SINGLE CELL

Workshop

COLOMBIA 2024

MISTI
MIT Global Experiences

Massachusetts Institute of Technology

Universidad de los Andes
Colombia

CeMoP
Grupo de Investigación celular y molecular de Microorganismos Patógenos

SHALEKLAB
from cells to systems

gen core
Centro de Secuenciación

Overall Goals

- Introduction to the rapidly expanding world of single-cell genomics
- Focus less on specific tools but more on underlying concepts to help you make informed choices
- Create a fun, collaborative, and interactive learning environment
- Help establish partnerships and build networks

Course Structure

- Combination of short lectures to understand the underlying principles of single-cell methods followed immediately by practical application

- Interrupt with questions, and feel free to let us know if there is anything you would like us to go in more depth

Plans for Bangkok With You All!

Overview:

- 2 cohorts – computational and wet lab
- 5 full days of training

Experimental Program:

- Introduction to experimental design
- Tissue dissociation, cell and nuclei isolation
- Cell loading
- 10x Genomics, HoneyComb
- Library preparation RNA and Hashing
- Introduction to data analysis

Thanks for your partnership!

Computational Program:

- Introduction to single-cell transcriptomics and experimental sequencing methods
- Introduction to computational analysis methods (R studio prep, introduction to R, R lab cell subsets; local & cloud)
- Data pre-processing & technical QC
- Data exploration & biological QC
- Identification and biological interpretation of cell types/states/NMF factors
- Data integration and batch correction
- Upstream regulator inference
- Trajectory Inference
- Immune repertoire scVDJ and CITE-Seq
- Cell-cell communication

Organizers




Trainers

Elliot Boblitt

Operations & Project Manager
Shalek lab

Manuel Guzman

MIT Chemistry Graduate student

Sarah Quinn

MIT Chemistry Graduate student

Víctor Manuel Duarte

M.Sc. student Universidad de los Andes

Angie Fonseca Fernandez

M.Sc., Ph.D. Candidate

Maritza Torres Moreno

M.Sc., Ph.D. Candidate

David fisher

Post-doctoral fellow

Tammy Ian

PhD student Harvard

An Rapid Introduction To Our Workshop

1. Experimental Track
    a. Sample processing
    b. Single-cell processing
        i. SMART-Seq2 + BCR/TCR
        ii. Drop-Seq + BCR/TCR
2. Computational Track
    a. Preprocessing
        i. Read quality
        ii. Alignment
        iii. Quality control (QC)
        iv. Normalization
        v. Feature Selection
        vi. Visualization
    b. Downstream Analyses
        i. Cluster
        ii. Annotate
        iii. Differential expression
        iv. Advanced Topics
3. Frontiers

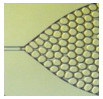
Single-Cell Expression Profiling Pipelines

Sample preparation → Cell capture and barcoding → Library prep → Sequencing

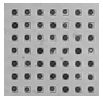
FACS

Micro-capture

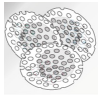
Droplet



Nanowell



Split/pool barcoding



PolyA vs random priming

3'/5' end tagging vs full-length

PCR vs *in vitro transcription*

…

MiSeq

~20M reads

NextSeq

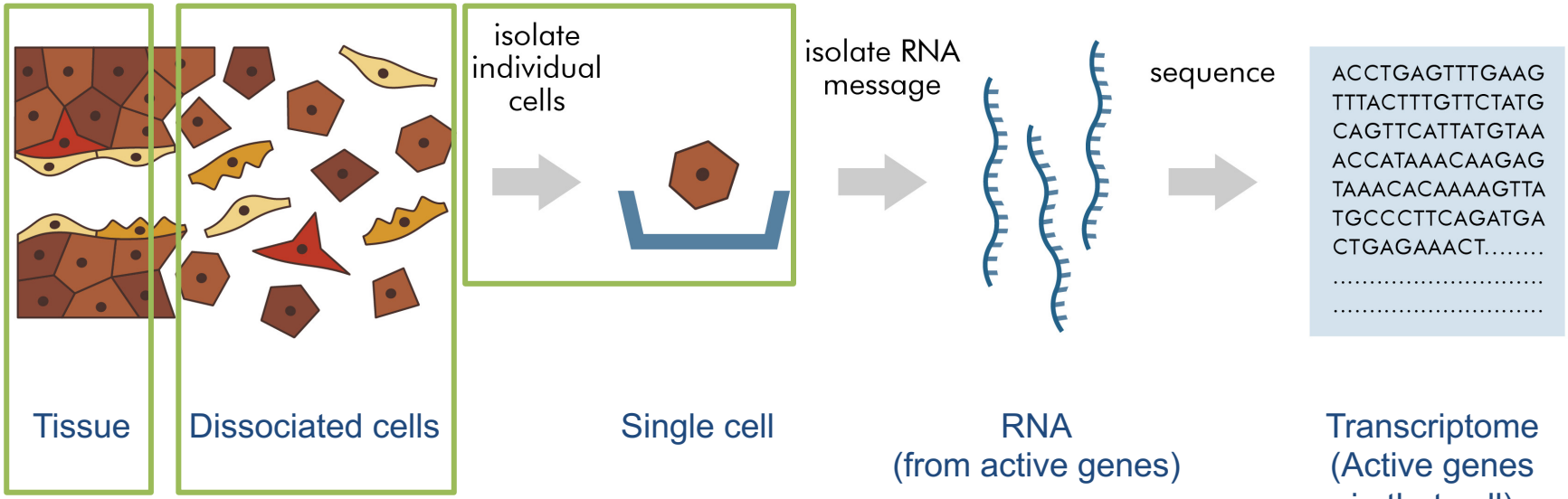
~ 500M Reads

NovaSeq:
~ 6B Reads

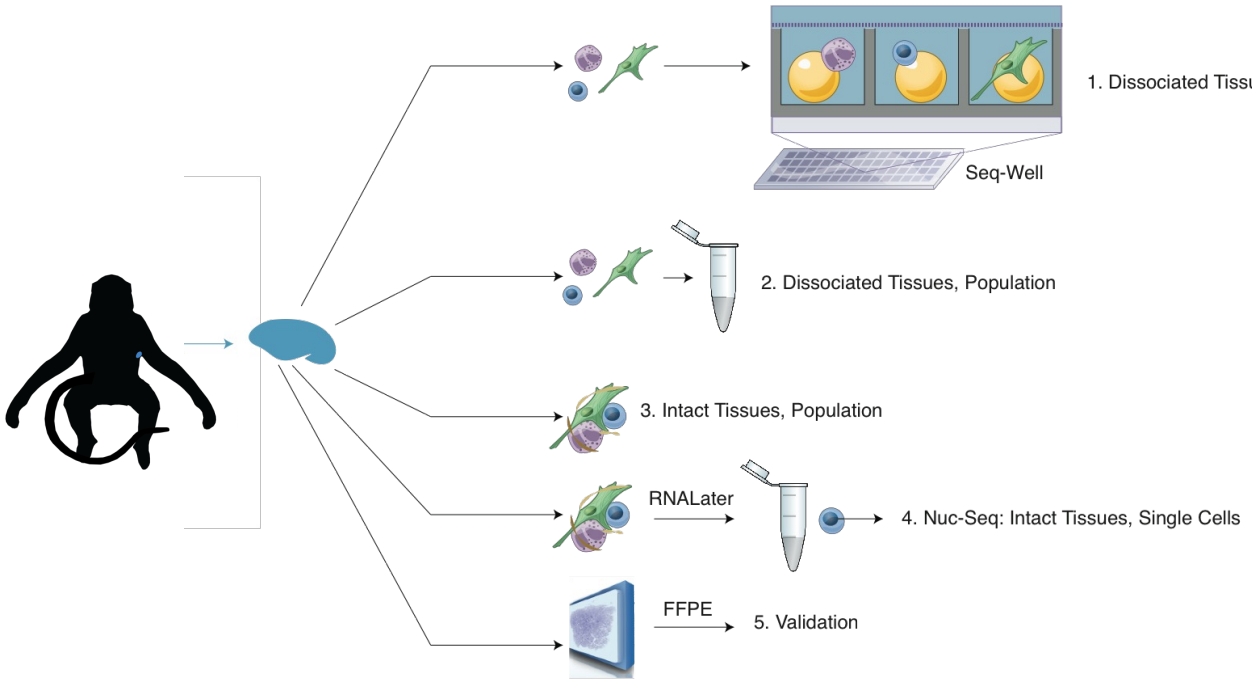
SHALEKLAB
from cells to systems

Sample Preparation

Sample preparation



Hoarding

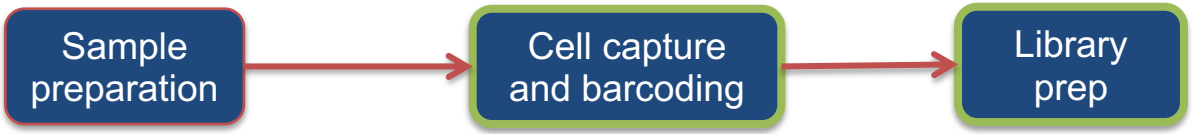


Sampling Bias

Controls for Dissociation

Follow up experiments (e.g. emerging technologies, spatial validation, live cells)

A Very Quick Introduction To Two Common Assays



1. Smart-Seq2: Plate-based RNA-Seq (2011-)
2. Drop-Seq: Droplet-based RNA-Seq (2015-)

The Challenge Of Single-Cell RNA-Seq

Ultra-low input RNA-Seq is problematic

Ribosomal RNA contamination

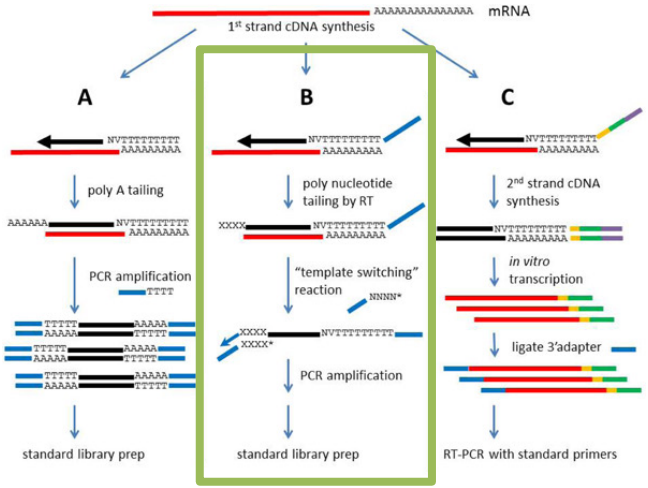
- rRNA overwhelms mRNA (~98%)
- polyA-selection is too inefficient

Bias in early PCR stages when using random priming

- “PCR Jackpotting” of some RNA molecules
- Suppression of some RNA molecules

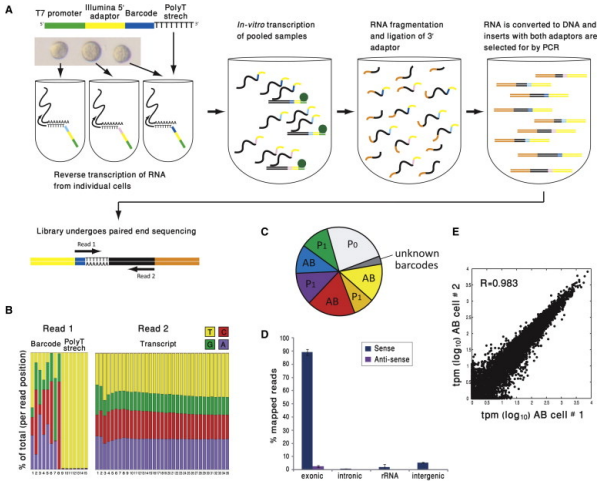
Overcoming rRNA

Tang/Quartz-Seq | Smart-Seq | CEL-Seq



From www.biotechniques.com

CEL-Seq



Hashimshony et al, Cell Rep, 2012

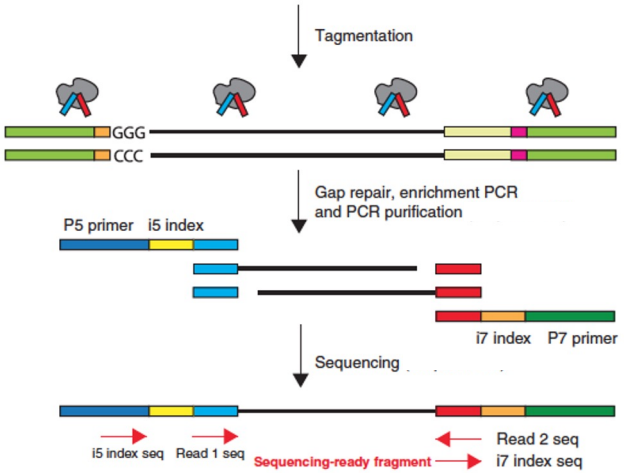
1. Smart-Seq2: Assay Overview

SMART oligos

Long: $TTTTT_{(30)}$[SMART]

Short: [SMART]rGrGrG

PCR p



amplification

Improvements

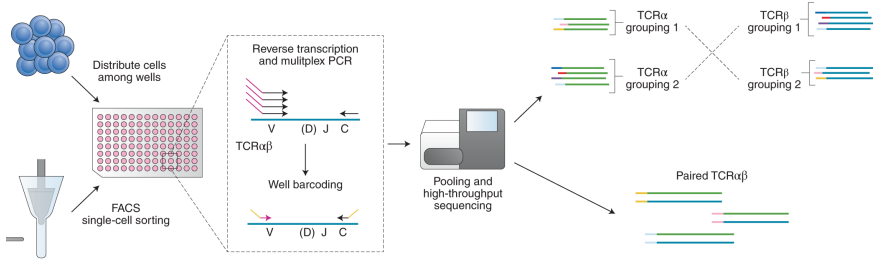
- mRNA only
- PCR more uniform a

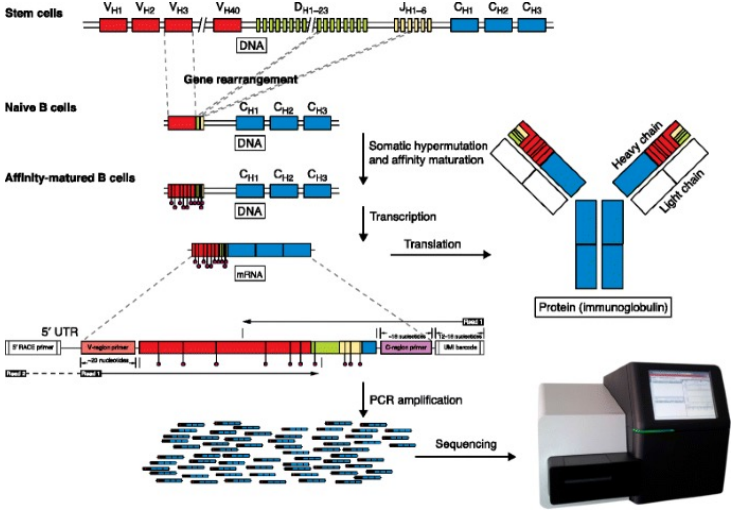
Problems

- Needs to transcribe full le
  - Falls off when encoun
  - Results in significant 3
- Long oligo can anneal at

ats, extremely long molecules

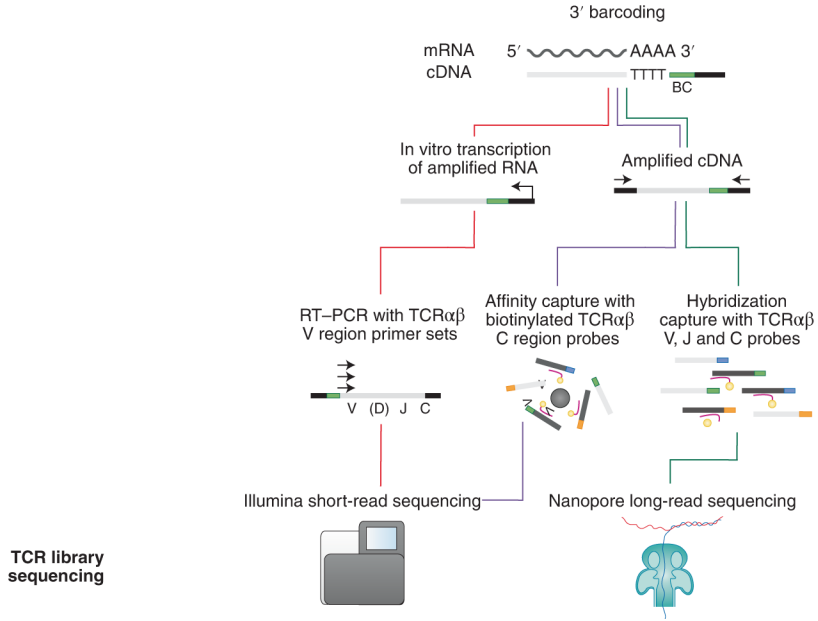
Capturing Additional Information – TCR/BCR



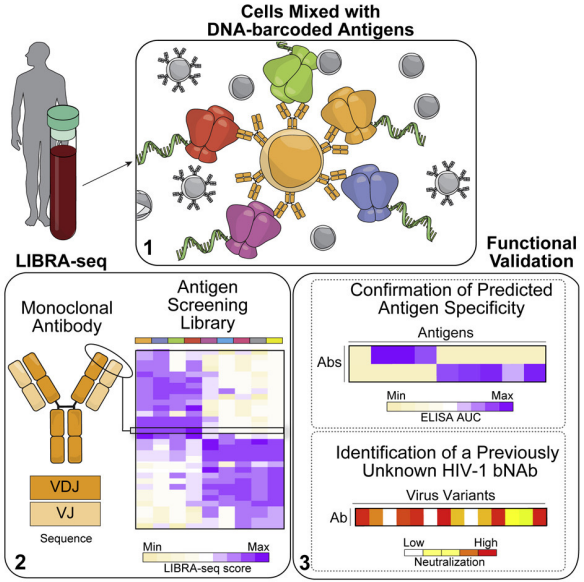


Pai et al, *Nature Methods* (2021); Yaari et al, *Genome Biology* (2015)

Capturing Additional Information – TCR/BCR



And More



Stoeckius et al, *Nature Methods* (2017); Setliff et al, *Cell* (2019)

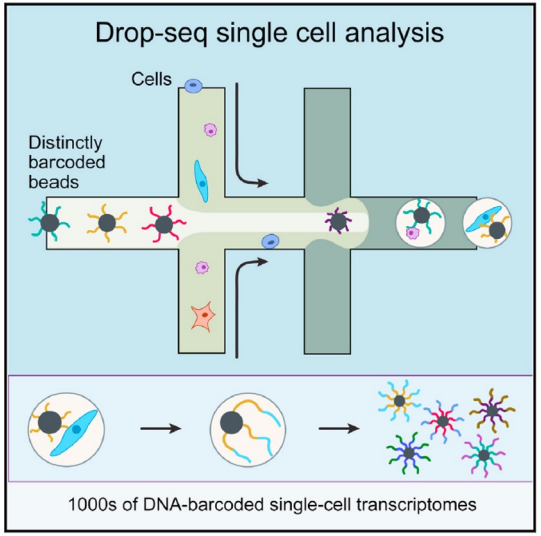
2. Drop-Seq

Resource

Cell

Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets

Graphical Abstract



Authors

Evan Z. Macosko, Anindita Basu, ..., Aviv Regev, Steven A. McCarroll

Correspondence

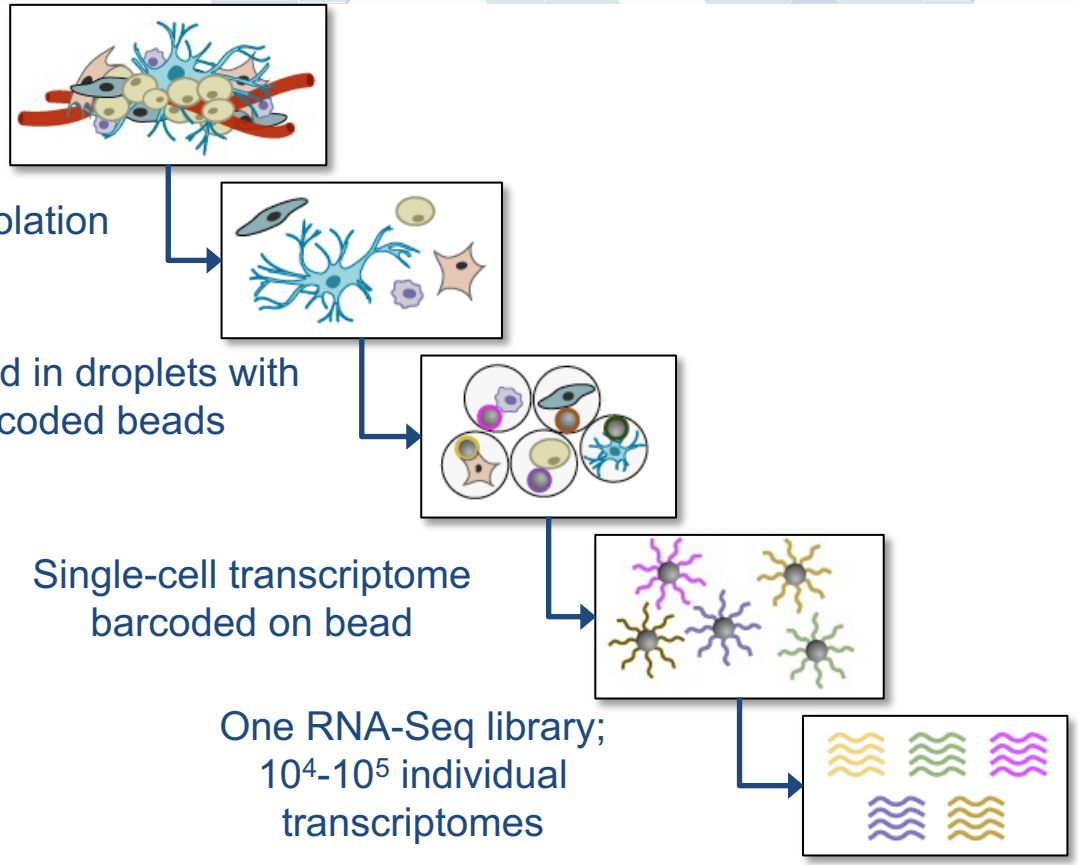
emacosko@genetics.med.harvard.edu (E.Z.M.), mccarroll@genetics.med.harvard.edu (S.A.M.)

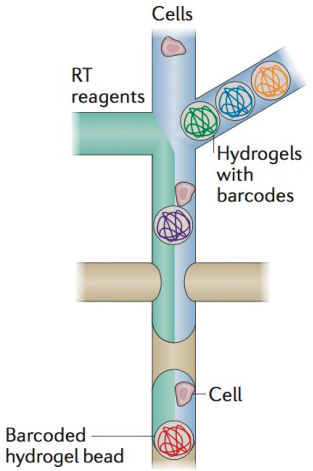
In Brief

Capturing single cells along with sets of uniquely barcoded primer beads together in tiny droplets enables large-scale, highly parallel single-cell transcriptomics. Applying this analysis to cells in mouse retinal tissue revealed transcriptionally distinct cell populations along with molecular markers of each type.

Macosko et al, *Cell* (2015)

Cells In Drops With Barcodes

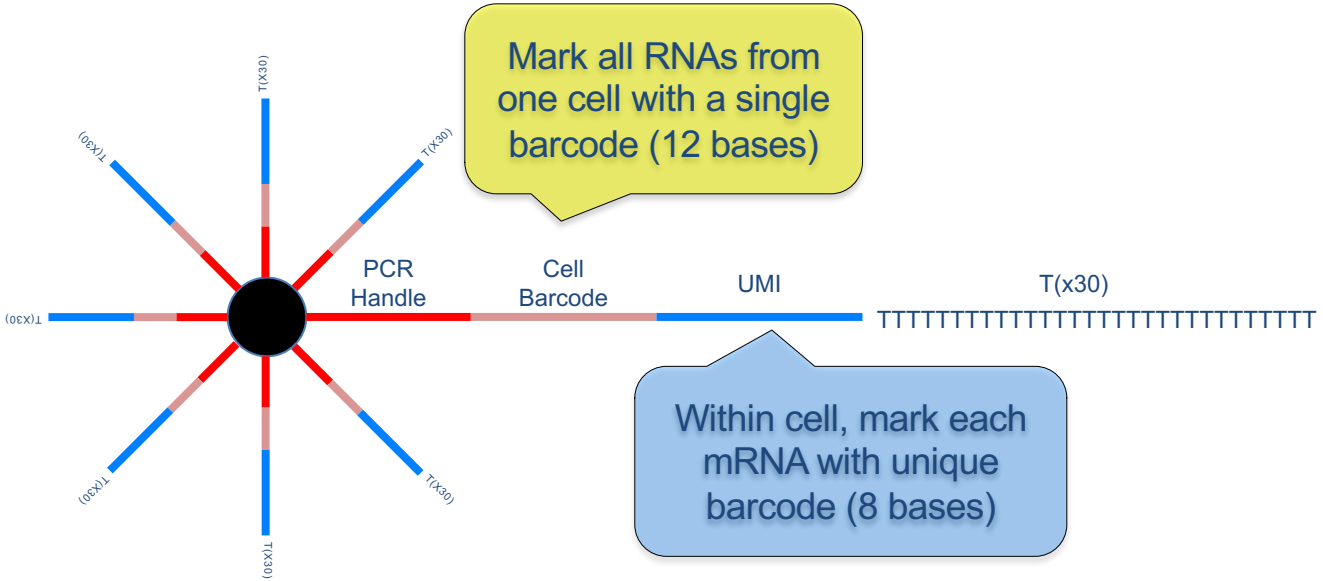




Klein et al, *Cell* 2015

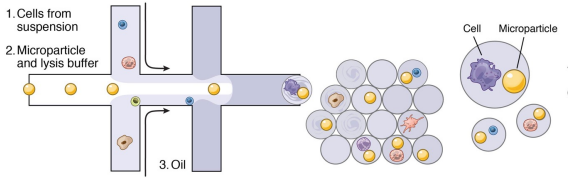
Macosko et al, *Cell* (2015)

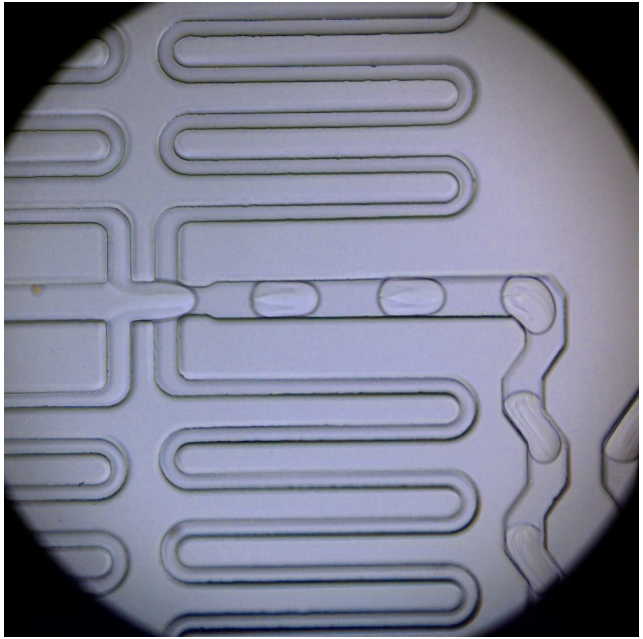
Cellular & Molecular Barcodes On Beads



Macosko et al, *Cell* (2015)

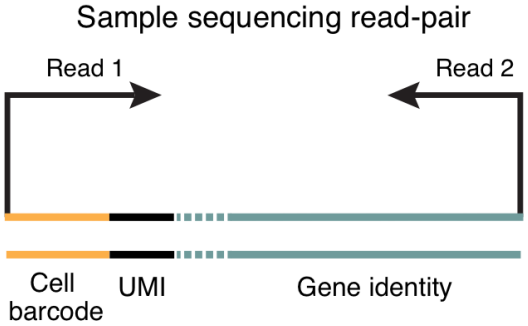
Co-Encapsulate Cells & Beads In Drops

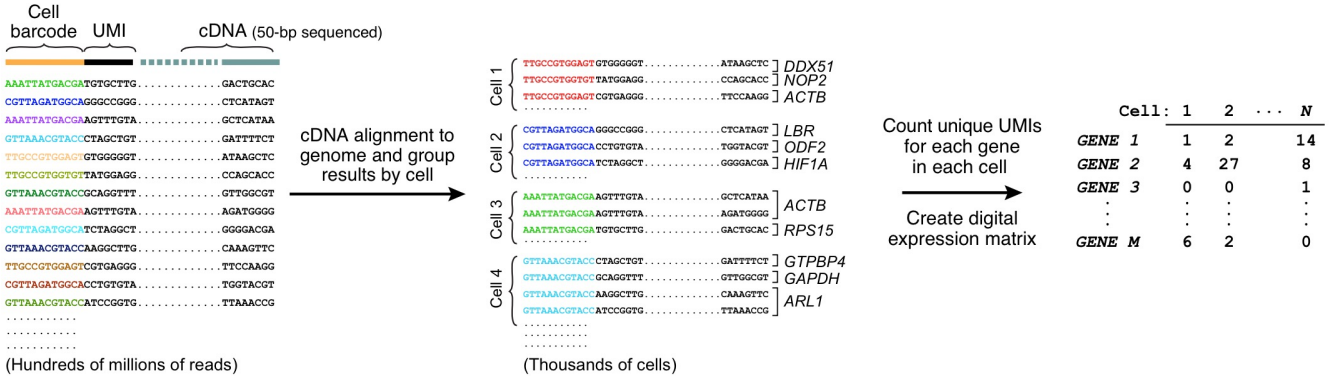




Macosko et al, *Cell* (2015)

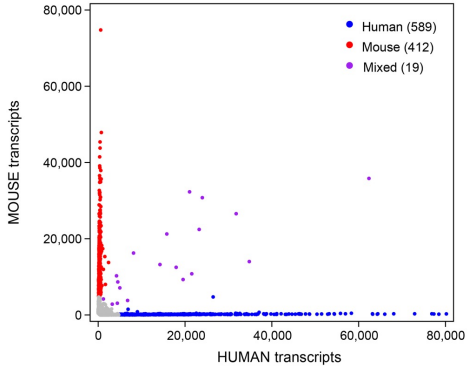
Drop-Seq: Schematic



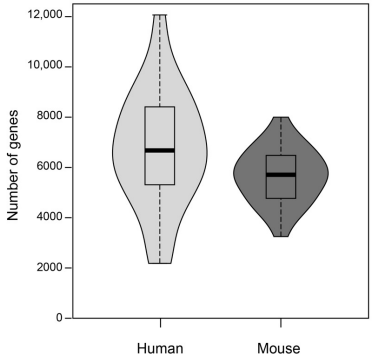


Macosko et al, *Cell* (2015)

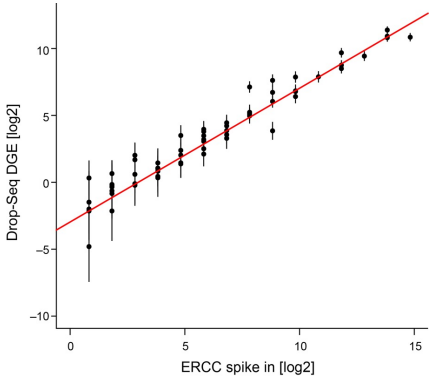
Drop-Seq Is Specific And Sensitive



✓ >95% cell-of-origin specificity



✓ ~7,000 human genes detected

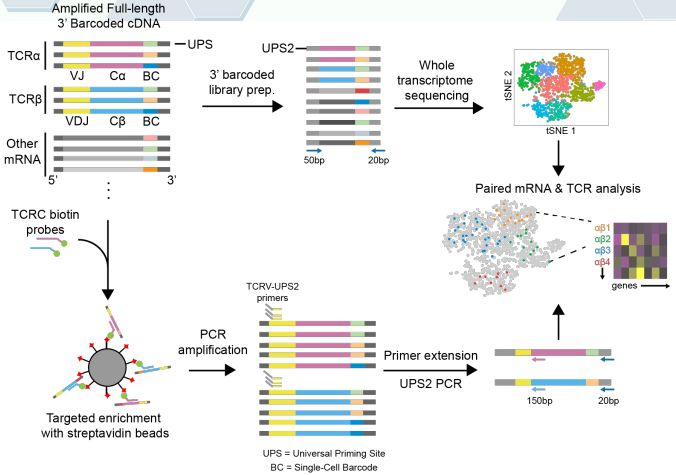


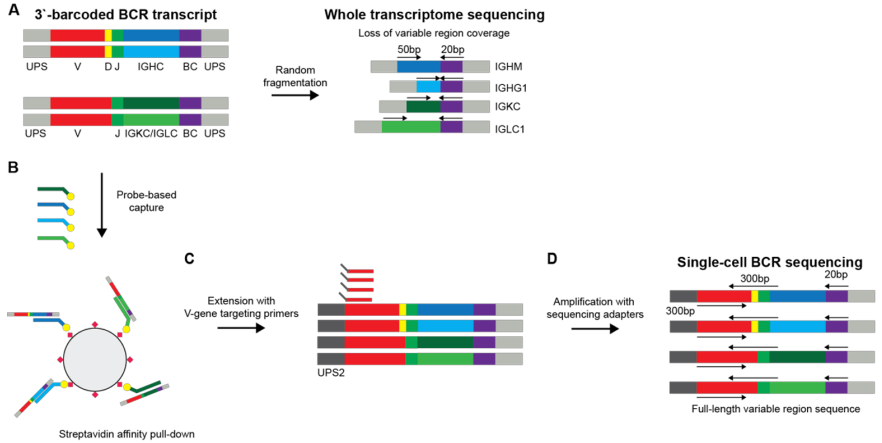
✓ Captures ~12% of single cell mRNA

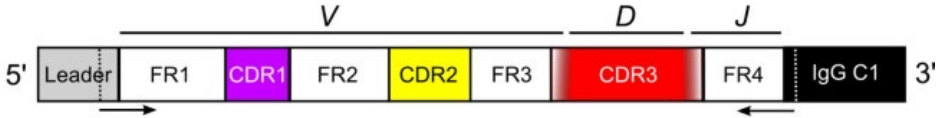
SHALEKLAB
from cells to systems

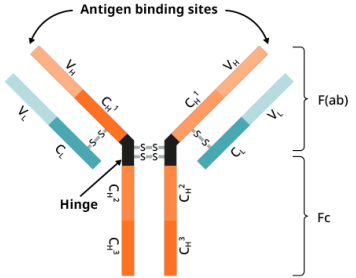
Macosko et al, *Cell* (2015)

Adding TCRs and BCRs



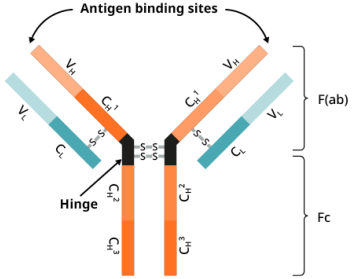


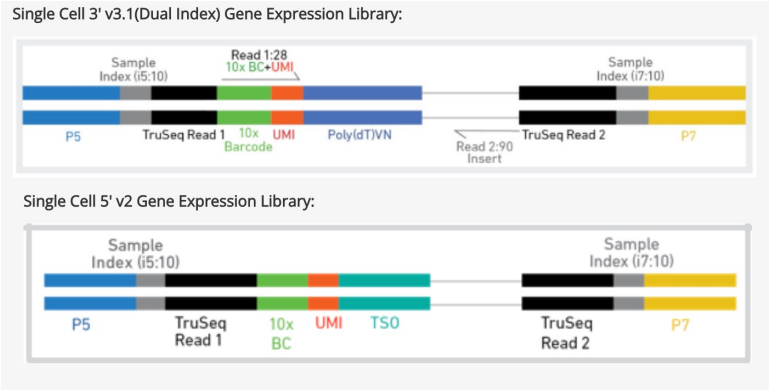


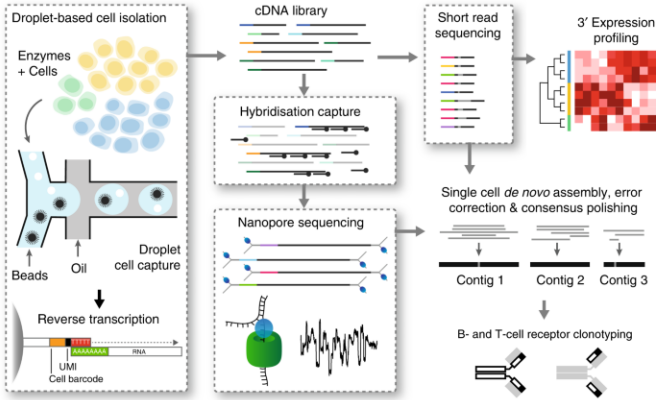


Tu et al, *Nat Immunol* (2019); Duncan et al, *BioRxiv* (2023)

Adding TCRs and BCRs + More







Larsen et al, *BMC Immunology* (2012); Abcam website;
10x Genomics website; Singh et al, *Nat Commun* (2019);

Considerations For Single-Cell RNA-Seq

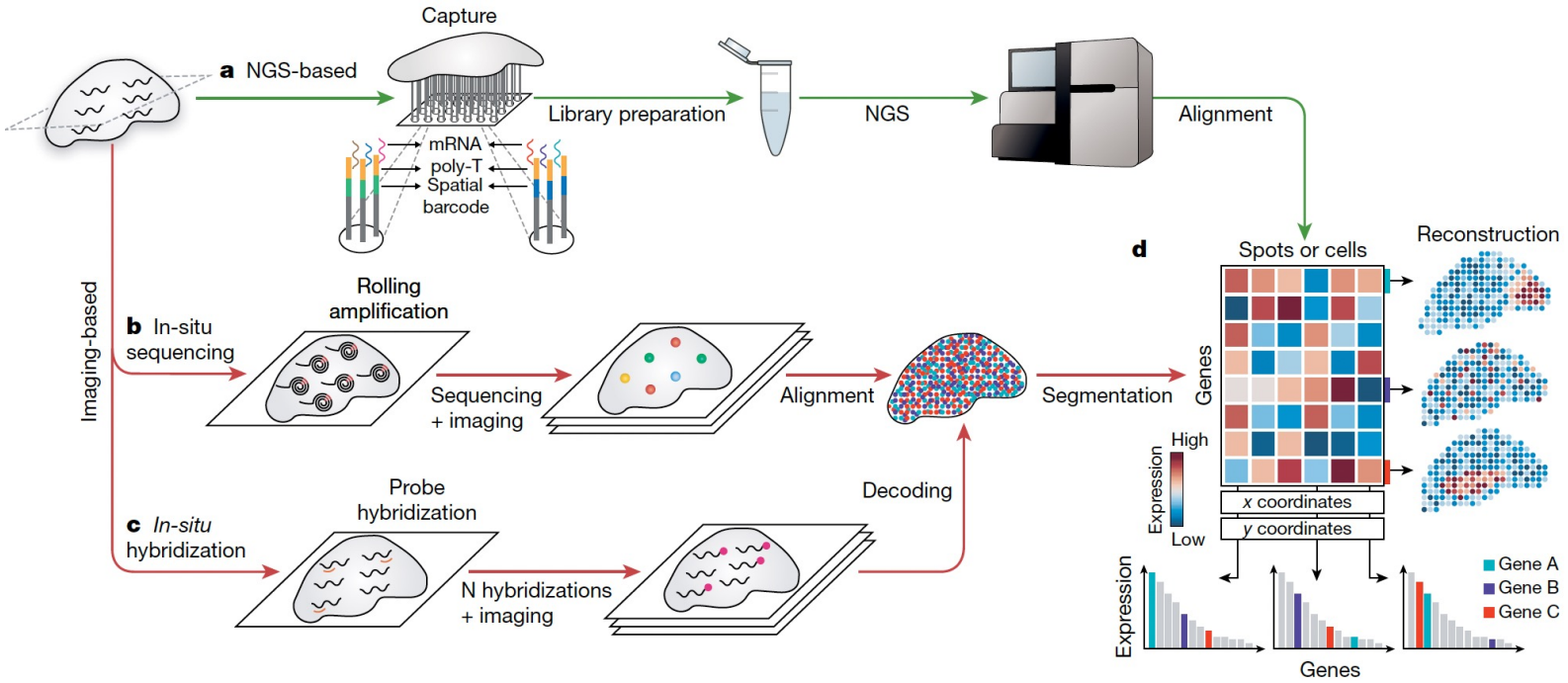
Choose protocol based on:

- Throughput (# of cells / reaction)
- Sample of origin
- Cost / Labor / Time limitations
- Gene body coverage - 5', 3' biased, or full-length?
- UMI vs no-UMI
- Sequencing depth / cell

For example:

- To classify all cell types in a diverse tissue (e.g., brain), I need high throughput
- To re-annotate the transcriptome or find new isoforms, I need full-length coverage
- I only have archival human samples so I need to use a method that permits fixed cells (or nuclei)

An Aside: Spatial Transcriptomics



Rao et al, *Nature* (2021)

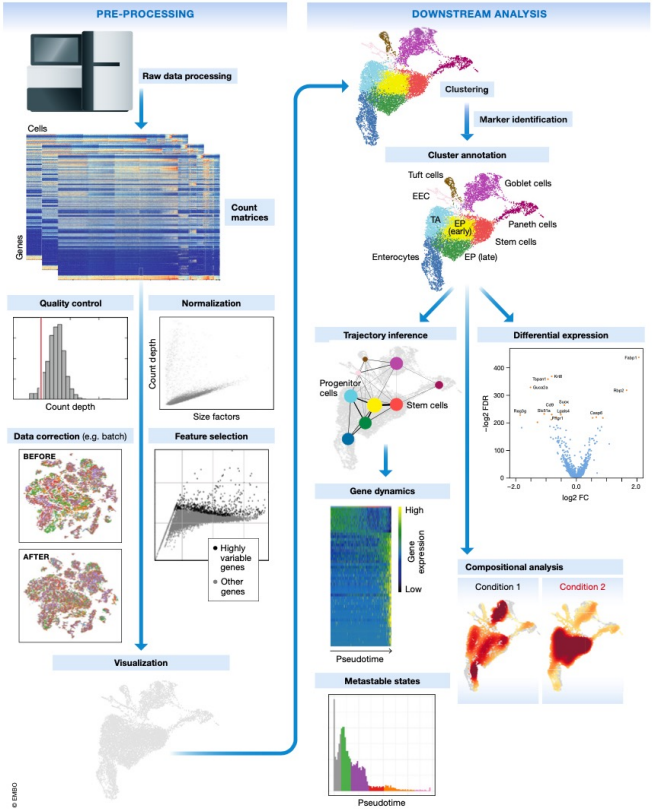
What Are The Main Tasks In scRNA-seq Analysis?

Preprocessing

- Read quality
- Alignment
- Quality control (QC) analysis
- Normalize the data
- Feature Selection
- Visualization

Downstream Analysis

- Cluster
- Annotate
- Differential expression
- Many other things… (Iterative Clustering, Trajectory Inference, RNA Velocity, Identify Regulatory Networks, etc.)

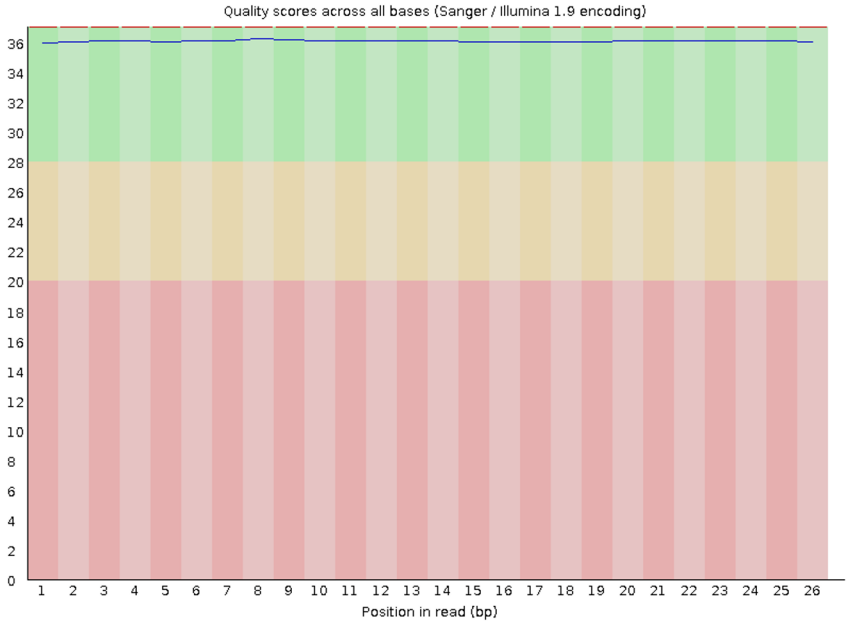


Luecken & Theis, *MSB* (2019)

SHALEKLAB
from cells to systems

Read Quality

Quality at each base for R1, across a sample.



But, sometimes things can go wrong….

- A particular spot on the machine might be bad.
- The machine may have trouble calling a specific base (“Is this an A or a T?”).
- Unusual RNA/cDNA contamination.

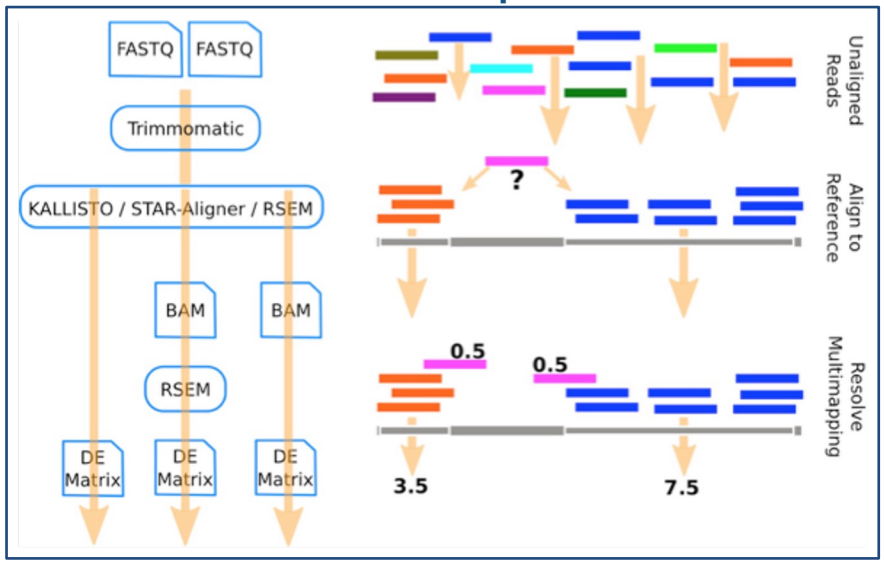
We typically run FastQC to examine the raw reads from the machine to detect issues.

Read Alignment

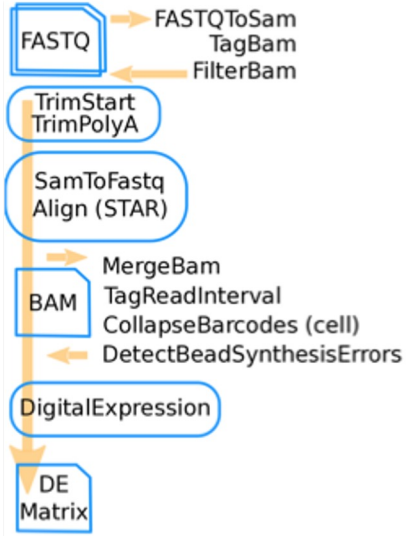
Single-Cell RNA-Seq assays are different and produce different sequences

- The sequence pipelines must be tailored to the sequence of interest.
- Many pipelines are NOT compatible but many show similarities.

Smart-Seq2



Drop-Seq/10x/Seq-Well



Common functionality:

- Trimming
- Alignment
- Generating count matrix

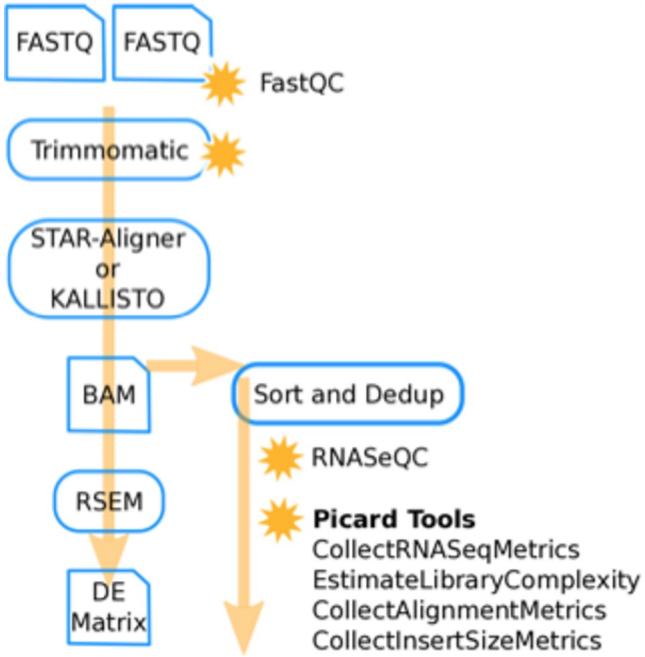
Adds:

- Bookkeeping for cell barcodes and UMIs
- Bead error detection
- Cell barcode collapsing
- UMI collapsing

http://mccarrolllab.org/dropseq/; https://software.broadinstitute.org/firecloud/

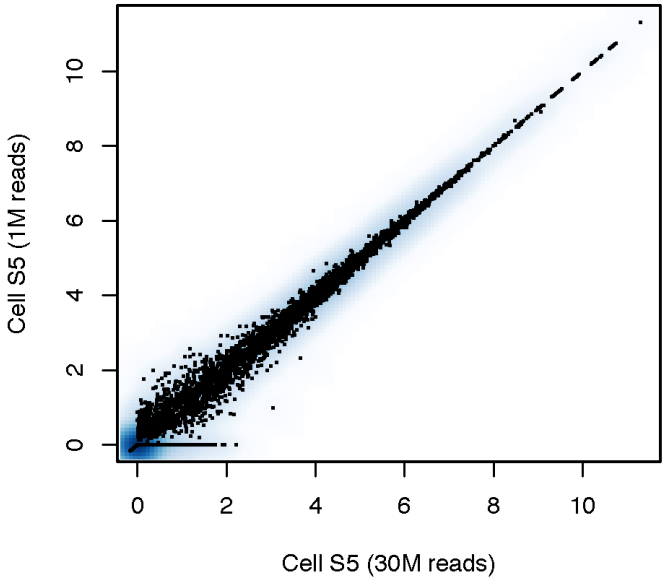
QC: Sequence Level Quality Control

Much of the QC is performed is using traditional tools.

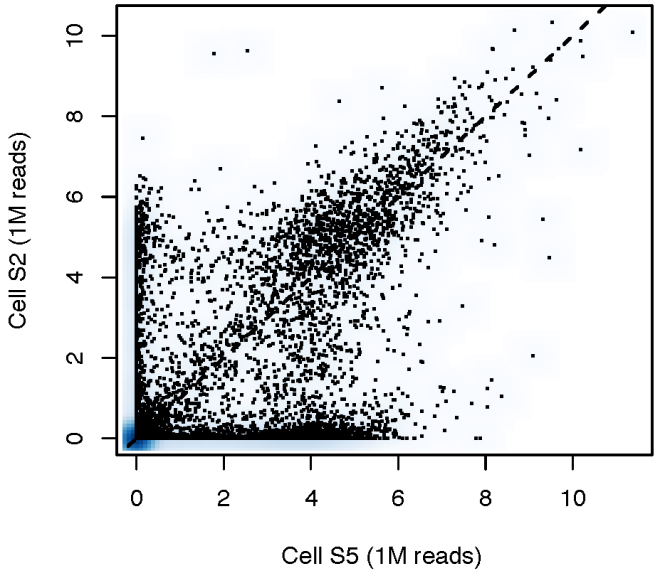


QC: How Deep To Sequence

Sampling Bias



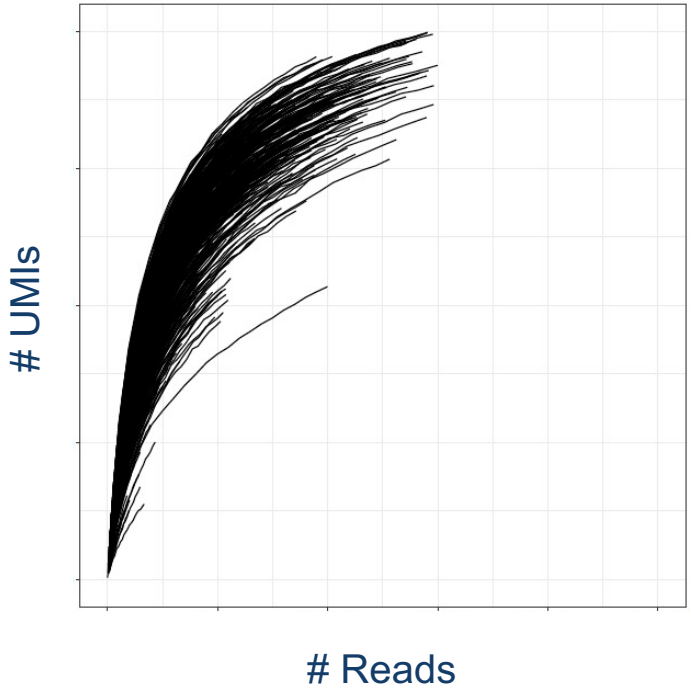
Cellular Variation



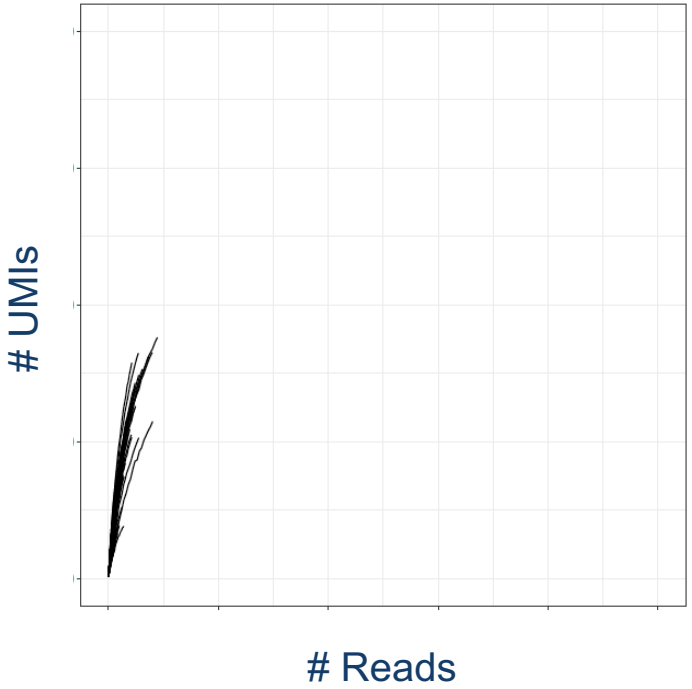
Shalek*, Satija* et al., *Nature* (2013); Svensson et al, *Nature Methods* (2017)

QC: Checking Sequencing Depth

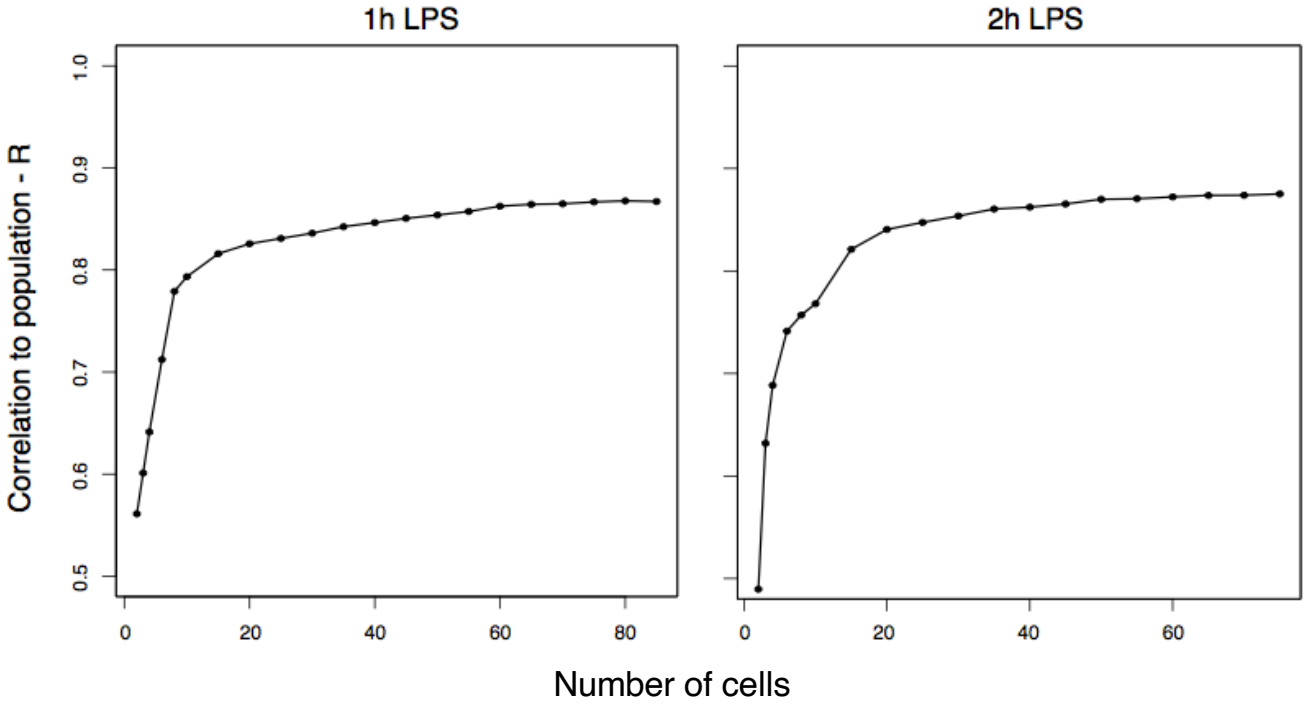
Close to saturated



Likely not saturated



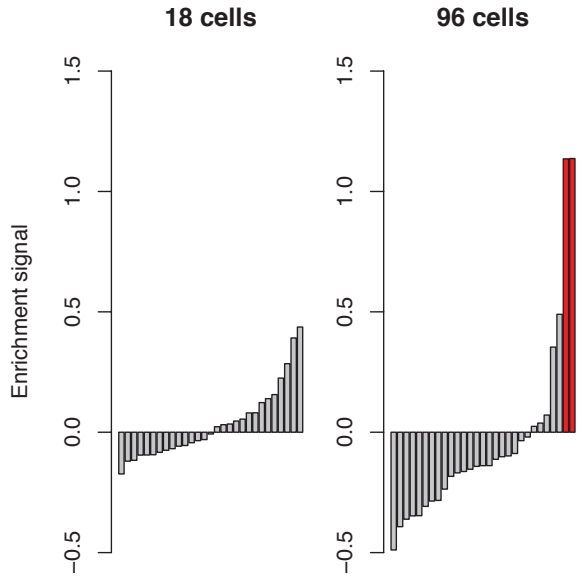
QC: How Many Cells To Sequence?



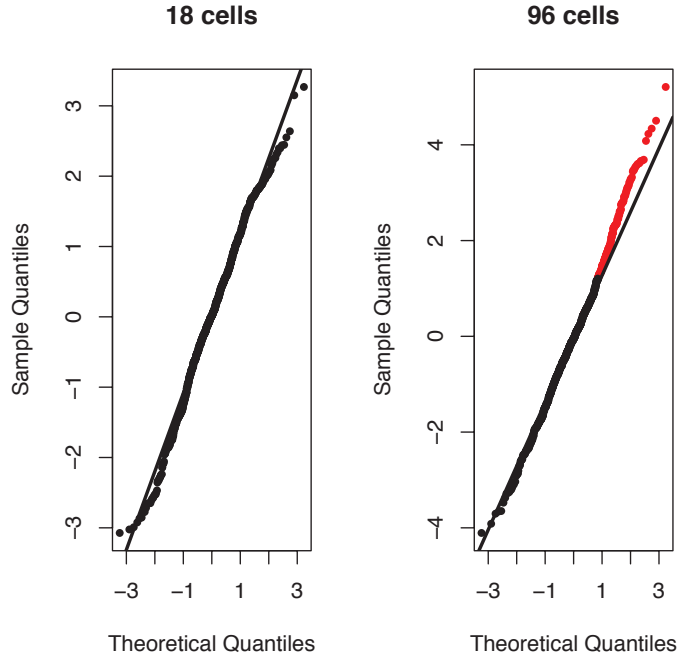
Shalek et al, *Nature* (2014)

QC: How Many Cells To Sequence?

Functional enrichments



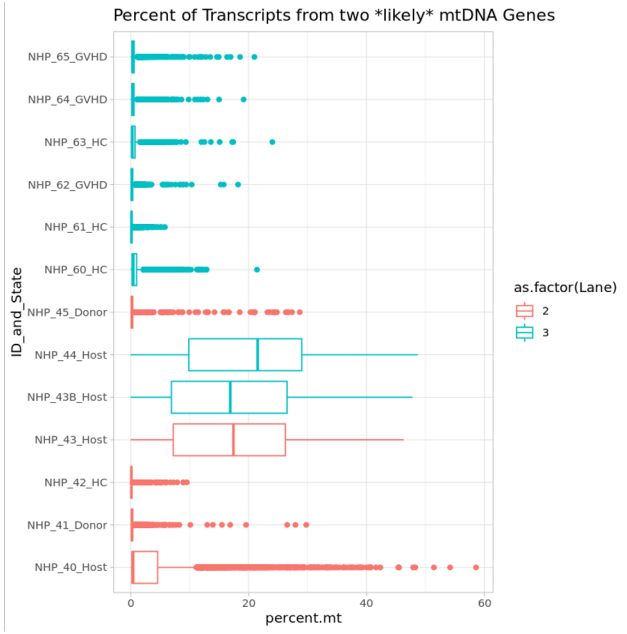
Gene signatures



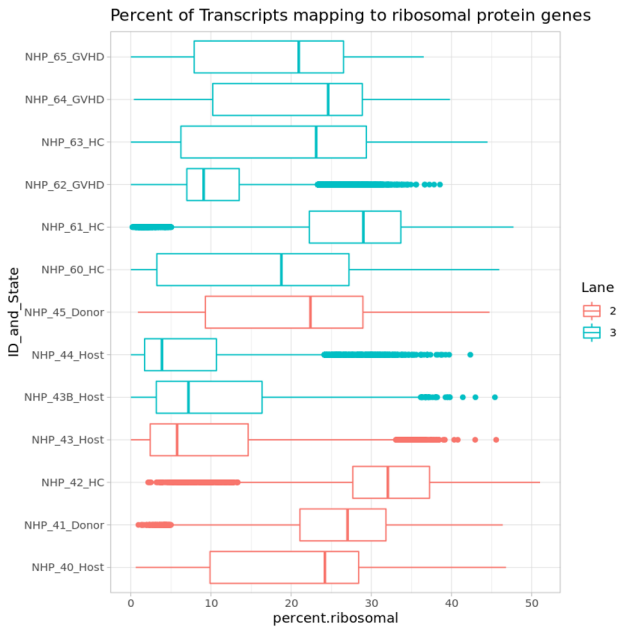
Shalek*, Satija* et al, *Nature* (2013); Shalek*, Satija* et al, *Nature* (2014)

QC: Fraction Of Transcripts From mtDNA/rRNA

High mtDNA alignment rates can suggest low quality cells (but not always).

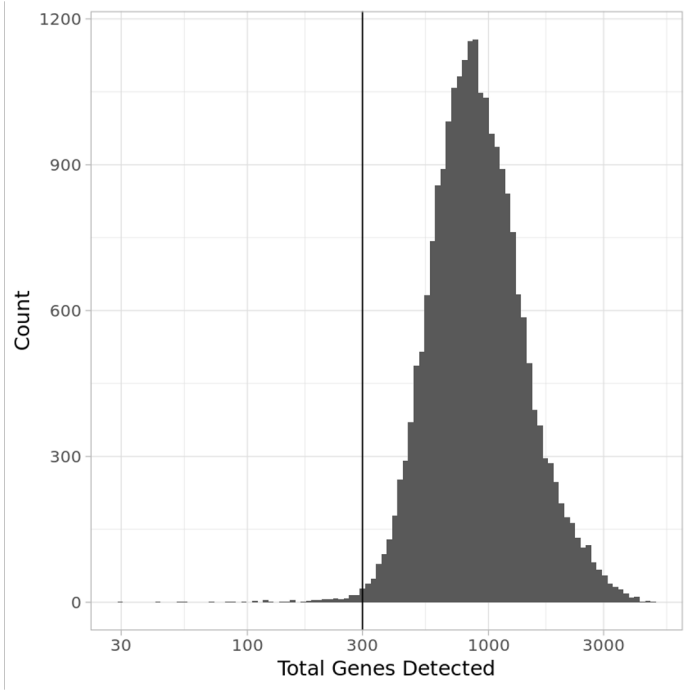


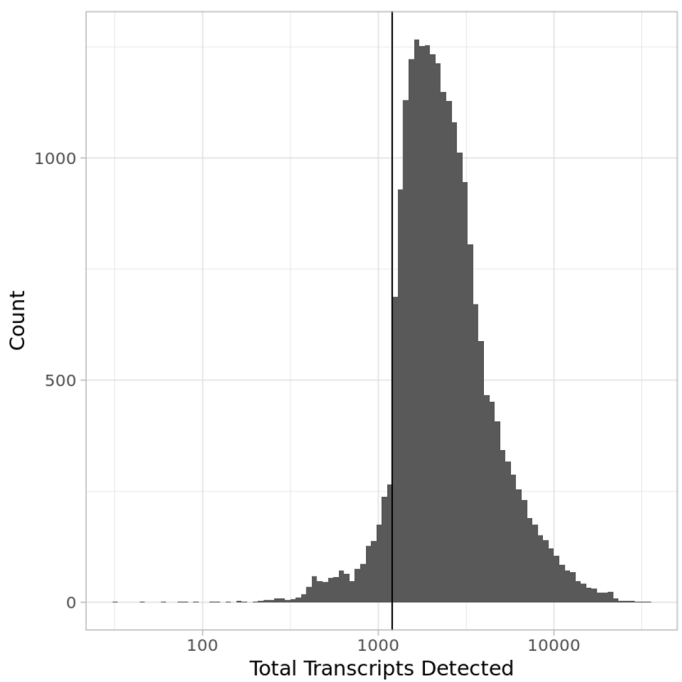
Ditto for rRNA



SHALEKLAB
from cells to systems

QC: Selecting A Bounds For Complexity?





Select something that makes sense & test robustness.

QC: Other Things To Consider

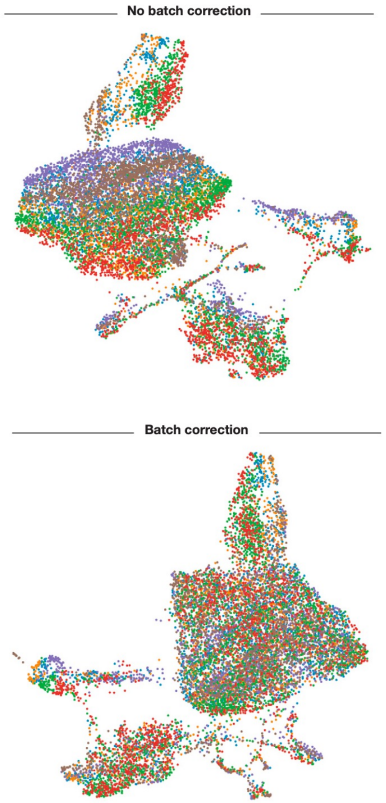
Ambient RNA contamination

Several programs like EmptyDrops (Lu et al, *Genome Bio*, 2019) exist to help remove ambient contamination in 10x/Drop-Seq/Seq-Well

Doublet Removal

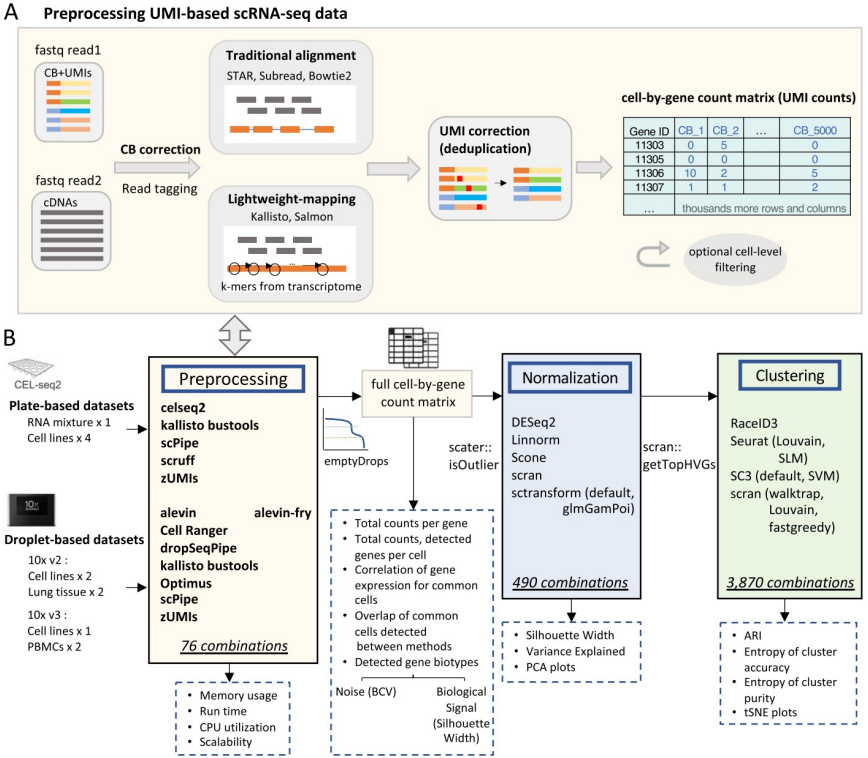
Several programs like Scrublet (Wolock et al, *Cell Systems*, 2019) exist to help identify and remove doubles

In some instances, additional corrections are needs (e.g., batch correction)



Luecken & Theis, *MSB* (2019)

Data Normalization



BE CAREFUL TO USE RAW VALUES FOR STATISTICAL TESTING!

You et al, *Genome Biology* (2021)

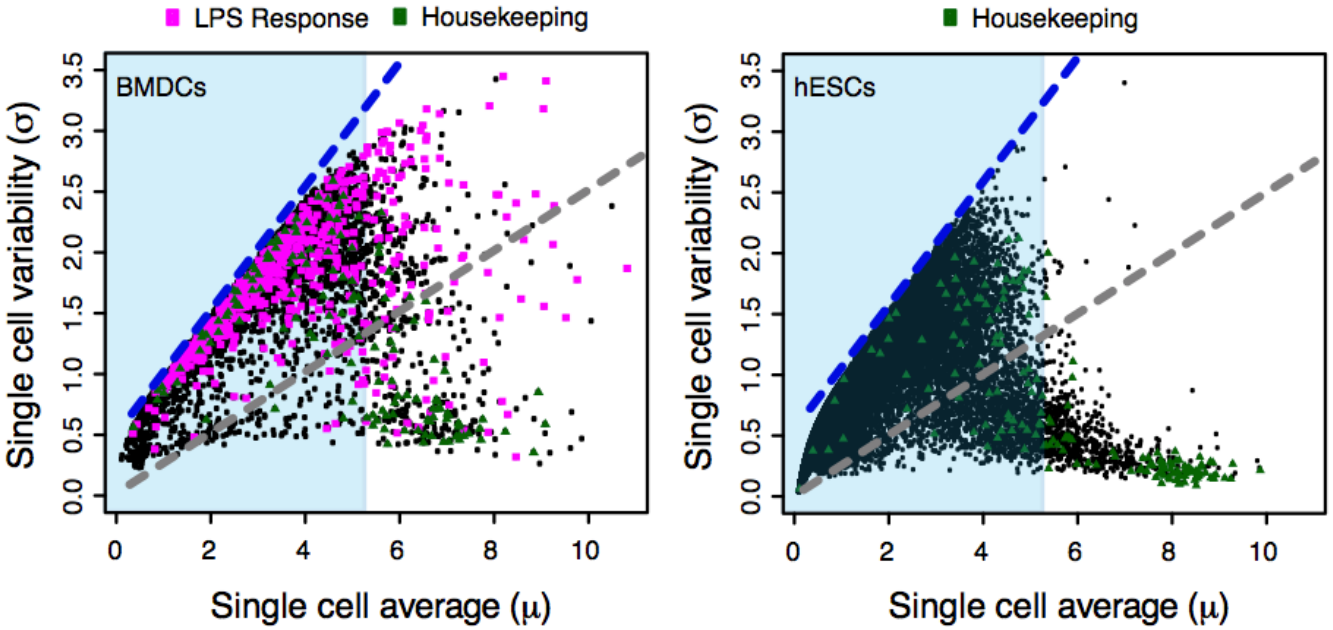
Feature Selection: The ‘Curse’ Of Dimensionality

- **Imaging/Flow cytometry**
  - Low-dimensional (~ 5 parameters)
  - Bi-axial gating plots

- **CYTOF:**
  - ‘High’-dimensional (~ 30 parameters)
  - Machine learning algorithms to identify structure

- **Single-cell genomics:**
  - Thousands of dimensions
  - Machine learning – but subset the data first

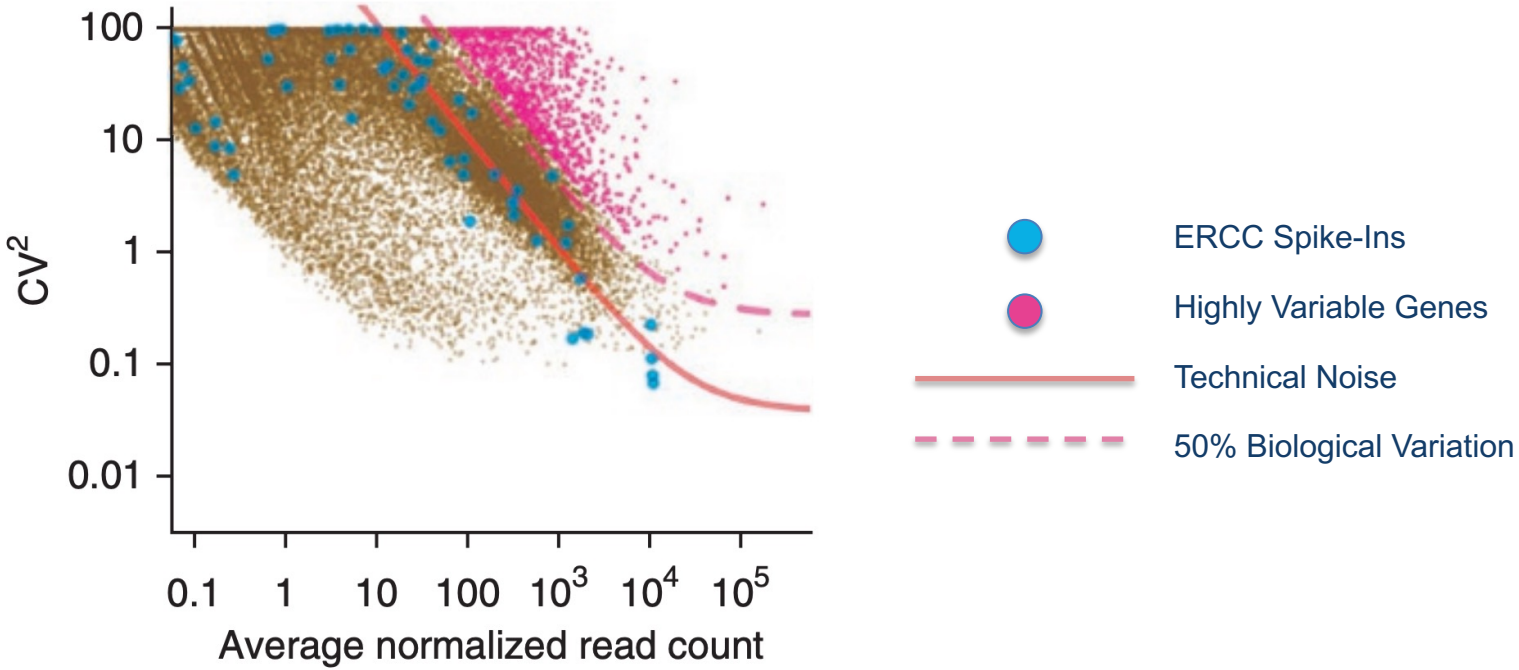
Feature Selection: ‘Variable’ Genes



SHALEKLAB
from cells to systems

Shalek*, Satija* et al, *Nature*, 2013

Drop-Out Impacts Variability In Gene Expression



Brennecke et al, *Nature Methods* (2014)

Feature Selection: How To Best Subset Data

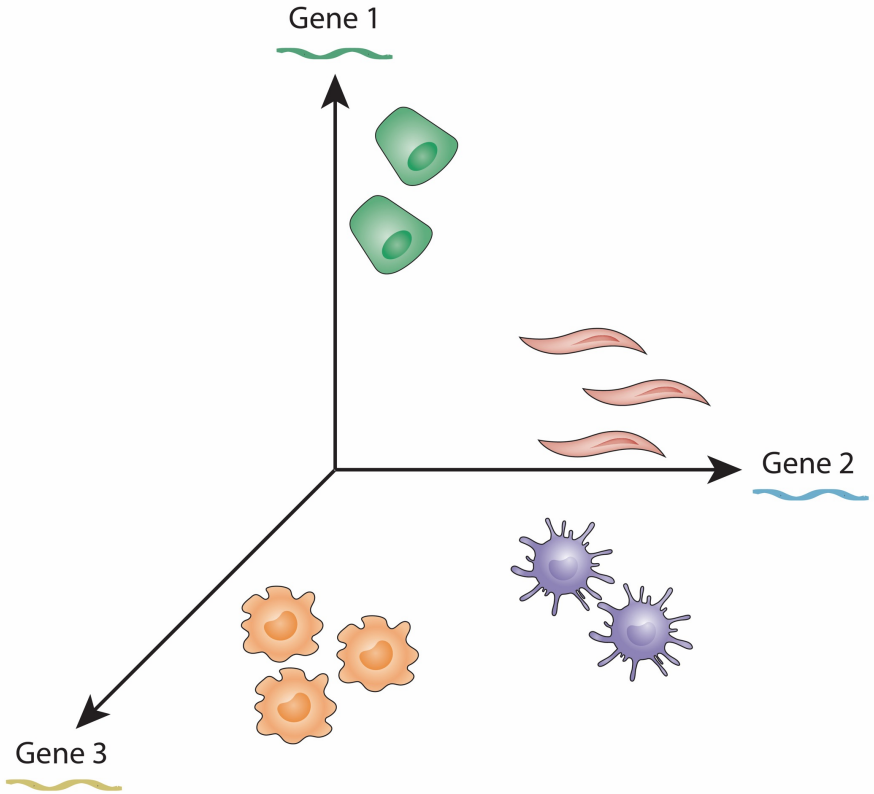
- **Choose genes you care about:**
  - Relevant to system/response of interest (i.e., induced genes)
  - Utilizes prior knowledge to help discover patterns in data
  - Afterwards - go back and identify new genes that fit the pattern

- **‘High-confidence’ measurements:**
  - Focus on highly-expressed genes
  - Have to select threshold, may focus only on housekeeping

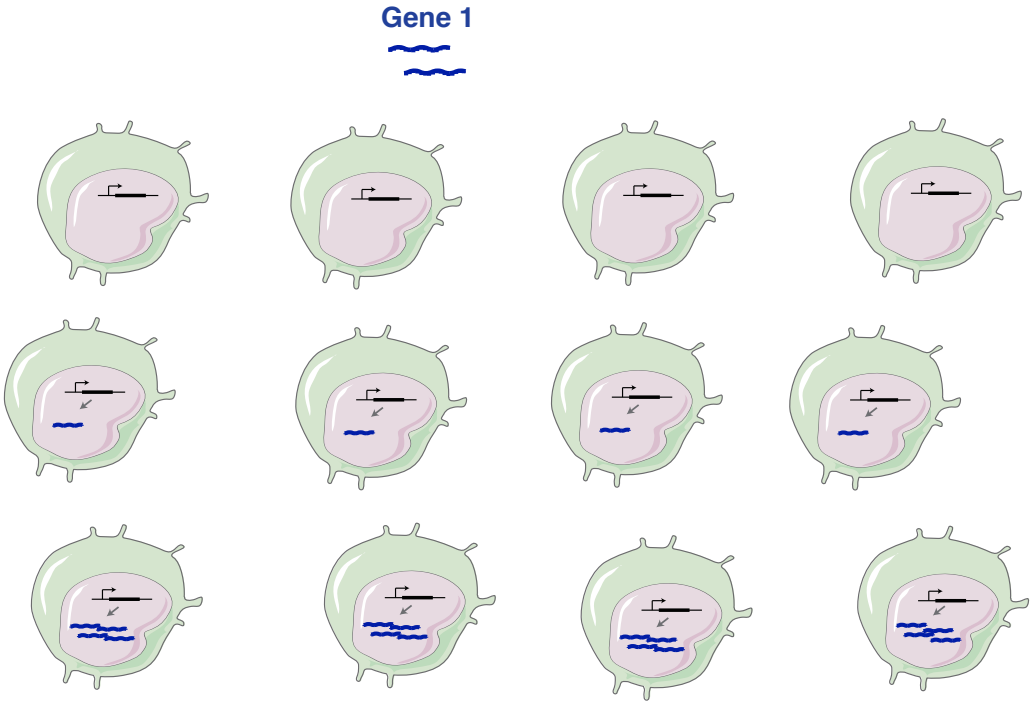
- **‘High-variability’ measurements:**
  - Discover unknown sources of heterogeneity
  - How to control for technical effects?

- **Fewer features means faster computation!**

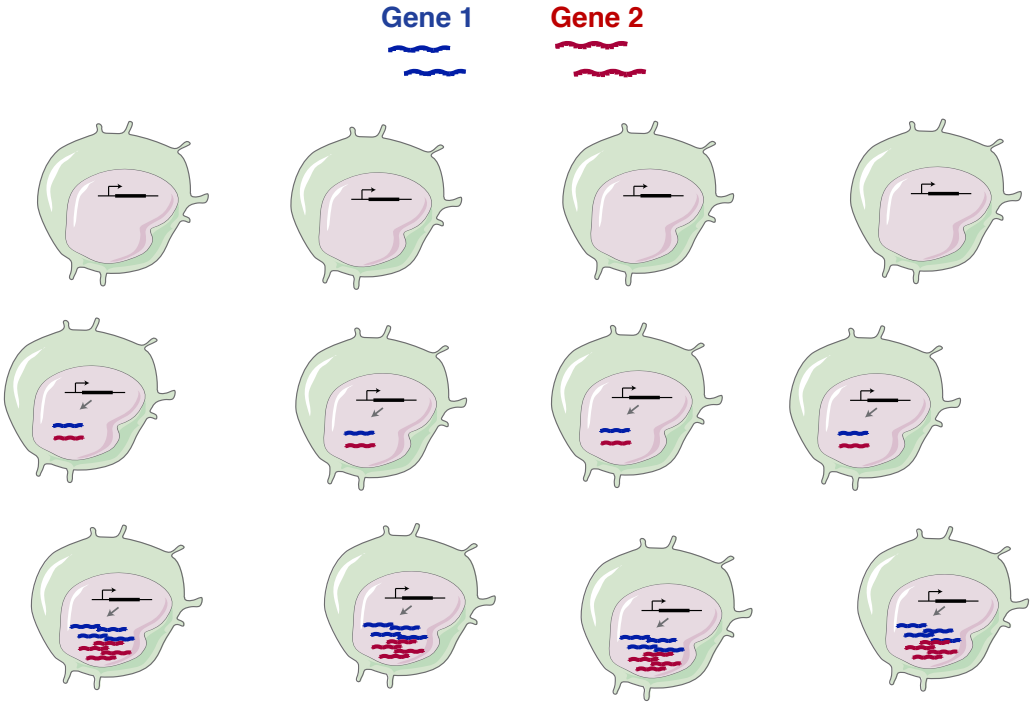
Visualization:
A Cell As A Point In Many, Many Dimensional Gene Expression Space



Visualization: Variation Is Interesting

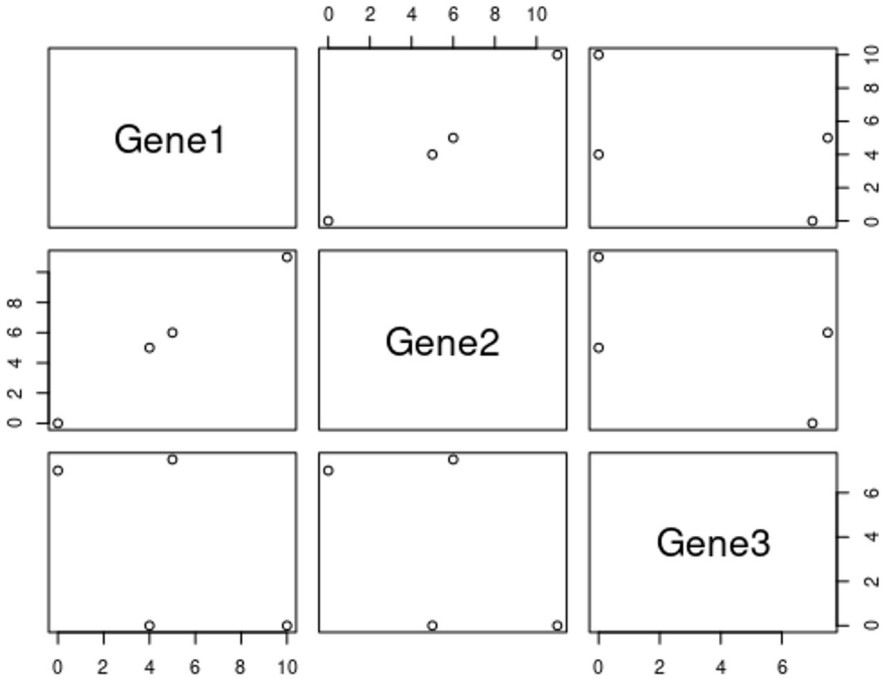


Visualization: Co-variation Implies Co-regulation



Visualization: Gene-Gene Correlation

With three genes we could do something like this...



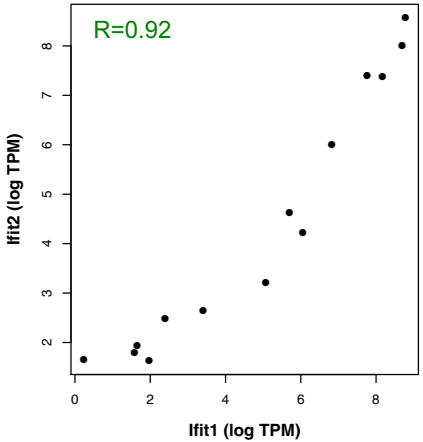
Will not work for ~20,000 genes.

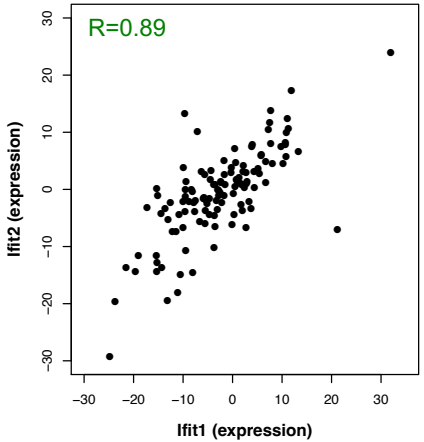
We need a way to summarize this big matrix.

To condense this information, and preserve the interesting parts, we perform dimensionality reduction.

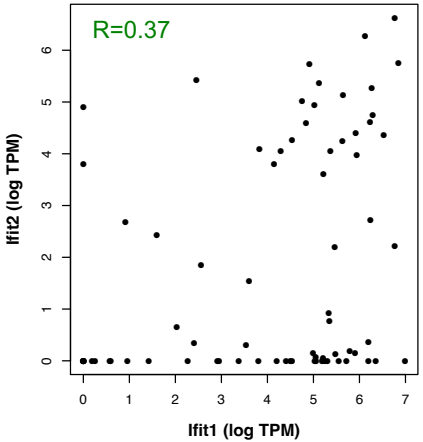
Visualization: Identifying Correlations At The Single-Cell Level?

POPULATIONS



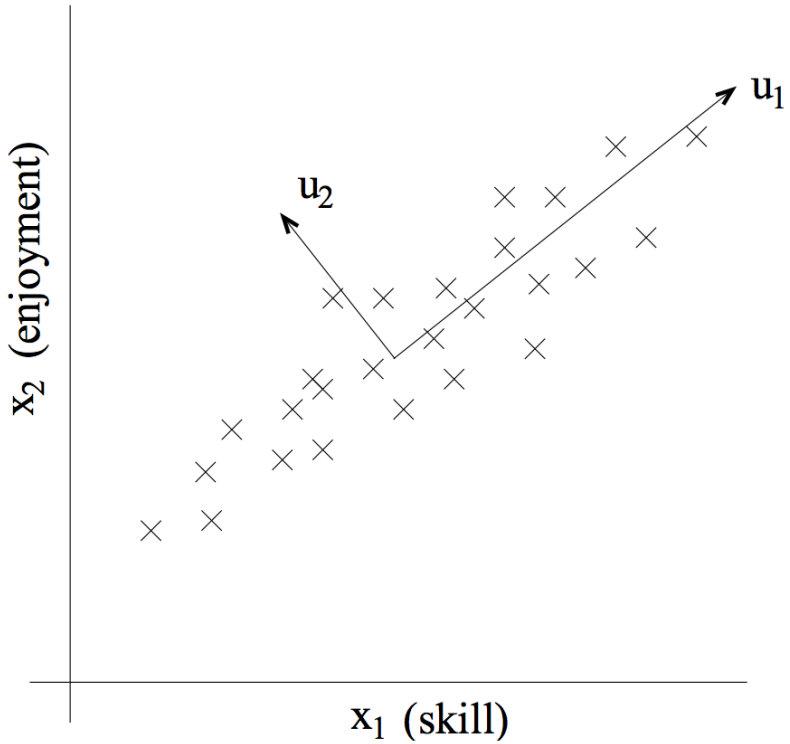


SINGLE CELLS

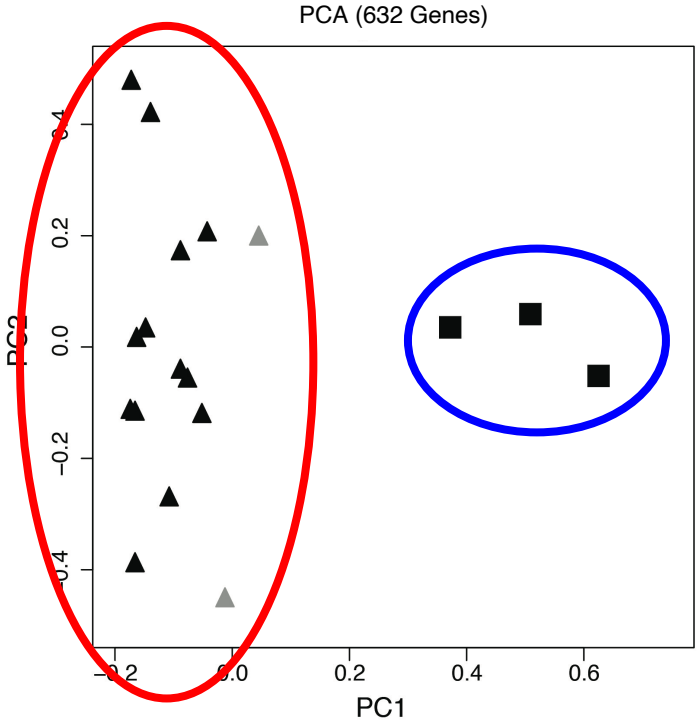


Dimensional reduction methods analyze multiple factors simultaneously

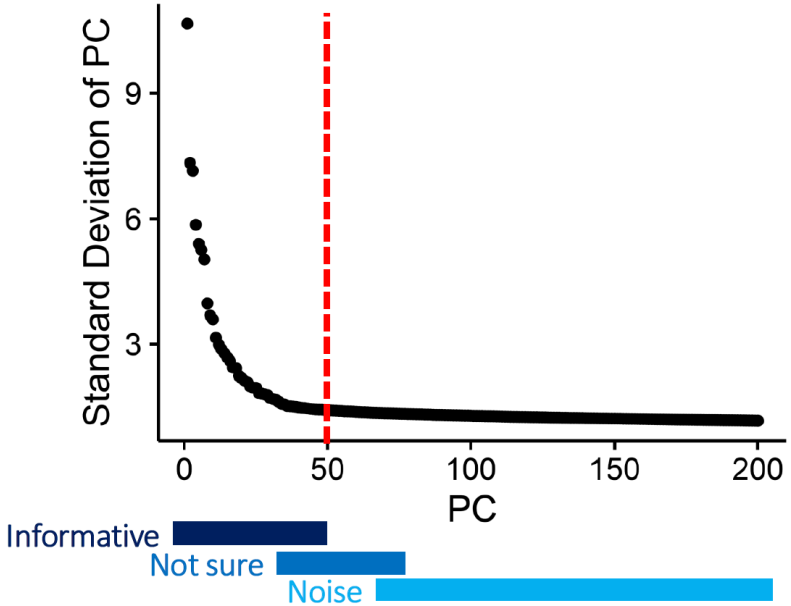
Visualization: PCA – A Toy Example



Visualization: PCA – 18 Cell Dataset



Visualization: Important Consideration For PCA



Input gene list

Can dramatically alter output

Interpretation:

‘Assigning ‘biology’ or function requires prior knowledge

PCs often correlate with technical quality

Not all PCs are significant (Chung, Storey, arXiv.org)

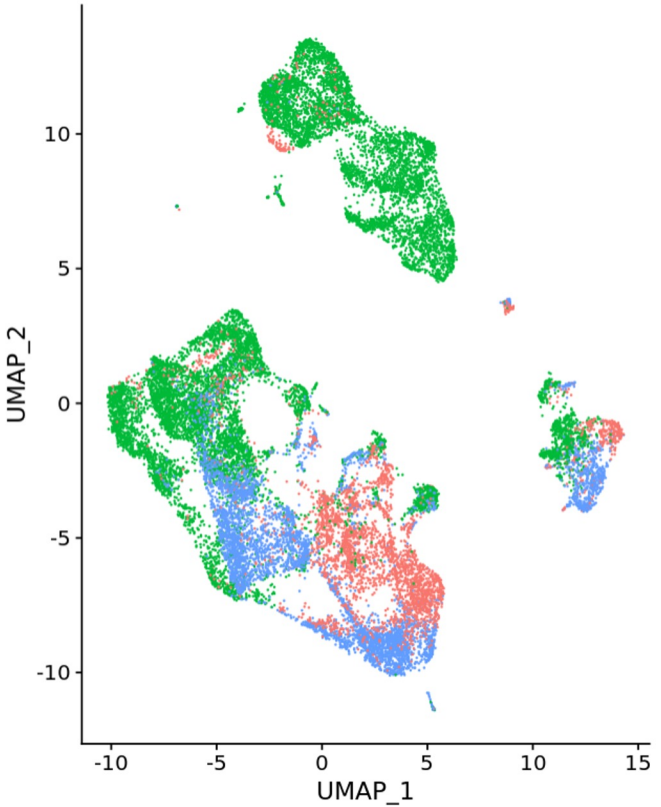
Limitations/extensions:

PCs represent **linear** combination of individual features

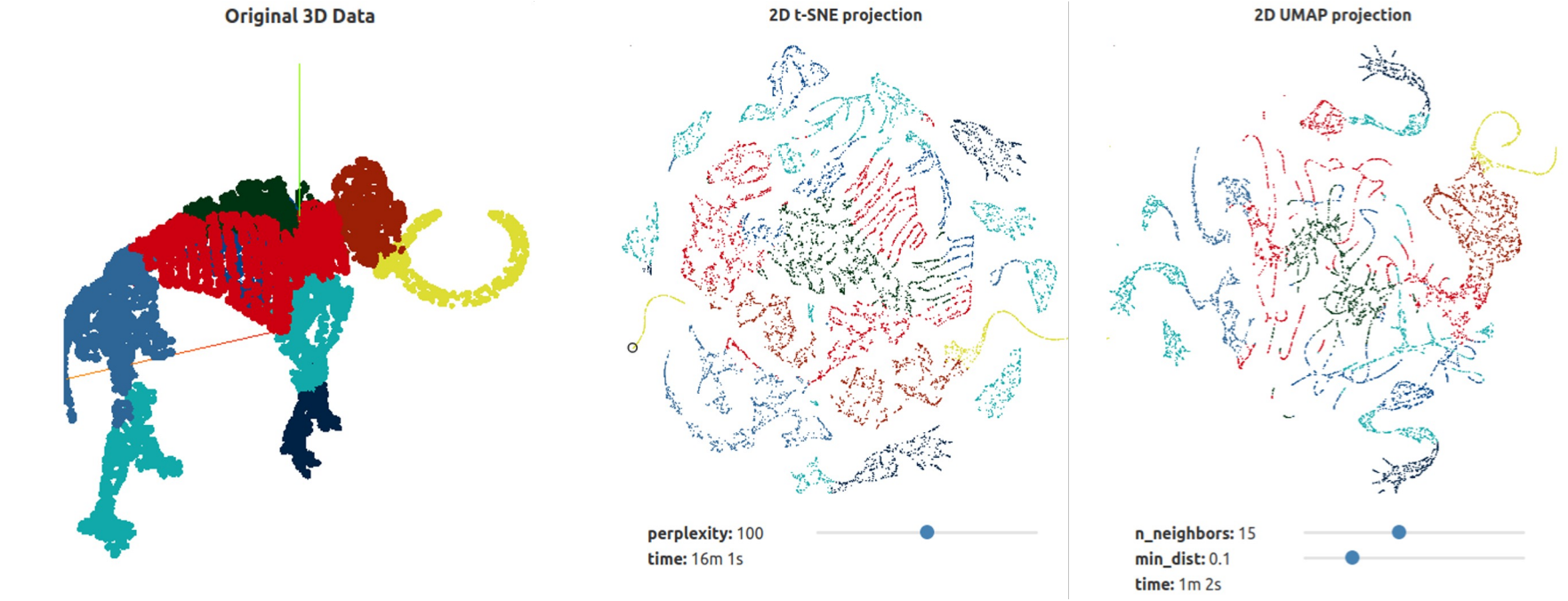
Visualization: UMAP - A Way To Visualize High-dimensional Data

PCA takes us from ~20,000 dimensions (genes) to ~20 - 50 dimensions (“metagenes”).

30-50 is still a lot to visualize. We use methods like tSNE and UMAP to summarize this information into 2 dimensions.



Visualization: UMAP Vs tSNE

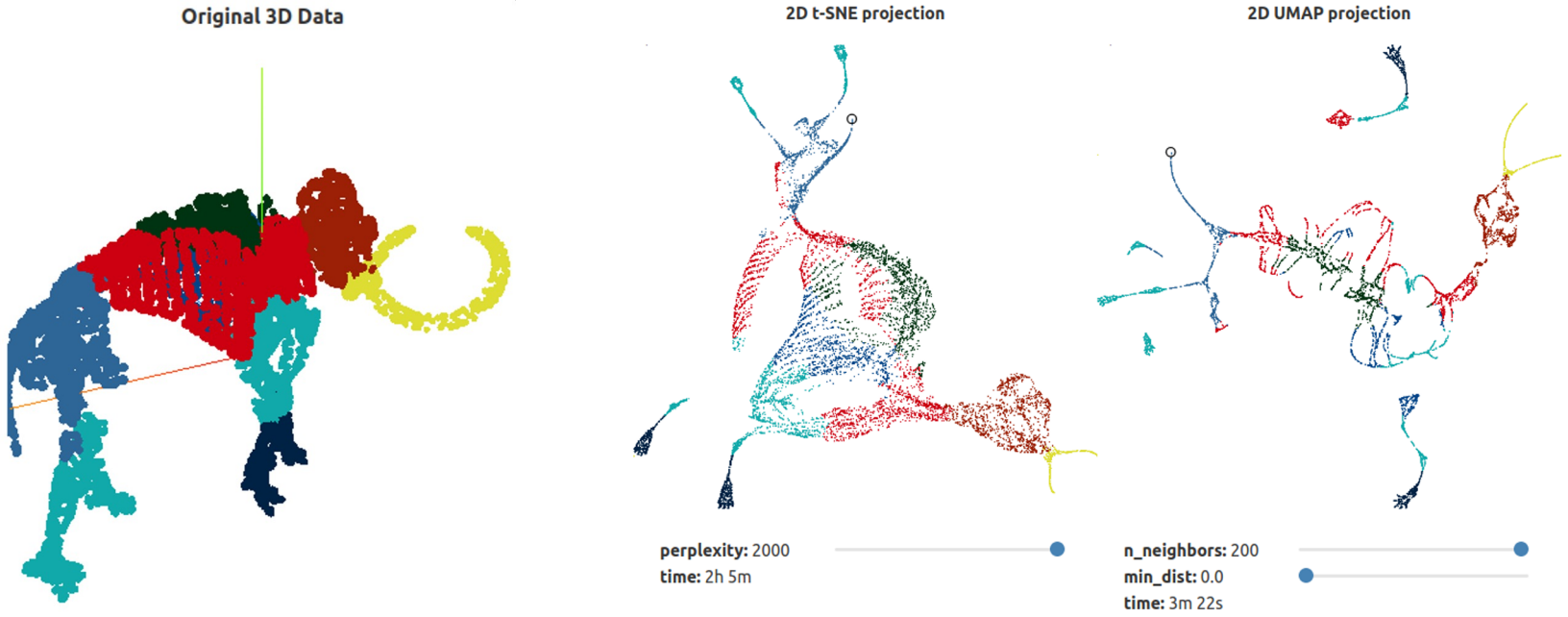


Both have good “local structure” (points generally near their neighbors), but UMAP has better “global structure” (tusks together, near head).

UMAP is also faster.

Examples from Google PAIR, https://pair-code.github.io/understanding-umap/ [Excellent tutorial!]

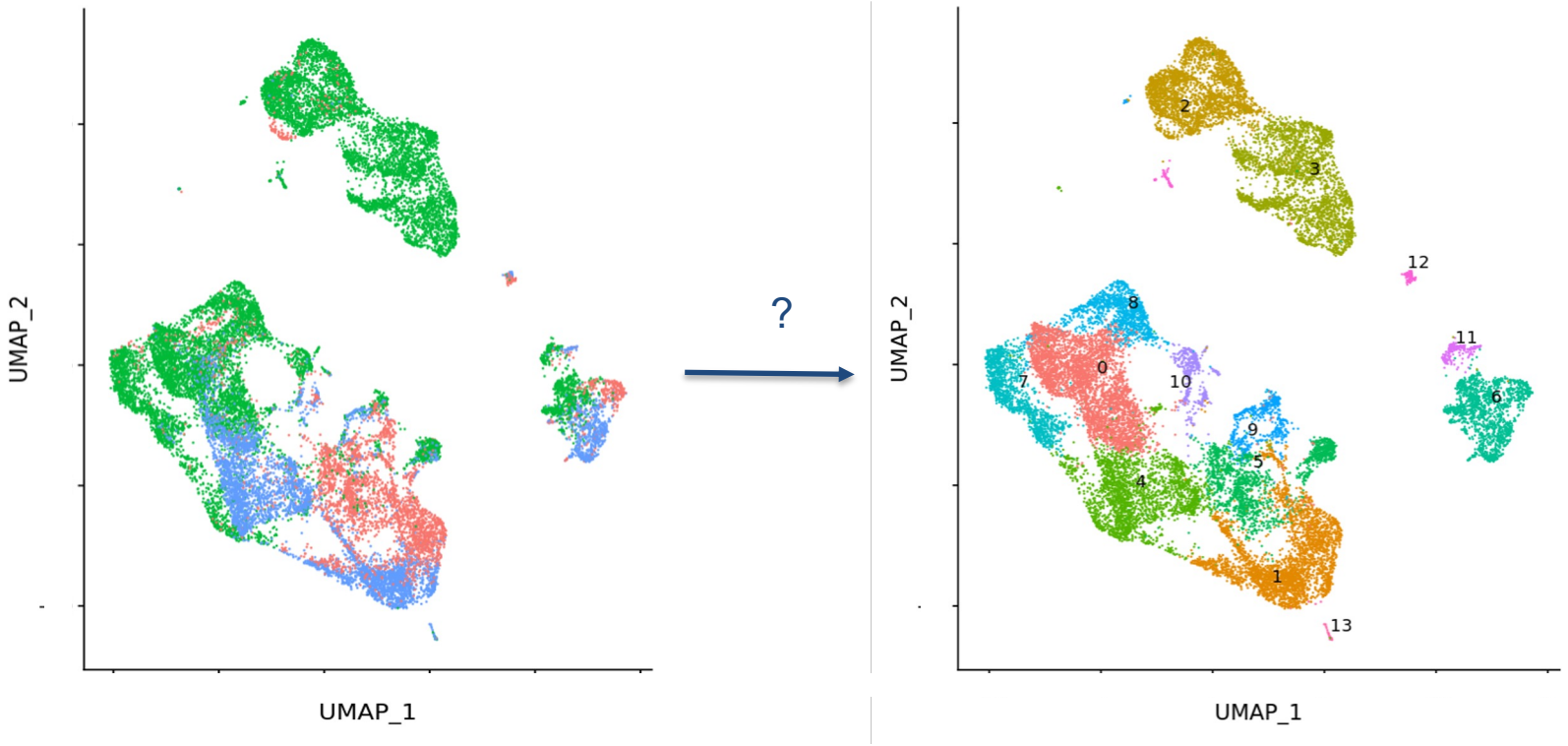
Visualization: UMAP Vs tSNE



Changing the parameters affects projections.
t-SNE looks good in this example here, but takes ~37x as much time to run.

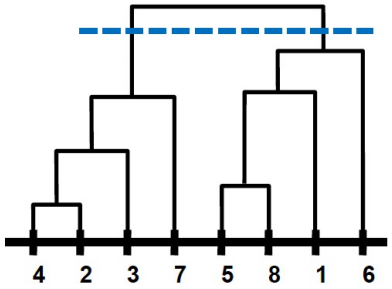
Examples from Google PAIR, https://pair-code.github.io/understanding-umap/ [Excellent tutorial!]

Clustering: Grouping Similar Cells Together

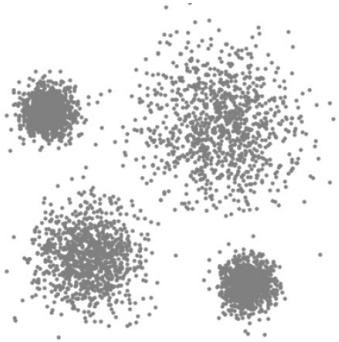


We want some objective way to identify groups of similar cells in the data.

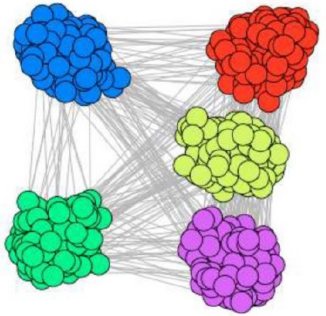
Clustering: Approaches To Clustering



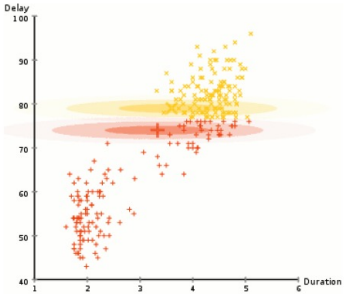
Hierarchical Clustering



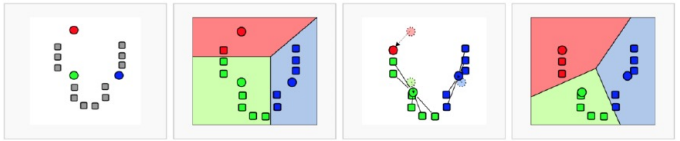
Mean shift clustering



Graph-based clustering



Gaussian mixture modeling

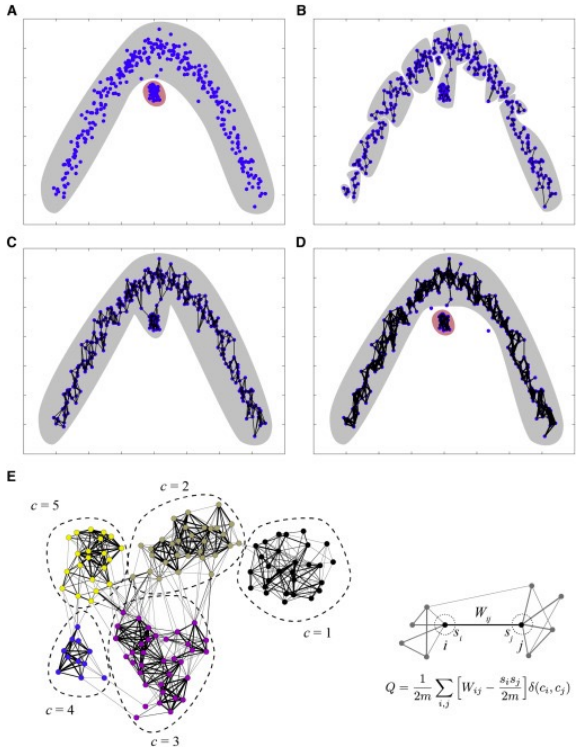


k-means clustering

https://nbisweden.github.io/excelerate-scRNAseq/ -- Ahmed Mahfouz

Clustering: KNN In Seurat

1. Construct KNN (k-nearest neighbor) graph based on the Euclidean distance in PCA space.
2. Refine the edge weights between any two cells based on the shared overlap in their local neighborhoods (Jaccard distance).
3. Cluster cells by optimizing for modularity (Louvain algorithm)



Levine et al, *Cell* (2015)

See Kiselev et al, *Nature Reviews Genetics* (2019) & Traag et al, *Scientific Reports (*2019)

Clustering: A Short List Of Methods

Name	Year	Method type	Strengths	Limitations
scanpy[4]	2018	PCA + graph-based	Very scalable	May not be accurate for small data sets
Seurat (latest)[3]	2016			
PhenoGraph[32]	2015			
SC3[22]	2017	PCA + *k*-means	High accuracy through consensus, provides estimation of *k*	High complexity, not scalable
SIMLR[24]	2017	Data-driven dimensionality reduction + *k*-means	Concurrent training of the distance metric improves sensitivity in noisy data sets	Adjusting the distance metric to make cells fit the clusters may artificially inflate quality measures
CIDR[25]	2017	PCA + hierarchical	Implicitly imputes dropouts when calculating distances	
GiniClust[75]	2016	DBSCAN	Sensitive to rare cell types	Not effective for the detection of large clusters
pcaReduce[27]	2016	PCA + *k*-means + hierarchical	Provides hierarchy of solutions	Very stochastic, does not provide a stable result
Tasic et al.[28]	2016	PCA + hierarchical	Cross validation used to perform fuzzy clustering	High complexity, no software package available
TSCAN[41]	2016	PCA + Gaussian mixture model	Combines clustering and pseudotime analysis	Assumes clusters follow multivariate normal distribution
mpath[45]	2016	Hierarchical	Combines clustering and pseudotime analysis	Uses empirically defined thresholds and a priori knowledge
BackSPIN[26]	2015	Biclustering (hierarchical)	Multiple rounds of feature selection improve clustering resolution	Tends to over-partition the data
RaceID[23], RaceID2[115], RaceID3	2015	*k*-Means	Detects rare cell types, provides estimation of *k*	Performs poorly when there are no rare cell types
SINCERA[5]	2015	Hierarchical	Method is intuitively easy to understand	Simple hierarchical clustering is used, may not be appropriate for very noisy data
SNN-Cliq[80]	2015	Graph-based	Provides estimation of *k*	High complexity, not scalable

DBSCAN, density-based spatial clustering of applications with noise; PCA, principal component analysis; scRNA-seq, single-cell RNA sequencing.

Kiselev et al, *Nature Review Genetics* (2019)

Clustering: Words Of Warning

It's basically a choose your own adventure. You pick:

1. Similarity measure
2. Algorithm

Different choices leads to different results and it's hard to know when to stop.

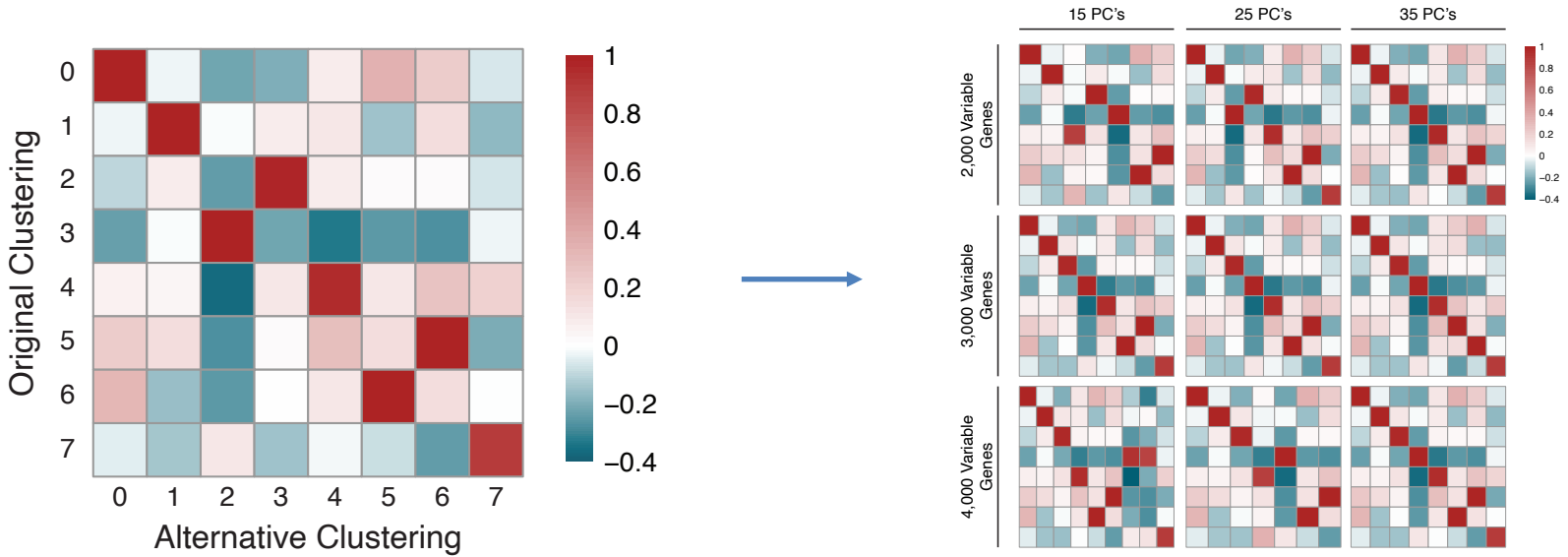
It's critical to examine yours, test what's driving them (QC metrics?) and test stability

And remember, your results provide hypotheses not answers!

Clustering: Robustness via Correlations between Average Expression Profiles

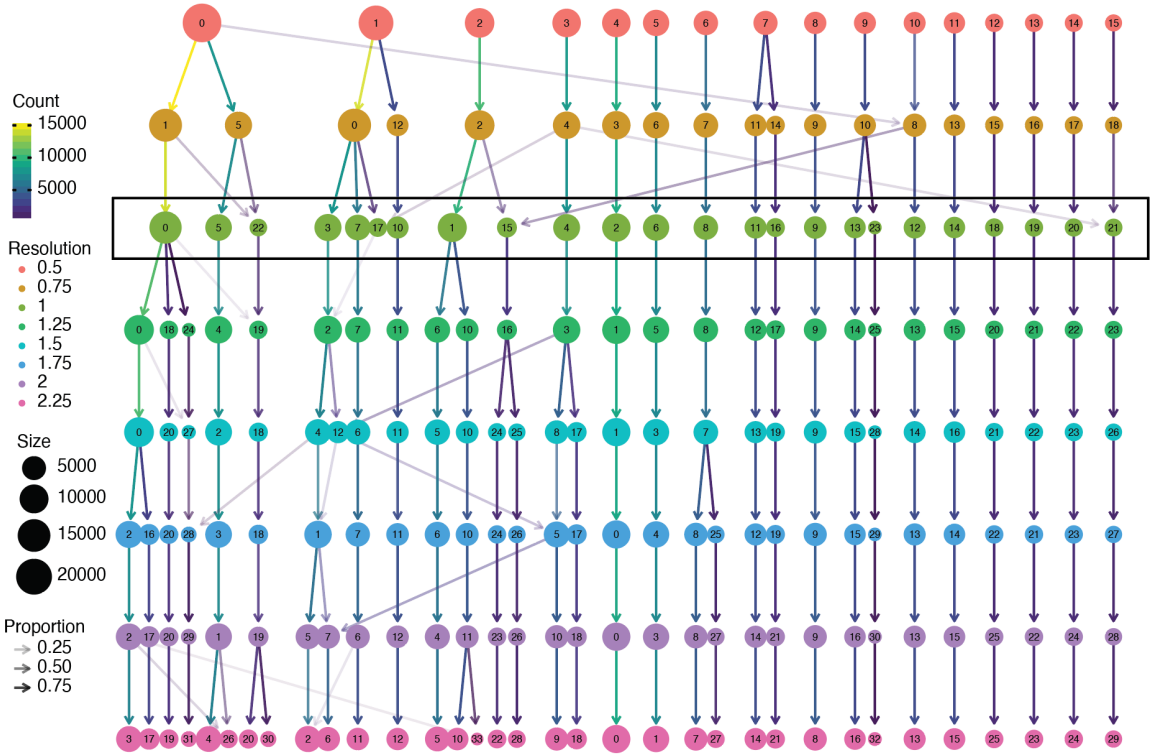
Idea:

- Original clustering: Annotate clusters based on a rationally-chosen set of parameters
- Alternative clustering: Repeat with other sets of parameters
- Find correlations between average expression profiles of clusters across the two objects



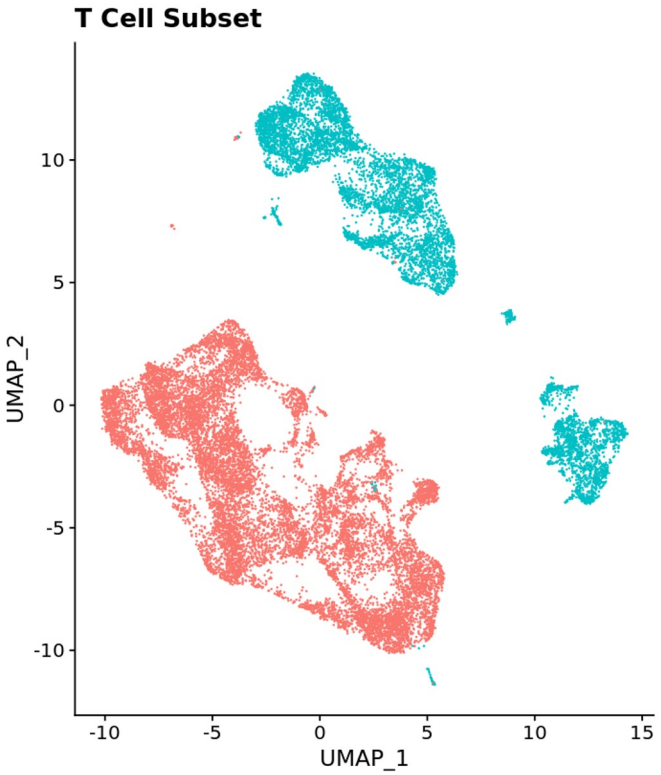
Thanks to Constantine Tzouanas

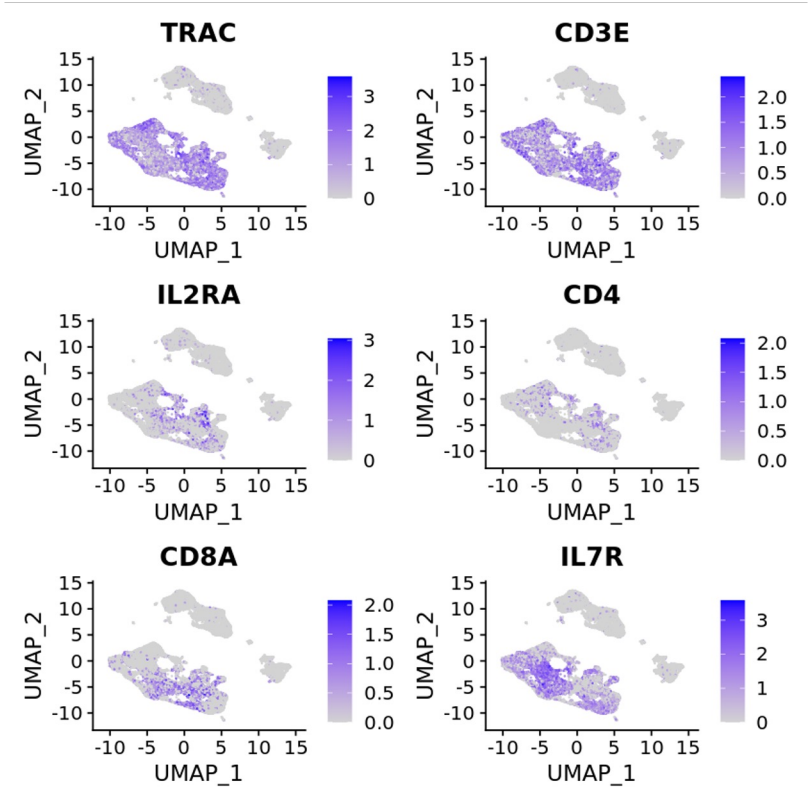
Clustering: Clustree



Zappia et al, *GigaScience* (2018)

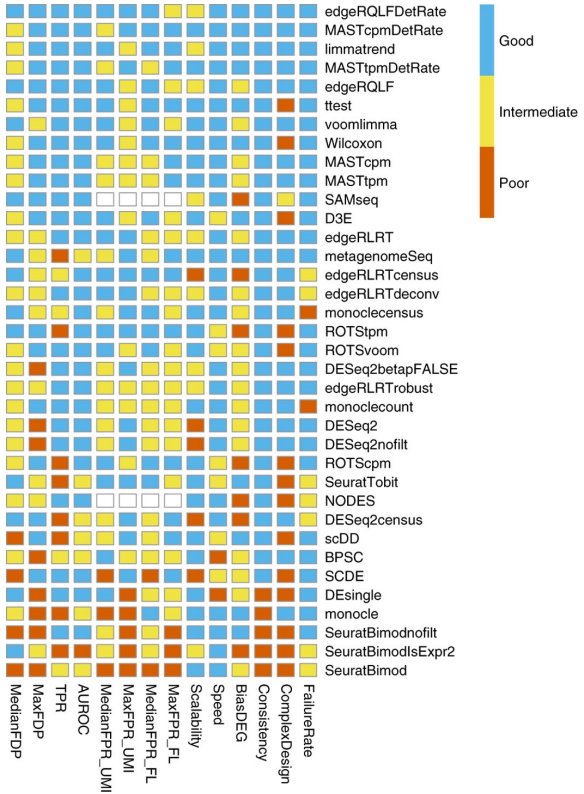
Marker Identification





SHALEKLAB
from cells to systems

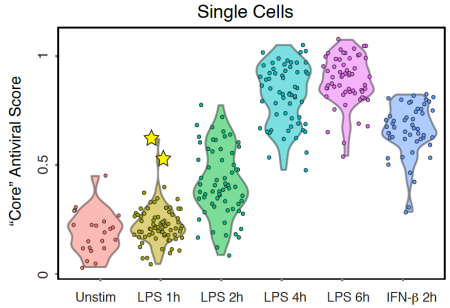
Differential Expression (DE): Many Methods, Complex Isn't Always Better



Soneson et al, *Nature Methods* (2018)

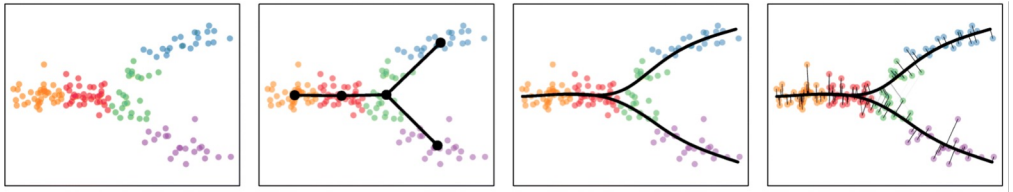
Many Other Things: What Else Can We Do With scRNA-Seq?

Pathway Analysis



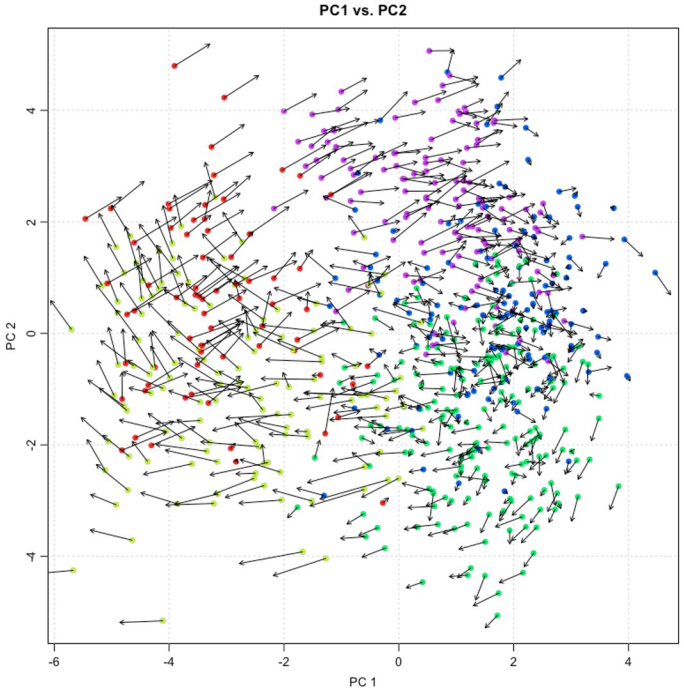
Shalek*, Satija* et al, *Nature*, 2014

Trajectory Analysis



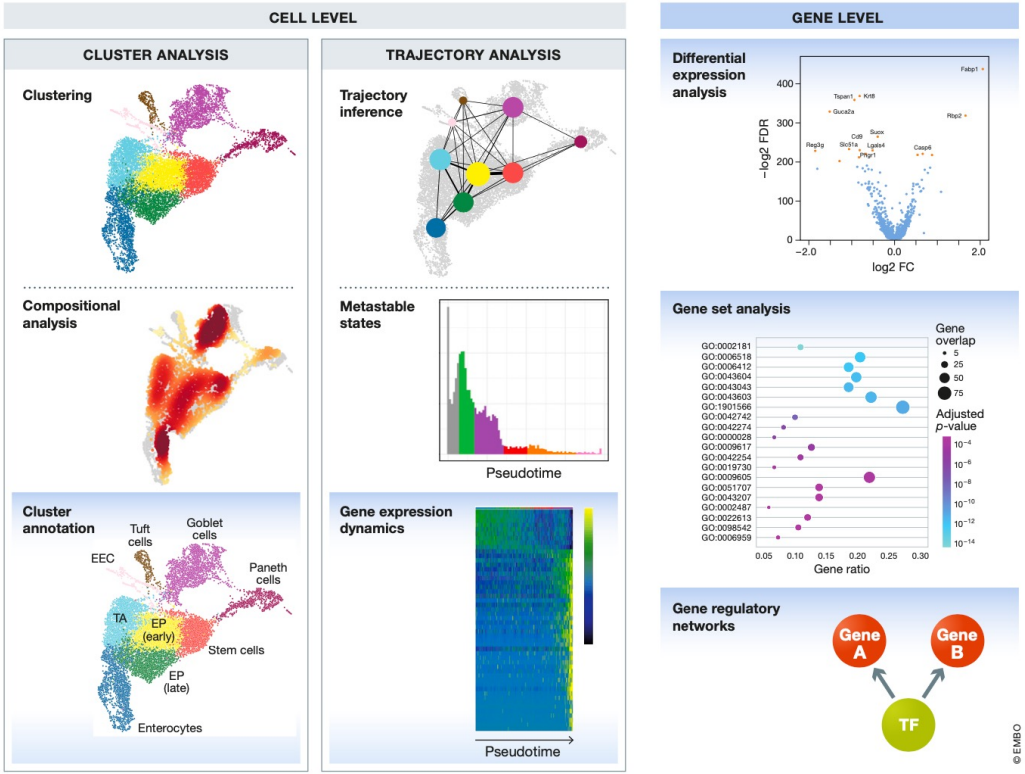
Street et al, *BMC Genomics* (2018)

RNA Velocity



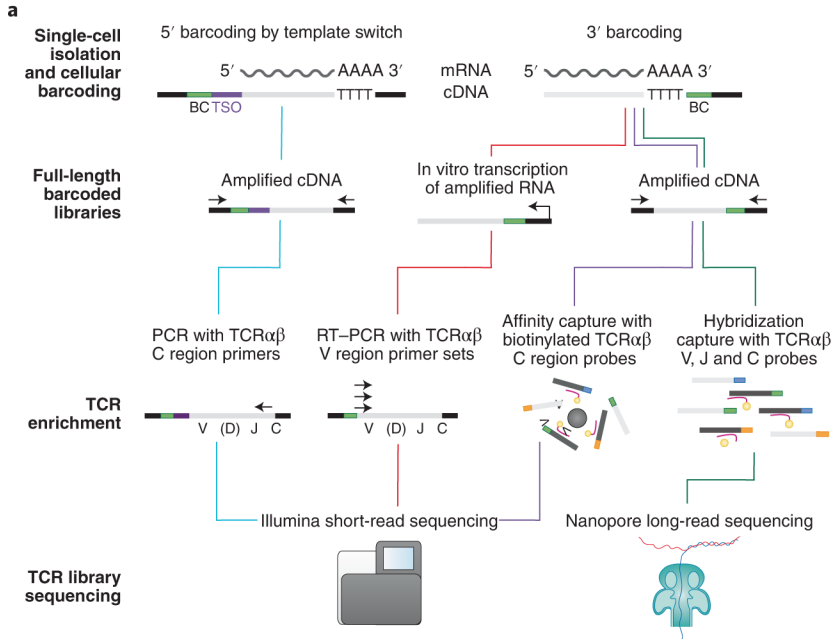
https://jef.works/blog/2020/01/14/rna_velocity_analysis_tutorial_tips/

Many Other Things: What Else Can We Do With scRNA-Seq?



Luecken & Theis, *MSB* (2019)

Capturing Additional Information – TCR/BCR

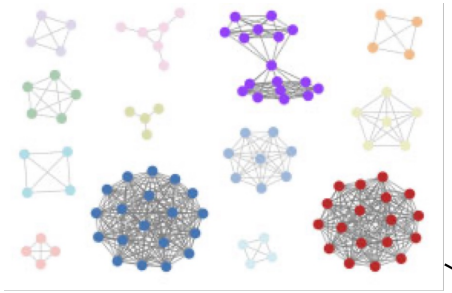


Pai et al, *Nature Methods* (2021)

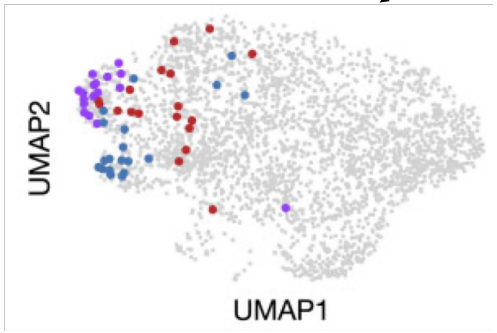
Computational Analyses Of VDJ Sequencing In Single Cells

Example workflow:

- Analyze single-cell profiles stratified by clonotypes.
- Relate clonal expansion to infection or vaccination
- Choose BCR sequences from expanded clones to generate monoclonal antibodies for diagnostics or therapeutics



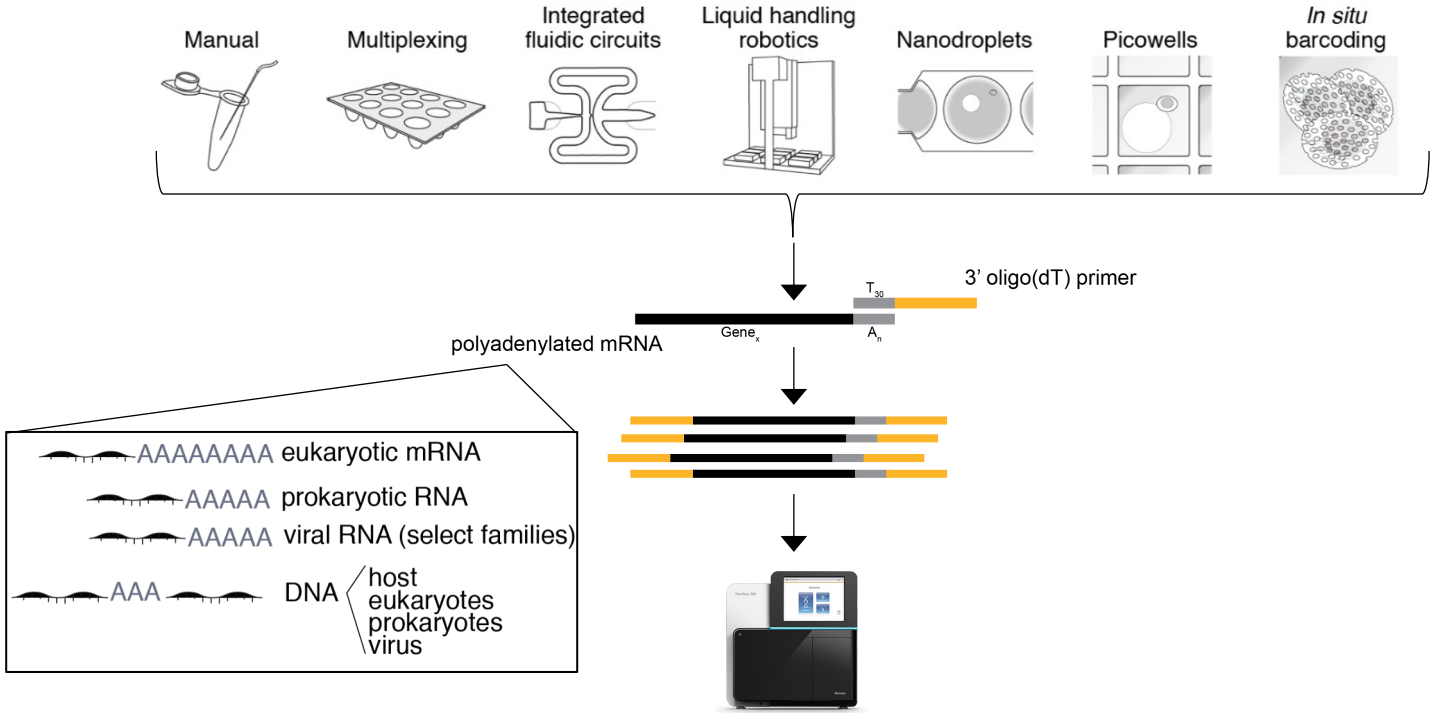
clonotypes group cells by shared BCR sequences



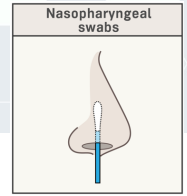
clonotypes add information to gene expression embeddings

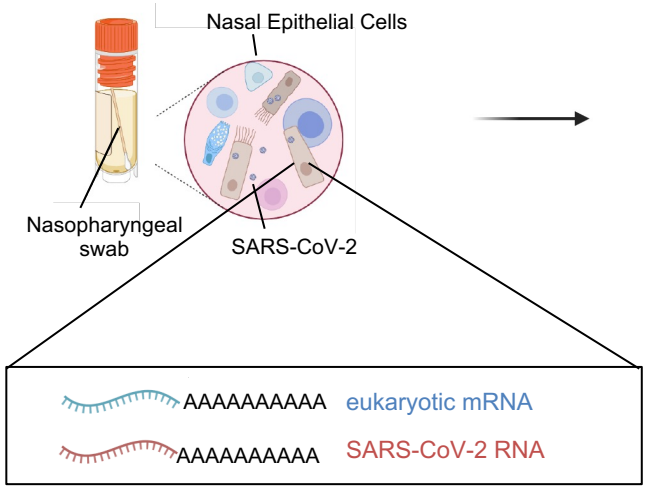
Sturm et al, *Bioinformatics* (2020)

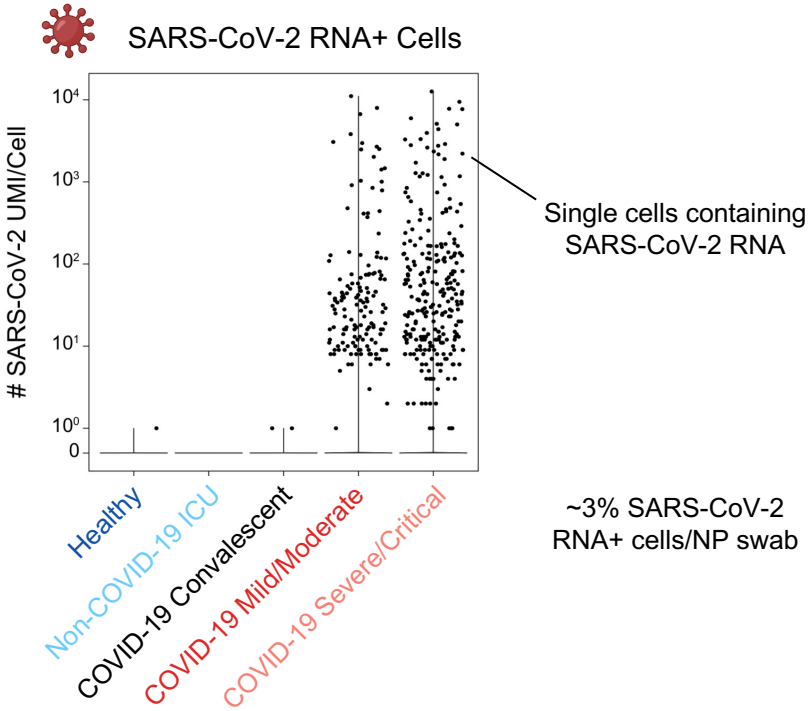
Detecting Viral Transcripts With Single-Cell RNA-Seq



Co-Detection of SARS-CoV-2 and Host RNA



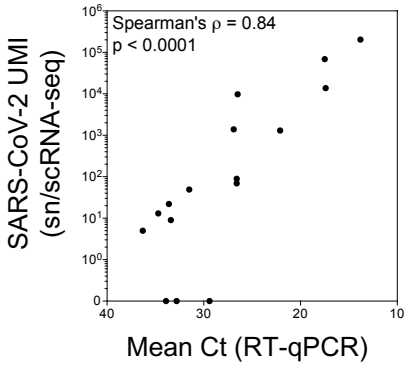




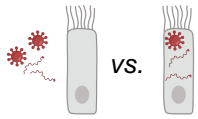
Zielger et al, *Cell* (2021)

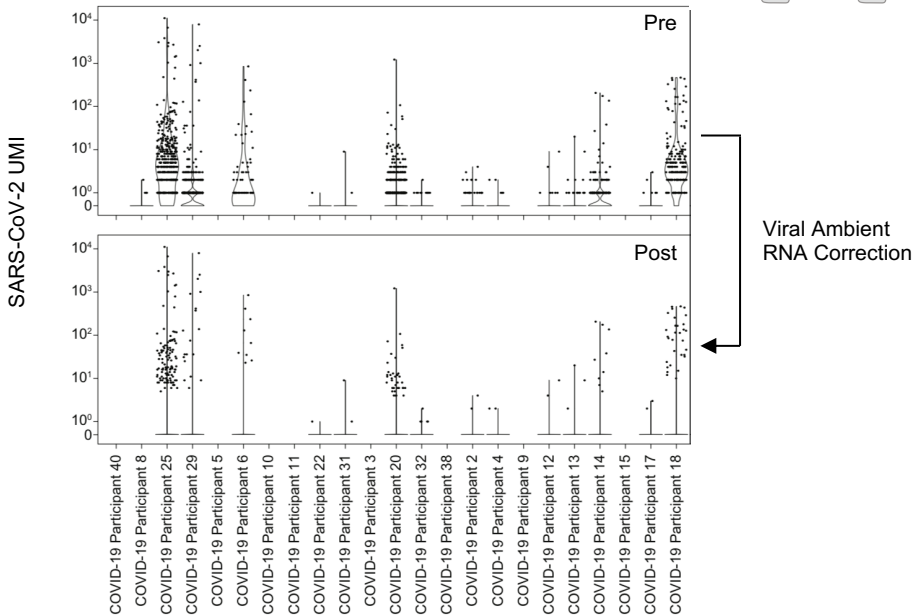
Co-Detection of SARS-CoV-2 and Host RNA

1. **Sensitivity** of scRNA-seq capture of viral RNA



2. **Correct** for **ambient** SARS-CoV-2 RNA



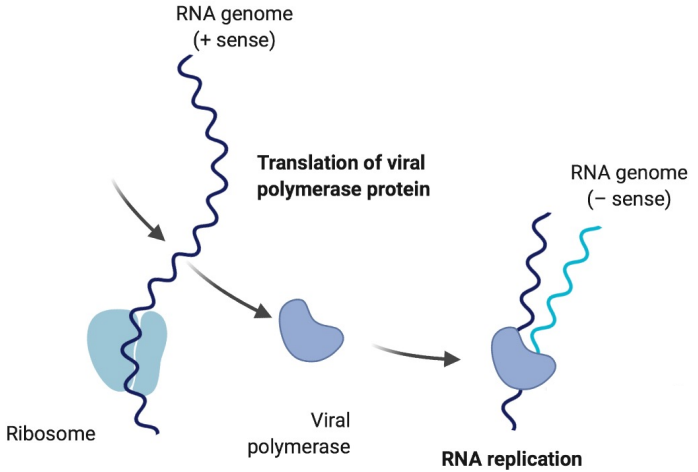


SHALEKLAB
from cells to systems

Kotilar et al, *Cell* (2020), Fleming et al, *bioRxiv* (2020),
Delorey et al, *Nature* (2021), Cao et al, *bioRxiv* (2020)

Co-Detection of SARS-CoV-2 and Host RNA

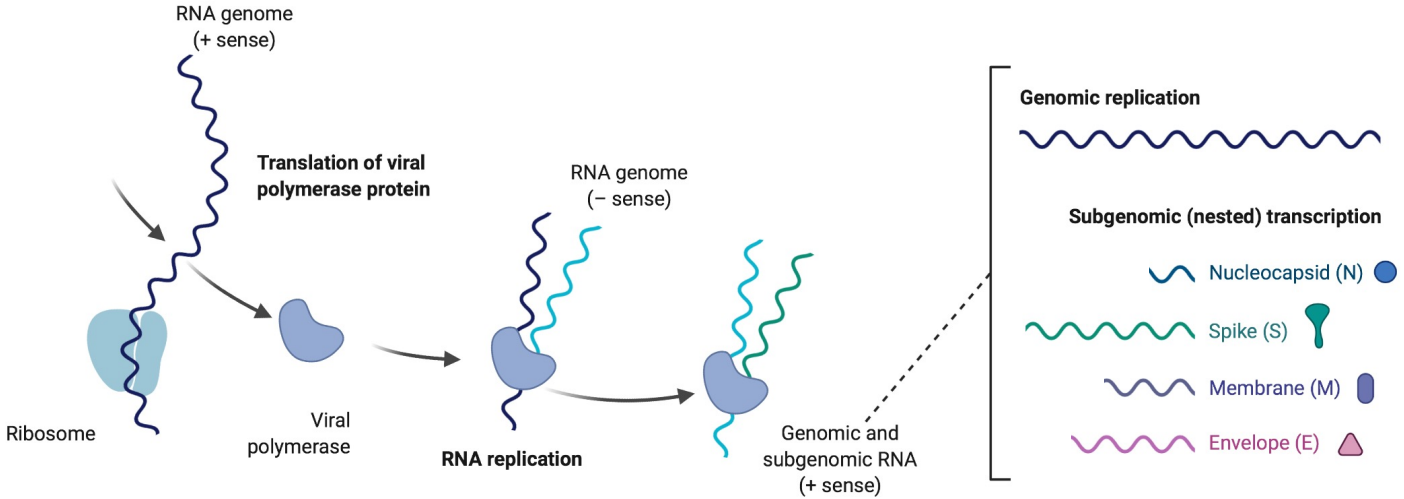
3. Identify RNA molecules **indicative** of **replicating** virus



Sawicki et al, *J Virology* (2008), BioRender

Co-Detection of SARS-CoV-2 and Host RNA

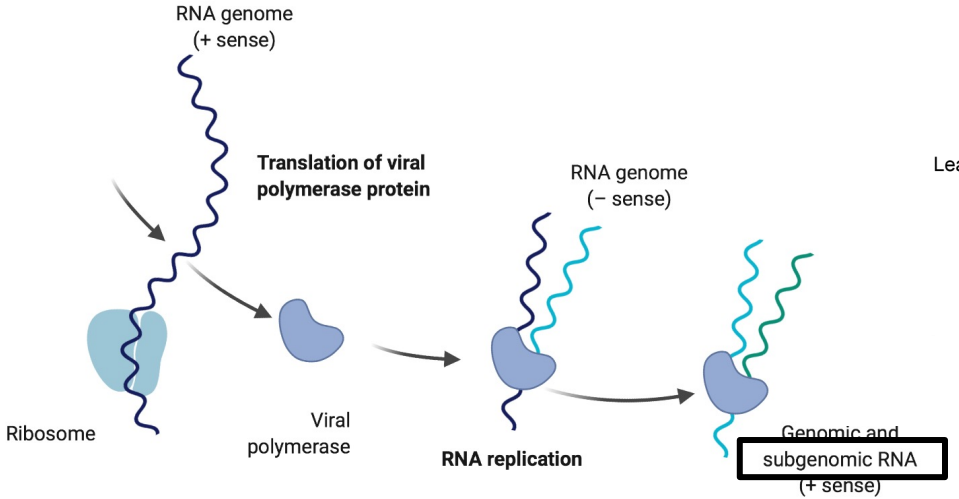
3. Identify RNA molecules **indicative** of **replicating** virus

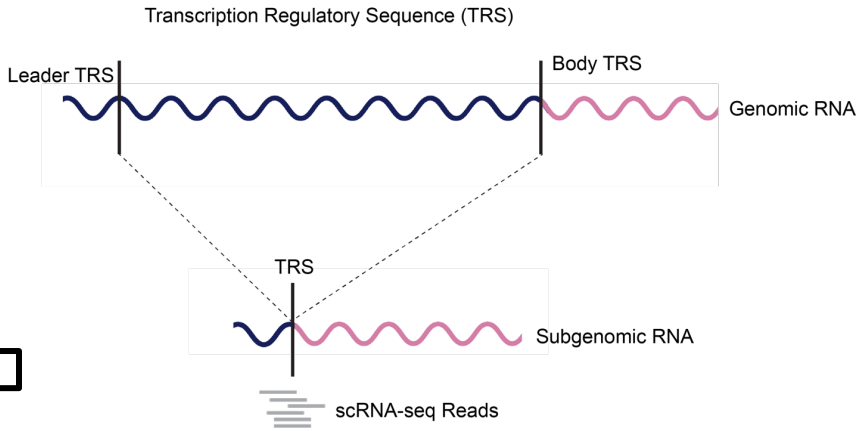


Sawicki et al, *J Virology* (2008), BioRender

Co-Detection of SARS-CoV-2 and Host RNA

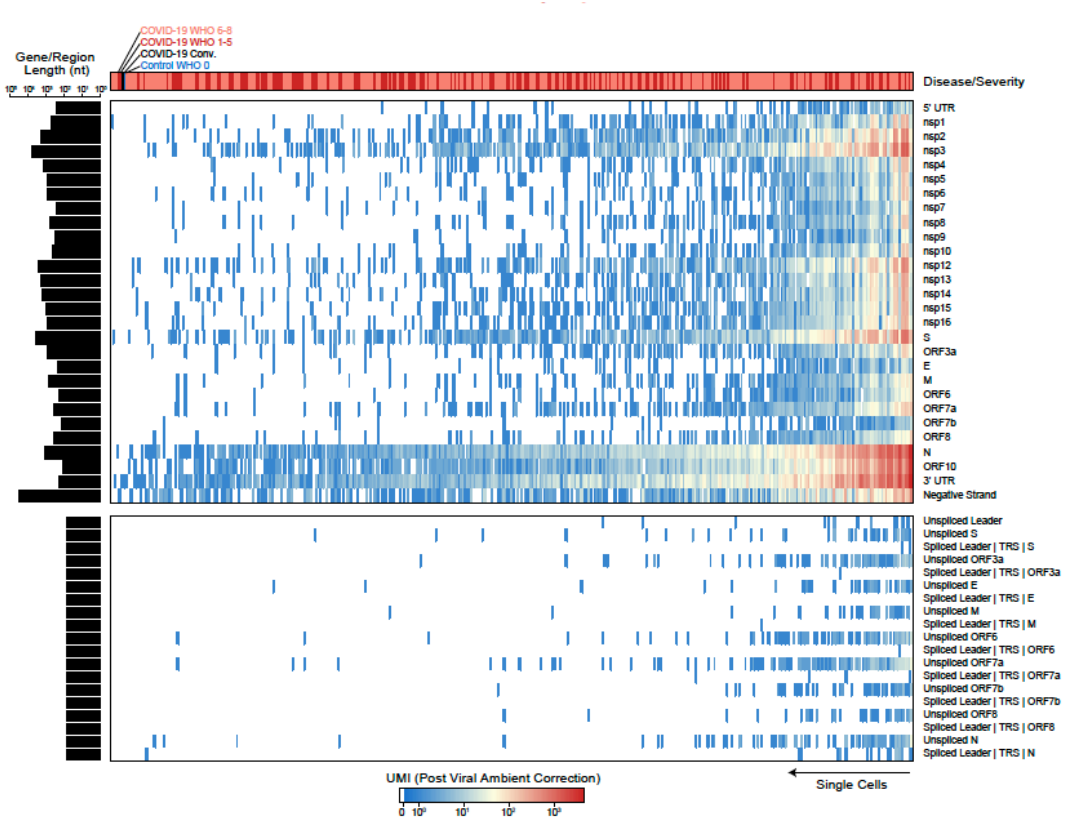
3. Identify RNA molecules **indicative** of **replicating** virus





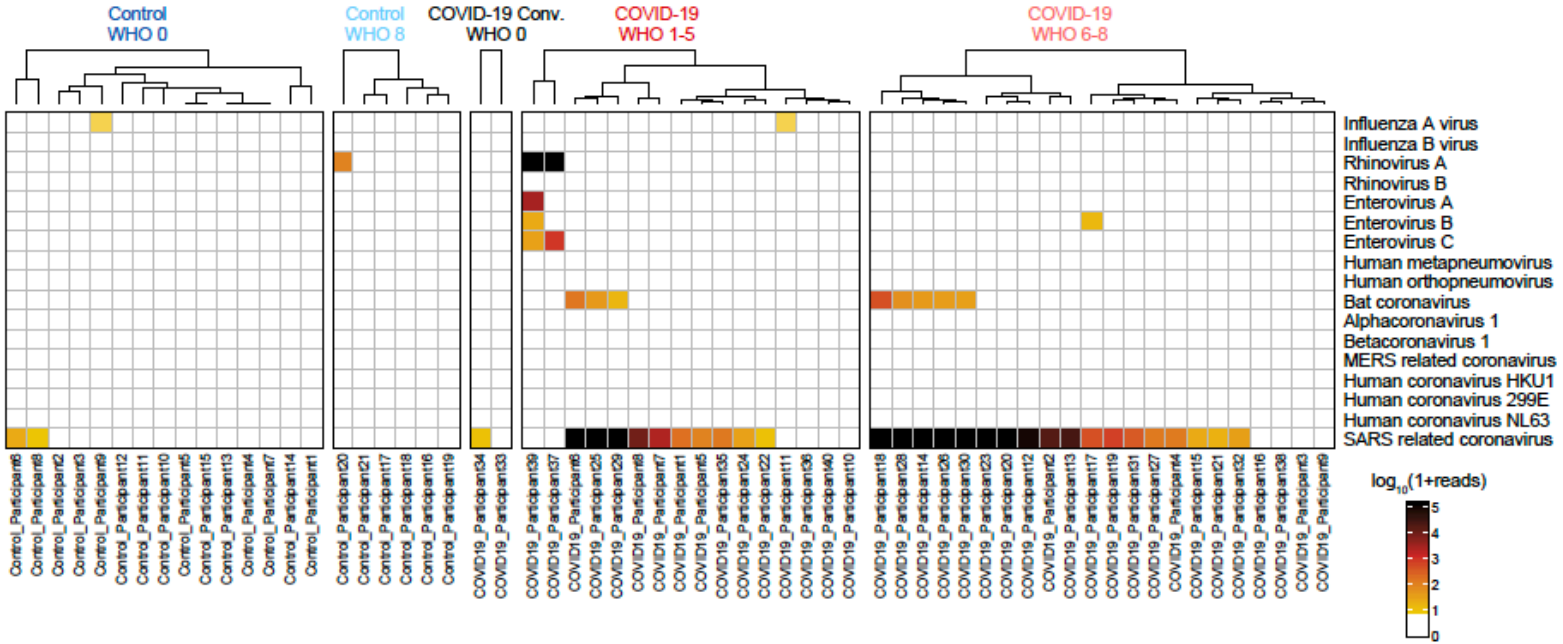
Sawicki et al, *J Virology* (2008), BioRender

Codetection Of Host And Viral RNA



Ziegler et al, *Cell* (2021).

COVID-19: Co-detection of host and viral RNA



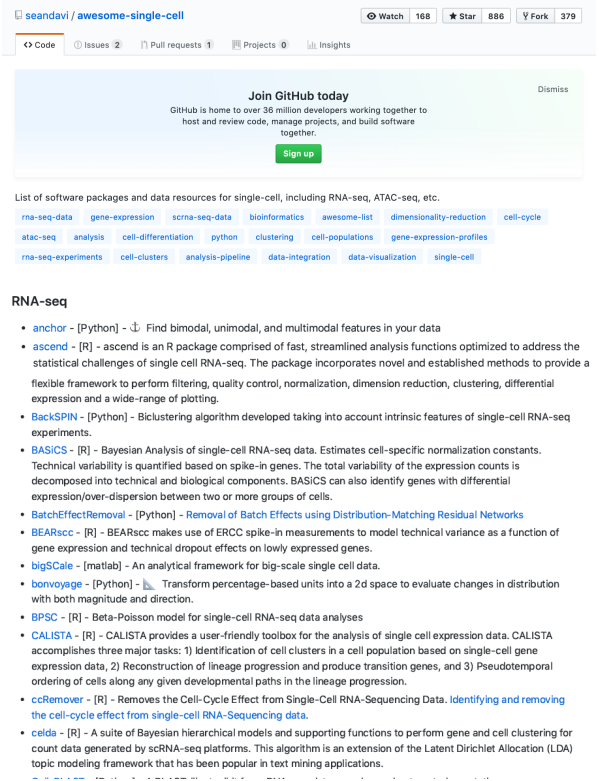
SARS-CoV-2 reads reliably detected in most Covid-19 participants

Kraken2 method adapted from Lemieux et al, *Science* (2020).

In Summary: There Are A LOT Of Analysis Tools Available

- There are over a hundred options out there of which many are good
- Each method or package makes certain assumptions about the data
- The most important thing is to carefully consider whether the tool is appropriate given your system and experimental design

*See Seurat, SCANPY and similar for some basic tutorials



seandavi / awesome-single-cell

Watch 168 ★ Star 886 Fork 379

Code Issues 2 Pull requests 1 Projects 0 Insights

Dismiss

Join GitHub today

GitHub is home to over 36 million developers working together to host and review code, manage projects, and build software together.

Sign up

List of software packages and data resources for single-cell, including RNA-seq, ATAC-seq, etc.

rna-seq-data gene-expression scrna-seq-data bioinformatics awesome-list dimensionality-reduction cell-cycle atac-seq analysis cell-differentiation python clustering cell-populations gene-expression-profiles rna-seq-experiments cell-clusters analysis-pipeline data-integration data-visualization single-cell

RNA-seq

- anchor - [Python] - ⚓ Find bimodal, unimodal, and multimodal features in your data
- ascend - [R] - ascend is an R package comprised of fast, streamlined analysis functions optimized to address the statistical challenges of single cell RNA-seq. The package incorporates novel and established methods to provide a flexible framework to perform filtering, quality control, normalization, dimension reduction, clustering, differential expression and a wide-range of plotting.
- BackSPIN - [Python] - Biclustering algorithm developed taking into account intrinsic features of single-cell RNA-seq experiments.
- BASiCS - [R] - Bayesian Analysis of single-cell RNA-seq data. Estimates cell-specific normalization constants. Technical variability is quantified based on spike-in genes. The total variability of the expression counts is decomposed into technical and biological components. BASiCS can also identify genes with differential expression/over-dispersion between two or more groups of cells.
- BatchEffectRemoval - [Python] - Removal of Batch Effects using Distribution-Matching Residual Networks
- BEARscc - [R] - BEARscc makes use of ERCC spike-in measurements to model technical variance as a function of gene expression and technical dropout effects on lowly expressed genes.
- bigSCale - [matlab] - An analytical framework for big-scale single cell data.
- bonvoyage - [Python] - 📐 Transform percentage-based units into a 2d space to evaluate changes in distribution with both magnitude and direction.
- BPSC - [R] - Beta-Poisson model for single-cell RNA-seq data analyses
- CALISTA - [R] - CALISTA provides a user-friendly toolbox for the analysis of single cell expression data. CALISTA accomplishes three major tasks: 1) Identification of cell clusters in a cell population based on single-cell gene expression data, 2) Reconstruction of lineage progression and produce transition genes, and 3) Pseudotemporal ordering of cells along any given developmental paths in the lineage progression.
- ccRemover - [R] - Removes the Cell-Cycle Effect from Single-Cell RNA-Sequencing Data. Identifying and removing the cell-cycle effect from single-cell RNA-Sequencing data.
- celda - [R] - A suite of Bayesian hierarchical models and supporting functions to perform gene and cell clustering for count data generated by scRNA-seq platforms. This algorithm is an extension of the Latent Dirichlet Allocation (LDA) topic modeling framework that has been popular in text mining applications.

+14 more pages

https://github.com/seandavi/awesome-single-cell

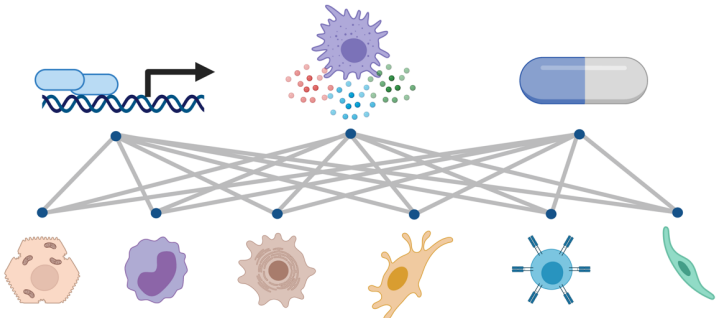
The Big Picture And The Future: Minding The Gap
From Passive Descriptions To Actionable Mechanisms

scRNA-seq can be very effective at dissecting out the key cellular and molecular features of complex biological systems

However, we need ways to move from correlative analyses to mechanistic insights that can impact human health and disease.

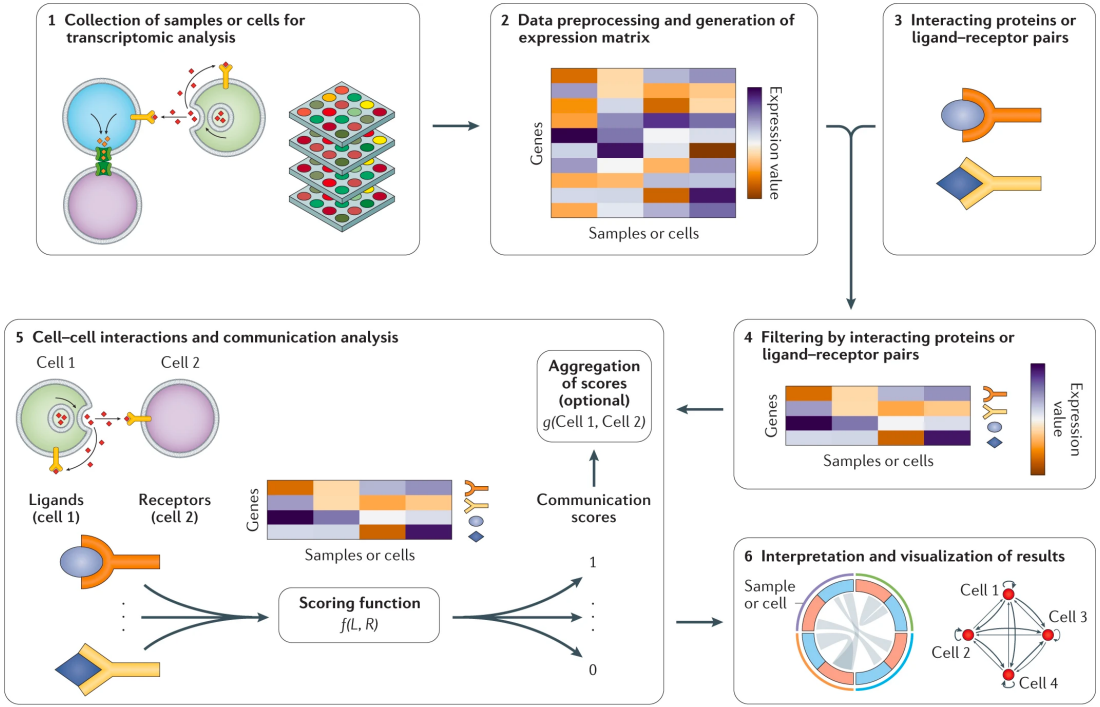
This means computing:

- Intercellular communication networks
- Gene regulatory networks
- And more…



Schematic made via biorender.com

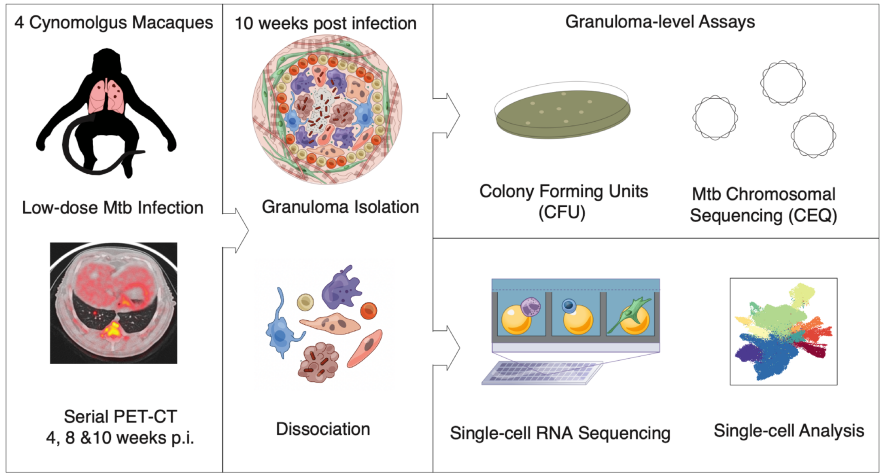
From Passive Descriptions To Actionable Mechanisms: Deciphering How Cells Function *In Situ* Via Intercellular Communication Networks

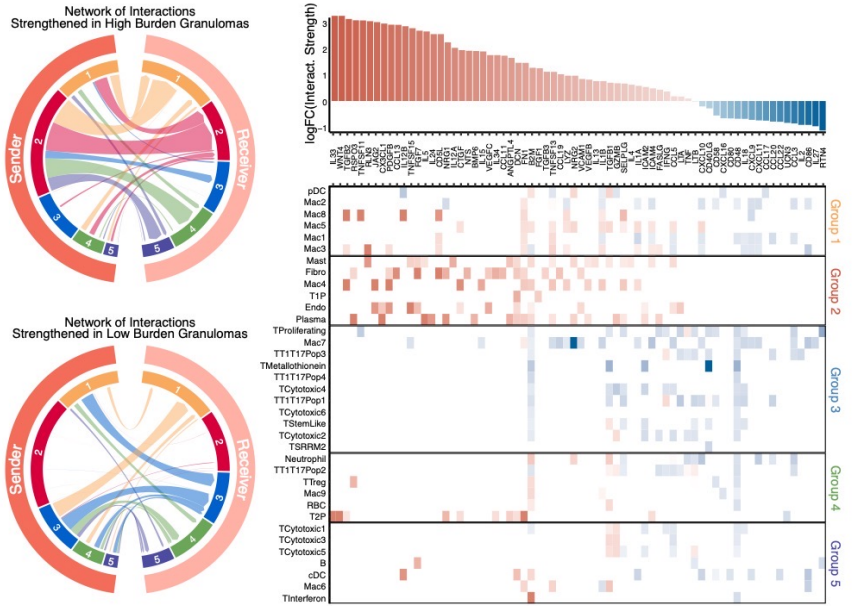


Many frameworks and tools for inferring intercellular communication; this is just one broad class

Armingol et al., *Nature Reviews Genetics* (2021)

From Passive Descriptions To Actionable Mechanisms: Intercellular Communication Networks

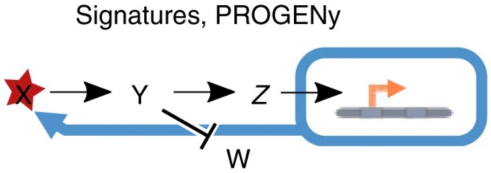




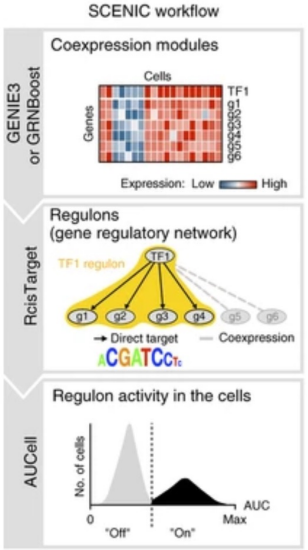
Inferring intercellular signaling → disease-associated interaction networks → prioritized, testable ligands

From Passive Descriptions To Actionable Mechanisms: Gene Regulatory Networks

Simple: Build dictionaries of stereotypical pathway/TF signatures

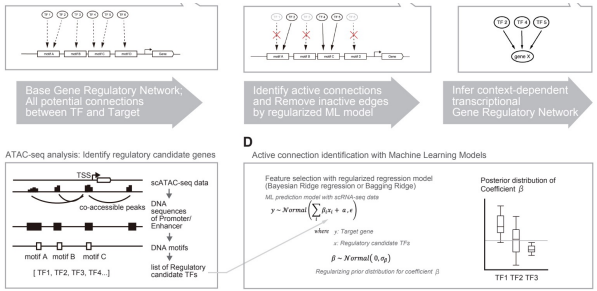


Medium: Infer networks via TF-vs-downstream gene correlations



Complex: Infer networks from linked single-cell ATAC & RNA data

CellOracle

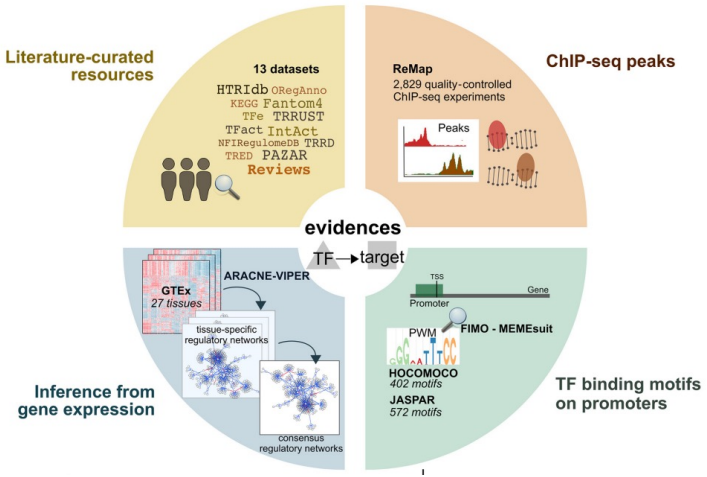


Many mathematical frameworks for inferring gene regulatory networks…

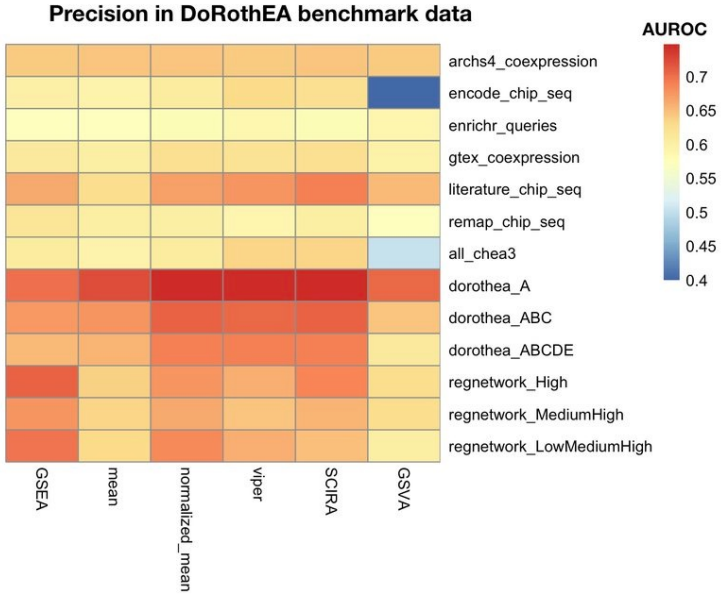
Schubert et al, *Nature Communications* (2018); Aibar et al, *Nature Methods* (2017); Kanimoto et al, *bioRxiv* (2020)

From Passive Descriptions To Actionable Mechanisms: Gene Regulatory Networks

Many options for training data

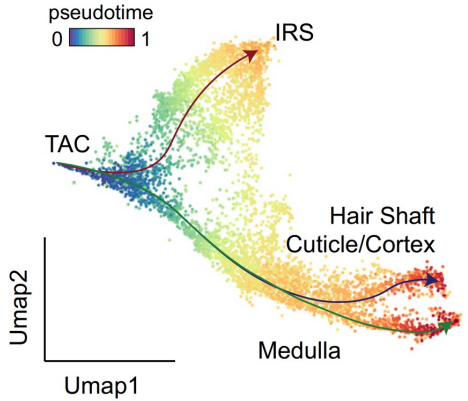


Math approach vs. training data

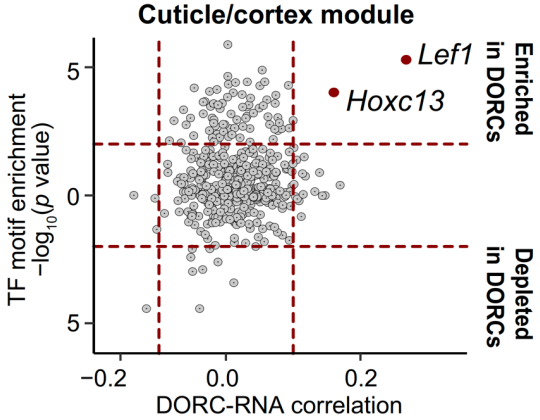


From Passive Descriptions To Actionable Mechanisms: Gene Regulatory Networks

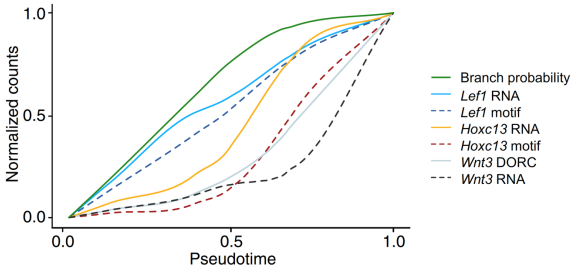
Start: Stem cell differentiation trajectory



Middle: Inferring TFs that regulate cell type determination genes



End: Sequential activation/priming of lineage-determining TFs



Inference of sets of transcription factors driving cell fate determination and lineage priming

Ma et al, *Cell* (2022)

Looking Beyond The Streetlamp: What Might We Be Missing?

So, we get actionable leads from:

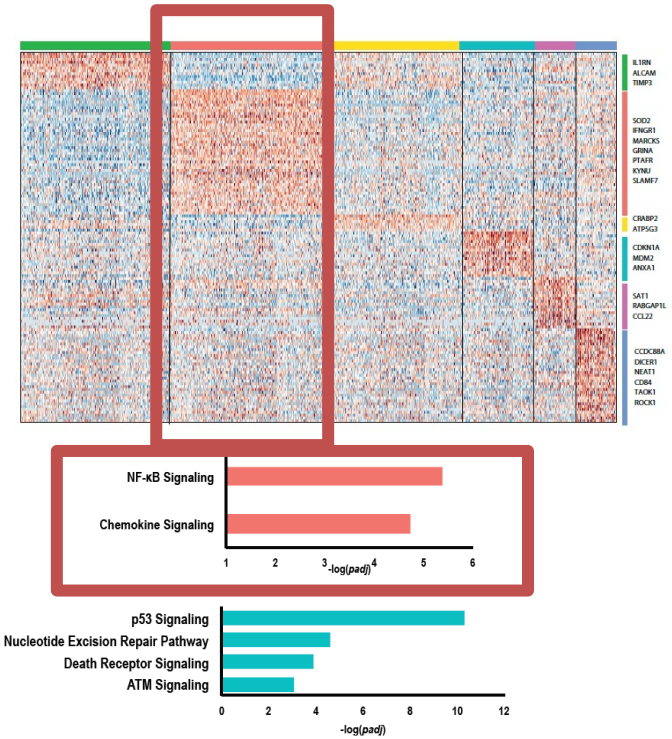
- Intercellular communication networks
- Gene regulatory networks
- And more: Drug repurposing, cross-disease convergence, etc.

However, we need experimental and computational frameworks to capture biological features not routinely measured in current assays

- Non-genetically-encoded signaling: Metabolites, DAMPs, neuronal innervation, etc.
- Long-range interactions: Immune infiltration, cross-organ links, microbiome, etc.
- Tissue-scale microenvironments: Mechanobiology, signaling gradients, cellular neighborhood structures, etc.

Long-term goal: Actionable handles that enable us to tune the knob of immune control

The Other Elephant In The Room



MSigDB
Molecular Signatures Database

Molecular Signatures Database v6.1

Overview

The Molecular Signatures Database (MSigDB) is a collection of annotated gene sets for use with GSEA software. From this web site, you can

- **Search** for gene sets by keyword.
- **Browse** gene sets by name or collection.
- **Examine** a gene set and its annotations. See, for example, the GO_NOTCH_SIGNALING_PATHWAY gene set page.
- **Download** gene sets.
- **Investigate** gene sets:
  - **Compute overlaps** between your gene set and gene sets in MSigDB.
  - **Categorize** members of a gene set by gene families.
  - **View the expression profile** of a gene set in a provided public expression compendia.

License Terms

GSEA and MSigDB are available for use under these license terms.

Please register to download the GSEA software, access our web tools, and view the MSigDB gene sets. After registering, you can log in at any time using your email address. Registration is free. Its only purpose is to help us track usage for reports to our funding agencies.

Current Version

MSigDB database v6.1 updated October 2017. Release notes.
GSEA/MSigDB web site v6.3 released January 2018

Collections

The MSigDB gene sets are divided into 8 major collections:

H **hallmark gene sets** are coherently expressed signatures derived by aggregating many MSigDB gene sets to represent well-defined biological states or processes.

C1 **positional gene sets** for each human chromosome and cytogenetic band.

C2 **curated gene sets** from online pathway databases, publications in PubMed, and knowledge of domain experts.

C3 **motif gene sets** based on conserved cis-regulatory motifs from a comparative analysis of the human, mouse, rat, and dog genomes.

C4 **computational gene sets** defined by mining large collections of cancer-oriented microarray data.

C5 **GO gene sets** consist of genes annotated by the same GO terms.

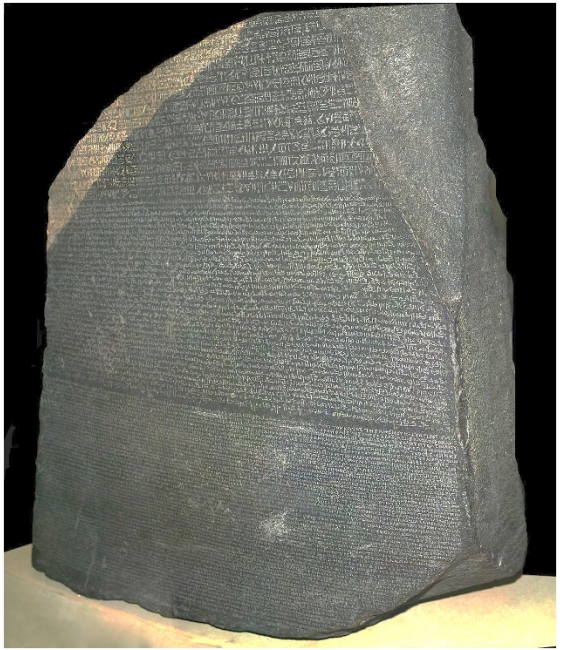
C6 **oncogenic gene sets** defined directly from microarray gene expression data from cancer gene perturbations.

C7 **immunologic gene sets** defined directly from microarray gene expression data from immunologic studies.

GOLDRATH_EFF_VS_MEMORY_CD8_TCELL_DN
GOLDRATH_EFF_VS_MEMORY_CD8_TCELL_UP
GOLDRATH_NAIVE_VS_EFF_CD8_TCELL_DN
GOLDRATH_NAIVE_VS_EFF_CD8_TCELL_UP
GOLDRATH_NAIVE_VS_MEMORY_CD8_TCELL_DN
GOLDRATH_NAIVE_VS_MEMORY_CD8_TCELL_UP
GSE10094_LCMV_VS_LISTERIA_IND_EFF_CD4_TCELL_DN
GSE10094_LCMV_VS_LISTERIA_IND_EFF_CD4_TCELL_UP
GSE10147_IL3_AND_HIVP17_VS_IL3_AND_CPG_STIM_PDC_DN
GSE10147_IL3_AND_HIVP17_VS_IL3_AND_CPG_STIM_PDC_UP
GSE10147_IL3_VS_IL3_AND_CPG_STIM_PDC_DN
GSE10147_IL3_VS_IL3_AND_CPG_STIM_PDC_UP
GSE10147_IL3_VS_IL3_AND_HIVP17_STIM_PDC_DN
GSE10147_IL3_VS_IL3_AND_HIVP17_STIM_PDC_UP
GSE10211_UV_INACT_SENDAI_VS_LIVE_SENDAI_VIRUS_TRACHEAL_EPITHELIAL_CELLS_DN
GSE10211_UV_INACT_SENDAI_VS_LIVE_SENDAI_VIRUS_TRACHEAL_EPITHELIAL_CELLS_UP
GSE10239_KLRG1INT_VS_KLRG1HIGH_EFF_CD8_TCELL_DN
GSE10239_KLRG1INT_VS_KLRG1HIGH_EFF_CD8_TCELL_UP
GSE10239_MEMORY_VS_DAY4.5_EFF_CD8_TCELL_DN
GSE10239_MEMORY_VS_DAY4.5_EFF_CD8_TCELL_UP
GSE10239_MEMORY_VS_KLRG1HIGH_EFF_CD8_TCELL_DN
GSE10239_MEMORY_VS_KLRG1HIGH_EFF_CD8_TCELL_UP
GSE10239_MEMORY_VS_KLRG1INT_EFF_CD8_TCELL_DN
GSE10239_MEMORY_VS_KLRG1INT_EFF_CD8
GSE21546_SAP1A_KO_VS_SAP1A_KO_AND_ELK1_KO_DP_THYMOCYTES_UP
GSE21546_UNSTIM_VS_ANTI_CD3_STIM_DP_THYMOCYTES_DN
GSE21546_UNSTIM_VS_ANTI_CD3_STIM_DP_THYMOCYTES_UP
GSE21546_UNSTIM_VS_ANTI_CD3_STIM_ELK1_KO_DP_THYMOCYTES_DN
GSE21546_UNSTIM_VS_ANTI_CD3_STIM_ELK1_KO_DP_THYMOCYTES_UP
GSE21546_UNSTIM_VS_ANTI_CD3_STIM_SAP1A_KO_AND_ELK1_KO_DP_THYMOCYTES_DN
GSE21546_UNSTIM_VS_ANTI_CD3_STIM_SAP1A_KO_AND_ELK1_KO_DP_THYMOCYTES_UP
GSE21546_UNSTIM_VS_ANTI_CD3_STIM_SAP1A_KO_DP_THYMOCYTES_DN
GSE21546_UNSTIM_VS_ANTI_CD3_STIM_SAP1A_KO_DP_THYMOCYTES_UP
GSE21546_WT_VS_ELK1_KO_ANTI_CD3_STIM_DP_THYMOCYTES_DN
GSE21546_WT_VS_ELK1_KO_ANTI_CD3_STIM_DP_THYMOCYTES_UP
GSE21546_WT_VS_ELK1_KO_DP_THYMOCYTES_DN
GSE21546_WT_VS_ELK1_KO_DP_THYMOCYTES_UP
GSE21546_WT_VS_SAP1A_KO_AND_ELK1_KO_ANTI_CD3_STIM_DP_THYMOCYTES_DN
GSE21546_WT_VS_SAP1A_KO_AND_ELK1_KO_ANTI_CD3_STIM_DP_THYMOCYTES_UP
GSE21546_WT_VS_SAP1A_KO_AND_ELK1_KO_DP_THYMOCYTES_DN
GSE21546_WT_VS_SAP1A_KO_AND_ELK1_KO_DP_THYMOCYTES_UP
GSE21546_WT_VS_SAP1A_KO_ANTI_CD3_STIM_DP_THYMOCYTES_DN
GSE21546_WT_VS_SAP1A_KO_ANTI_CD3_STIM_DP_THYMOCYTES_UP
GSE21546_WT_VS_SAP1A_KO_DP_THYMOCYTES_DN
GSE21546_WT_VS_SAP1A_KO_DP_THYMOCYTES_UP
GSE21670_IL6_VS_TGFB_AND_IL6_TREATED
GSE36476_YOUNG_VS_OLD_DONOR_MEMORY_CD4_TCELL_72H_TSST_ACT_UP
GSE36476_YOUNG_VS_OLD_DONOR_MEMORY_CD4_TCELL_DN
GSE36476_YOUNG_VS_OLD_DONOR_MEMORY_CD4_TCELL_UP
GSE36527_CD62L_HIGH_CD69_NEG_VS_CD62L_LOW_CD69_POS_TREG_KLRG1_NEG_DN
GSE36527_CD62L_HIGH_CD69_NEG_VS_CD62L_LOW_CD69_POS_TREG_KLRG1_NEG_UP
GSE36527_CD62L_HIGH_VS_CD62L_LOW_TREG_CD69_NEG_KLRG1_NEG_DN
GSE36527_CD62L_HIGH_VS_CD62L_LOW_TREG_CD69_NEG_KLRG1_NEG_UP
GSE36527_CD69_NEG_VS_POS_TREG_CD62L_LOS_KLRG1_NEG_DN
GSE36527_CD69_NEG_VS_POS_TREG_CD62L_LOS_KLRG1_NEG_UP
GSE36826_NORMAL_VS_STAPH_AUREUS_INF_IL1R_KO_SKIN_DN
GSE36826_NORMAL_VS_STAPH_AUREUS_INF_IL1R_KO_SKIN_UP
GSE36826_NORMAL_VS_STAPH_AUREUS_INF_SKIN_DN
GSE36826_NORMAL_VS_STAPH_AUREUS_INF_SKIN_UP
GSE36826_WT_VS_IL1R_KO_SKIN_DN
GSE36826_WT_VS_IL1R_KO_SKIN_STAPH_AUREUS_INF_UP
GSE36826_WT_VS_IL1R_KO_SKIN_UP
GSE36826_WT_VS_IL1R_KO_SKIN_STAPH_AUREUS_INF_DN
GSE36888_STAT5_AB_KNOCKIN_VS_WT_TCELL_IL2_TREATED_17H_DN
GSE36888_STAT5_AB_KNOCKIN_VS_WT_TCELL_IL2_TREATED_17H_UP
GSE36888_STAT5_AB_KNOCKIN_VS_WT_TCELL_IL2_TREATED_2H_DN
GSE36888_STAT5_AB_KNOCKIN_VS_WT_TCELL_IL2_TREATED_2H_UP
GSE36888_STAT5_AB_KNOCKIN_VS_WT_TCELL_IL2_TREATED_6H_DN
GSE36888_STAT5_AB_KNOCKIN_VS_WT_TCELL_IL2_TREATED_6H_UP
GSE36888_UNTREATED_VS_IL2_TREATED_5

SHALEKLAB
from cells to systems

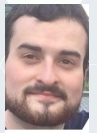
Next Steps: Scaling To Enabling Rapid Translation



Large queryable libraries and signature sets

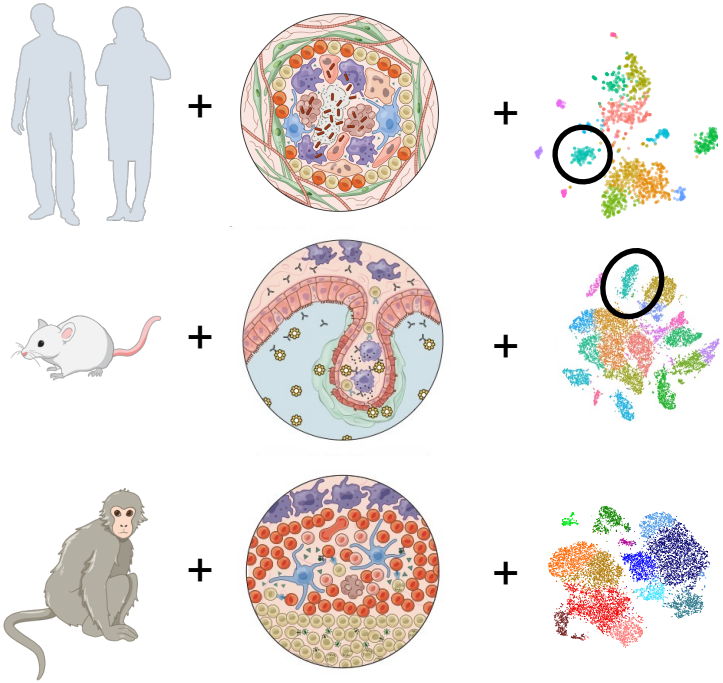
Left: www.businessinsider.com; Right: www.wikipedia.com

A Queryable Multi-Species Library Empowers New Analyses

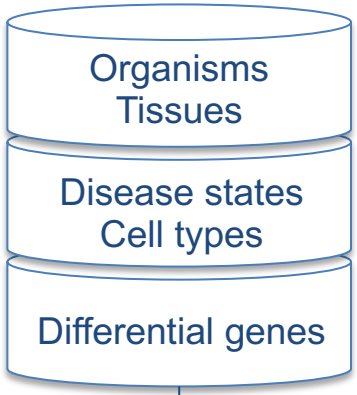



James Gatter Sarah Nyquist

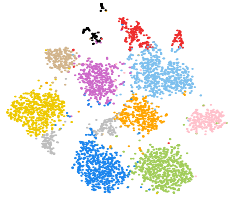
Previously analyzed datasets



Alexandria Data Resource



New datasets



Query for:
Analogous cell types
Similar differential genes

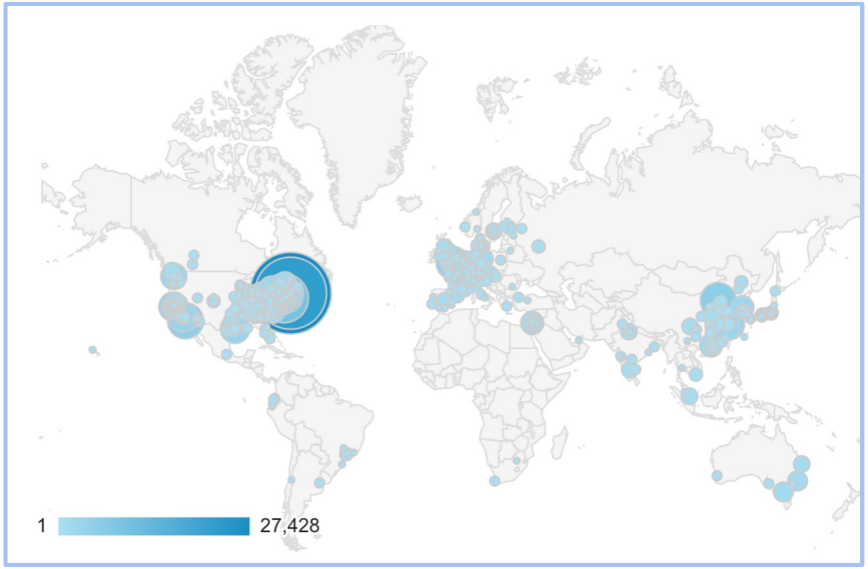
Meta-analyses of existing datasets

Easily access data for:
- Same tissue in different organisms or diseases
- Cell types with similar genes

SHALEKLAB
from cells to systems

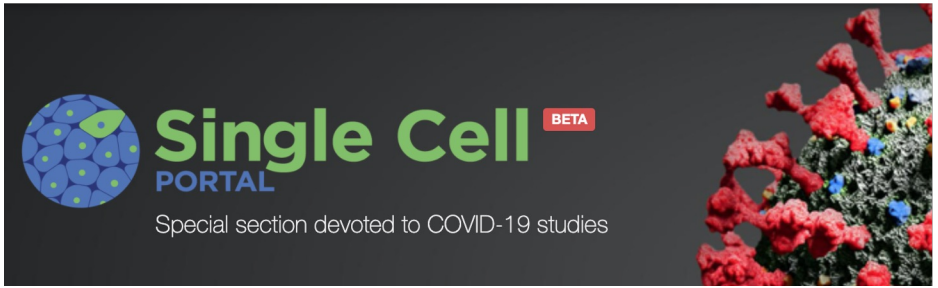
Ziegler et al, *Cell* (2020)

Alexandria & The SCP Ecosystem



BICCN Anatomy and Morphology Project

Featuring
63 studies
1,141,824 cells

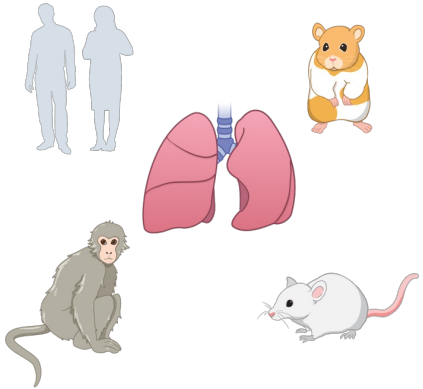


Single Cell BETA
PORTAL
Reducing barriers and accelerating single-cell research

Featuring
944 studies
25,342,321 cells

SHALEKLAB
from cells to systems

Making “Rosetta Stones”



In vivo systems

DATA!!!

scRNA-Seq

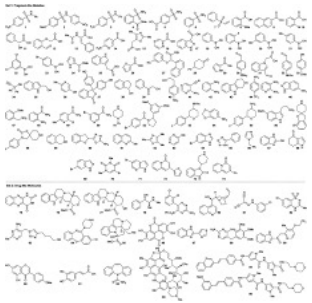
RNA-Seq

ATAC-Seq

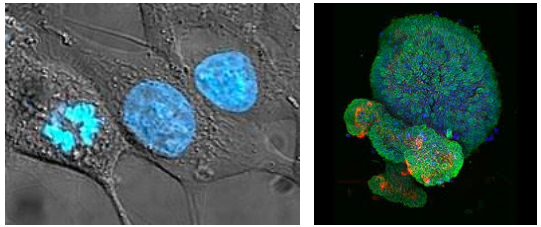
Imaging

Your Favorite Assay Here

…

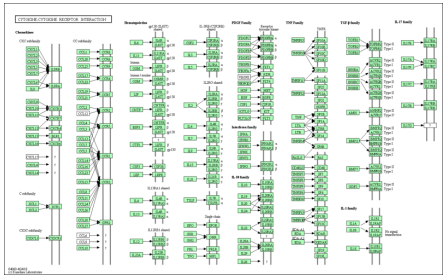


Compound Libraries



In vitro systems

Discovery!



Cytokines, Chemokines

SHALEKLAB
from cells to systems

Applying single-cell genomics to understand the pathogenesis of the yeast *Malassezia* in skin keratinocytes

Prof. Alex Shalek Ph.D-MIT
Prof. Adriana Marcela Celis Ramírez Ph.D-Universidad de los Andes

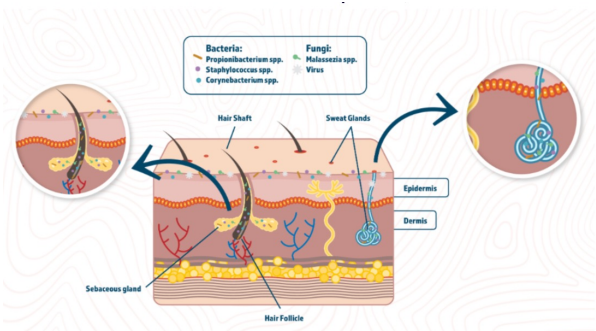
Presenting: Sergio Triana Ph.D-Postdoc MIT

Malassezia

Malassezia is a lipophilic unicellular fungus

- Human and animal microbiota member (is a major component of the skin microbiome)
- Switch to opportunistic pathogens

21 species (*M. globosa, M. furfur, M. sympodialis* and *M. restricta* are the most frequent)



Dornell. J, 2021. Technology Networks. (2024, February 27). The Skin Microbiome. https://www.technologynetworks.com/immunology/articles/the-skin-microbiome-356627[1]

J.Fungi 2022, 8, 708. https://doi.org/10.3390/jof8070708

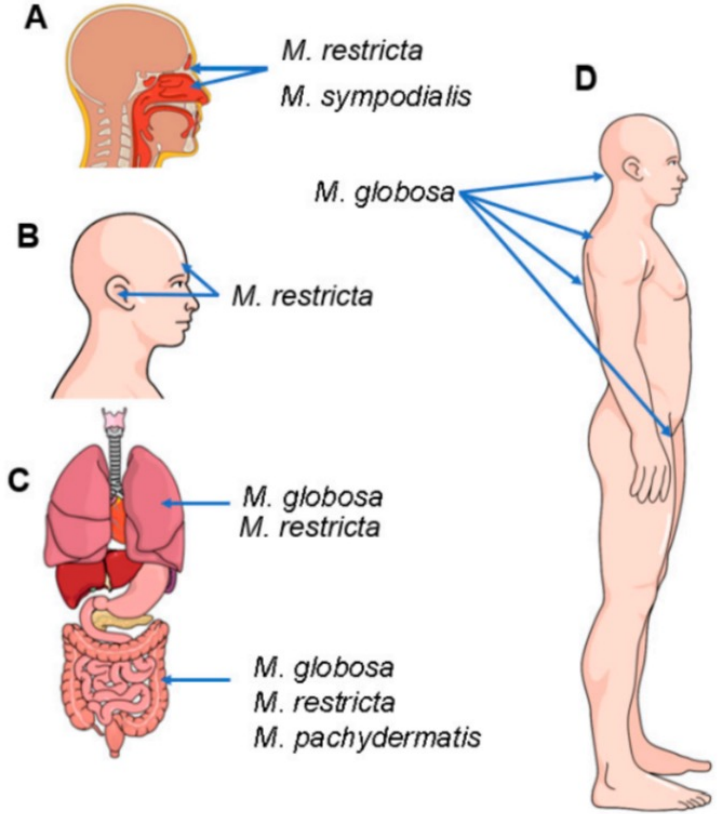


Figure 1. Predominant *Malassezia* species found as commensals in the sinonasal cavity (A), on the skin of the forehead and external ear canal (B), in the lungs and gastrointestinal tract (C) and on the skin of the occiput, back and groin (D).

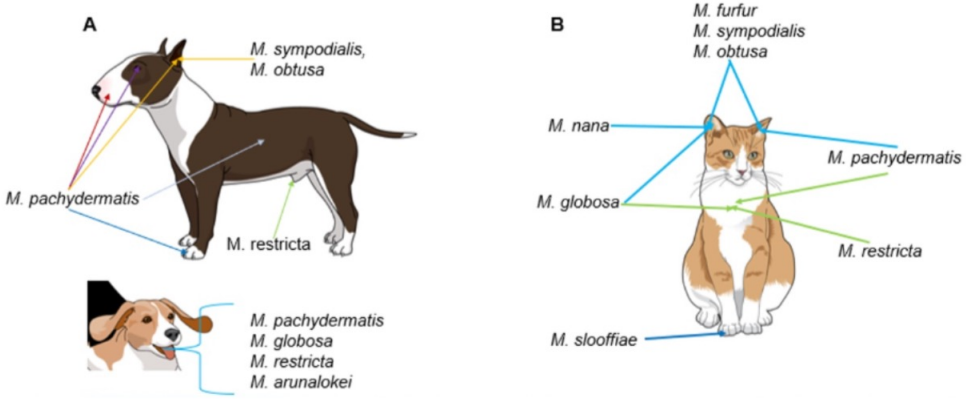
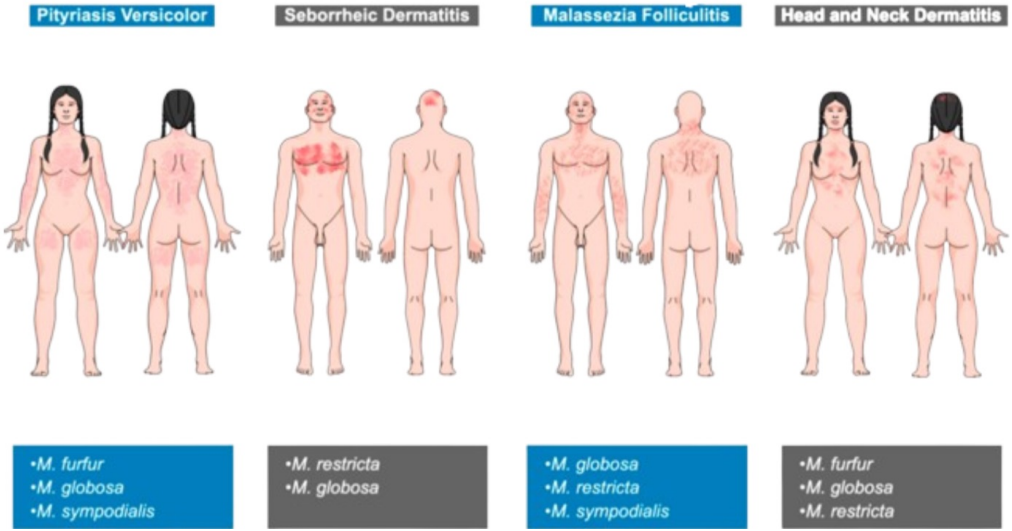


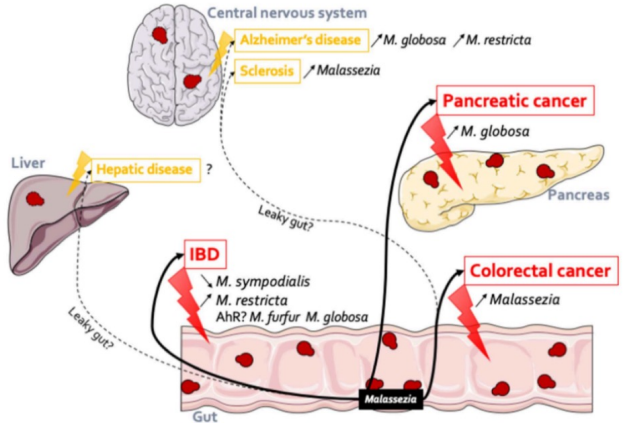
Figure 2. *Malassezia* species colonization in healthy dogs (species colonization in healthy dogs and cats. A) and cats (B)

Malassezia-related diseases

- Superficial
- Disseminated and proinflammatory

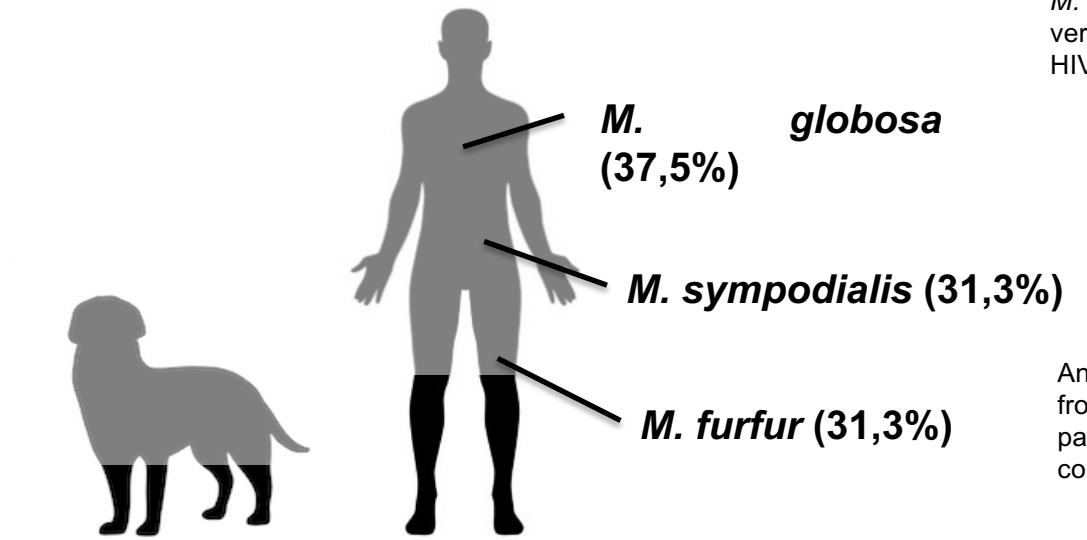


J. Fungi 2022, 8, 708. https://doi.org/10.3390/jof8070708



Front. Cell. Infect. Microbiol. Vol 10, No. 201, doi: https://doi.org/10.3389/fcimb.2020.00201

Malassezia in Colombia



Biomedica, 25(2), 189-195.

M. globosa predominates in patients with pityriasis versicolor (67%) and with seborrheic dermatitis in HIV-patients (85%).

Biomedica, 25(2), 189-195.

Antifungal sensitivity assays demonstrate that isolates from both HIV-positive and HIV-negative Colombian patients are resistant or have high minimal inhibitory concentrations to amphotericin B.

Contreras Carreño, A. G. (2021). In vitro susceptibility of *Malassezia furfur* to antifungal compounds from HIV-positive and negative patients.

32.1 % of the dogs had otitis due to *M. pachydermatis*.

Rev Inv Vet Perú 2021; 32(2): e20020
http://dx.doi.org/10.15381/rivep.v32i2.20020

Relevant data

+8k millions

2013 Direct US health care costs of skin disease

+14k million

is expected that dandruff market to reach by 2030

Global Dandruff Treatment Market – Industry Trends and Forecast to 2030

+3k million

are generated by skin care products

2017 by the American Academy of Dermatology, Inc.http://dx.doi.org/10.1016/j.jaad.2016.12.043

Survey by Data Bridge Market Research (https://www.databridgemarketresearch.com/reports/global-dandruff-treatment-market)

Objetives

Networking

1. To start a network across MIT and Universidad de Los Andes, a collaboration started by Sergio Triana, a Colombian postdoc in the Shalek lab, and Alumni from the CRG group.

2. Make use of the two research groups' vast knowledge to train researchers and students in state of the art technologies for the analysis of infectious diseases.

Funded by MISTI award

Research

3. To study host-*Malassezia* interaction in developing superficial diseases.

4. To gain a more comprehensive view of yeast and keratinocyte cells' interaction and the cause of skin diseases.

Currently looking for seed funding

Research groups

CeMoP
Grupo de Investigación Celular y Molecular de Microorganismos Patógenos

The CeMoP research group (CRG), Universidad de los Andes, Colombia, studies pathogenic fungi with afocus on the skin-pathogenic yeast *Malassezia*. Specifically, the group aims to study its virulence and therapeutic targetability using *in vitro* and *in vivo* models.

gen · core
Centro de Secuenciación

The Sequencing Center is a laboratory that provides high quality services in Next Generation Sequencing, Capillary Electrophoresis, and Molecular Biology techniques.

SHALEK LAB
from cells to systems

The Shalek lab at MIT has pioneered the use of single-cell genomics to study host-pathogen interactions, cancer progression, and immunology. Additionally, they have codeveloped Seq-Well, a portable, low-cost platform for massively parallel scRNA-seq.

Universidad de los Andes
Colombia

Massachusetts Institute of Technology

In summary

Thanks to this MISTI award we will

- Establish a scientific network between MIT and Universidad de Los Andes
- Perform a workshop in Colombia to train local scientist on the field of single-cell biology
- Set up a collaboration for studying host-pathogen interaction in *Malassezia* an generate preliminary data to secure funding from national and international entities.

Future research depending on funding :

- Single-cell RNA sequencing of Keratinocytes infected with different *Malassezia species and knockouts* to identify the role in pathogenesis of lipid synthesis yeast genes.
- Apply of scRNA-seq technology on skin samples from healthy and diseased samples from patients in Colombia.

Plans for Bangkok With You All!

Overview:

- 2 cohorts – computational and wet lab
- 5 full days of training

Experimental Program:

- Introduction to experimental design
- Tissue dissociation, cell isolation
- Cell loading
- 10x Genomics, Parse Bio
- Library preparation RNA and Hashing
- Introduction to data analysis

Thanks for your partnership!

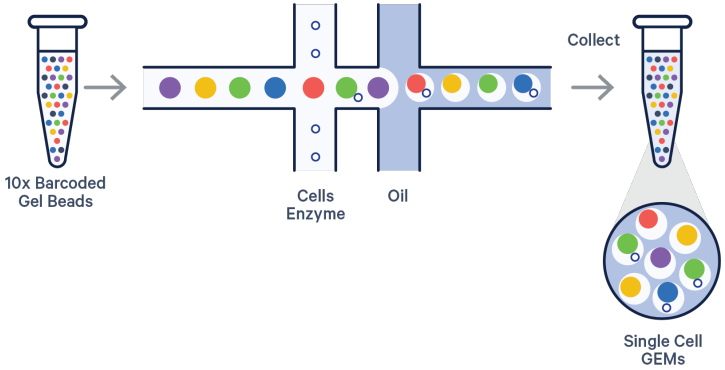
Computational Program:

- Introduction to single-cell transcriptomics and experimental sequencing methods
- Introduction to computational analysis methods (R studio prep, introduction to R, R lab cell subsets; local & cloud)
- Data pre-processing & technical QC
- Data exploration & biological QC
- Identification and biological interpretation of cell types/states/NMF factors
- Data integration and batch correction
- Upstream regulator inference
- Trajectory Inference
- Sc-ATAC-seq
- Immune repertoire scVDJ and CITE-Seq
- Cell-cell communication

Wetlab track Overview

Droplet Based

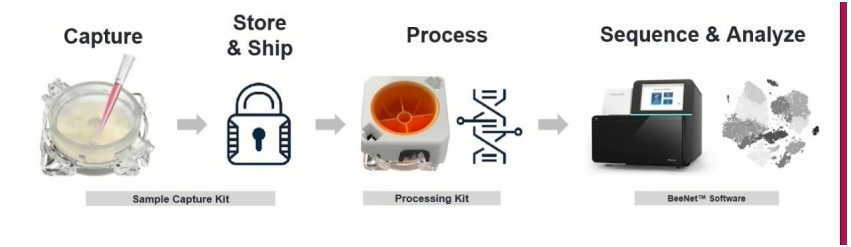
Chromium Single Cell FLEX (10x Genomics)



10x GENOMICS®

Nanowel based

Honeycomb Biotechnologies HIVEs CLX

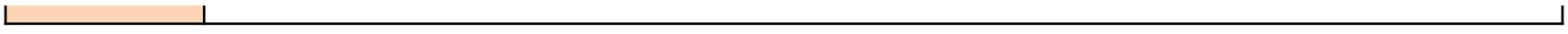


Parse BIOSCIENCES

Wetlab track schedule

Friday	**Data analysis**	
	Introduction to data analysis and biological interpretation	Computational track project presentations and closing remarks

Computational track schedule



Aims for Computational track projects

- You will get a subset dataset from the paper and some prompts to guide your analysis
- Discuss with the training team
- Brainstorm
- Try new things, explore the data and learn!
- Make a 5-10 min presentation

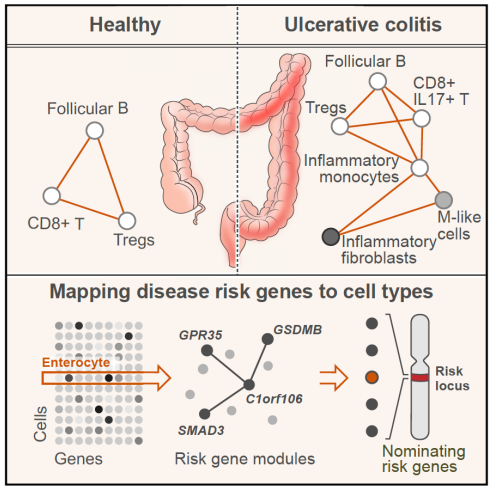
Project 1.

Resource

Cell

Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis

Graphical Abstract



Authors

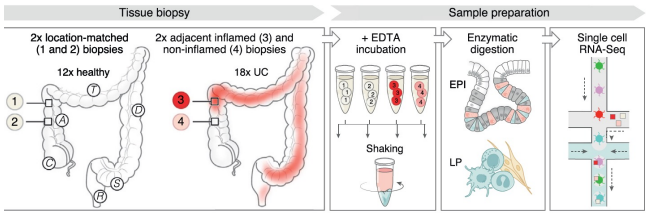
Christopher S. Smillie, Moshe Biton, Jose Ordovas-Montanes, ..., Alex K. Shalek, Ramnik J. Xavier, Aviv Regev

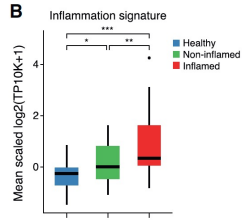
Correspondence

aananthakrishnan@mgh.harvard.edu (A.N.A.), shalek@mit.edu (A.K.S.), xavier@molbio.mgh.harvard.edu (R.J.X.), aregev@broadinstitute.org (A.R.)

In Brief

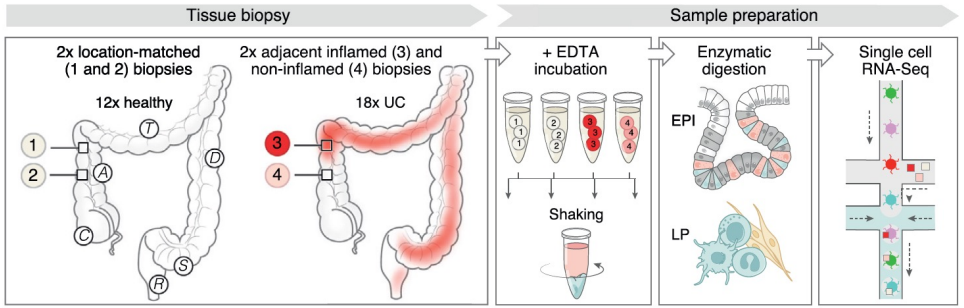
Single-cell analyses of colon biopsy specimens from patients with ulcerative colitis delineate how expression patterns and shifting cell populations may shape disease and drug resistance, and provide a framework for linking GWAS risk loci with specific cell types and functional pathways.

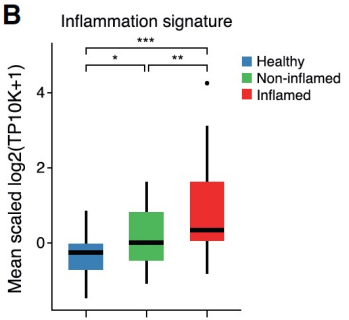




Project 1. GUT

Single-cell analyses of colon biopsy specimens from patients with ulcerative colitis delineate how expression patterns and shifting cell populations may shape disease





What do you get:
4500 *cells*
30 Donors
3 Types of samples (Healthy and Ulcerative colitis (Non-inflamed and Inflamed))

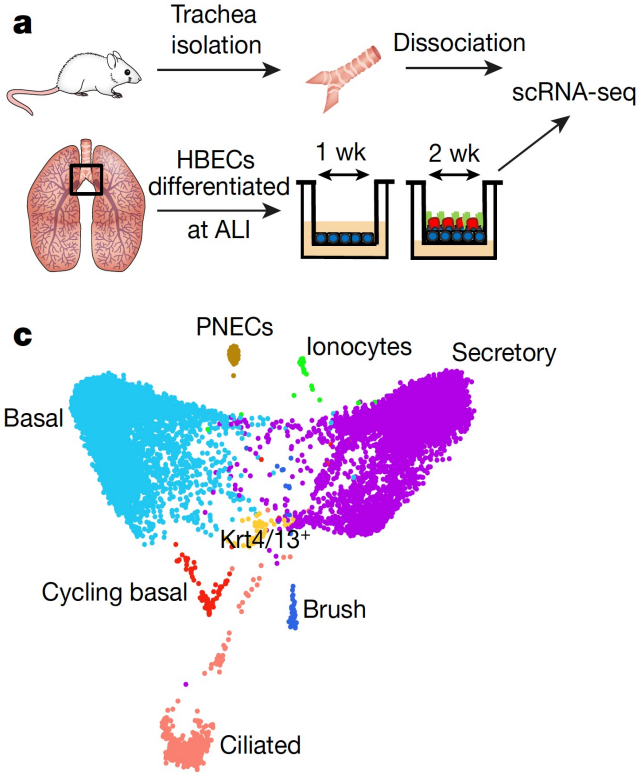
Questions to Answer.
What is the quality of the cell? Do you need to filter cells?
Is there any batch effect?
What are the main cell types present in the sample?
Do you see any differential expression between the 3 types of samples?

LETTER

https://doi.org/10.1038/s41586-018-0394-

A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte

Lindsey W. Plasschaert[1,5,7], Rapolas Žilionis[2,3,7], Rayman Choo-Wing[1,5], Virginia Savova[2,6], Judith Knehr[4], Guglielmo Roma[4], Allon M. Klein[2]* & Aron B. Jaffe[1,5]*

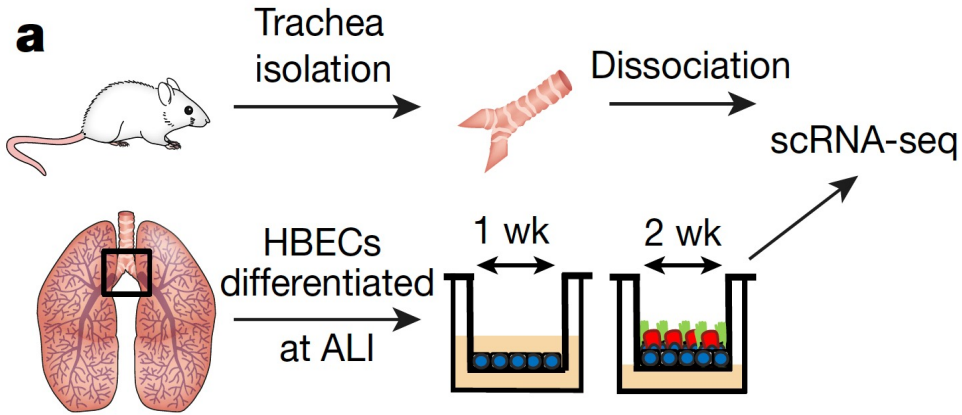


Project 2.

Single-cell analyses reveals cell states that represent known and novel cell populations of the lung, They identified a novel, rare cell type that they call the 'pulmonary ionocyte', which co-expresses FOXI1, H+-ATPase (V-ATPase) and CFTR, the gene that is mutated in cystic fibrosis.

What do you get:
From Human
2970 cells
3 Donors

From Mouse
2912 cells
10 Mouse
5 Time points!



Questions to Answer.
What is the quality of the cell? Do you need
Is there any batch effect?
What are the main cell types present in the
Do you see any differences between Mouse and Human data?

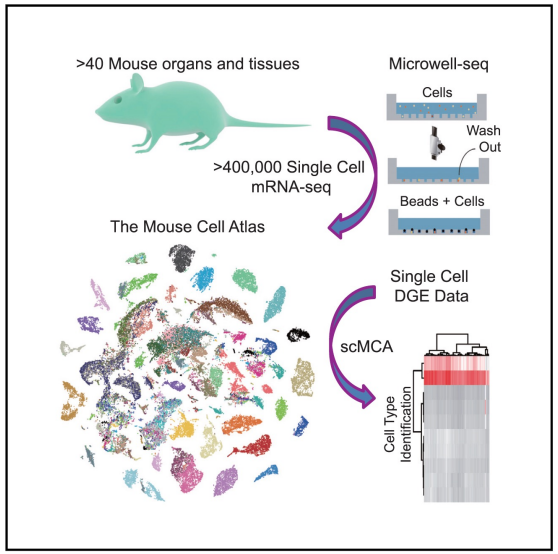
Project 3 MOUSE ATLAS

Resource

Cell

Mapping the Mouse Cell Atlas by Microwell-Seq

Graphical Abstract



Authors

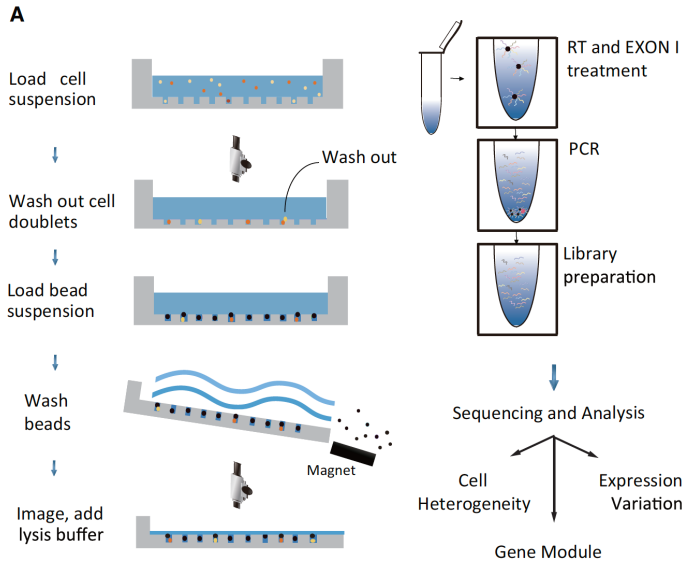
Xiaoping Han, Renying Wang, Yincong Zhou, ..., Guo-Cheng Yuan, Ming Chen, Guoji Guo

Correspondence

xhan@zju.edu.cn (X.H.),
ggj@zju.edu.cn (G.G.)

In Brief

Development of Microwell-seq allows construction of a mouse cell atlas at the single-cell level with a high-throughput and low-cost platform.

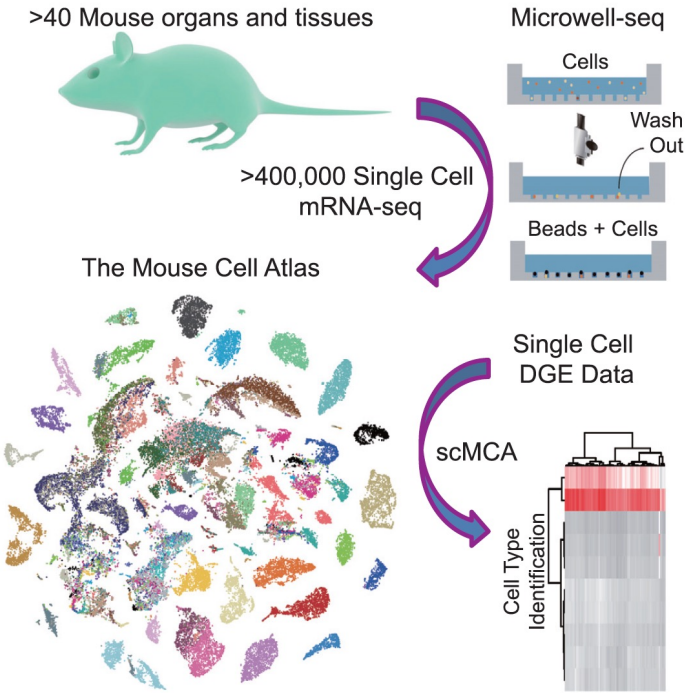


Project 3.

analyzed more than 400,000 single cells covering all of the major mouseorgans and constructed a basic scheme for a mouse cell atlas (MCA). We reveal a single-cell hierarchy for many tissues that have not been well characterized previously

What do you get:
4310 *cells*
6 tissues (5 immune and 1 epithelial)
18 Mouse

Questions to Answer.
What is the quality of the cells? Do you need to filter cells?
Is there any batch effect?
What are the main cell types present in the sample?
Can you detect Immune cells in multiple tissues?
Do you see any differences between the
same immune cell type between tissues?



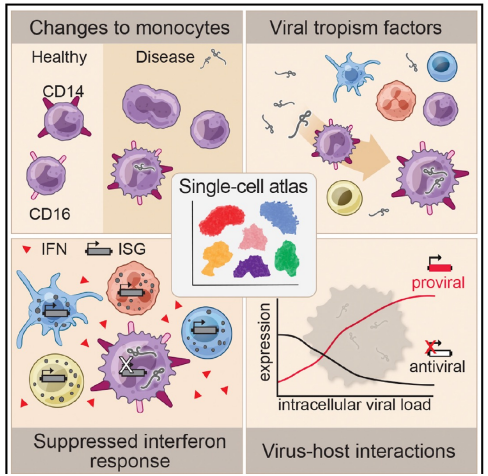
Project 4 EBOLA

Cell

Resource

Single-Cell Profiling of Ebola Virus Disease *In Vivo* Reveals Viral and Host Dynamics

Graphical Abstract



Authors

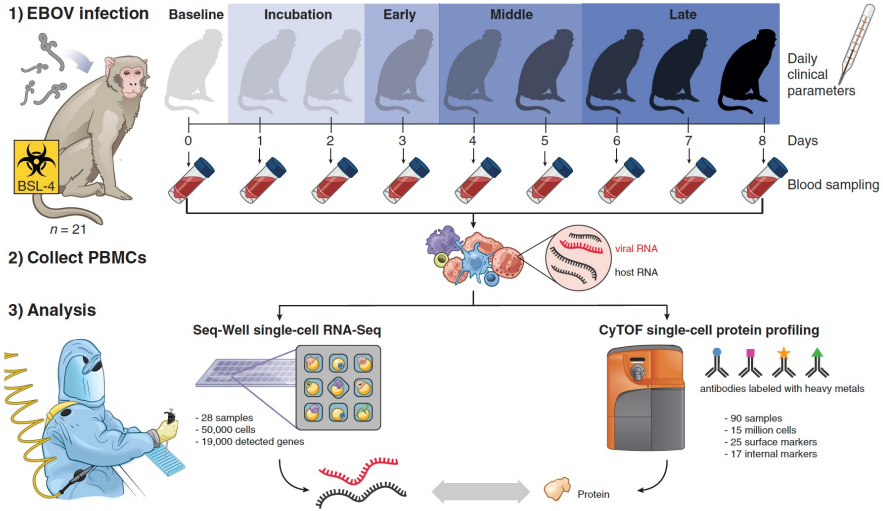
Dylan Kotliar, Aaron E. Lin, James Logue, ..., Alex K. Shalek, Pardis C. Sabeti, Richard S. Bennett

Correspondence

dylan_kotliar@hms.harvard.edu (D.K.), alin@broadinstitute.org (A.E.L.), lisa.hensley@nih.gov (L.E.H.), mcilwain@stanford.edu (D.R.M.)

In Brief

Single-cell profiling of circulating immune cells during Ebola virus (EBOV) infection in non-human primates resolves molecular correlates of viral tropism, characterizes replication dynamics within infected cells, and distinguishes expression changes that are mediated by viral infection from those due to cytokine signaling.

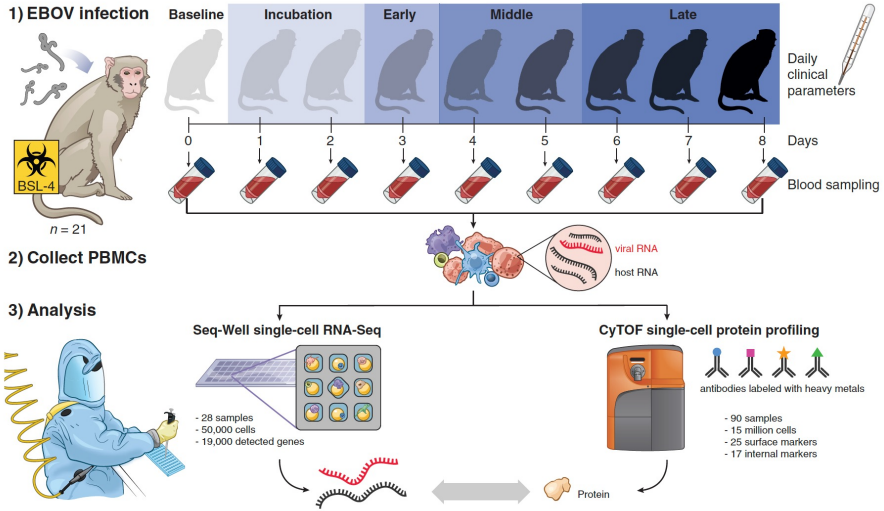


Project 4.

Single-cell profiling of circulating immune cells during Ebola virus (EBOV) infection in non-human primates resolves molecular correlates of viral tropism, characterizes replication dynamics within infected cells, and distinguishes expression changes that are mediated by viral infection from those due to cytokine signaling.

What do you get:
3607 *cells*
13 Monkeys
5 Timepoints (DPI)

Questions to Answer.
What is the quality of the cells? Do you need to filter cells?
Is there any batch effect between monkeys?
What are the main cell types present in the sample?
Where can we detect ebola transcripts?
Do you see any differences between timepoints?



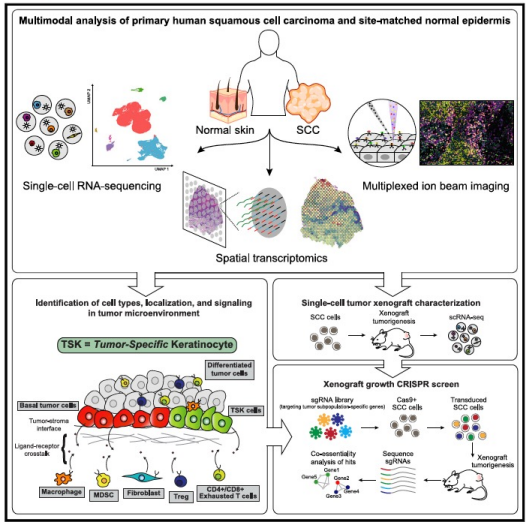
Project 5 CANCER

Cell

Resource

Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma

Graphical Abstract



Authors

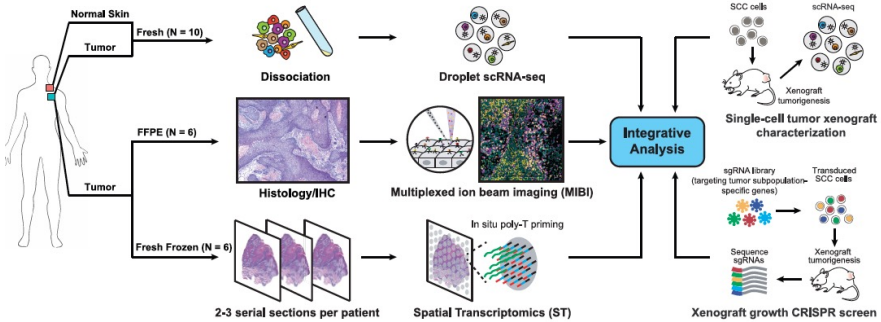
Andrew L. Ji, Adam J. Rubin, Kim Thrane, ..., Garry P. Nolan, Joakim Lundeberg, Paul A. Khavari

Correspondence

khavari@stanford.edu

In Brief

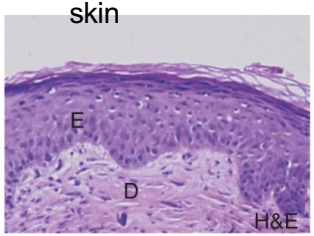
Integration of high-dimensional multi-omics approaches to characterize human cutaneous squamous cell carcinoma identifies a tumor-specific keratinocyte population as well as the immune infiltrates and heterogeneity at tumor leading edges.



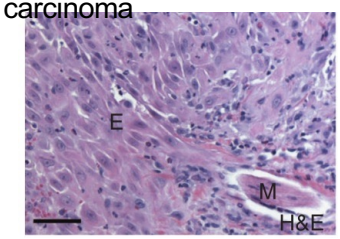
Project 5: comparison of T cells in healthy skin and skim tumors

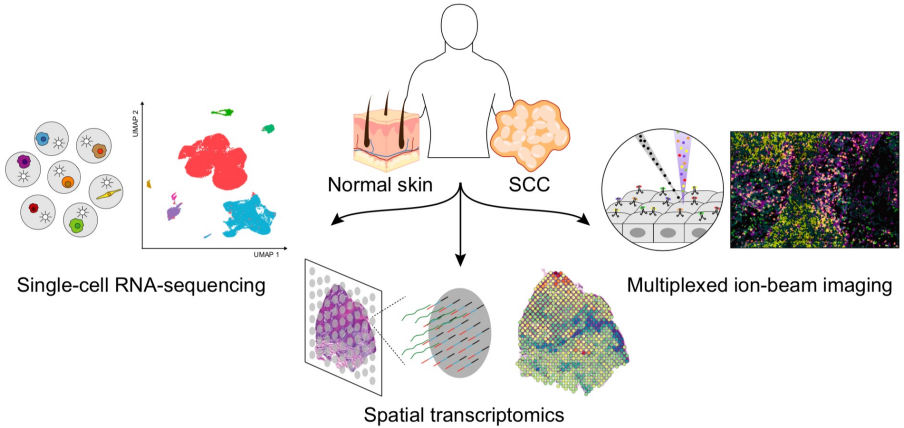
To define the cellular composition and architecture of cutaneous squamous cell carcinoma (cSCC), we combined single-cell RNA sequencing with spatial transcriptomics and multiplexed ion beam imaging from a series of human cSCCs and matched normal skin. These data define cSCC tumor and stromal cell subpopulations, the spatial niches where they interact, and the communicating gene networks that they engage in cancer.

Normal skin



Cutaneous squamous cell carcinoma





What do you get
1889 cells
11 donors
Tumor and normal samples

Questions
Do T cells in normal tissue look different than T cells in tumors?
Are tumor T cells similar across patients?
What are the types of T cells that we observe in tumors?
Do we see many exhausted or proliferating cells?

Ji et al., 2020

Computational track - to keep in mind

- We will be using google collab and Github **https://github.com/ShalekLab/MISTI_Bogota_2024_workshop**

- Please sign up to your prefer project by noon today!

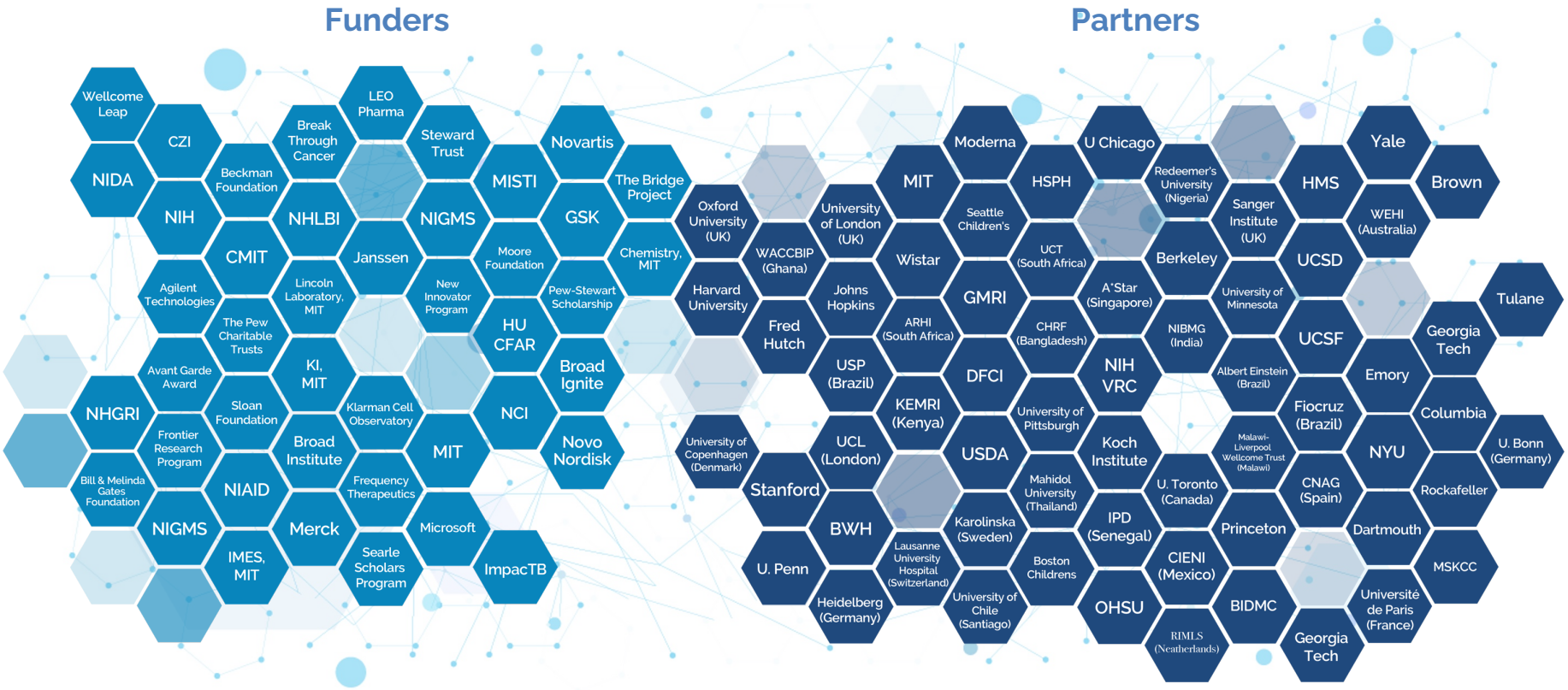
Please sign up to your project of interest!
Introduction to single-cell groups

Project 1 GUT	Group 1					
Project 2 LUNG	Group 2					
	Group 3					
Project 3 MOUSE	Group 4					
	Group 5					
Project 4 EBOLA	Group 6					
Project 5 CANCER	Group 7					

QR sign up sheet

Thank You!!!

Collaborators



SHALEKLAB
from cells to systems

Questions?