NASDC
NCI Awardee Skills Development Consortium

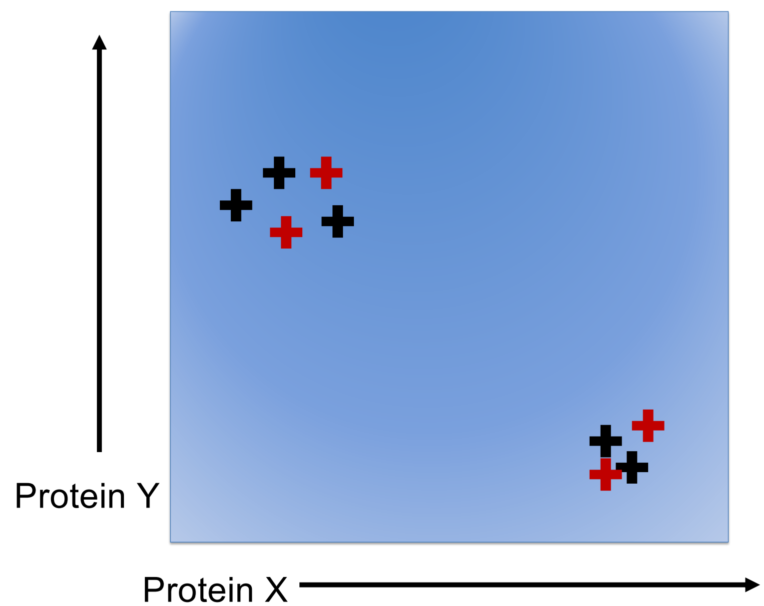
Single Cell Genomics

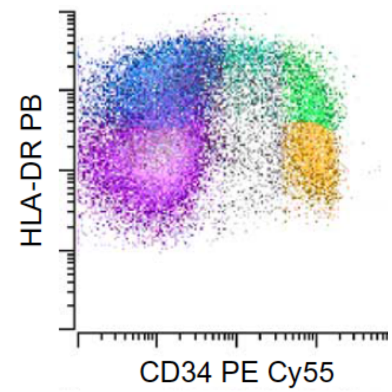
Dana Pe'er
Chair, Computational and Systems Biology
Memorial Sloan Kettering Cancer Center

A Geometric Approach to Phenotype

Cell Phenotype: A configuration of multidimensional cellular features

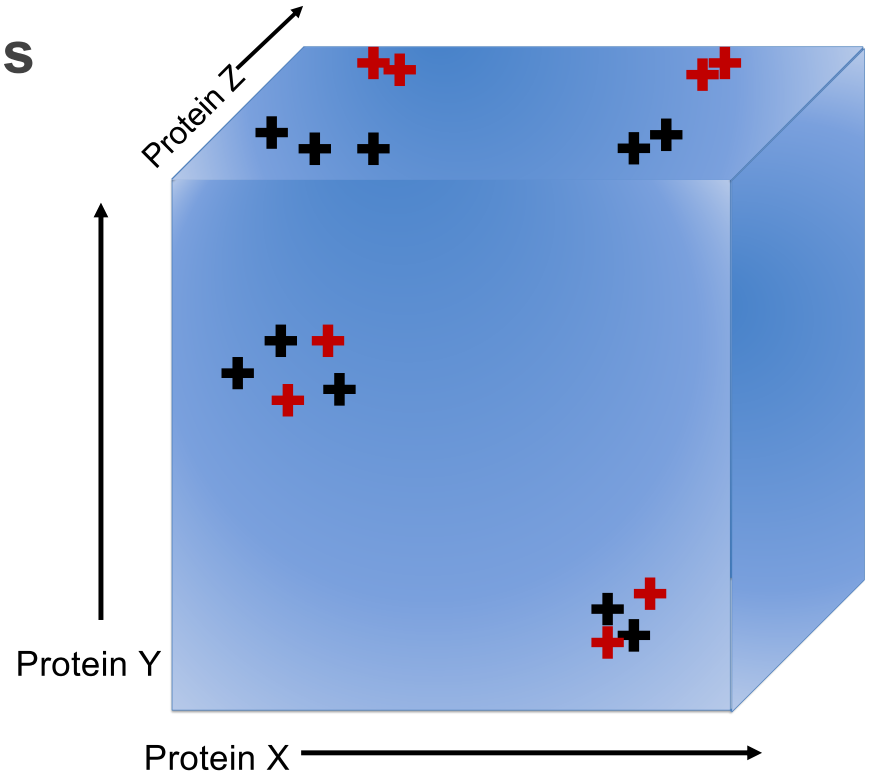
» Defines a region in "**phenotypic space**"





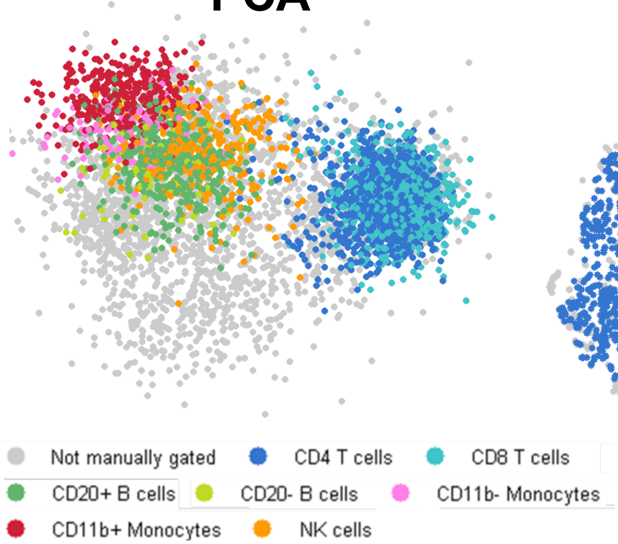
Adding dimensions reveals subtypes

High dimensional single cell technologies: CyTOF, single-cell RNA-seq now generate millions of multi-parameter cells.



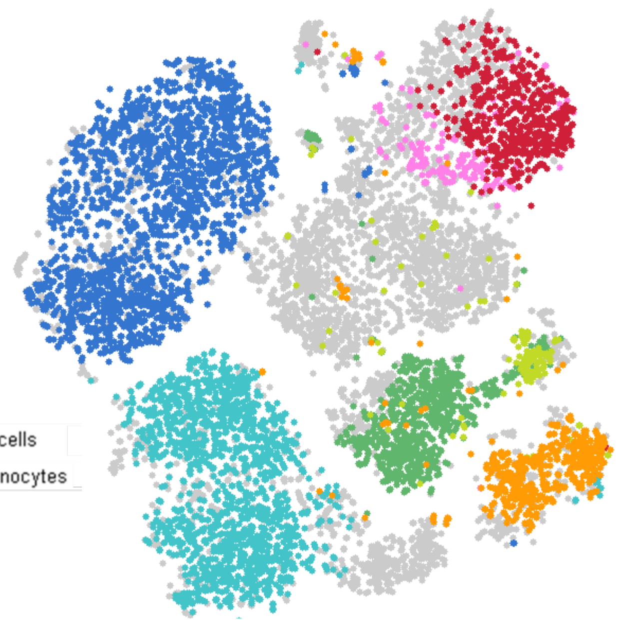
Cell phenotypes accumulate in complex non-linear manifolds

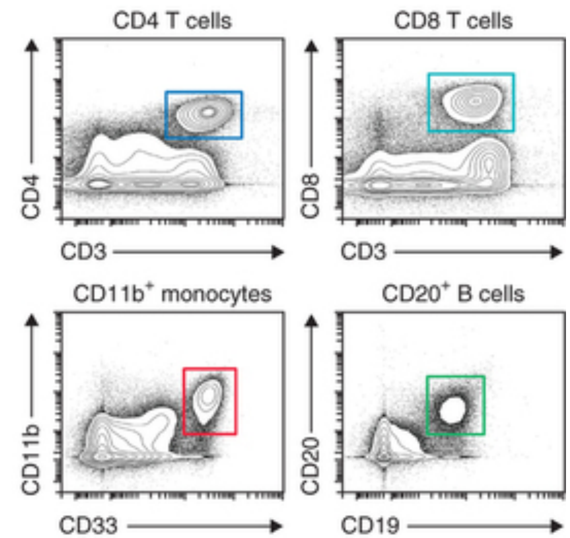
PCA



Amir et. al. Nature Biotech 2013

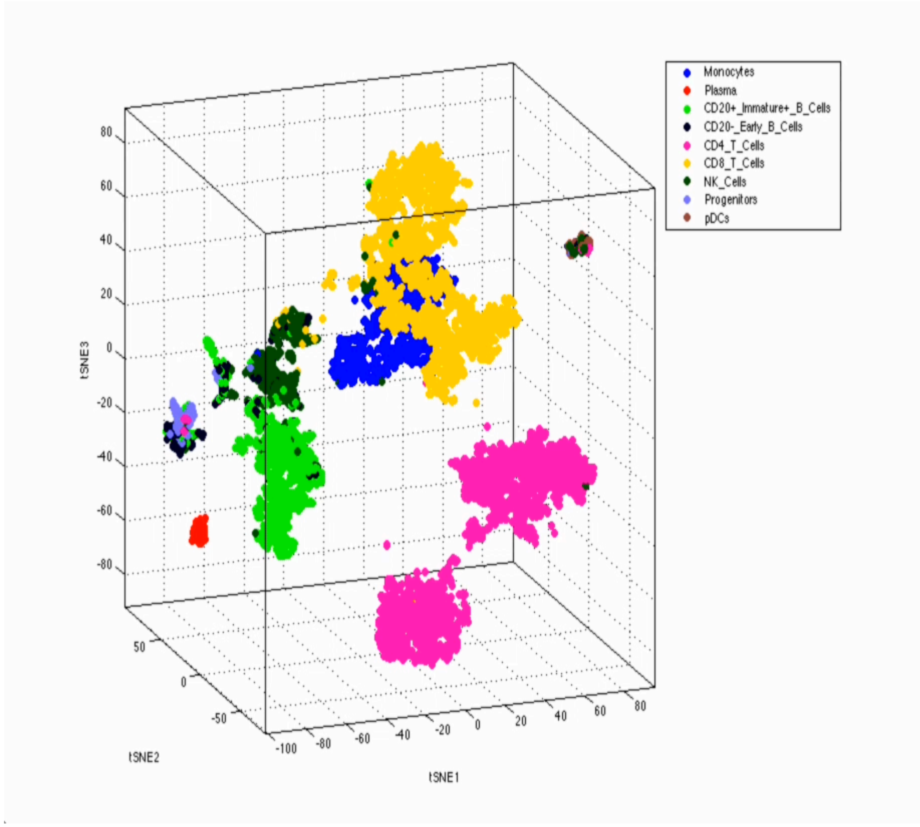
tSNE





Cell phenotypes accumulate in complex non-linear geometric shapes

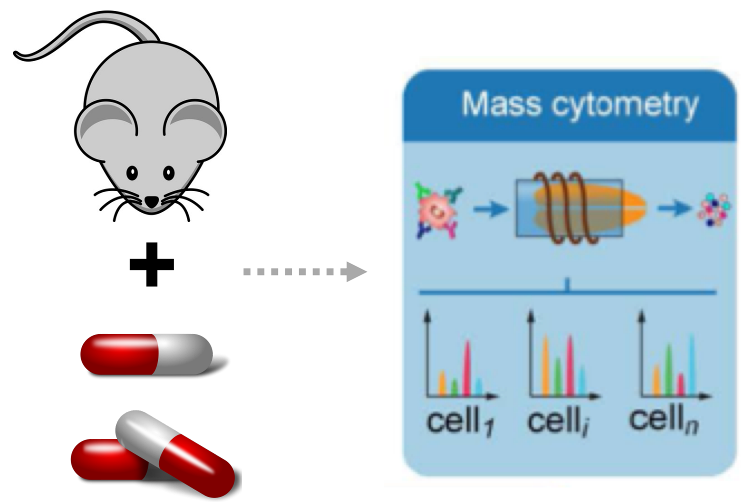
- » Cells accumulate in densities – robust states
- » High-dimensional data but low dimensional structure:
  - › Data lies on “Manifold”
- » This defines “cell types”
- » UMAP is more commonly used for immune subsets



Clustering: Dissecting response to checkpoint blockade

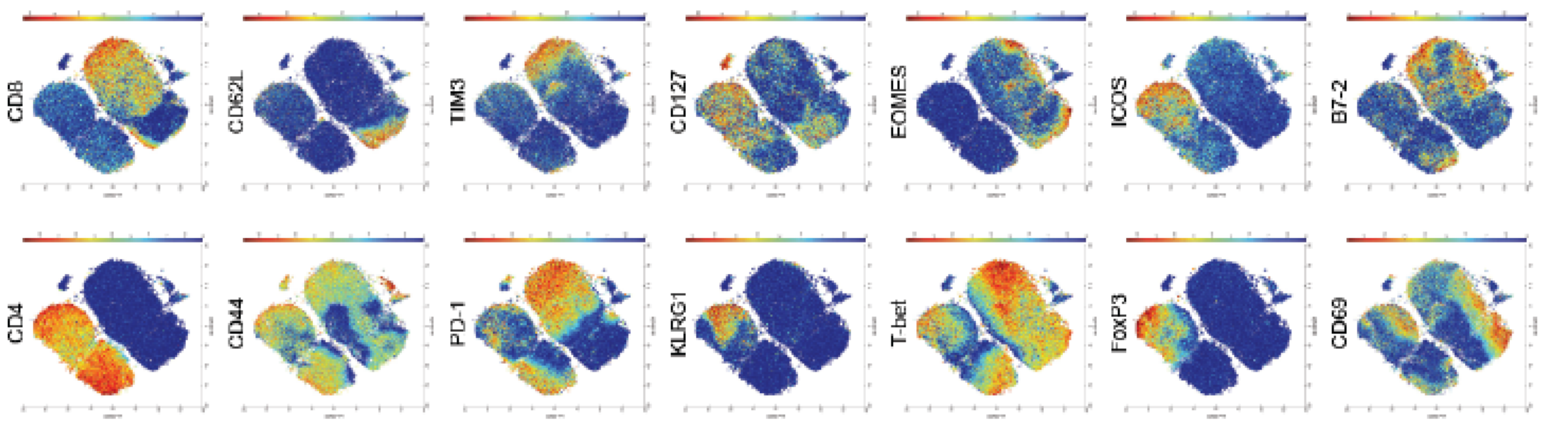
» Measured response to anti-PD1 and anti-CTLA4 in MC38 colorectal tumors.

» 33 surface and 10 intracellular markers: Cell type(e.g. CD8, CD4, CD11b, CD19), T cell differentiation and activation markers (e.g. PD-1, ICOS, TIM3, KLRG1, CD127), T cell lineage transcription factors (e.g. T-BET, EOMES, GATA3, BCL6). This gives 528 biaxial plots.



Wei et. al. Cell 2017

Rich Heterogeneity observed in data



How can we interpret this?

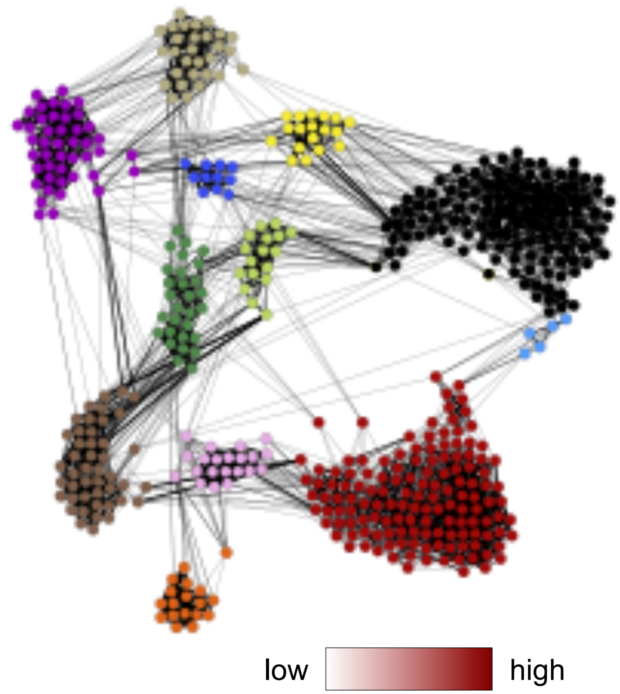
» It is hard for us to interpret the story looking one marker at at time

» Key mistake: NEVER over interpret 2D projections of the data. These can very misleading.

» I call this ”tSNE tea reading” and have often led to the wrong conclusion.

Clustering: unbiased characterization of subpopulations

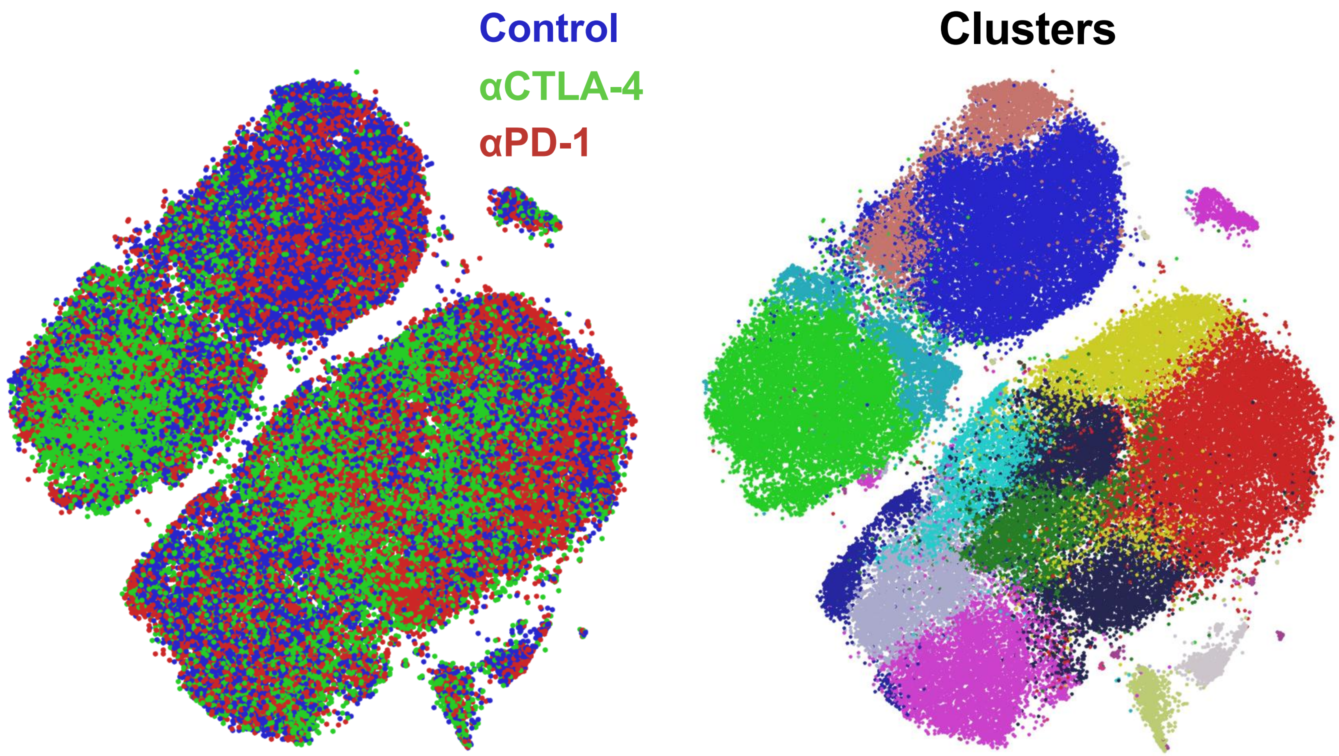
- » Instead of gating, a data-driven approach
- » There are many clustering approaches to decompose the entire dataset
- » Most common approach is graph-based clustering:
  - › Each node is a cell
  - › Each cell is connected to a neighborhood of “similar” cells
  - › Use methods derived from social networks



Levine*, Simonds* et. al. Cell 2014

Tumor infiltrating T cell subsets

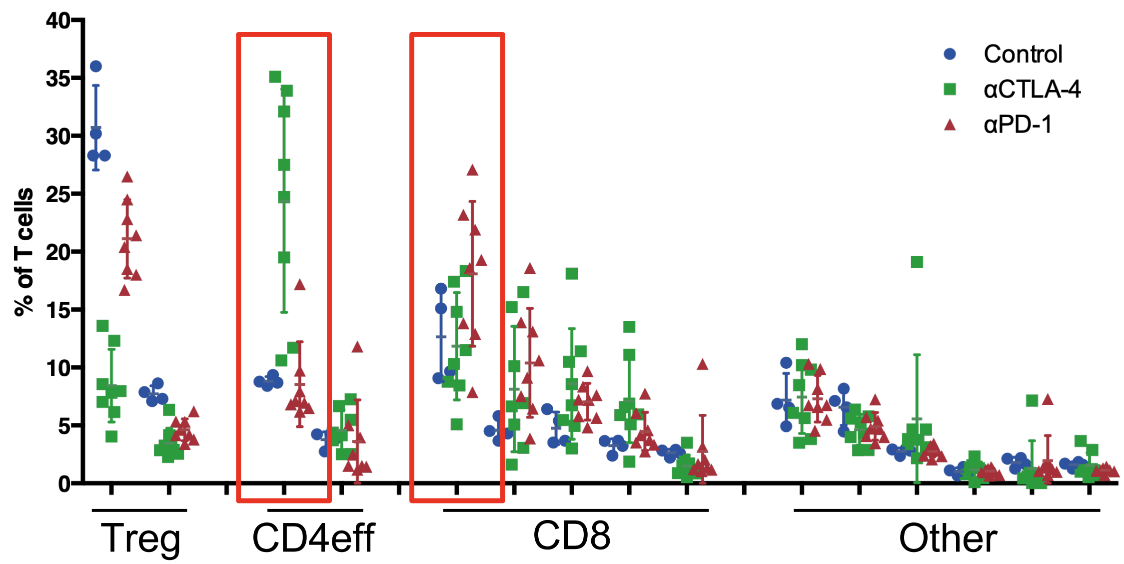
15 distinct tumor infiltrating T cell clusters found



Memorial Sloan Kettering Cancer Center

Population dynamics in response to therapy

- » CD8 expanded following both therapies, but not all CD8 clusters expanded equally
- » Expansion of CD4 phenotype only in response to CTLA-4, but does not change after PD-1
- » Note: Cluster proportion can be statistically unstable, check that your clustering is robust



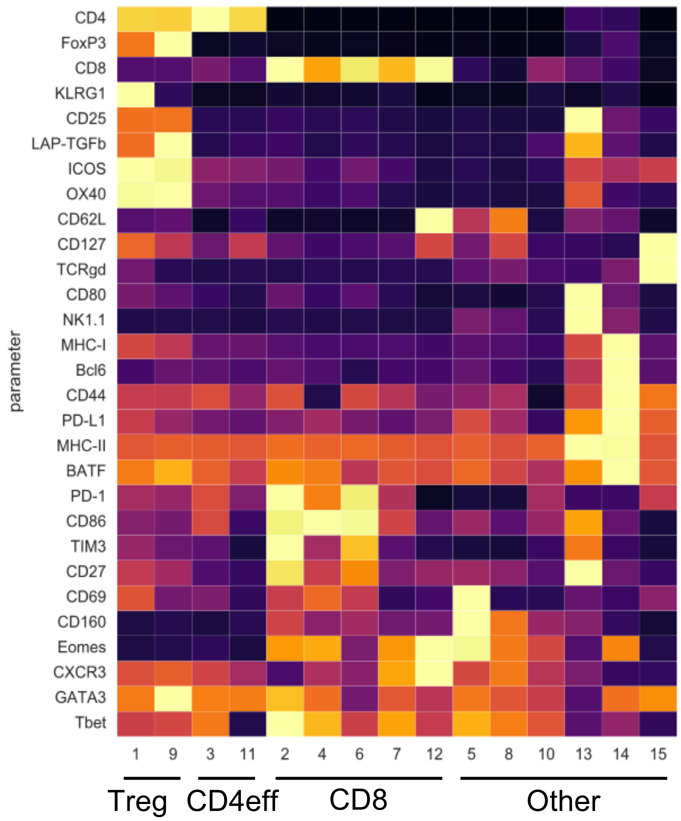
Interpreting clusters with heatmaps

» Heatmaps is a good way to interpret clusters

» TIM3+ PD-1high: Expansion of exhausted population (IdU staining) -> reinvigoration of exhausted phenotype

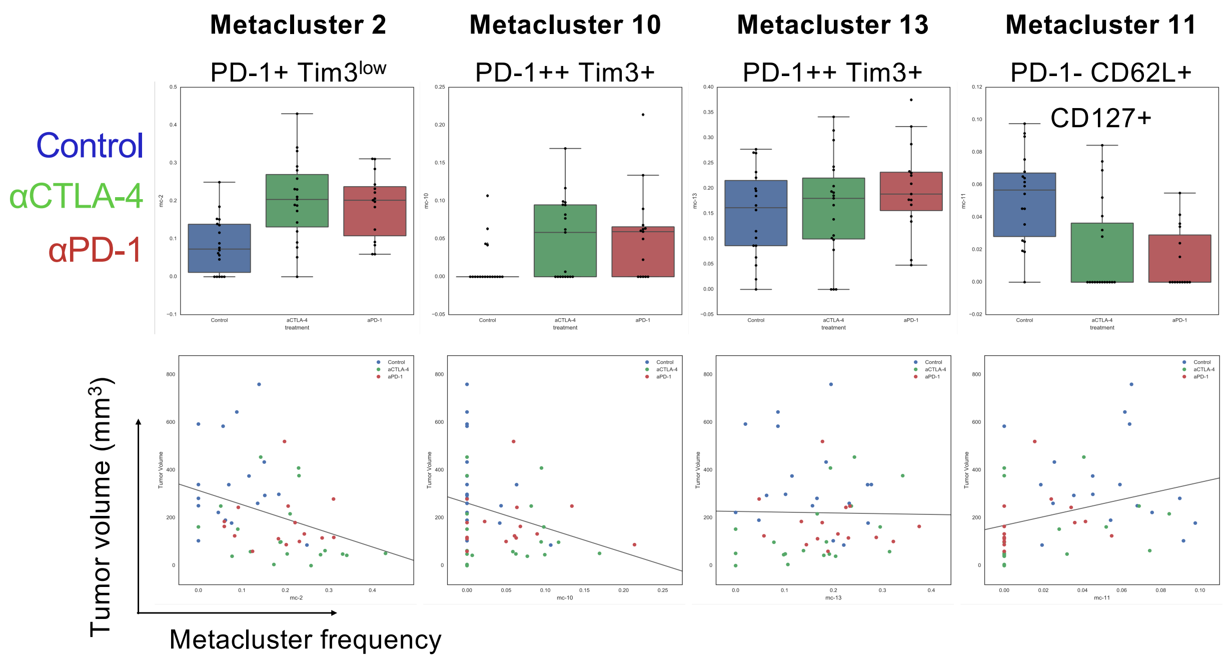
» CD4 Th1-like phenotype (PD-1high ICOSint CD86+ TBET+) expands only in response to CTLA4, low IdU staining suggests increased differentiation or infiltration.

» Data demonstrates different mechanism of action for anti-PD1 and anti-CTLA4

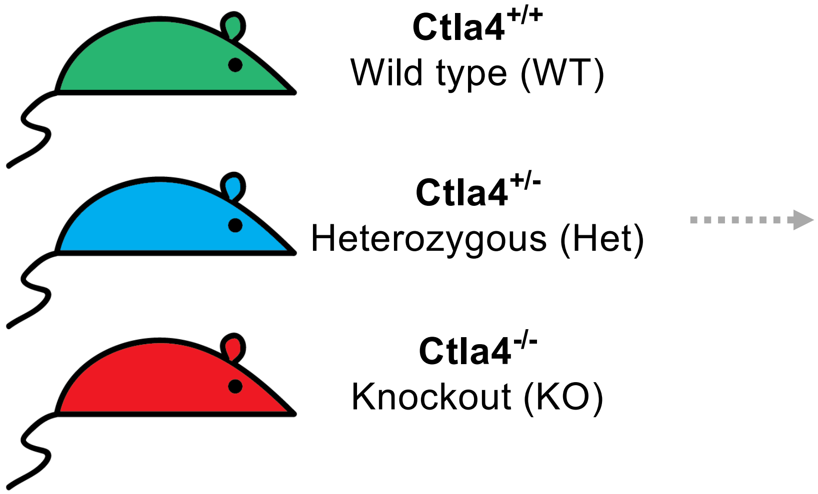


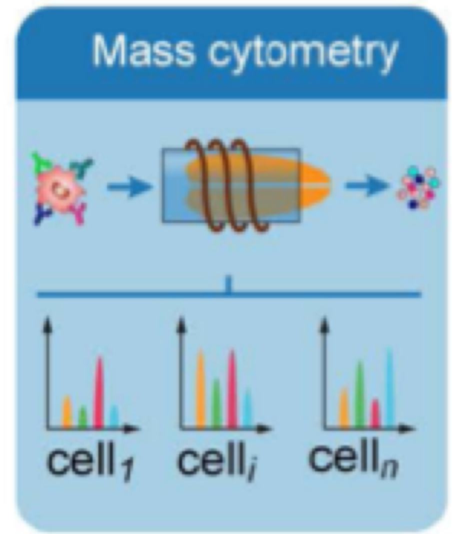
T cell subsets that correlate with tumor growth

- Only 2 CD8+ populations correlated negatively with tumor growth.
- Subtle multivariate phenotypic differences, distinguish T cell populations with dramatic functional differences



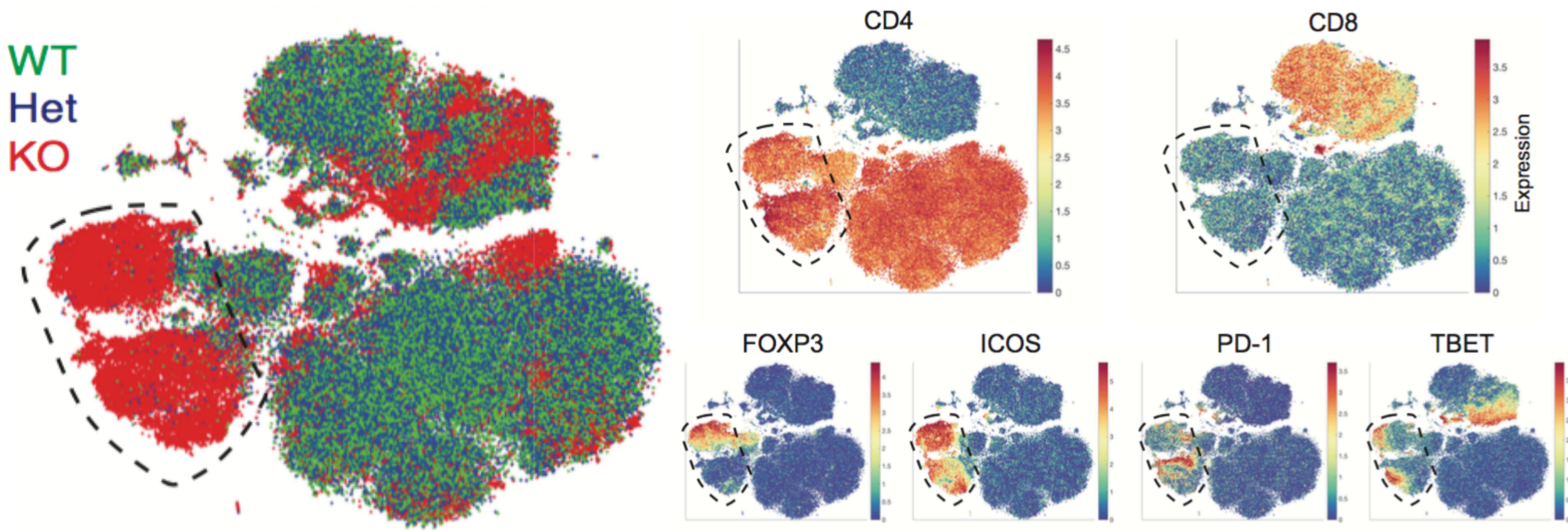
Archetypes: Effect of negative costimulation on T cell differentiation





Wei*, Sharma* et. al. Immunity 2019

Loss of CTLA-4 has dramatic effect on CD4 T cells



Archetypes verses clusters

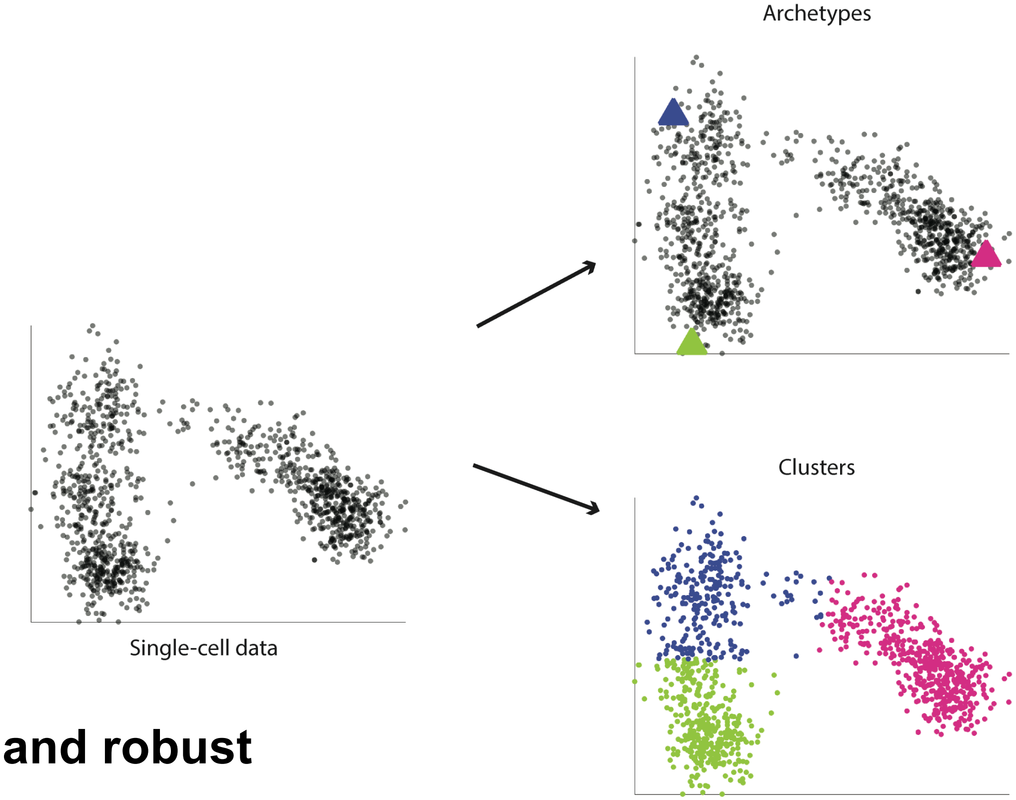
» Clusters identify centroids of discrete populations

» Immune phenotypes are often not well separated

» Archetypes characterize the "extreme" conditions in the data.

» They often relate to a biological process (e.g. exhaustion)

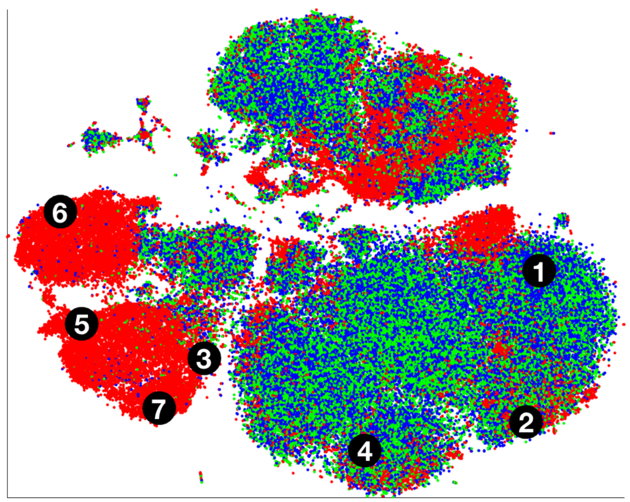
Archetype analysis is very stable and robust

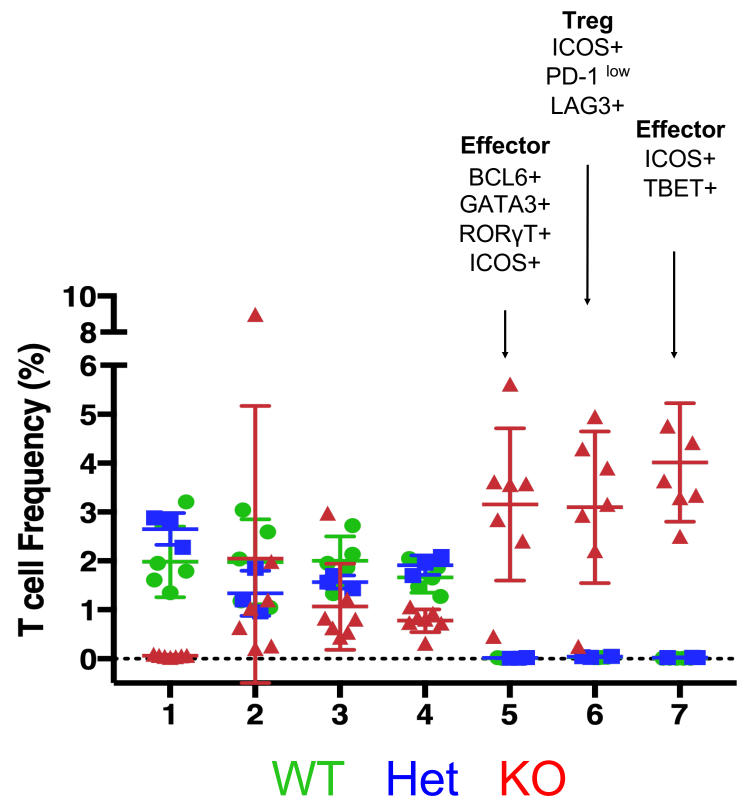


A Geometric Approach to Phenotype

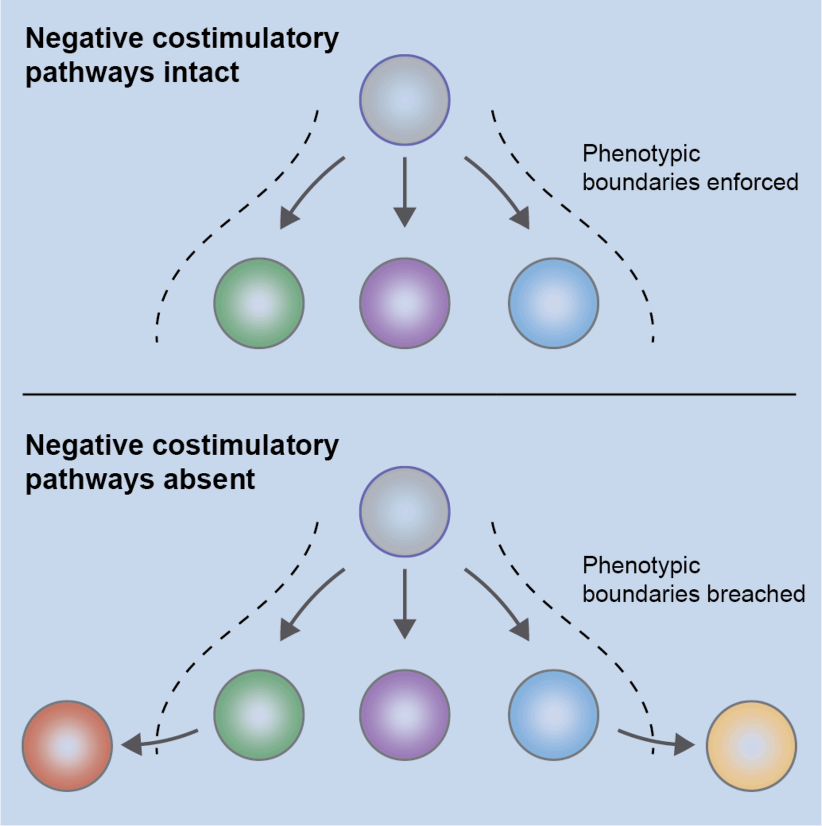
During development inhibitory signal from CTLA4 defines limits on T-cell phenotypes

CD4 archetypes



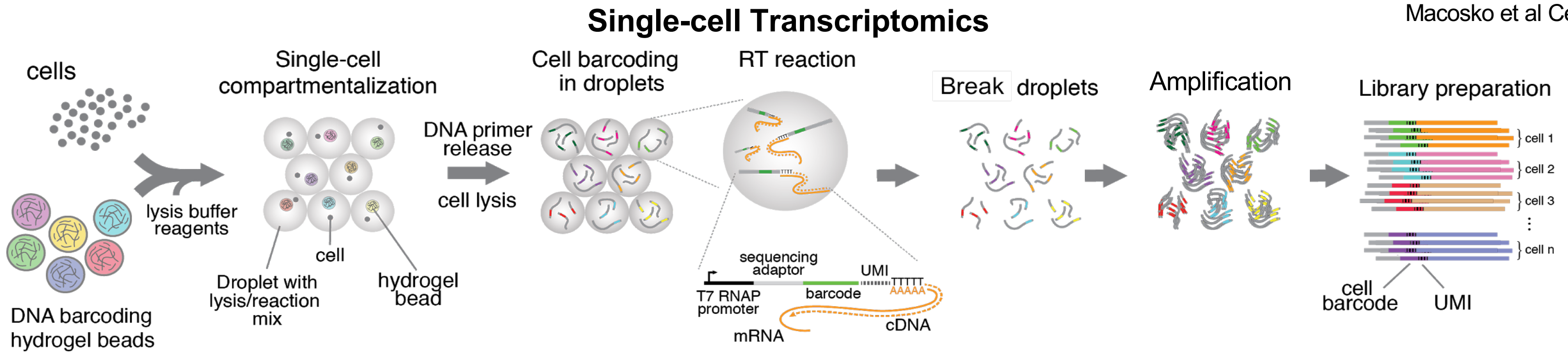


Negative costimulation constrains T cell differentiation



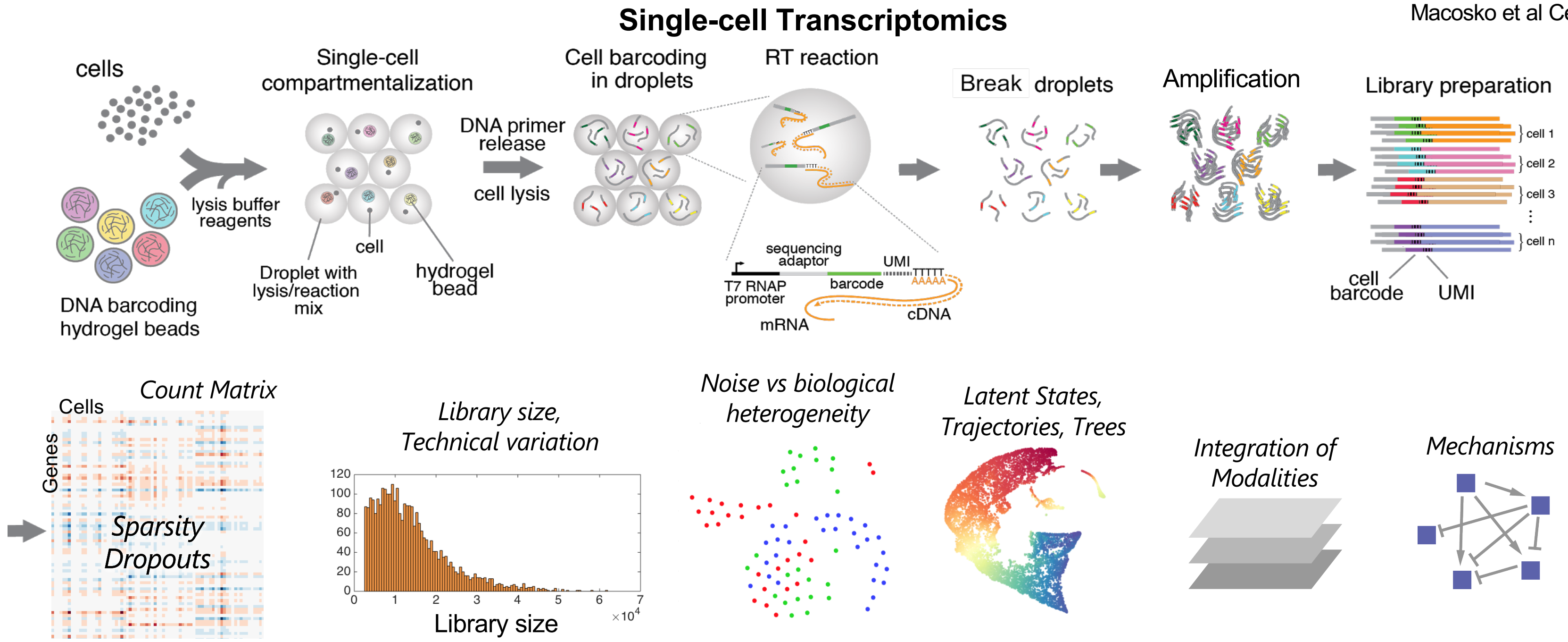
Microfluidics: Single-cell RNA-seq across thousands of cells

Klein et al Cell 2015
Macosko et al Cell 2015



Single-cell data analysis involves major computational challenges

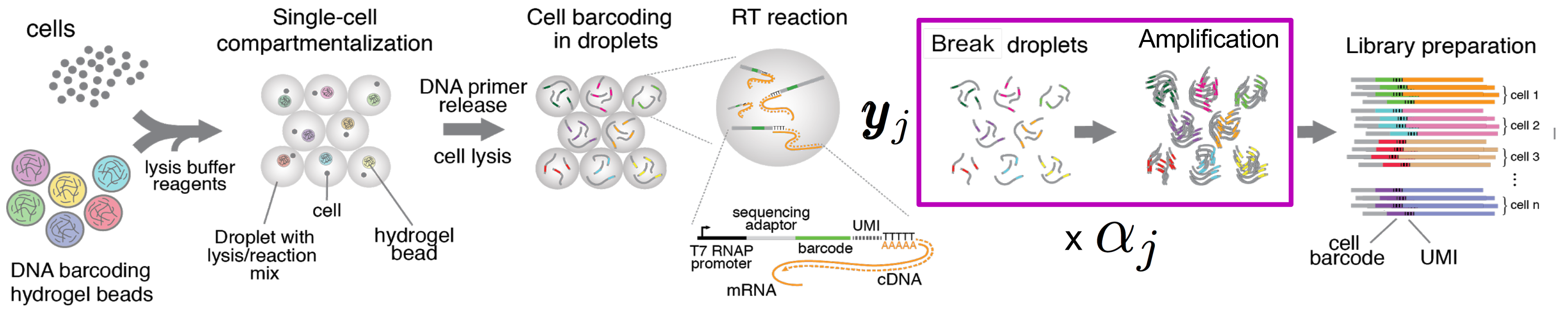
Klein et al Cell 2015
Macosko et al Cell 2015



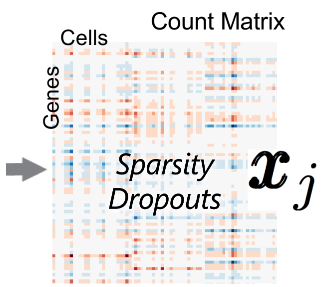
Single-cell data analysis involves major computational challenges

Klein et al Cell 2015
Macosko et al Cell 2015

Single-cell Transcriptomics



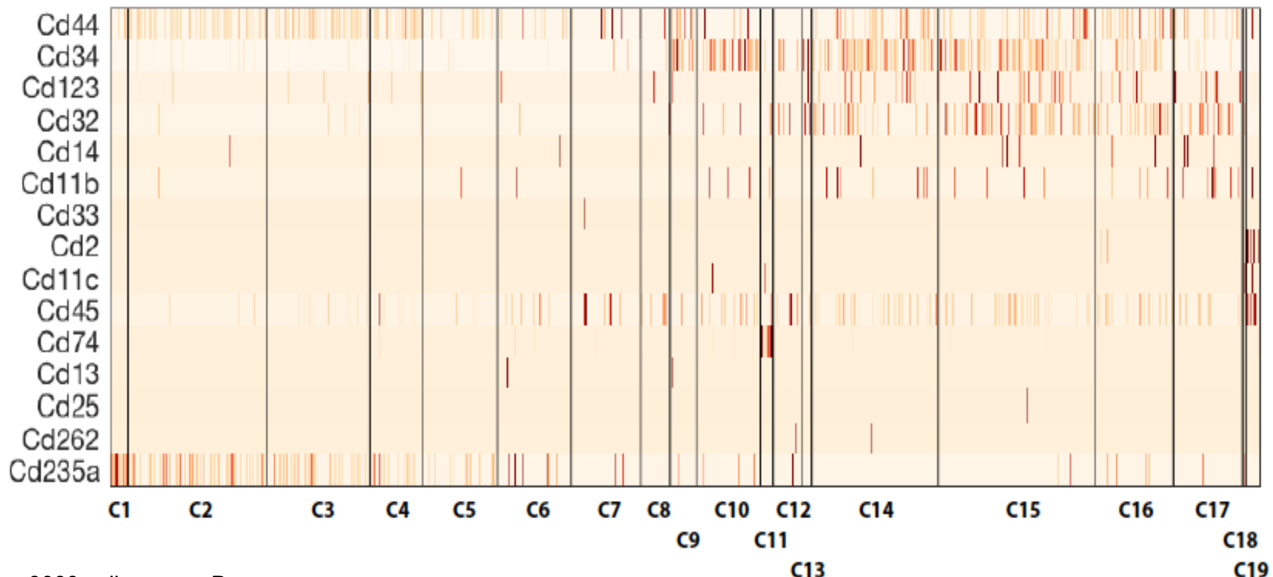
Each cell can have a different amplification rate, capture efficiency, RNA degradation, etc
→ different library size



Measured expression in cell ***j*** is noisy

Single-cell RNA-seq samples 5% of transcripts in each cell

- Surface markers used for gating typically have very low RNA-levels and are poorly captured in most immune cells.
- For example: Monocyte clusters have
  - 1.6% cells expressing CD14
  - 5.8% cells expressing CD11b
- The power comes from measuring many cells



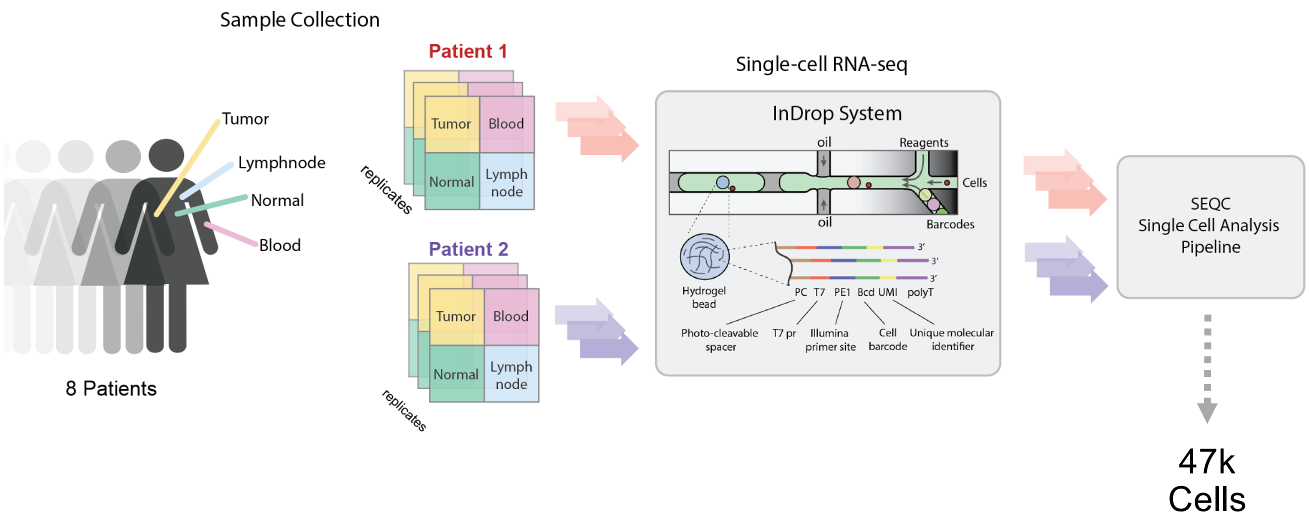
3000 cells mouse Bone marrow
Paul, et.al. Cell 2016

Characterization of tumor immune cells in breast cancer

Data-Driven approach:
> 3000-10,000 CD45+ collected per tumor

» What is the immune states and the structure of the tumor immune ecosystem?

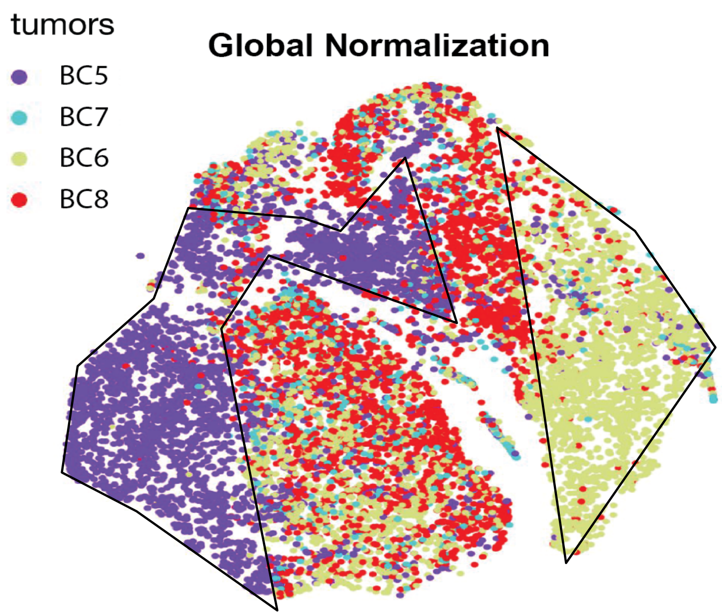
» How do cell subsets differ between tissue microenvironments?

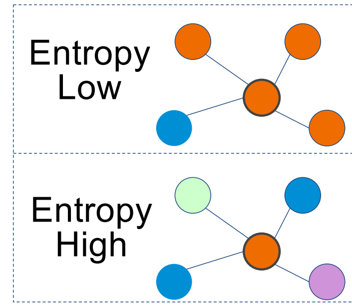


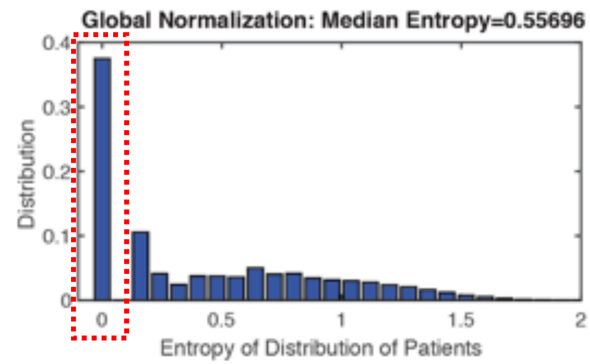
Azizi*, Plitas*, Carr*, Cornish* et. al. Cell 2018

Significant batch effects confound multi-tumor analysis

All batch correction algorithms make strong assumptions and have trade-offs: there is no free lunch

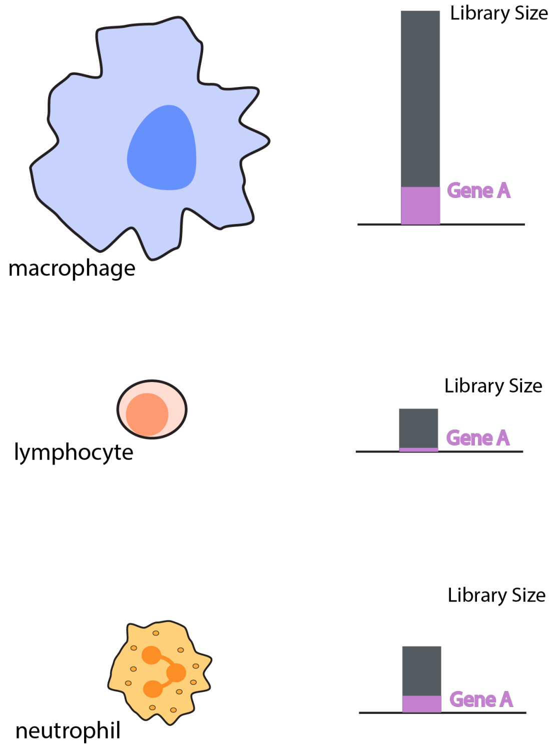






Normalization is a key unresolved problem in data analysis: Most common approach global normalization

Memorial Sloan Kettering
Cancer Center

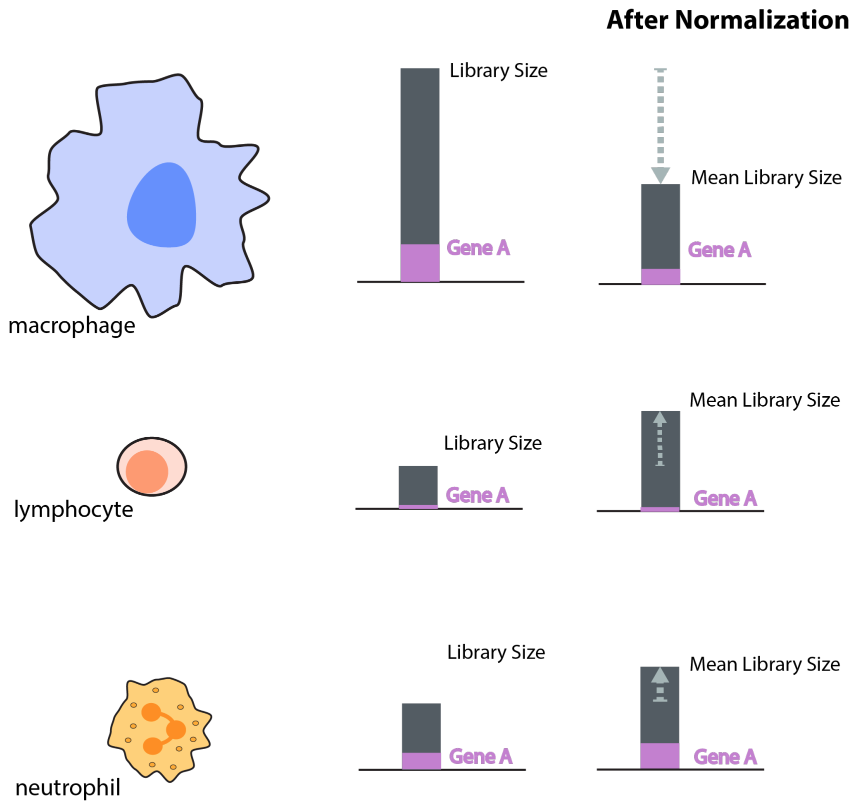


Cells with different sizes have very different total number of transcripts

Example Housekeeping Gene

High chance of Dropouts in smaller cells

Problem with Global Normalization

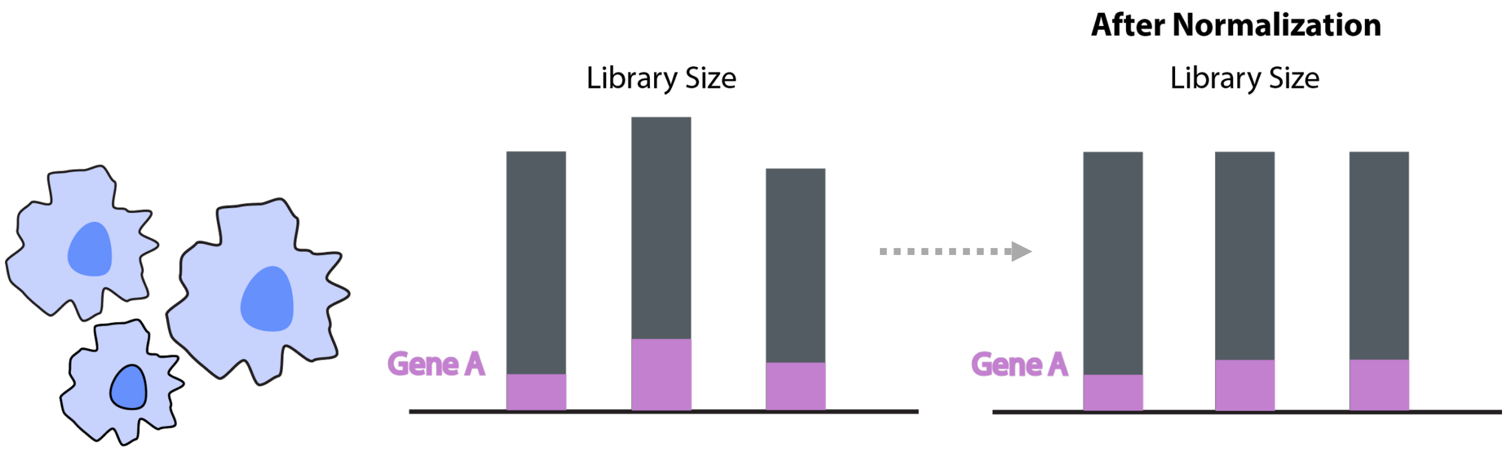


Note: Most bulk methods fail miserably on single cell RNA-seq

Dropout not resolved

Spurious Differential Expression

Different normalization for each cell type



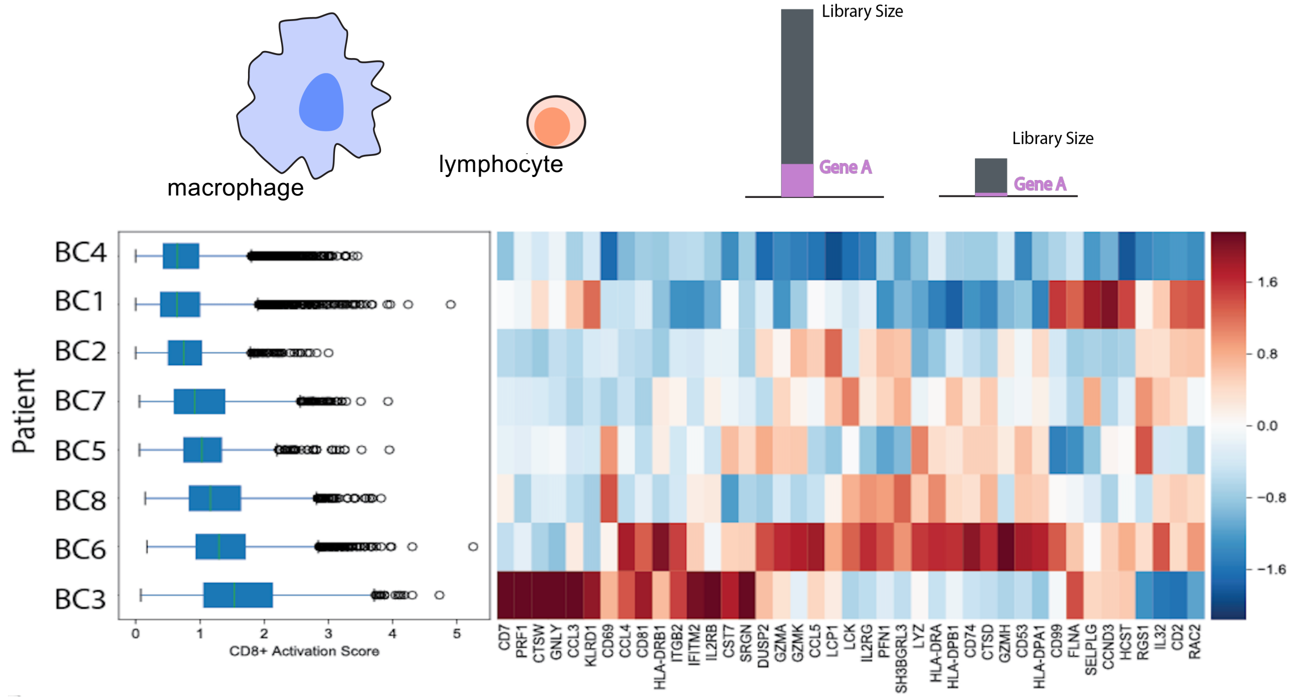
I recommend SCRAN for immune datasets

There are strong batch effects with and between samples

» Differences between cells / samples convolute both biological and technical reasons (e.g. active T-cell also have more transcripts)

» Normalization methods assume similarity and often remove biological differences as well

» There is no free lunch, but one should search for the best trade-off



Biscuit improves clustering

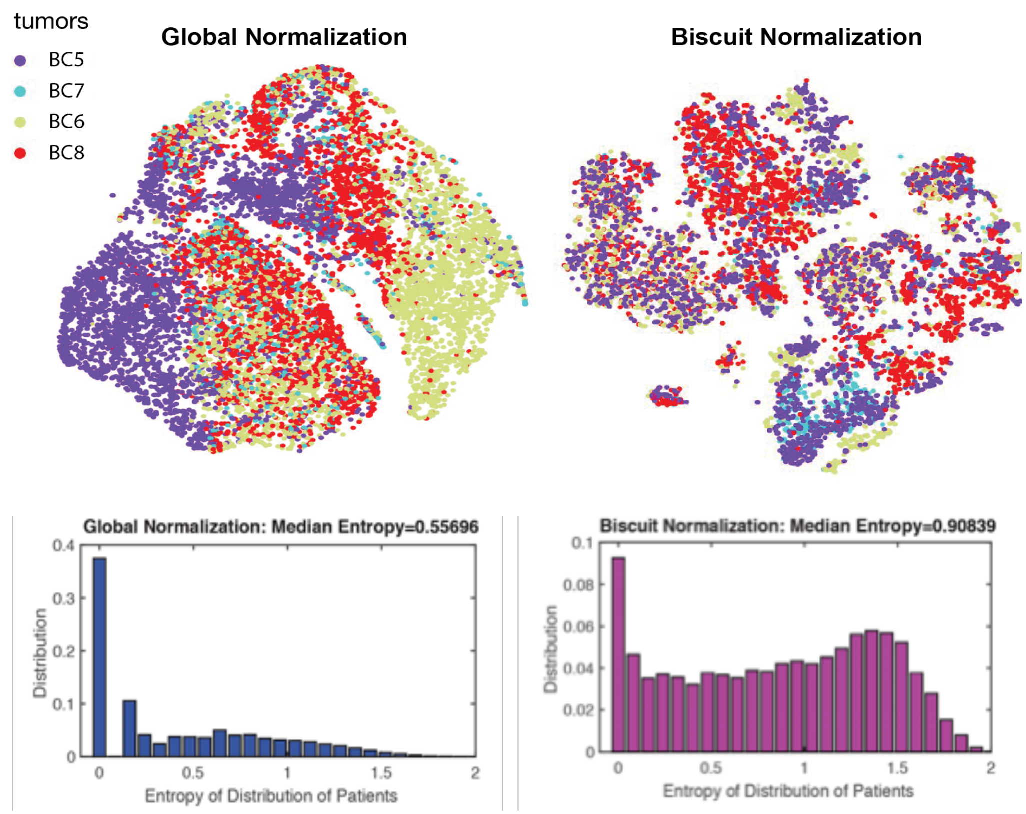
» Samples mix well after Biscuit: high entropy

» Biscuit is robust

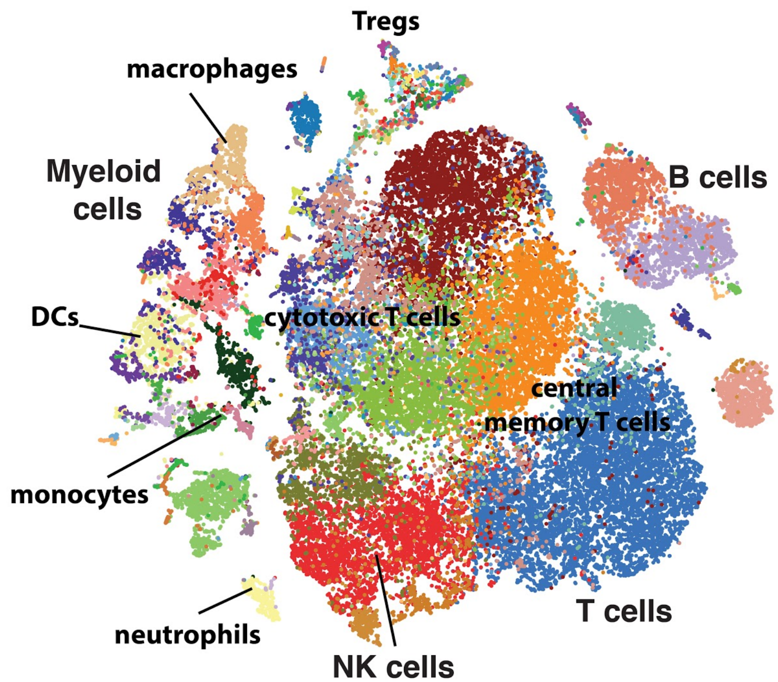
» Biscuit's parametric model can provide DEGs

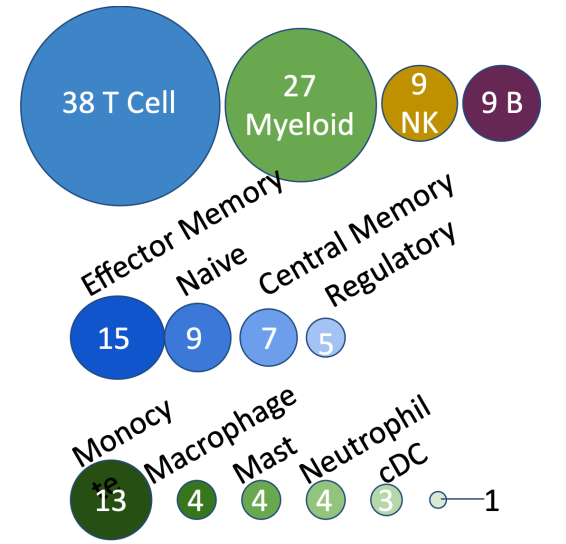
» However: Biscuit is computationally heavy

Prabhakaran*, Azizi* et. al. ICML 2016



Single-cell Immune atlas of ~50,000 cells showing 83 distinct states



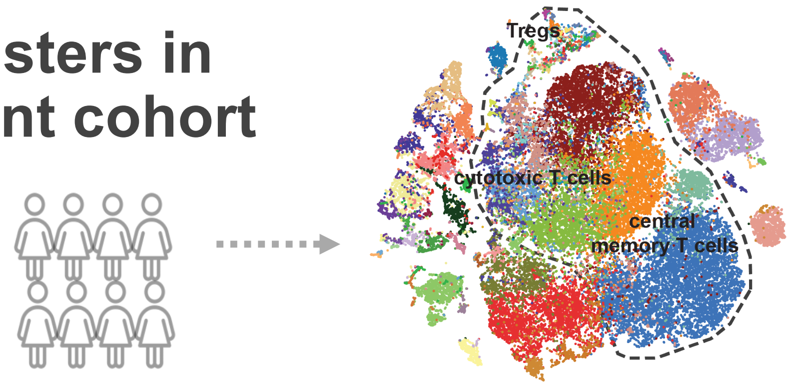


What are all these states and is the clustering reliable?

Azizi*, Plitas*, Carr*, Cornish* et. al. Cell 2018

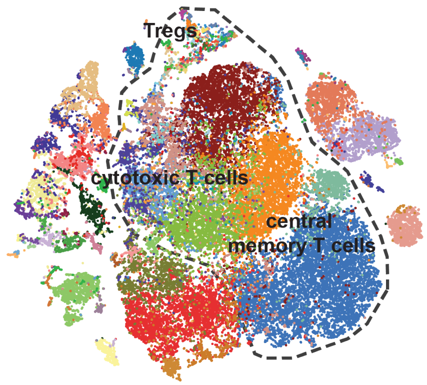
Validating clusters in an independent cohort

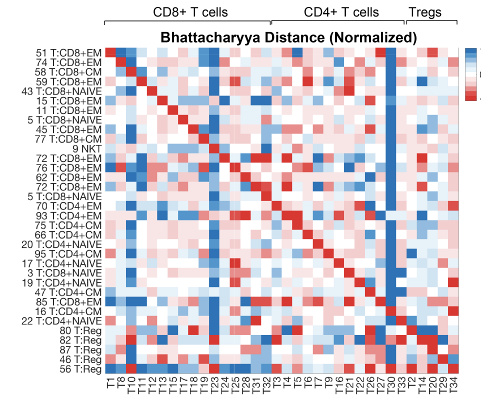
CD45+ cells in BC1-8 (inDrop)



Validating clusters in an independent cohort

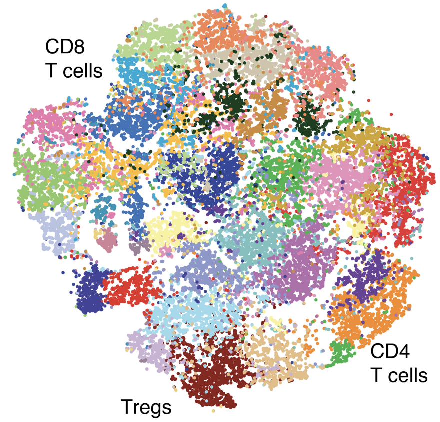
CD45+ cells in BC1-8 (inDrop)





Three new patients

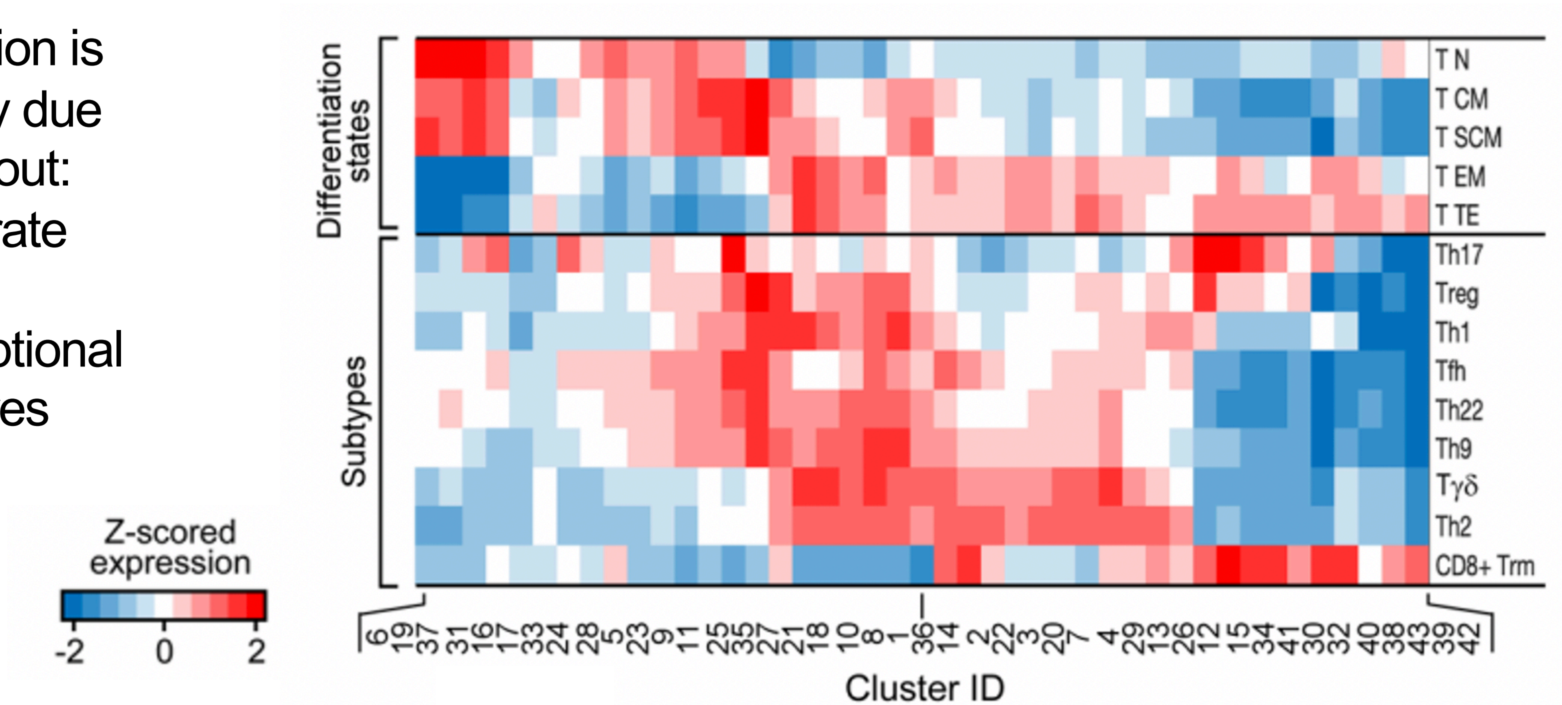
CD3+ cells from BC9-11 (10x)



Near one-to-one mapping to 34 T cell clusters

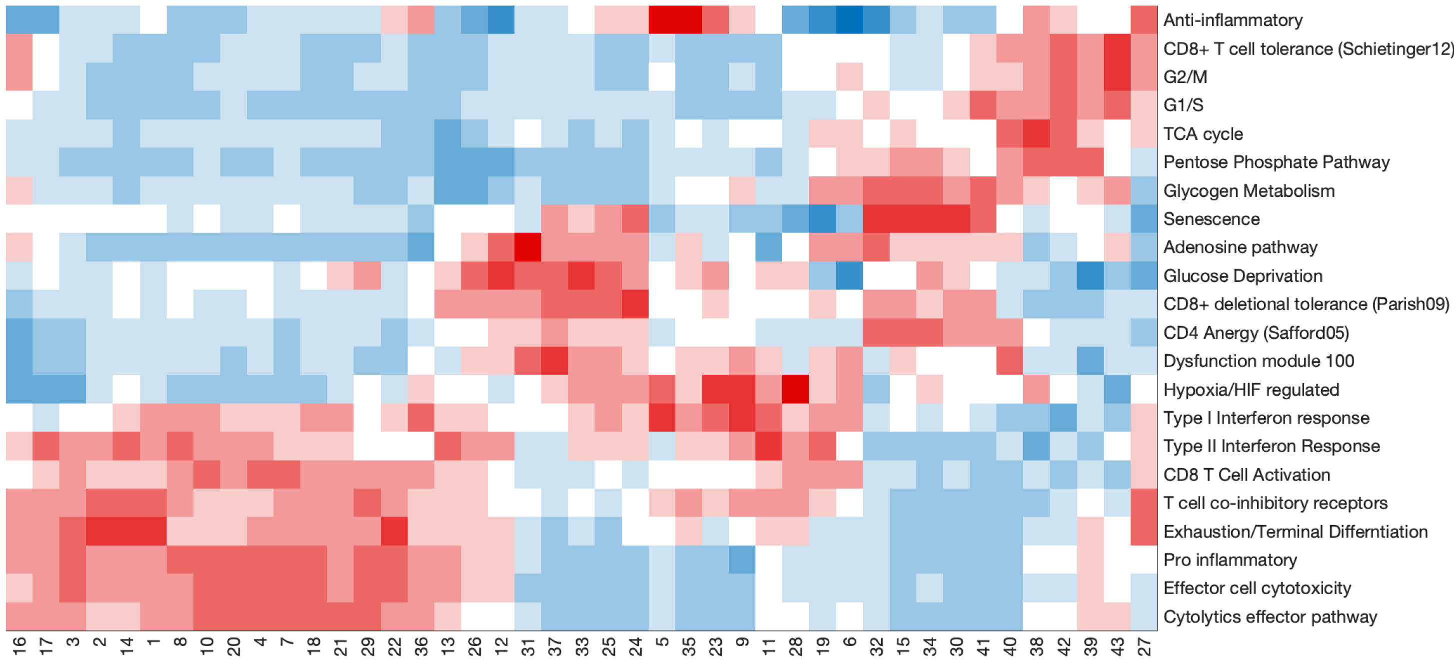
Clusters vary in subtype and differentiation state

Annotation is not easy due to drop-out: incorporate larger transcriptional signatures

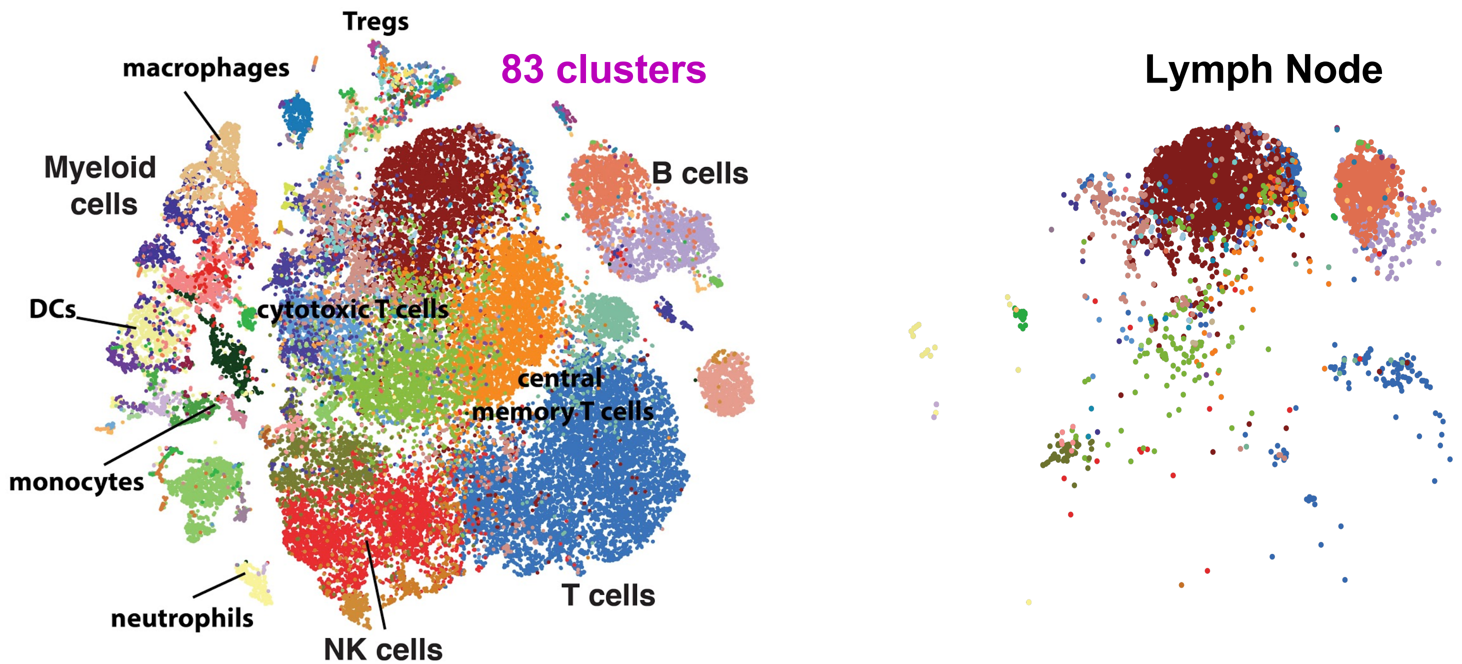


Metabolic and Immune Programs also vary across

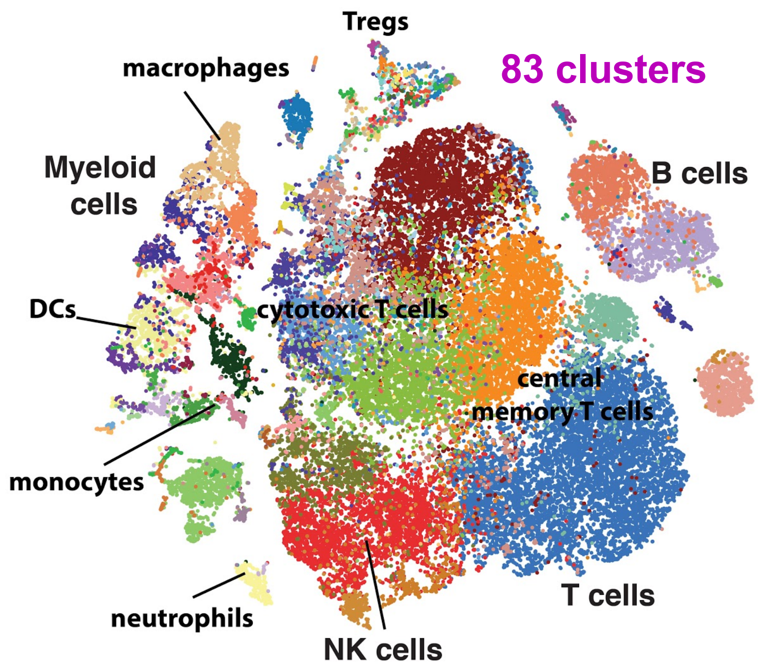
The quality of your gene signatures sets the quality of your annotation, curate these from good sources suitable for your biology

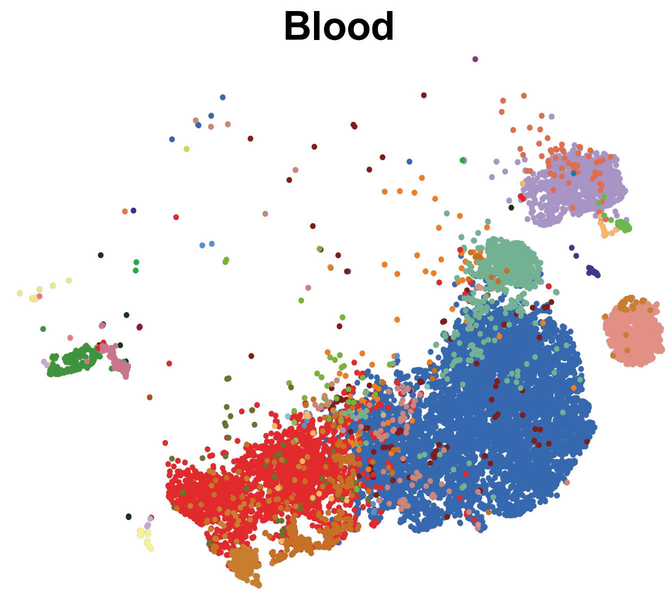


Atlas enables comparing immune states across microenvironments

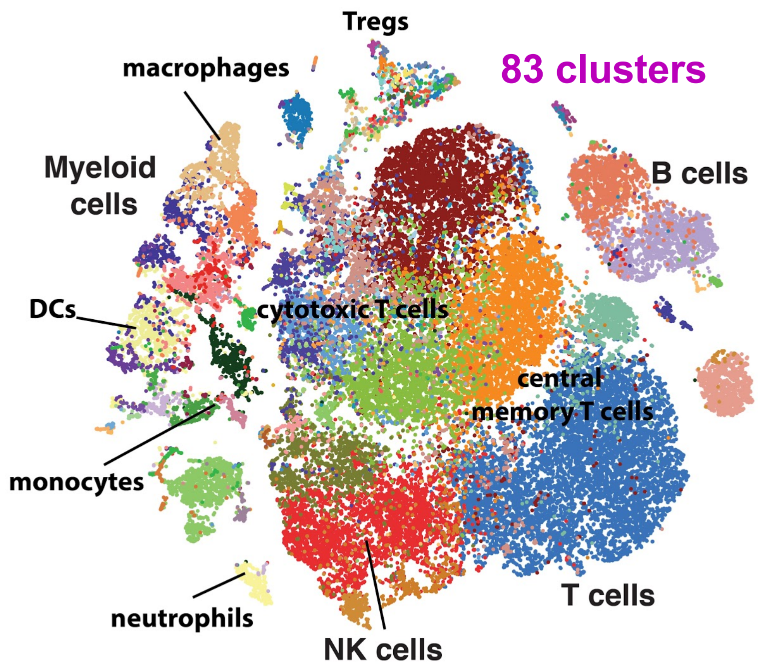


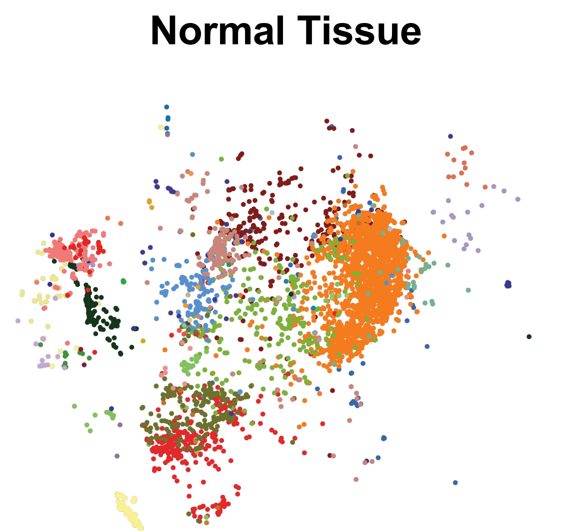
Atlas enables comparing immune states across microenvironments





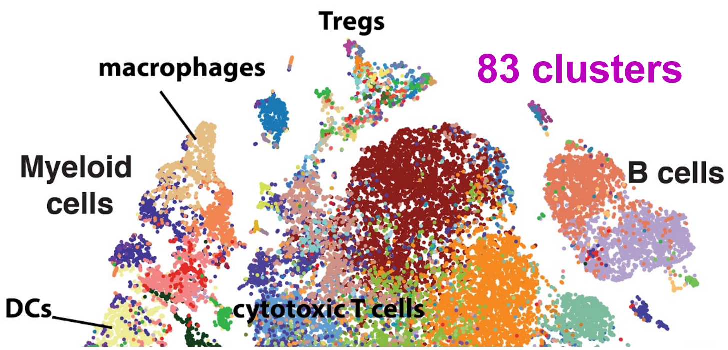
Atlas enables comparing immune states across microenvironments



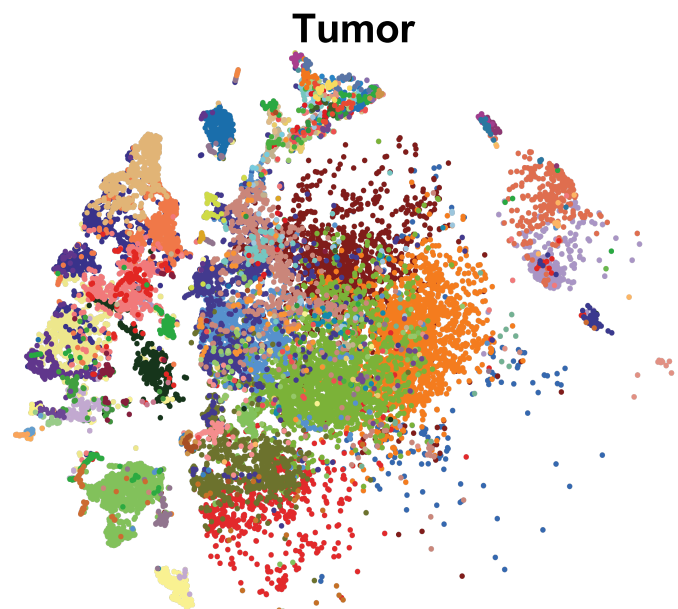


Tissue resident immune cells are dramatically different than those in immune organs

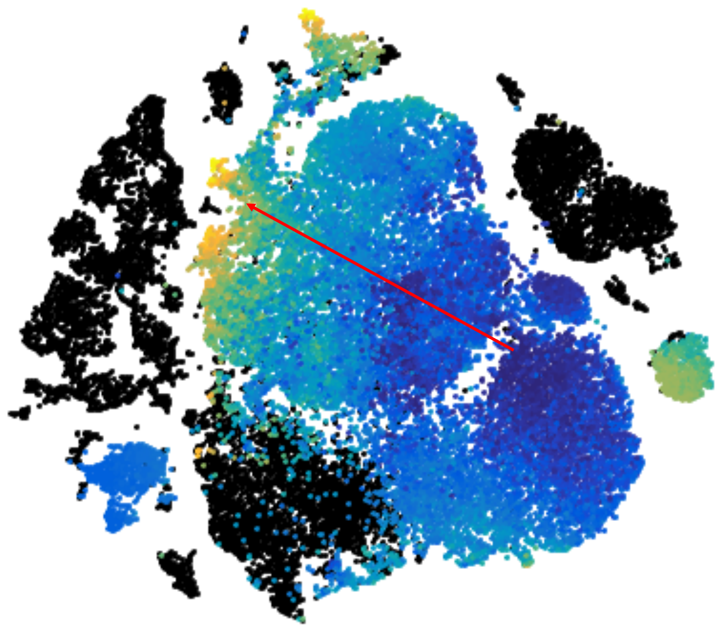
Atlas enables comparing immune states across microenvironments

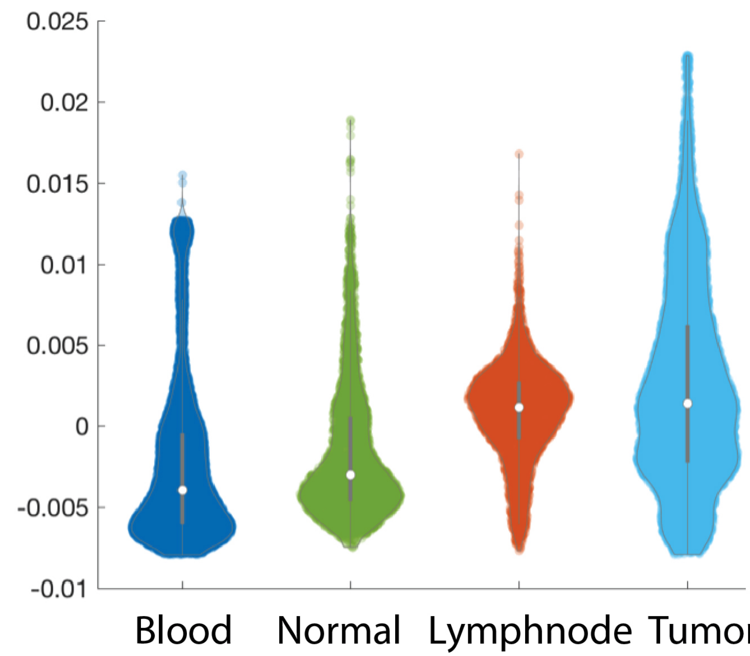


- » 19 T-cell clusters shared between tumor and breast-tissue
- » 17 T-cell clusters unique to tumor are more activated and more cytotoxic



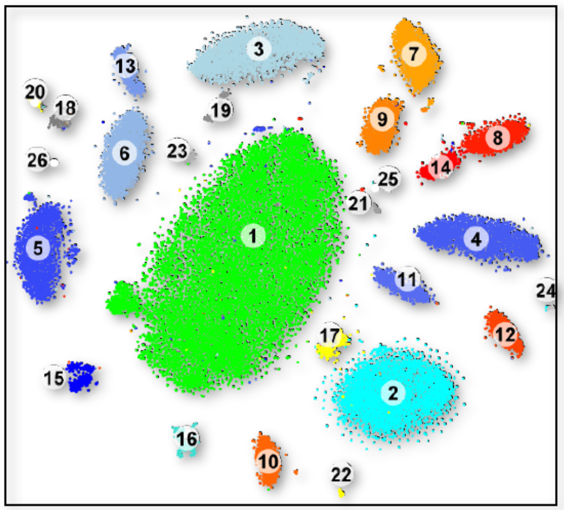
Intratumoral T cells reside on a continuous activation trajectory!



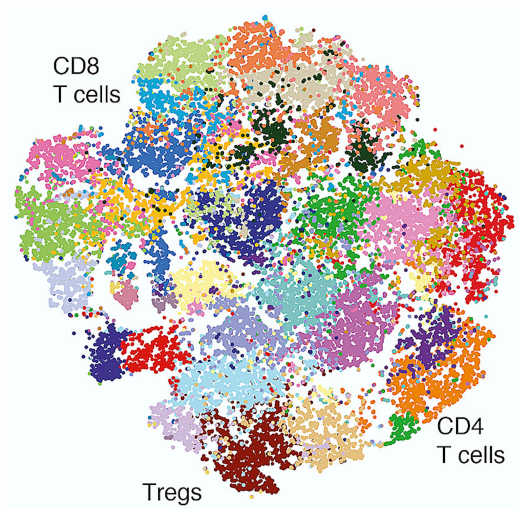


Some cell types are more well separated than others

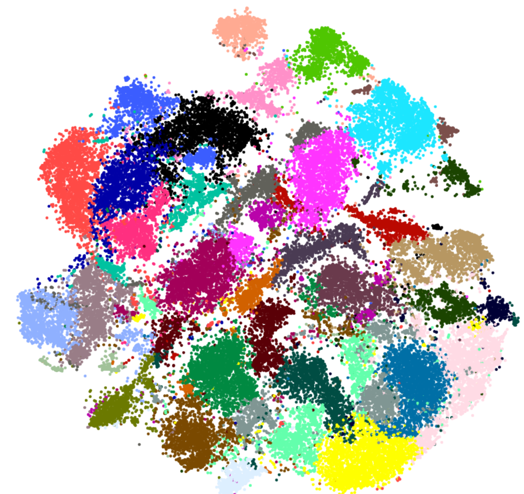
Biopolar Cells



T Cells in Breast Cancer



T Cells in Lung Cancer



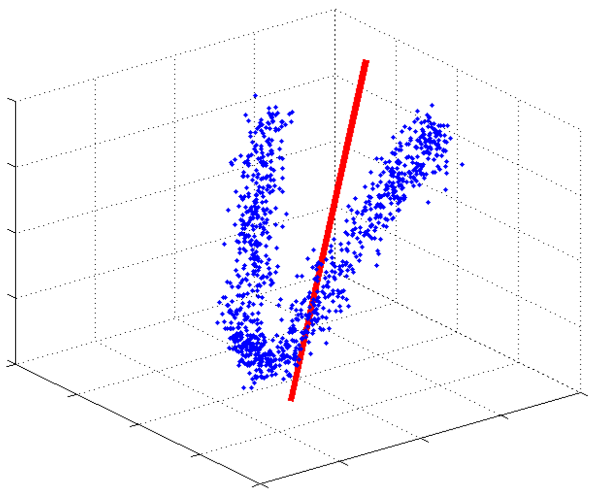
Shekhar et al. Cell 2016
Data from Azizi et al. Cell 2018
Data from Laughney et al. Nat Med 2020

Neuronal cells are far better separated than T-cells

Diffusion components capture continuous trends in data

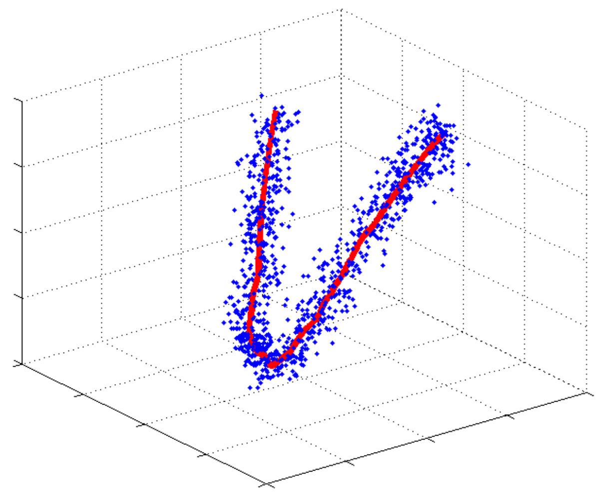
Diffusion maps are a "non-linear" version of PCA that follows the data density

PCA



Linear direction of variation

Diffusion maps

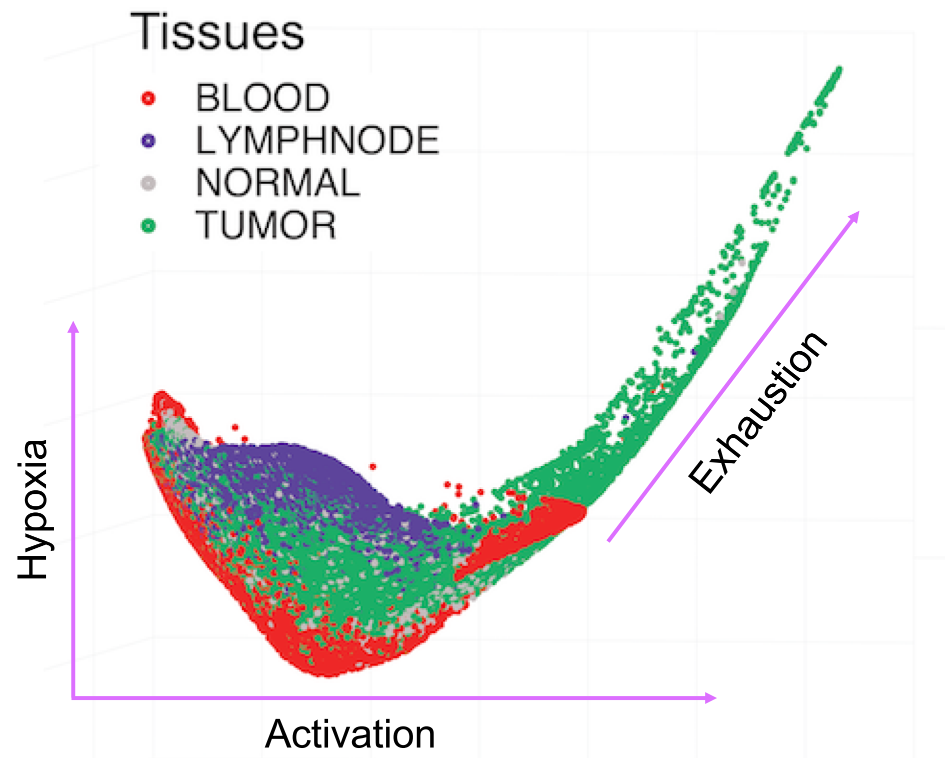


Trajectory through the manifold

Memorial Sloan Kettering Cancer Center

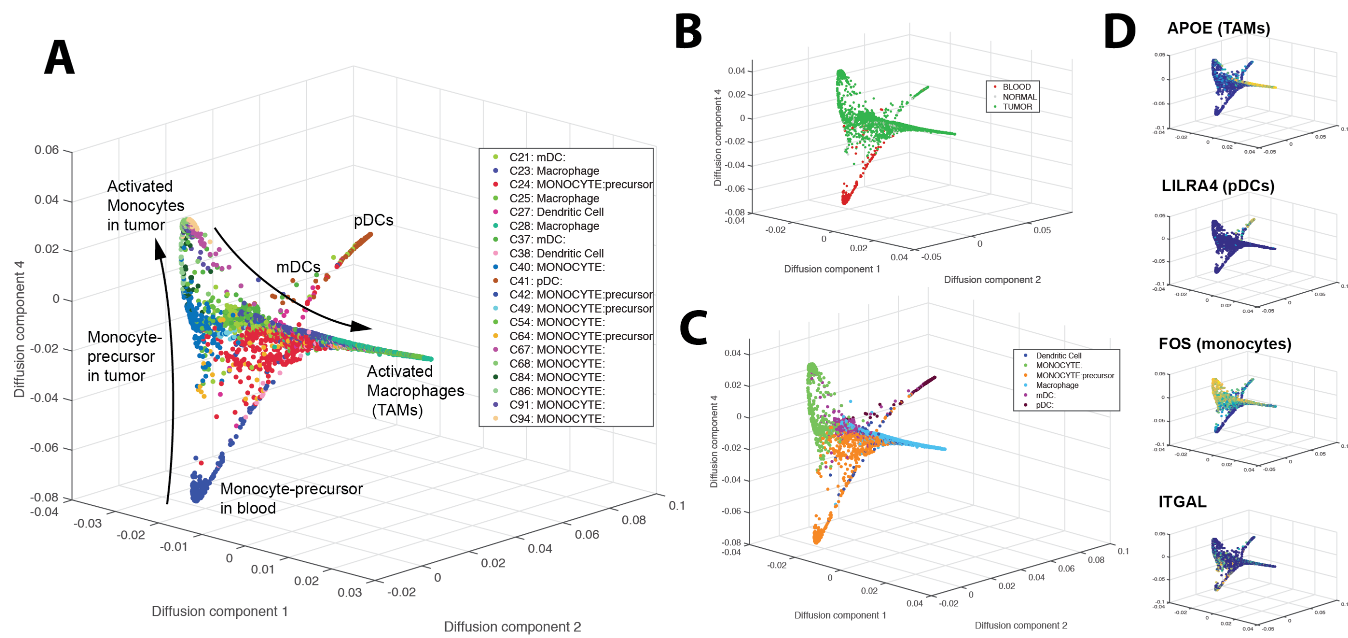
T-cell phenotypic space is continuous

- T-cells define a continuum of states
- Most of the variation can be captured by a few axes of variation
- First diffusion component that explains most of variation is T cell activation



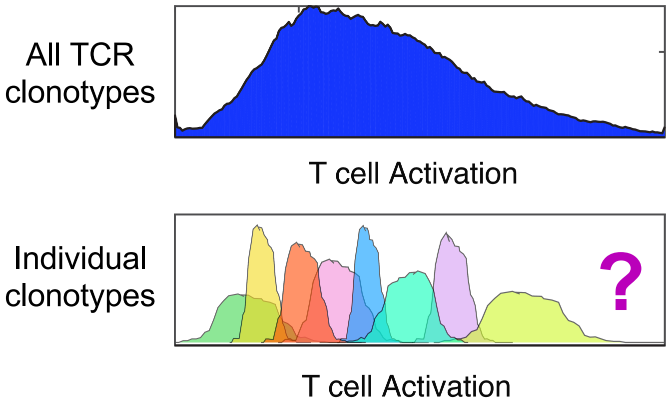
Activation and Differentiation Axes in Tumor Immune Atlas

- A large diversity of monocytic cells organize onto distinct axes of differentiation as they change environment / tissue context
- All cells have both M1 and M2 genes



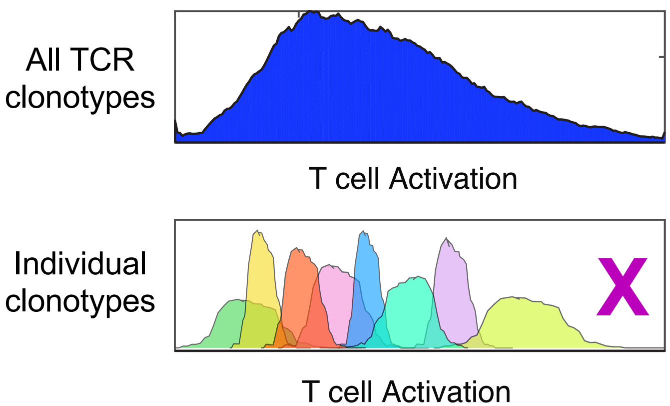
Does TCR repertoire diversity contribute to the continuous spectrum of T cell activation?

Paired single-cell TCR- and RNA-seq on 27K sorted T cells



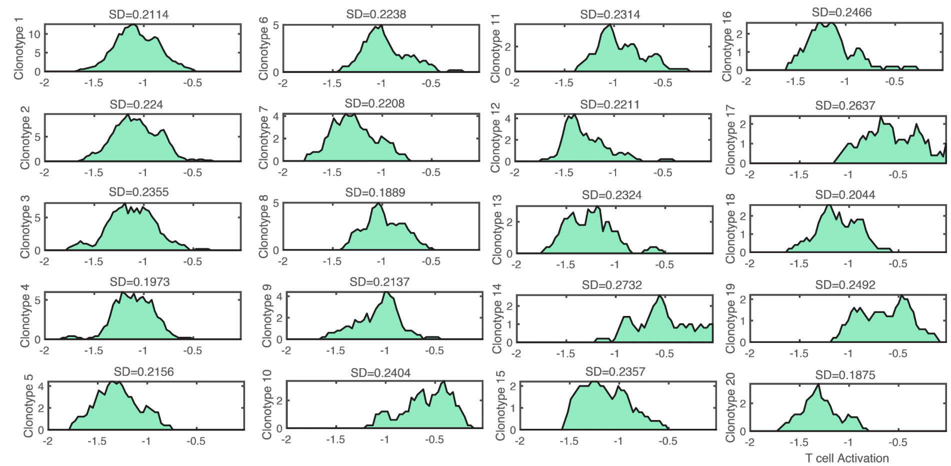
Does TCR repertoire diversity contribute to the continuous spectrum of T cell activation?

Paired single-cell TCR- and RNA-seq on 27K sorted T cells



TCR diversity is not the exclusive driver of the continuity of T cell activation

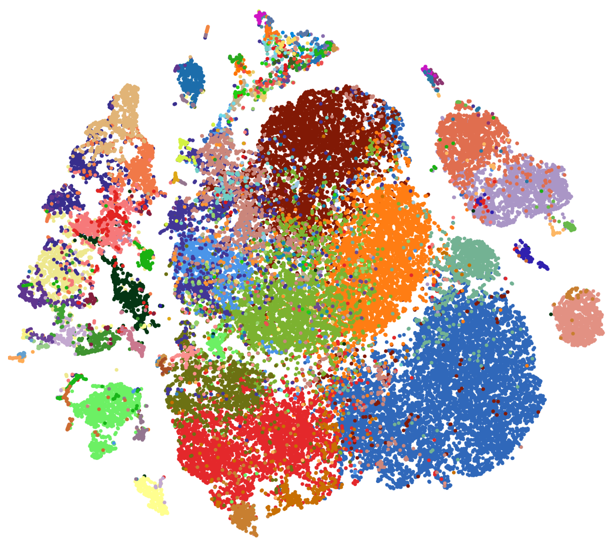
Dominant clonotypes from one tumor

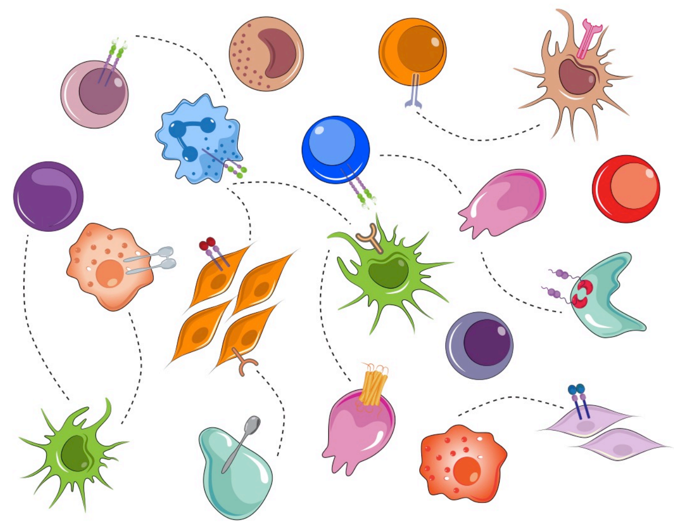


30-50% of the variation in activation explained by clonotypes (ANOVA test on 3 tumors)

We constructed a cell atlas of the tumor immune system:

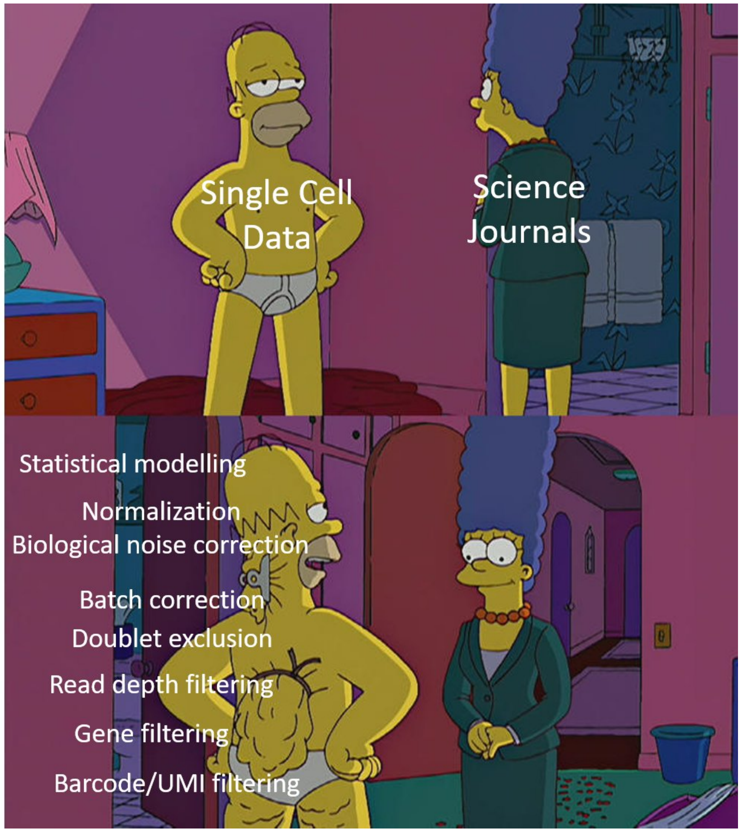
- » Captured a rich diversity of tumor immune cell types
- » Captured tissue specific differences in tumor-immune environment
- » Strongest axis of variation: continuum of activation states



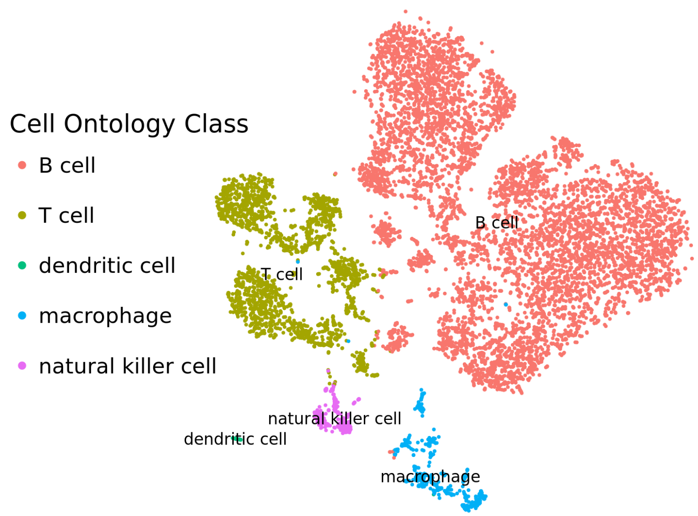


Data analysis of scRNA-seq is not straightforward

- » Every step in data-processing matters.
- » **Within sample normalization:** down-sampling, total count, log, z-score, scran, sc-transform
- » **Feature selection:** all genes, highly variable genes
- » Clustering
- » Batch-Correction



Sorting allows us to discover increasing heterogeneity

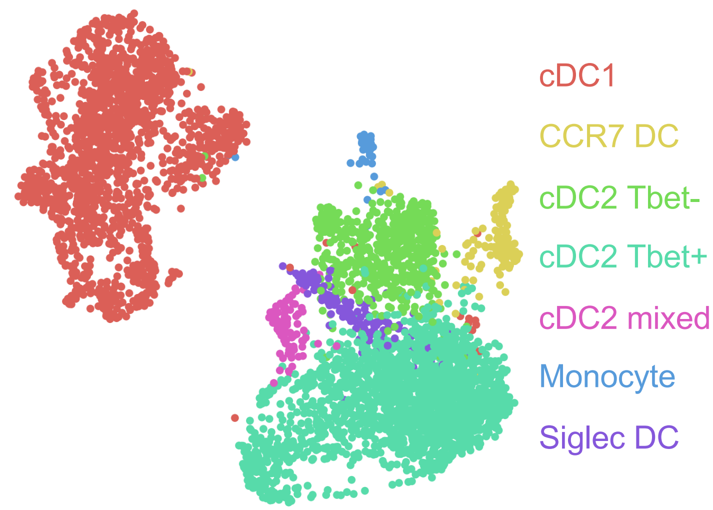


Dendritic cells are rare

Data from: CZI/Quake
Tabula Muris spleen

Cell types

Enrich for DCs using FACS



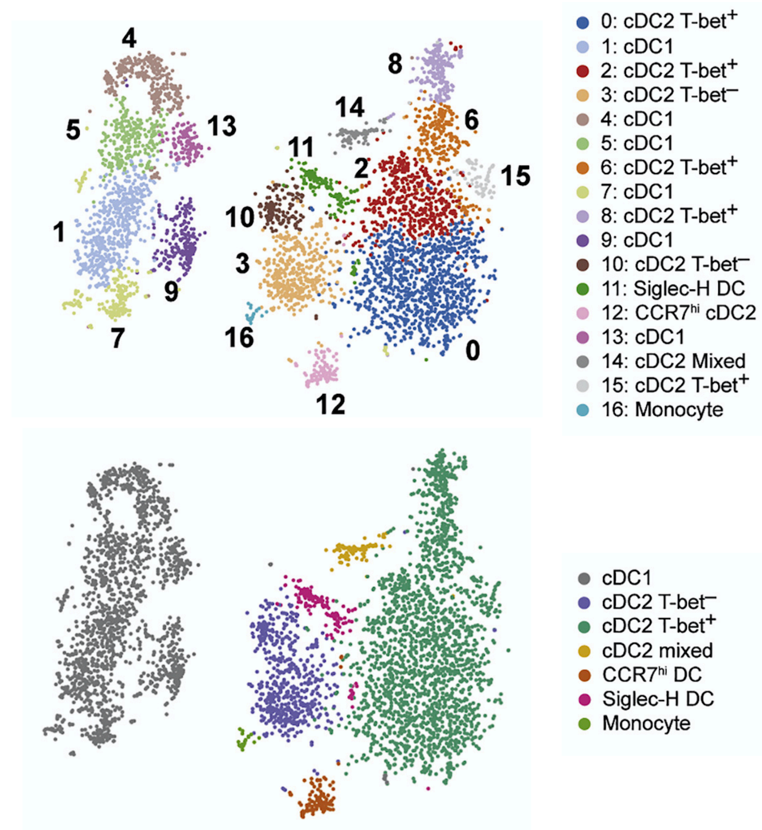
cDC2 Tbet+ are rare yet play important role in type 1 immune response

Data from Brown*, Gudjohnson*, et.al. Cell 2019

Sorting allows us to discover increasing heterogeneity

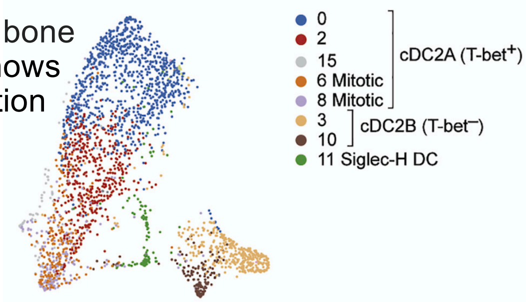
- » Clusters are not a one to one match with cell types
- » Annotating clusters is still one of the most laborious tasks in the analysis:
  - › Cite-seq will help
  - › Computational methods to automate are being developed
- » cDC2A and cDC2B Have Distinct Phenotypic and Functional Properties

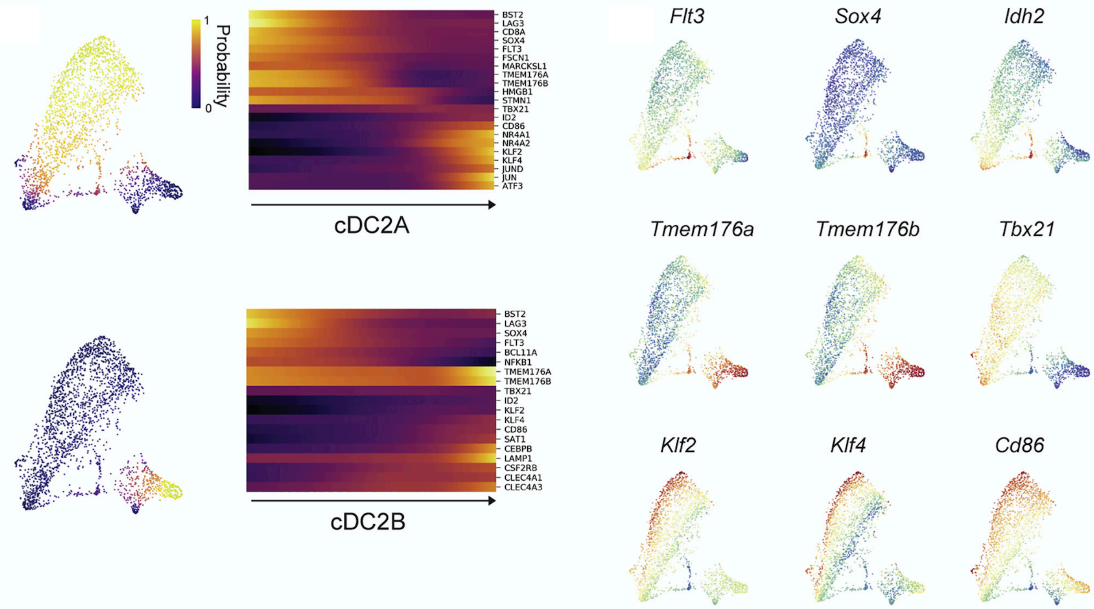
Brown*, Gudjohnson*, et.al. Cell 2019



Continuous trends show differentiation trajectories

Data from bone marrow shows differentiation





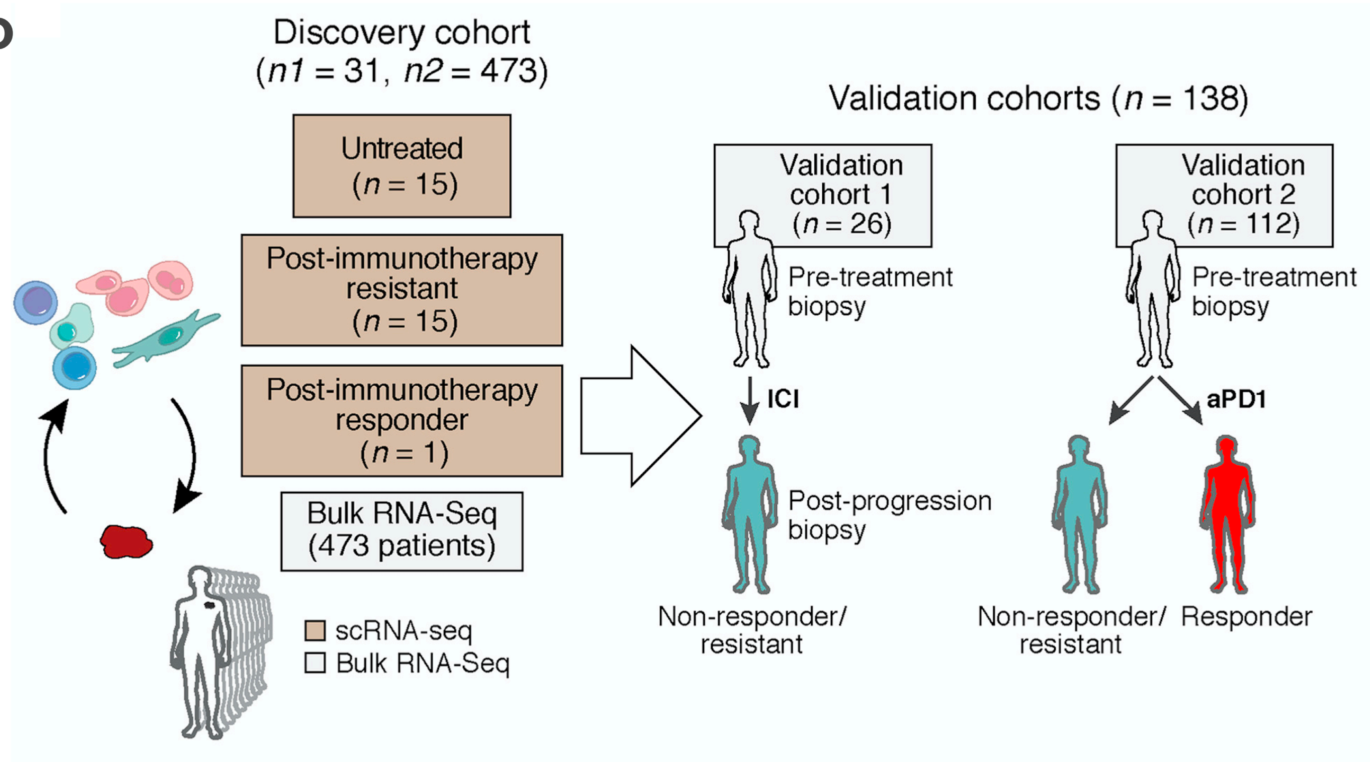
» Pseudo-time trajectory analysis could determine distinct paths of differentiation for cDC2A and cDC2B

» As well as the genes and TFs that change along these paths

Cohort design to find resistance program

» Search for the signature at the single cell level

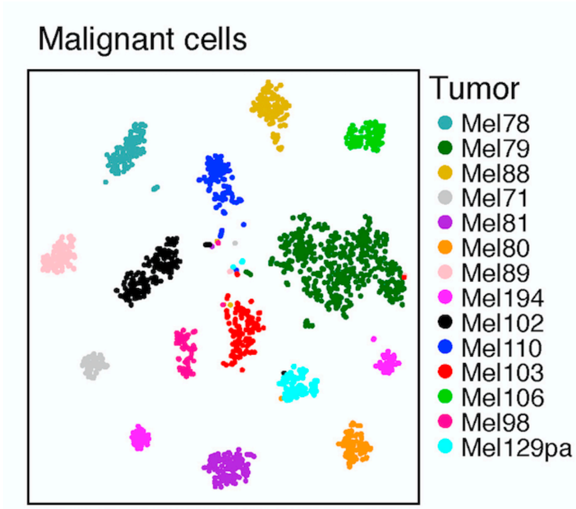
» Bulk to get larger cohort sizes

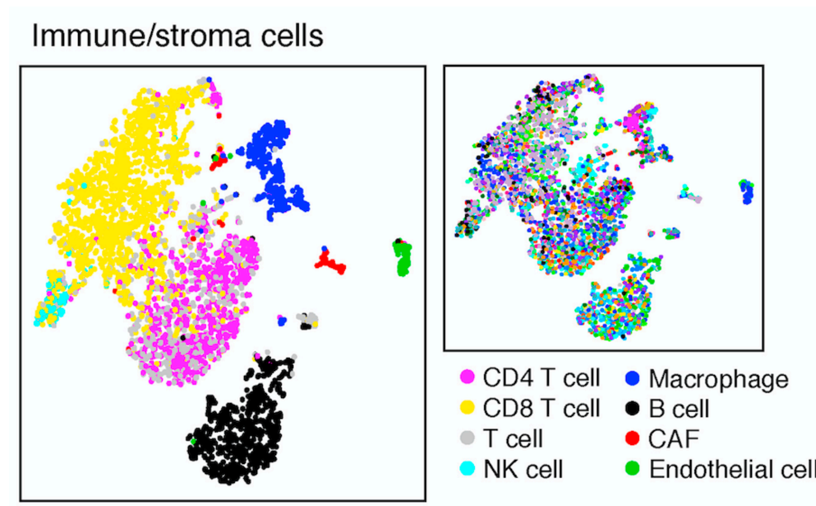


Jerby et al., Cell 2018

Single cell RNA-seq data of melanoma cohort

- Melamona cells are different and unique to each patient
- Immune subsets overlap between the patient
- But very much differ in abundances of different immune subtypes

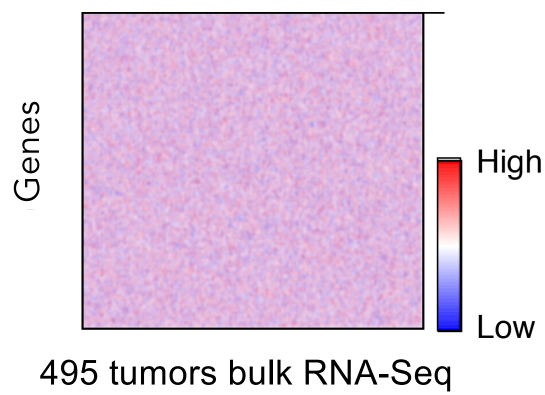




A program in malignant cells from T cell “cold” tumors

- Seed with correlated genes in scRNA-seq
- Search for tumor programs that correlate with coldness

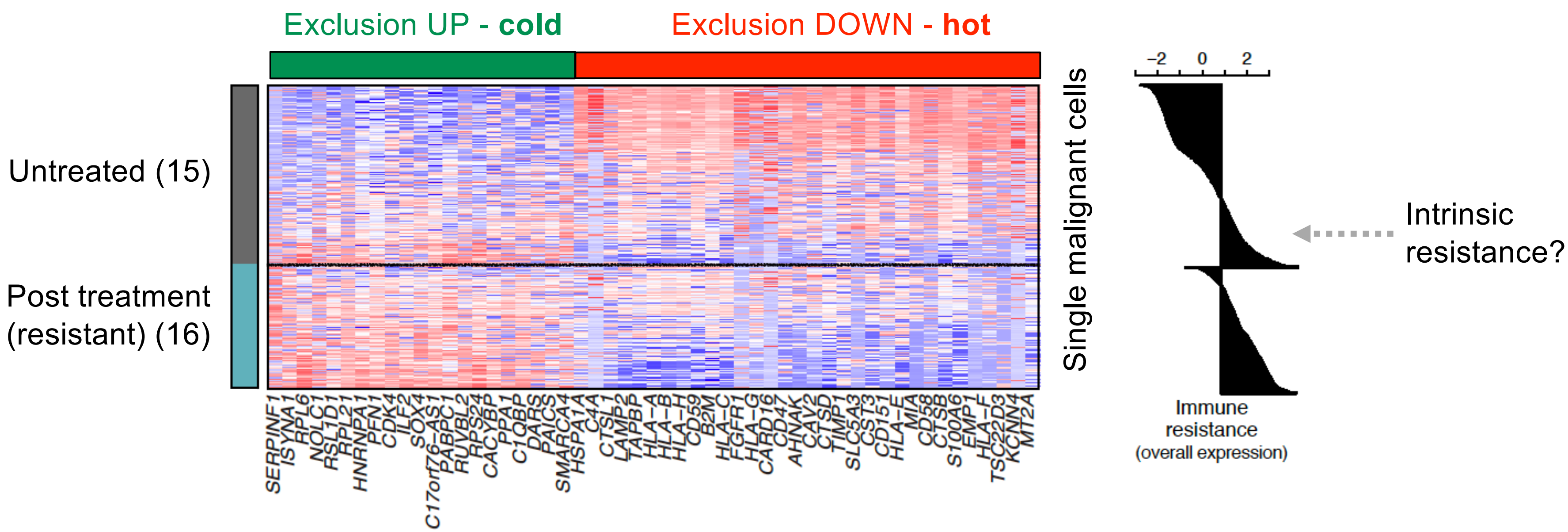
Single cell RNA-Seq



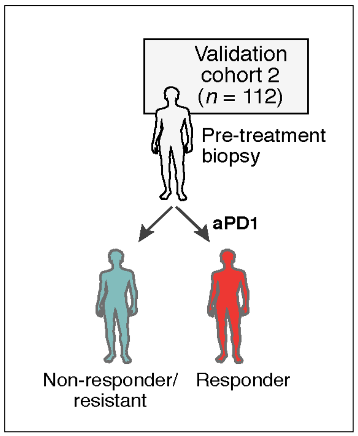
Myc targets
CDK4
CDK targets

MHC-I,
IFNg
SASP,

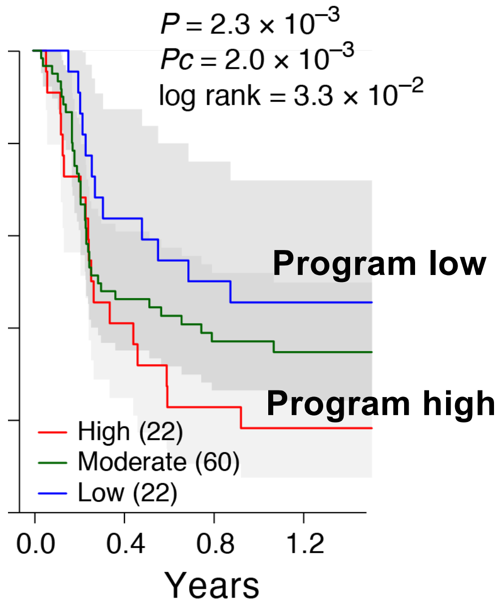
Exclusion program associated with resistance; but some cells express the program pre-treatment



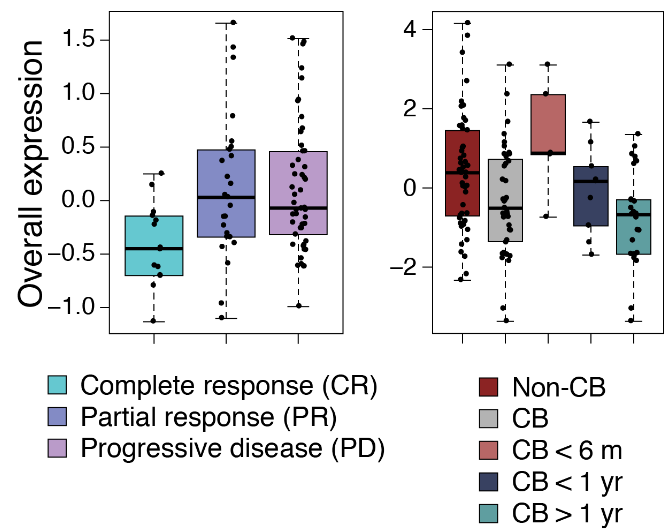
Validation cohort: Program predicts immunotherapy outcome



1. Progression-free survival



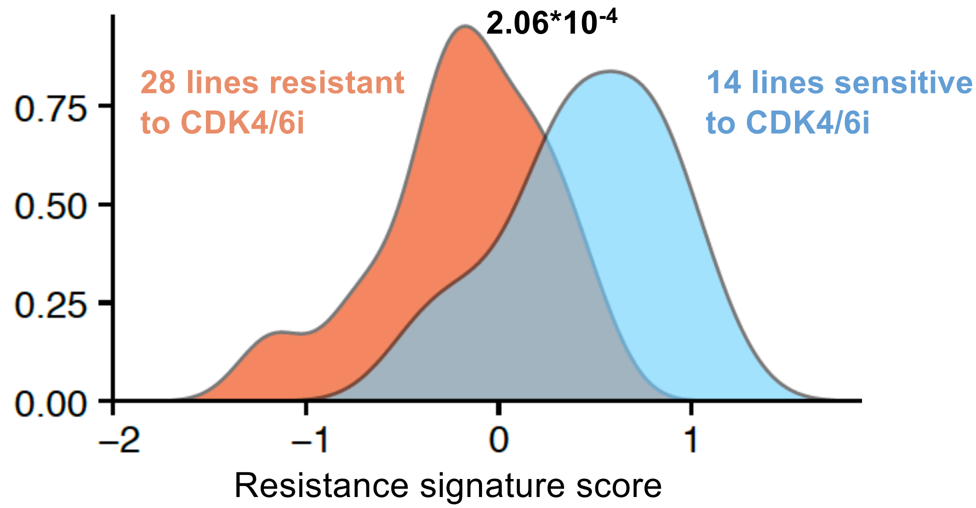
2. Clinical response and duration



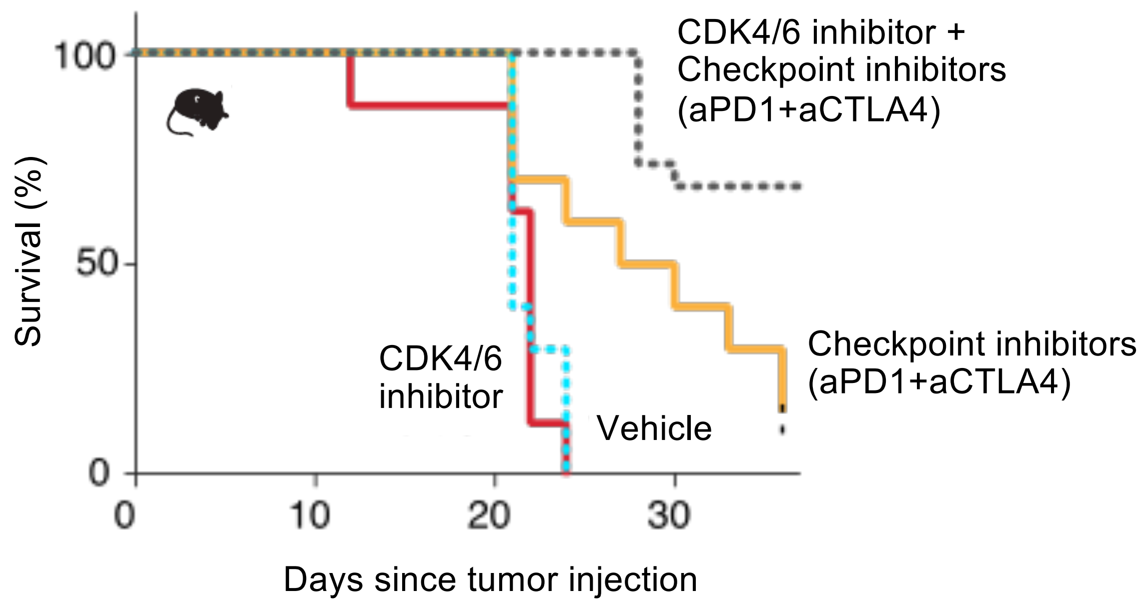
Computational search predicts CDK4/6 as program regulators

Query: which drugs are more toxic to cell lines overexpressing the program in a screen of 131 drugs across 639 human cell lines (Garnett et al., 2012)?

CDK4 and CDK4/6 target genes are induced in the exclusion program



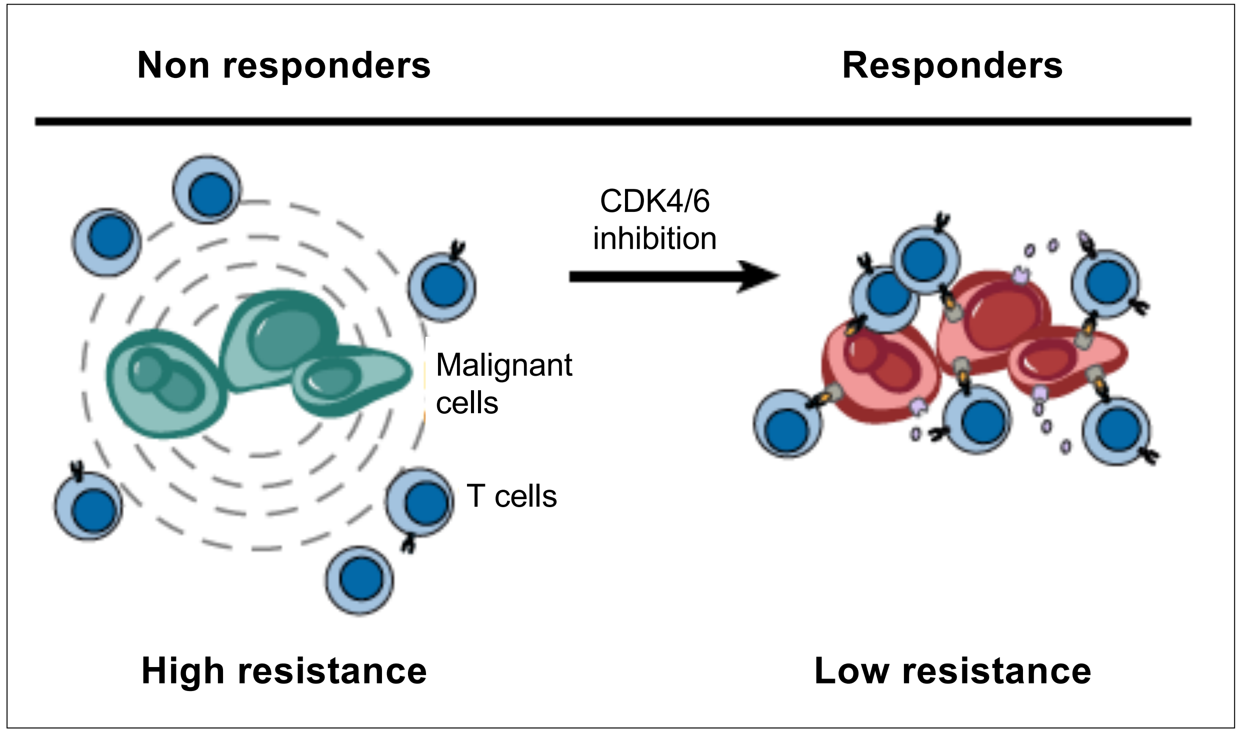
Abemaciclib sensitizes B16 melanoma tumors to checkpoint immunotherapy



Benefit depends on both Rb in malignant (B16) cells and on presence of CD8 cells

Model: The contribution of malignant cell programs to immune cell exclusion

Melanoma



Memorial Sloan Kettering
Cancer Center

NASDC
NCI Awardee Skills Development Consortium

Single Cell Genomics

Questions