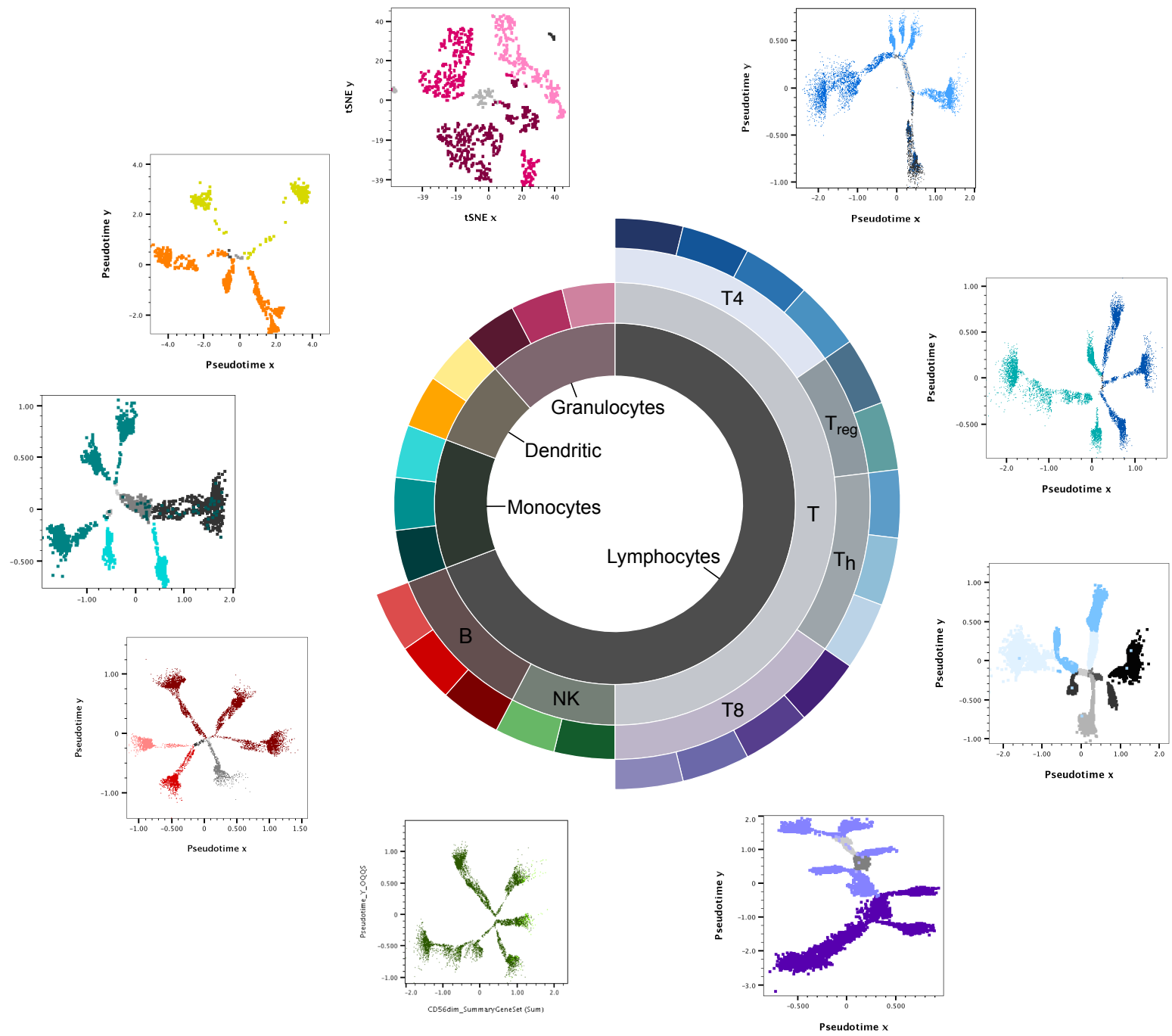
CD304
enriched,
CD14/19/3-

basophils (not
shown) 97A6
enriched, SSChigh
CD203c+



for research purposes only, not for use in diagnostic procedures

flow: Rieckmann et. al Nature Immunology 2017
single cell RNA-Seq: Team FlowJo



Explore Genomic
Cytometry by
interacting with
different species



H U M A N



M O U S E

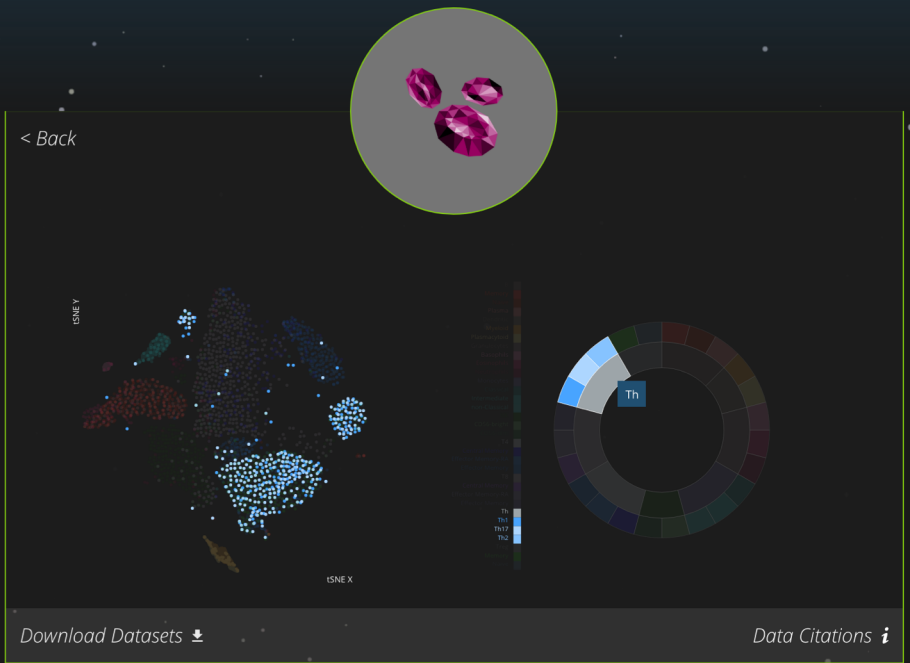
immune cell gene & proteome resource

genomiccytometry.com



launched 10.2.18
@ BOS UGM

- 01 HUMAN
- 02 BLOOD
- 03 PHENOTYPE
- 04



Immune Cell Social Network

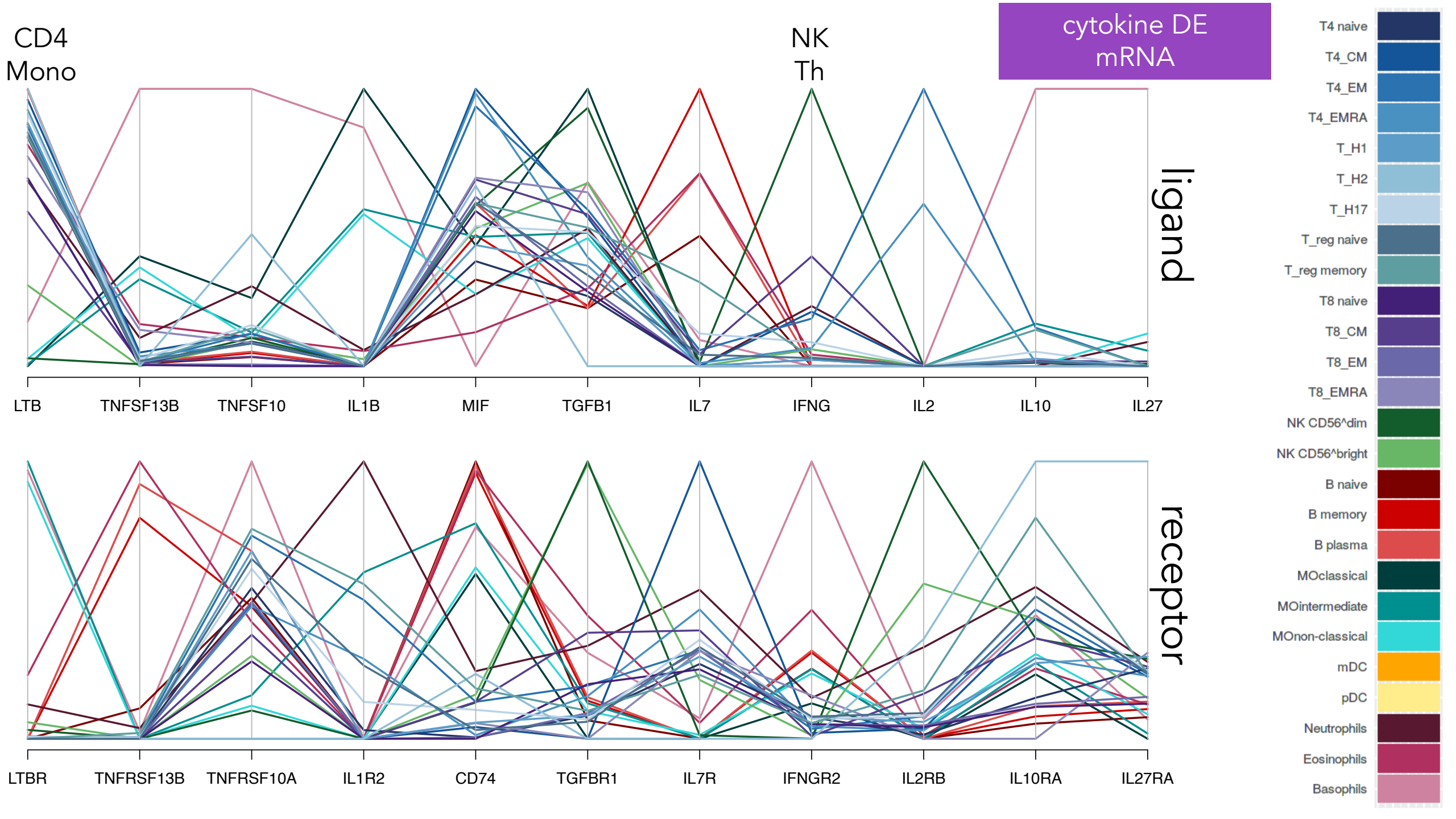
**experimental
setup**

**surface & secreted
interactome
analysis**

**26 steady state
immune cell models**

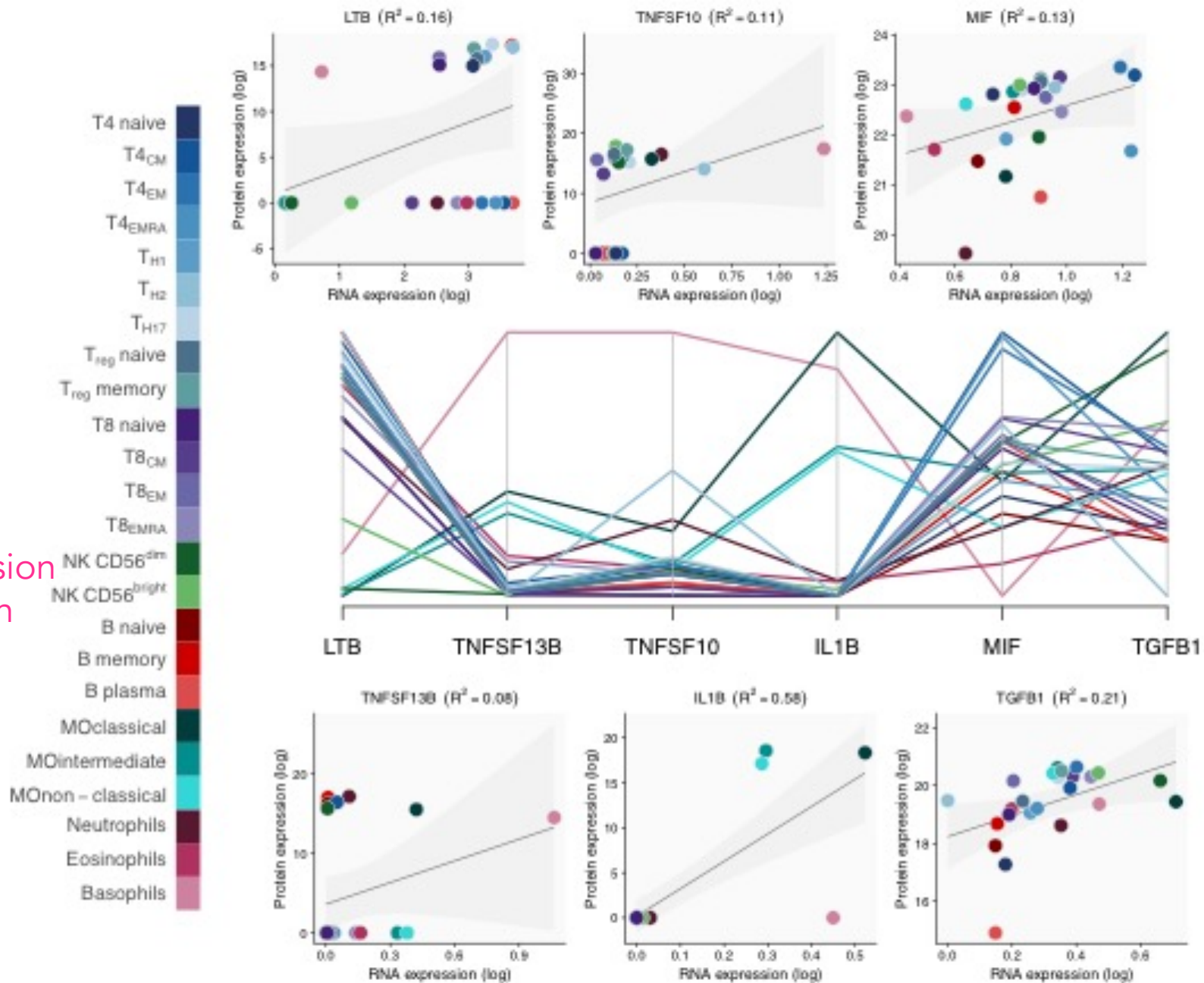
**predicting network
changes in disease**





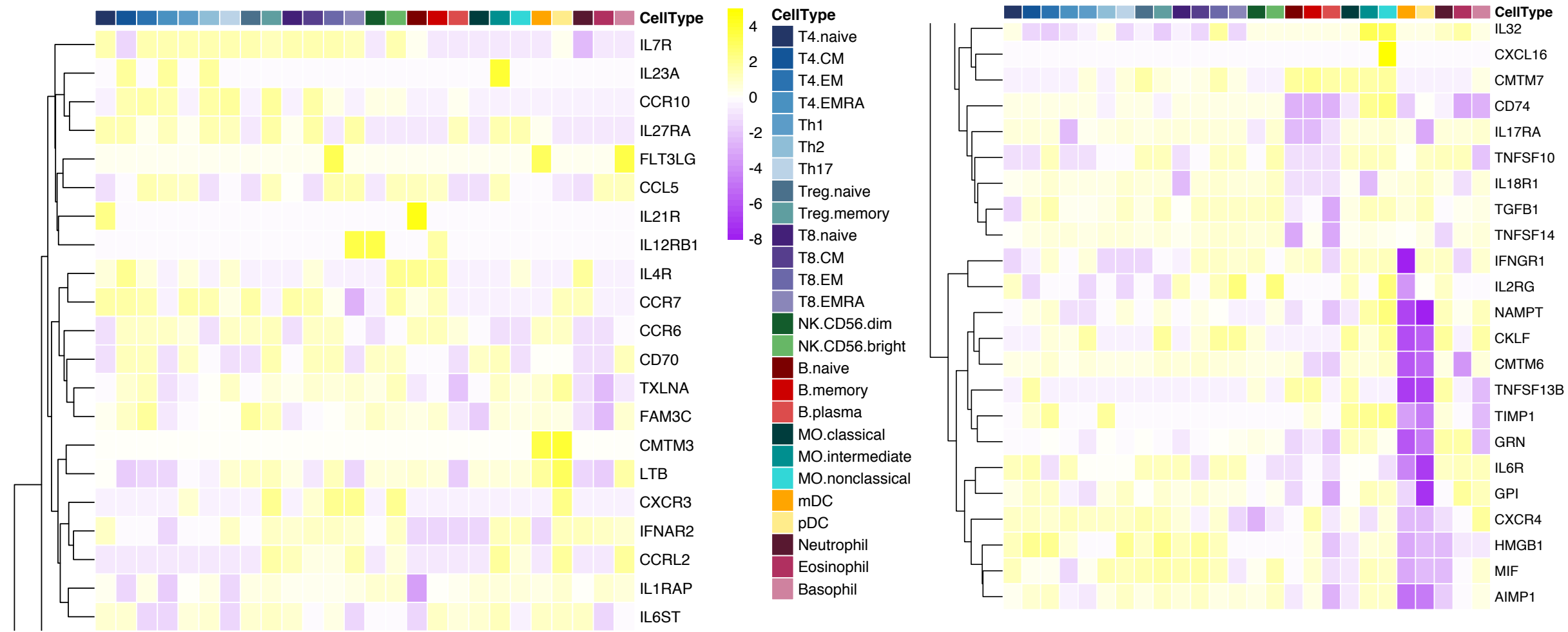
interactome

binary protein expression
HV gene expression

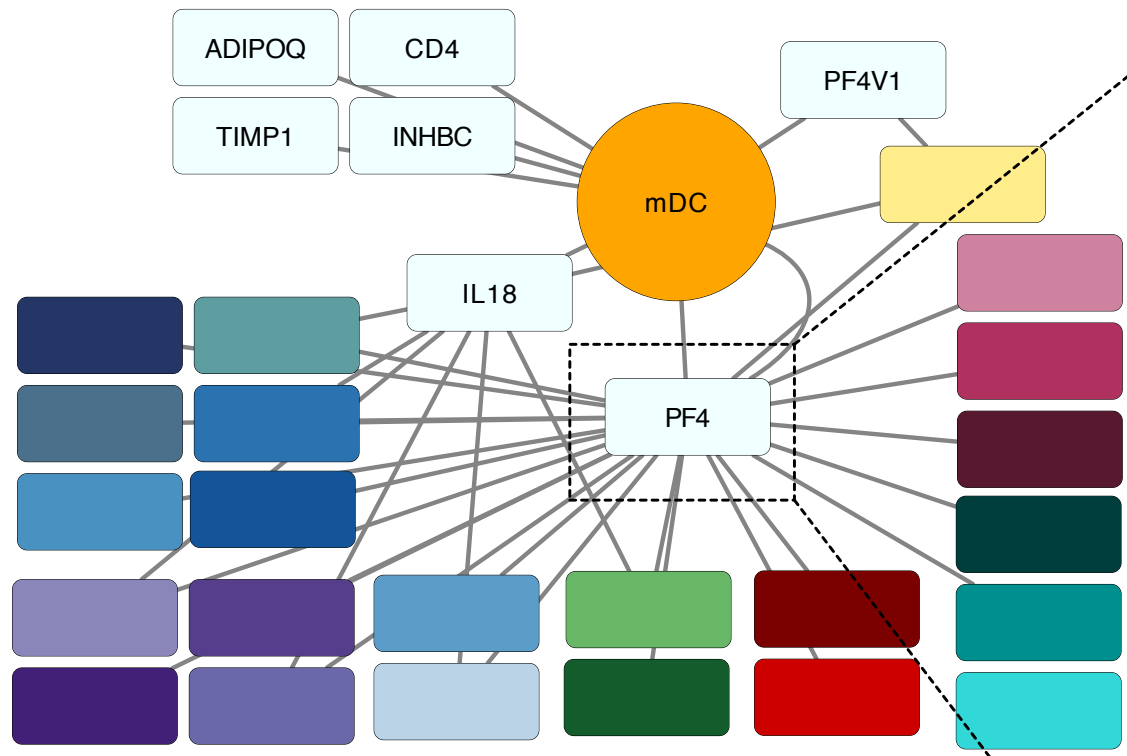


EXPRESSION DISCORDANCE

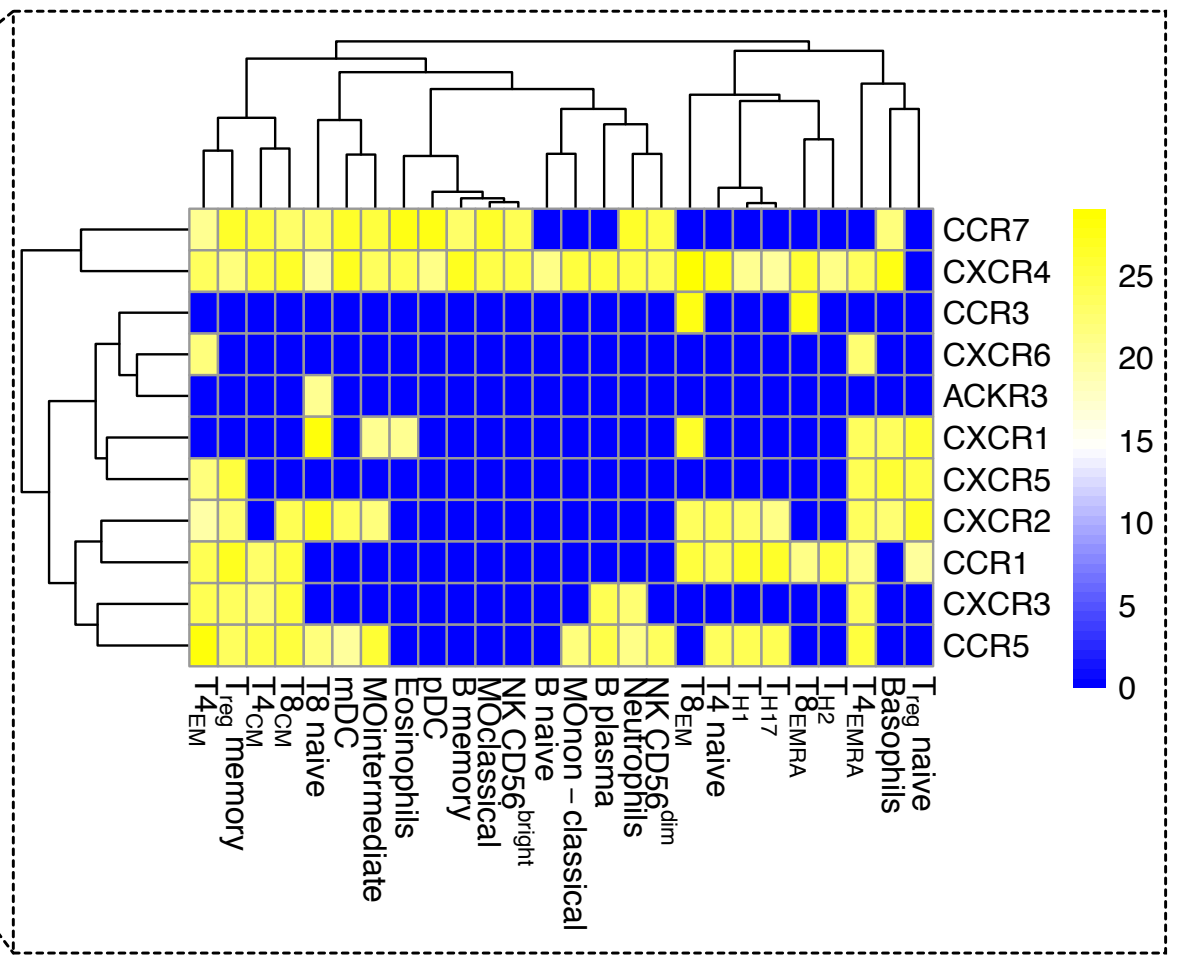
delta of protein Z-Score and RNA Z-score



NETWORK



PF4 RECEPTORS (protein expression)



can use to *rank* interactome

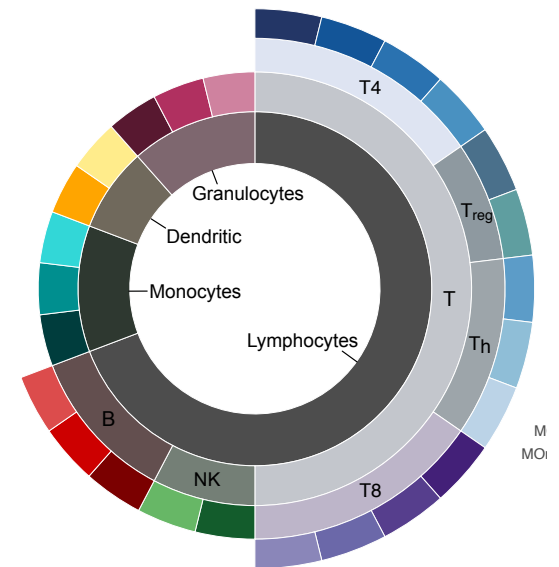
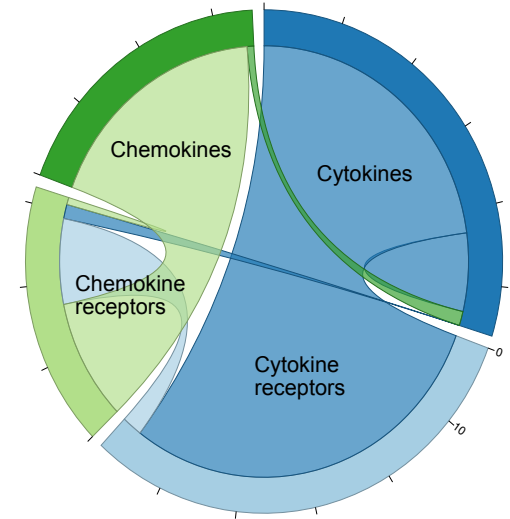
GENOMIC CYTOMETRY

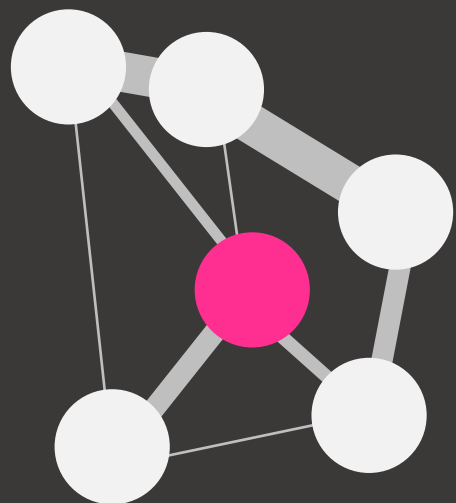
Observations:

- WTA lacks sensitivity to detect low expressing but identifying genes, challenging to do discovery immunology (*solution: targeted*)
- AbSeq/flow sorting is critical for doing immunology in the context of scRNA-Seq
- Field lacks *power ranked* interactome list (and temporal model)

New Resources:

- Interactome list (protein and RNA names)
- Identifying gene sets, differentially expressed genes and proteins for 26 canonical immune cell populations



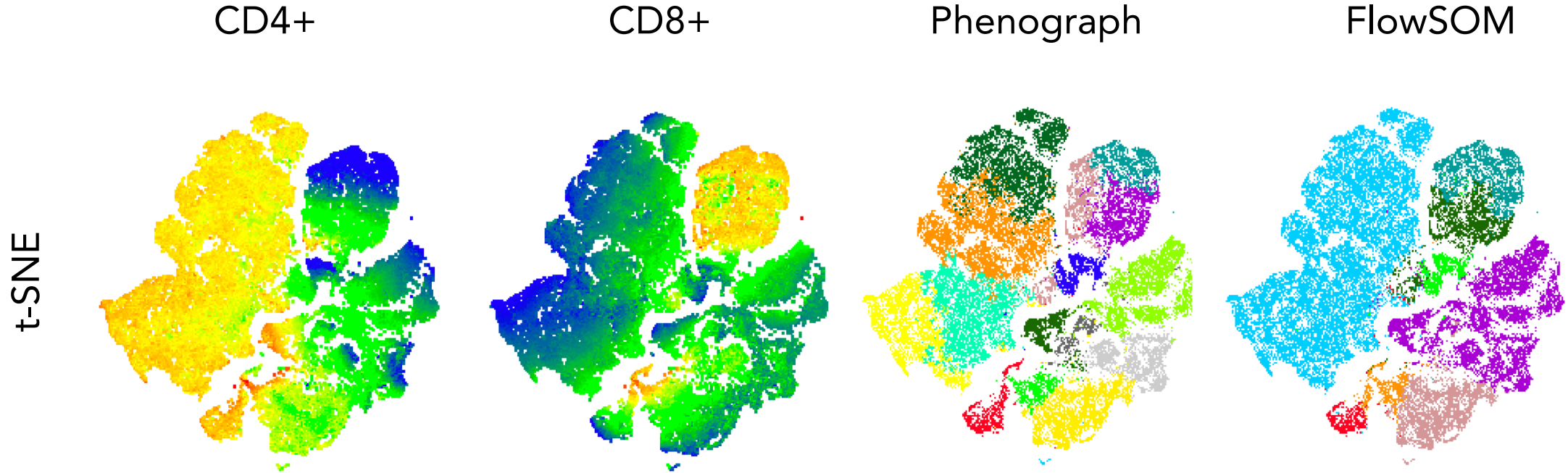


2

TRANSFORMATIVE CONNECTIONS

ONE
CELL
SOCIAL
NETWORK

TWO
REDUCTION
&
VISUALIZATION



wait... how did we get here?
what have we compromised?

WHEREFORE tSNE?

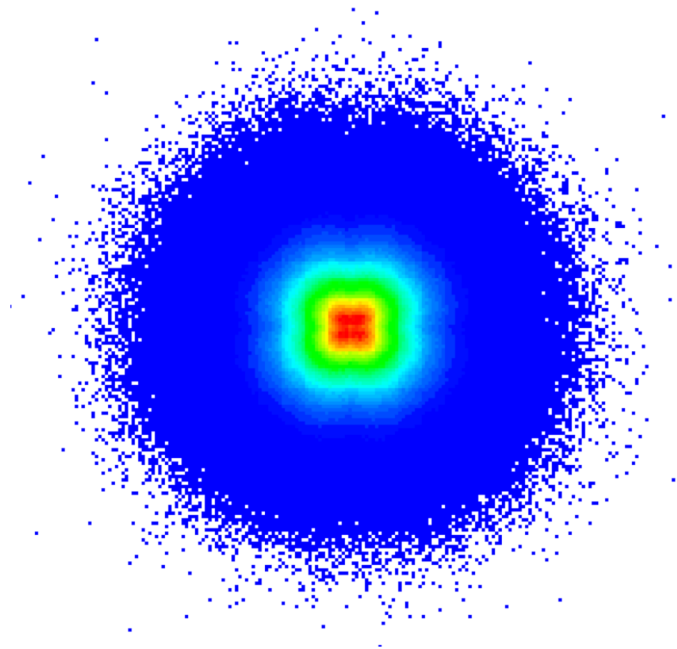
Combine/Concatenate

1. Hard to project new samples to an existing t-SNE space

Downsample

2. Computational complexity
(Large inputs take forever... especially with old FlowJo t-SNE plugin... hold that thought)

3. Large inputs yield useless t-SNE embeddings



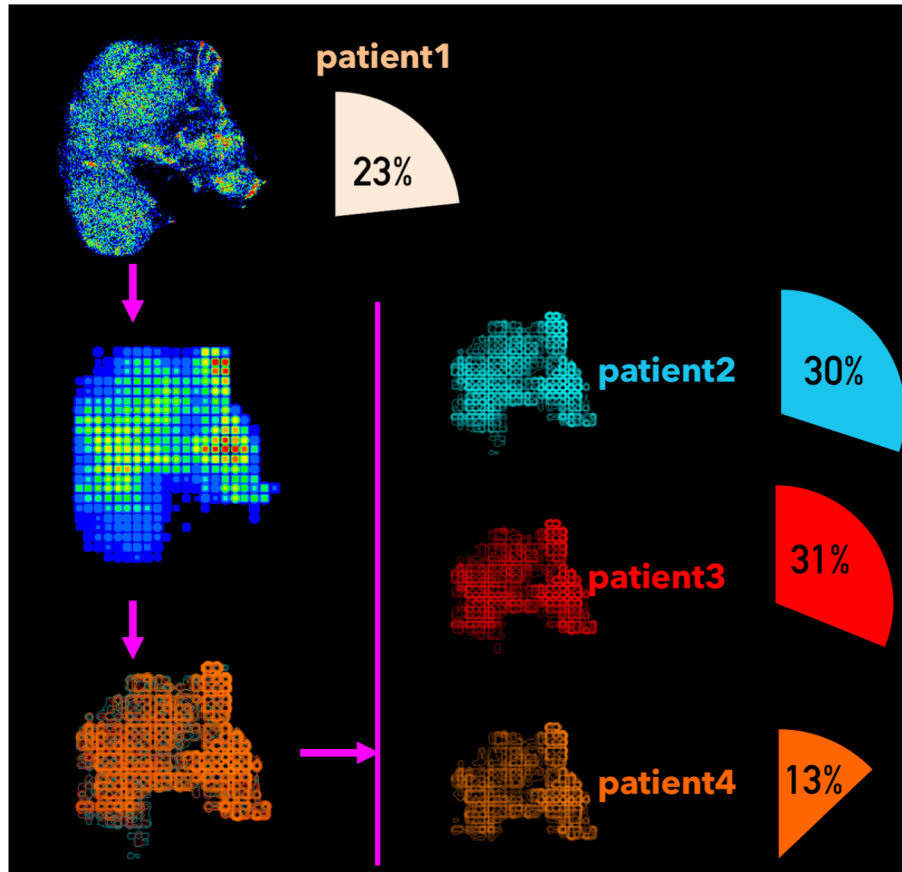
Standard t-SNE on 14 million cells

PROBLEM 1

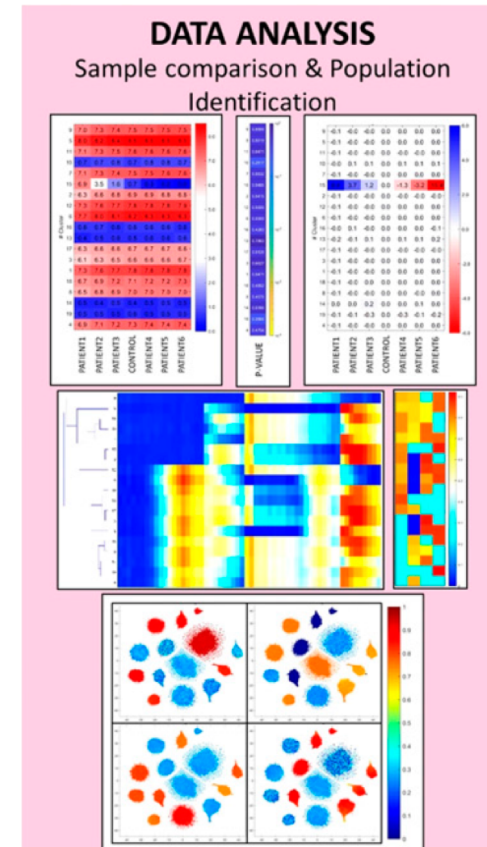
Hard to project new samples to an existing t-SNE space

Solved?

- Yes, somewhat
- Different approaches, typically based on nearest neighbors
- But you still need your t-SNE input to represent all cell populations



<-that guy, CYTO2016



Jimenez-Carretero et. al. J Immunol. 2018

what happens when we
downsample?

(experiment time!)

downsampled
20%



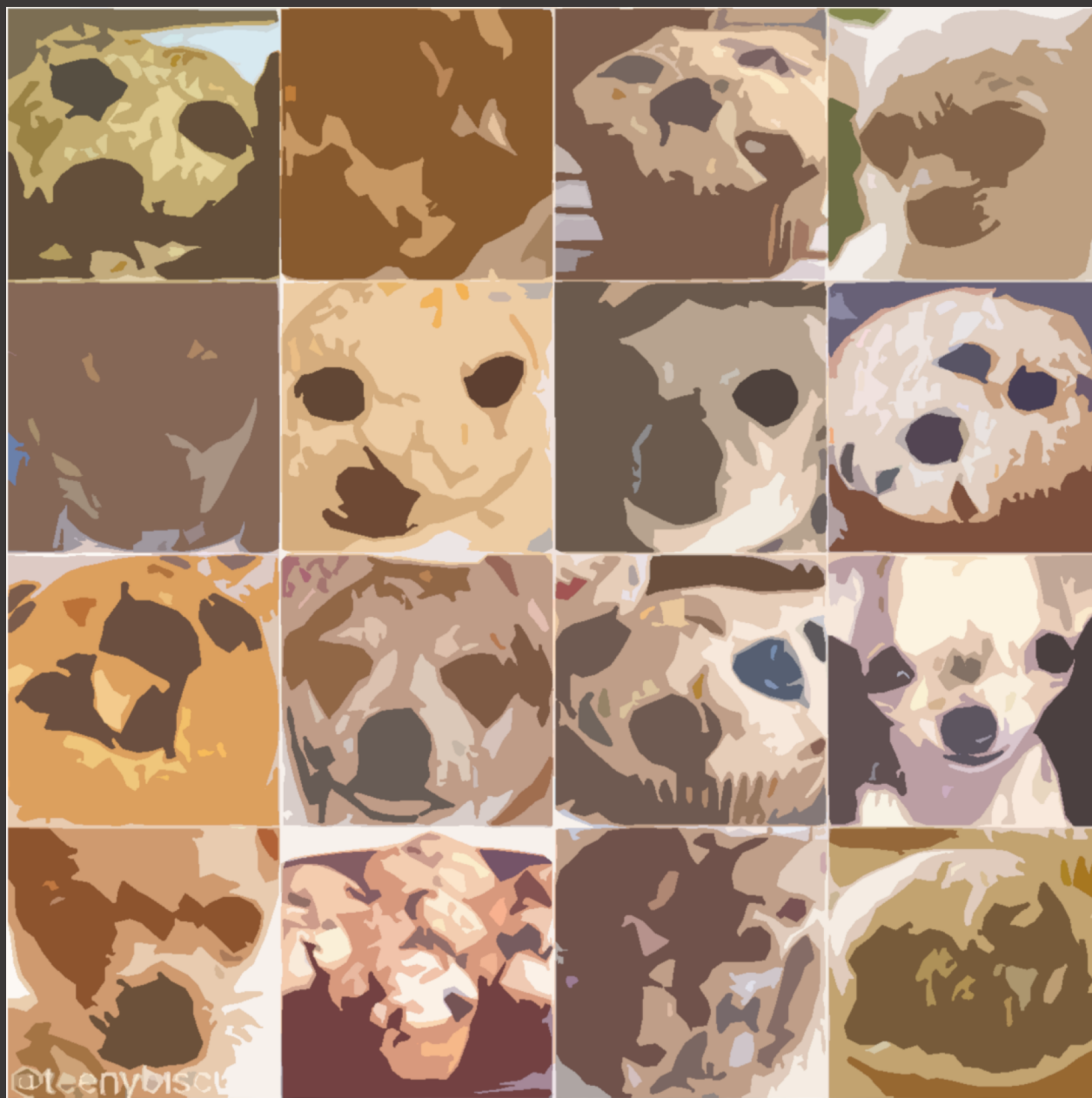
adapted from Yangyan Li
"A Brief Introduction to
Deep Learning"

downsampled
20%



adapted from Yangyan Li
"A Brief Introduction to
Deep Learning"

downsampled
60%



adapted from Yangyan Li
"A Brief Introduction to
Deep Learning"

downsampled
60%



adapted from Yangyan Li
"A Brief Introduction to
Deep Learning"

downsampled
90%



adapted from Yangyan Li
"A Brief Introduction to
Deep Learning"

downsampled
90%

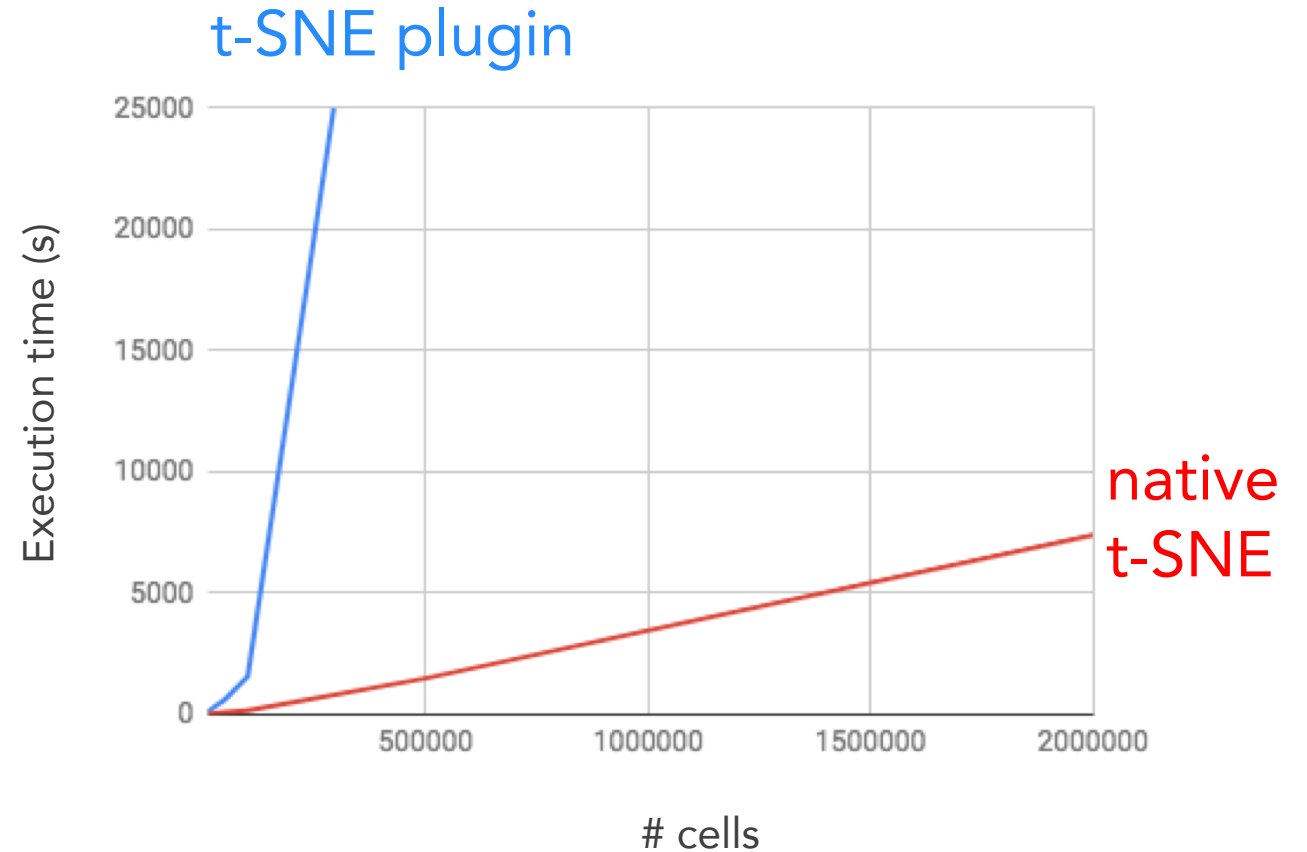


adapted from Yangyan Li
"A Brief Introduction to
Deep Learning"

Solved?

Yes! Now 3X faster than any available software

- FlowJo **native t-SNE** on 2,000,000 cells takes 2 hours
- Have tried up to 20M cells

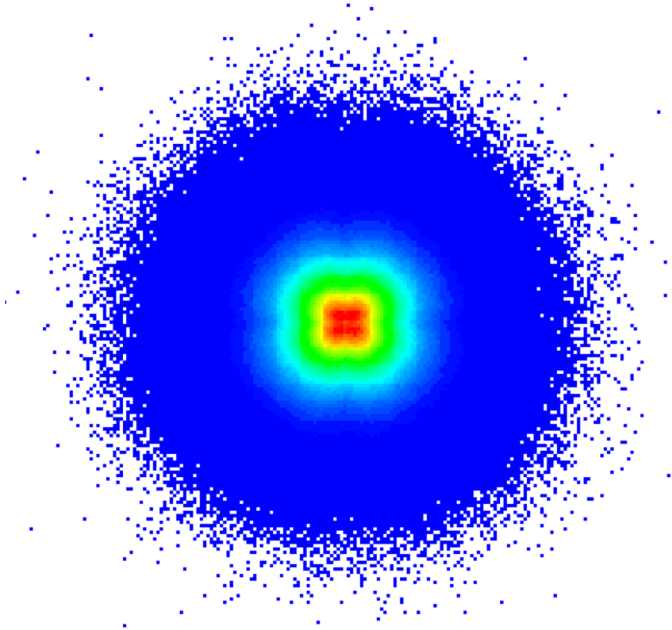


PROBLEM 3

Large inputs yield useless t-SNE embeddings

why do we get a ball?

HAVING A BALL?



repulsive forces
overcome
attractive forces



1. Set perplexity (#NN) to 3 → each cell attracted by only its nearest neighbors
2. but being repulsed by all other neighbors

Too much repulsing, cells cannot find their friends (region)

PROBLEM 3

Large inputs yield useless t-SNE embeddings

Solved?

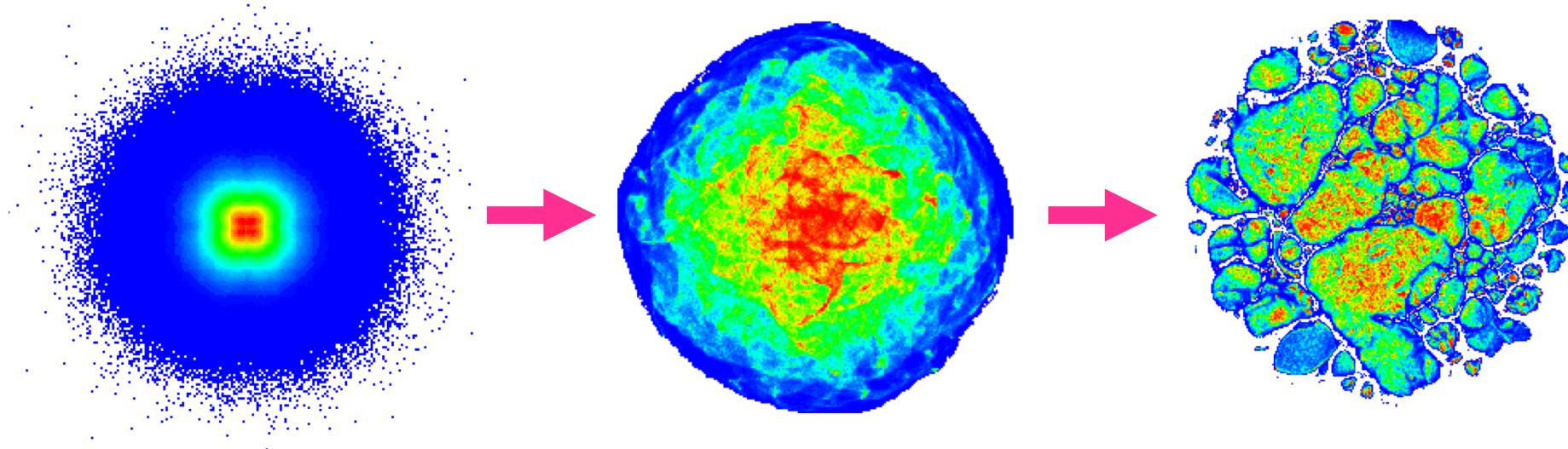
Yes!



opt-SNE

Belkina AC. et. al.

Automated optimal parameters for T-distributed stochastic neighbor embedding improve visualization and allow analysis of large datasets. Oct 24, 2018, bioRxiv id 451690, <https://doi.org/10.1101/451690>

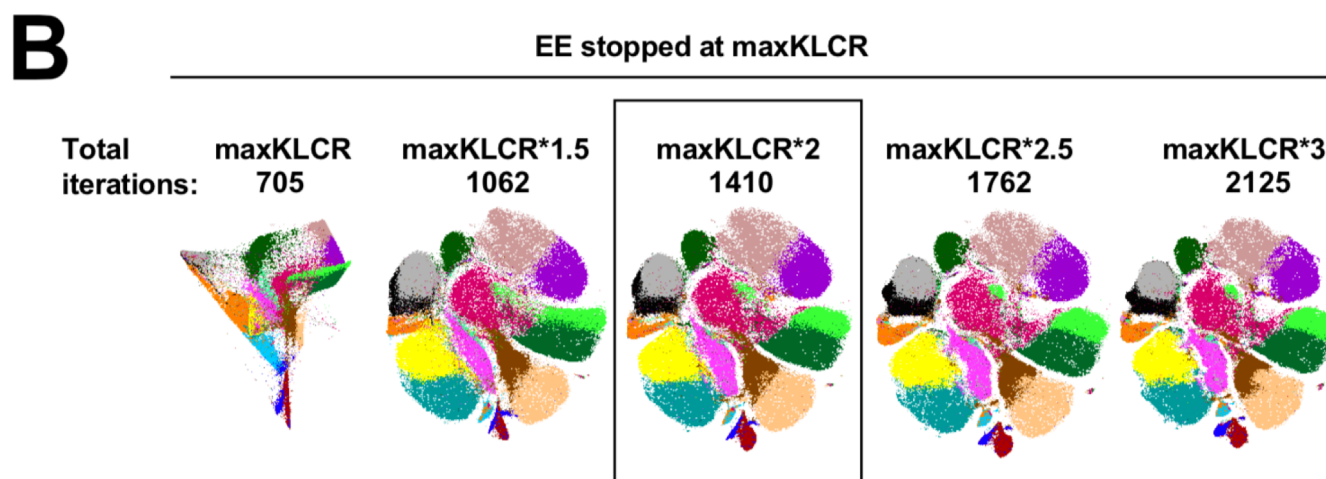
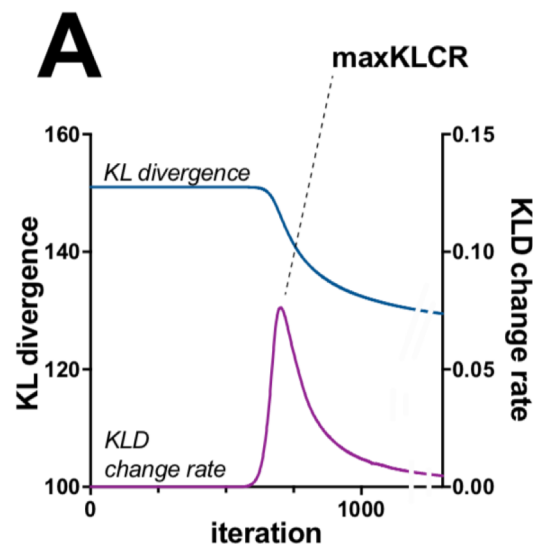


PROBLEM 3

Large inputs yield useless t-SNE embeddings

How?

1. Data-driven learning rate reflecting the input size
2. End early exaggeration (EE) by tracking the Kullback-Leibler Divergence (KLD)
3. Stop iterating when results are no longer improving



trial and error, lots of waiting:



PROBLEM 3

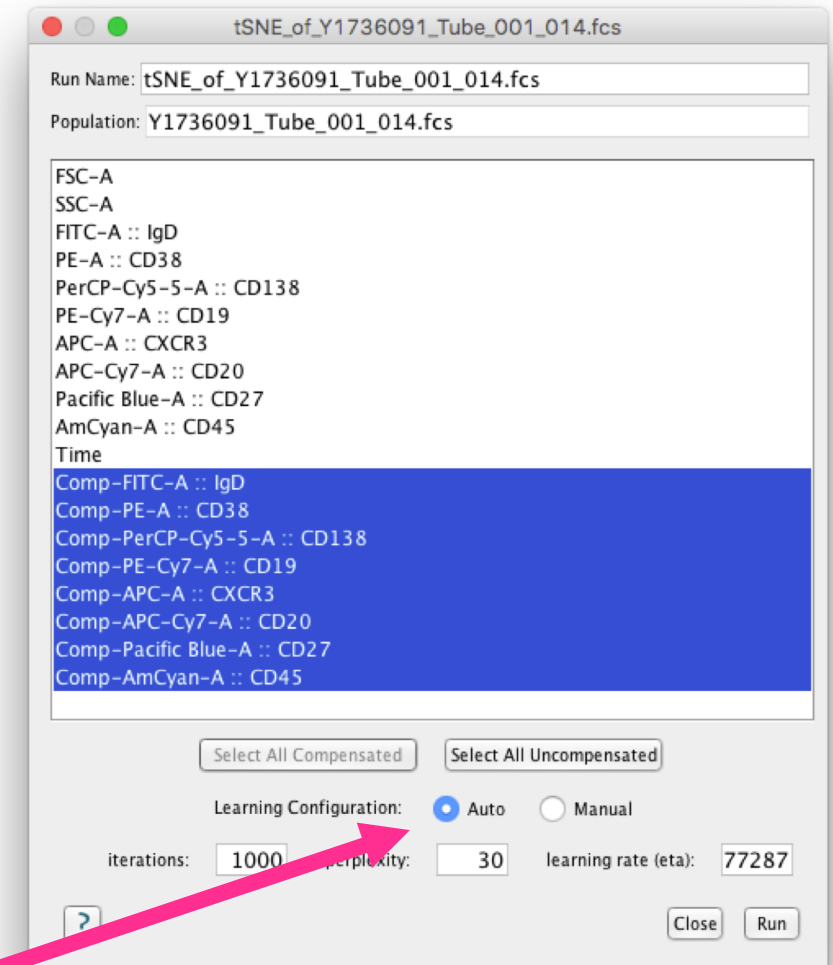
Large inputs yield useless t-SNE embeddings

I need it!

You can enable this now in
FlowJo 10.5.3 by putting

```
<DRPlatform  
showAutoLearning="1" />
```

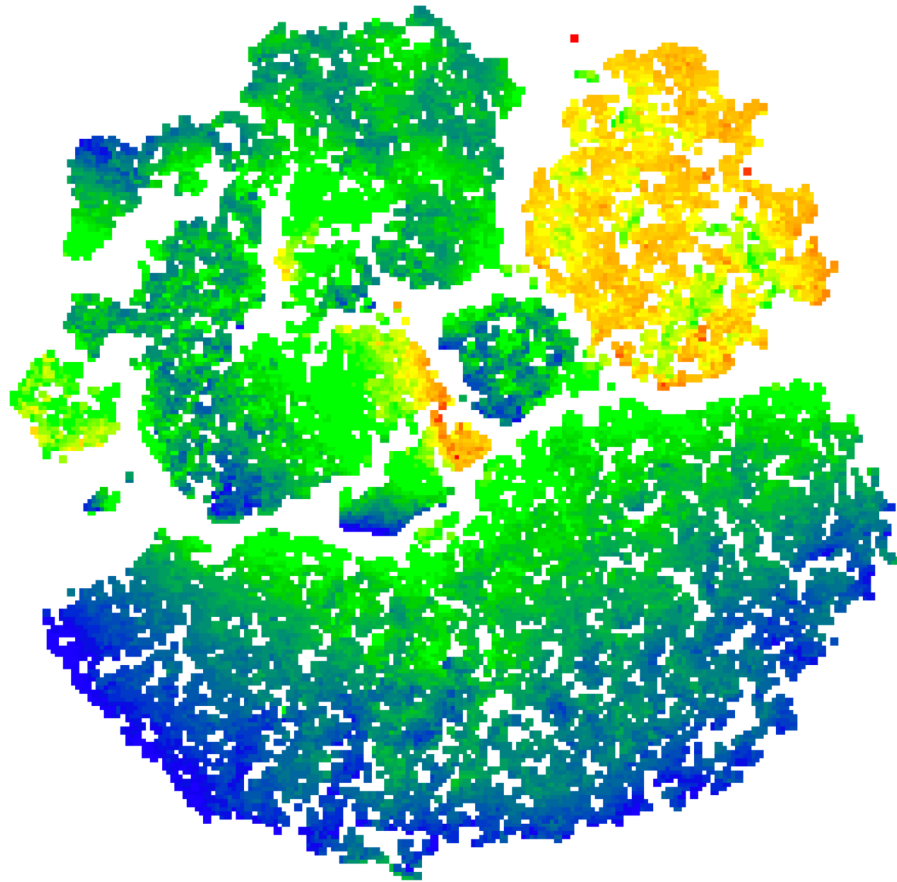
in your FlowJo10.prefs file
(that's it!)



this is NECyto!

what else you got?

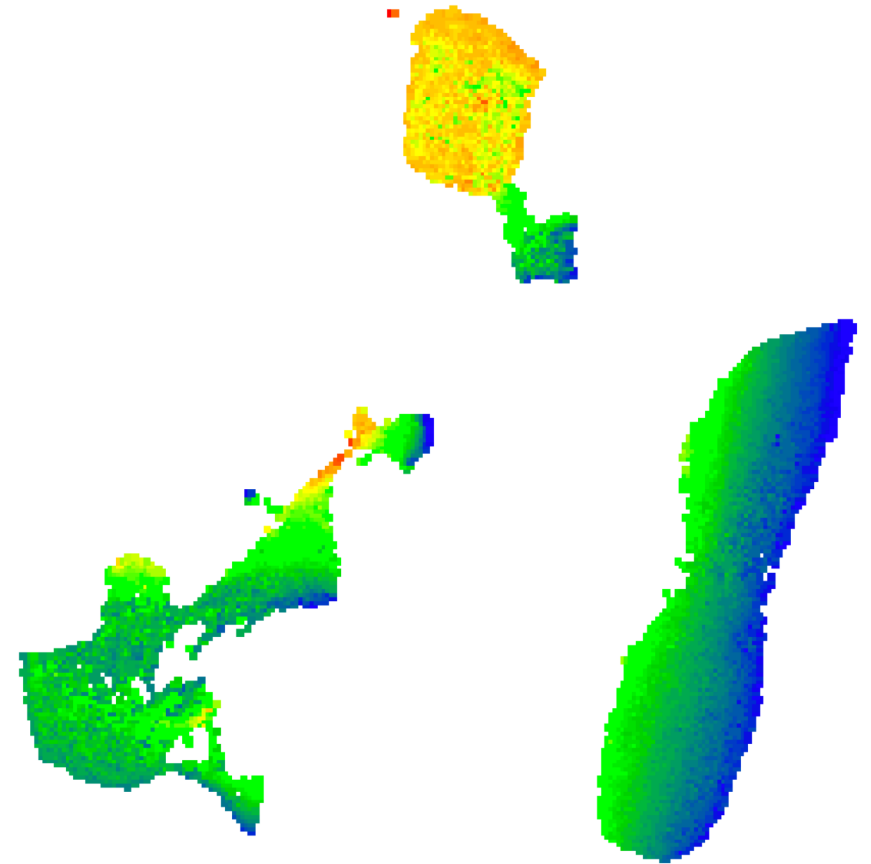
Dimensionality Reduction



Fit-SNE: Fast Fourier Transform-accelerated Interpolation-based t-SNE. 3X faster

Linderman, et al. *arXiv*. 2017

CD8+

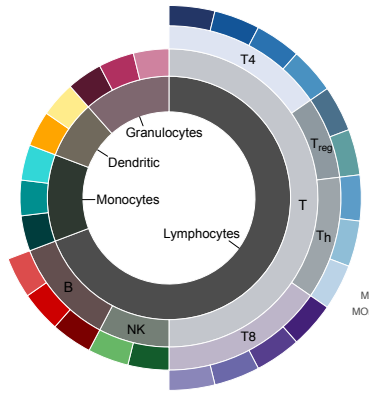


UMAP: Uniform Manifold Approximation and Projection

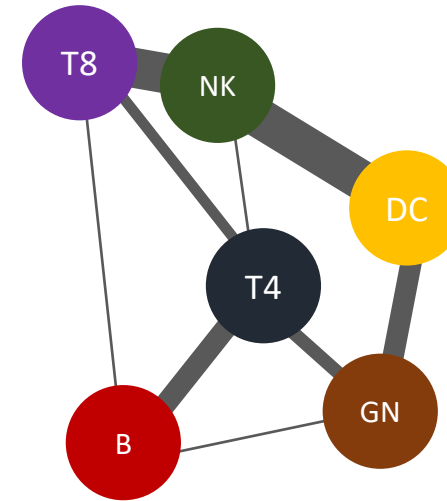
McInnes, et al. *arXiv*. 2018

SYSTEMS IMMUNOLOGY

1. Study/assay all or a (larger) targeted set of immune cells

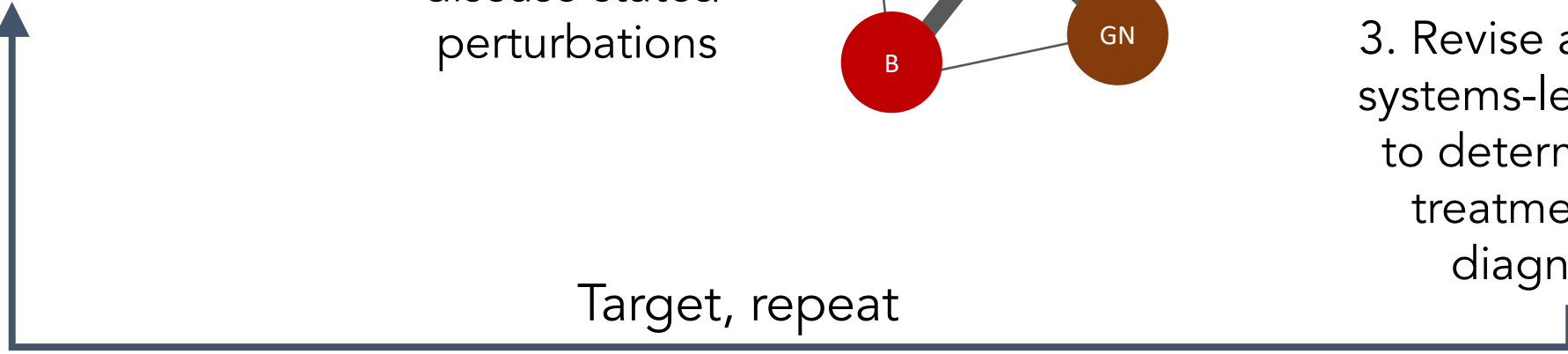


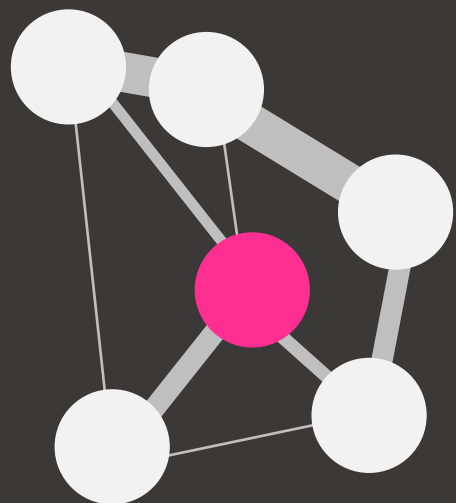
2. Study both cell activity & environment in disease states/ perturbations



3. Revise and curate systems-level model to determine best treatments and diagnostics

Target, repeat



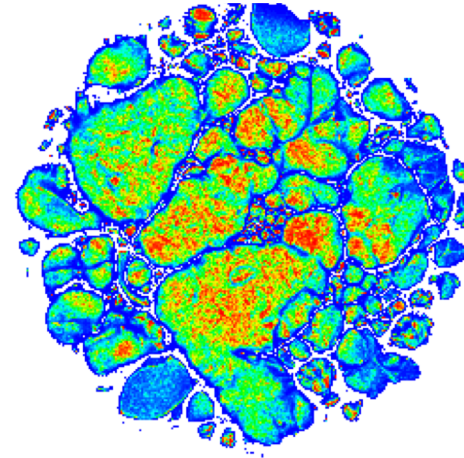
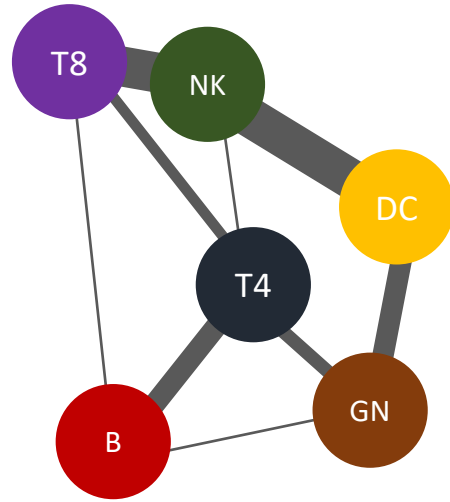


2

TRANSFORMATIVE CONNECTIONS

ONE
CELL
SOCIAL
NETWORK

TWO
REDUCTION
&
VISUALIZATION



mikes@flowjo.com
genomiccytometry.com