

Single Cell Transcriptomic Analysis

Acknowledgments:

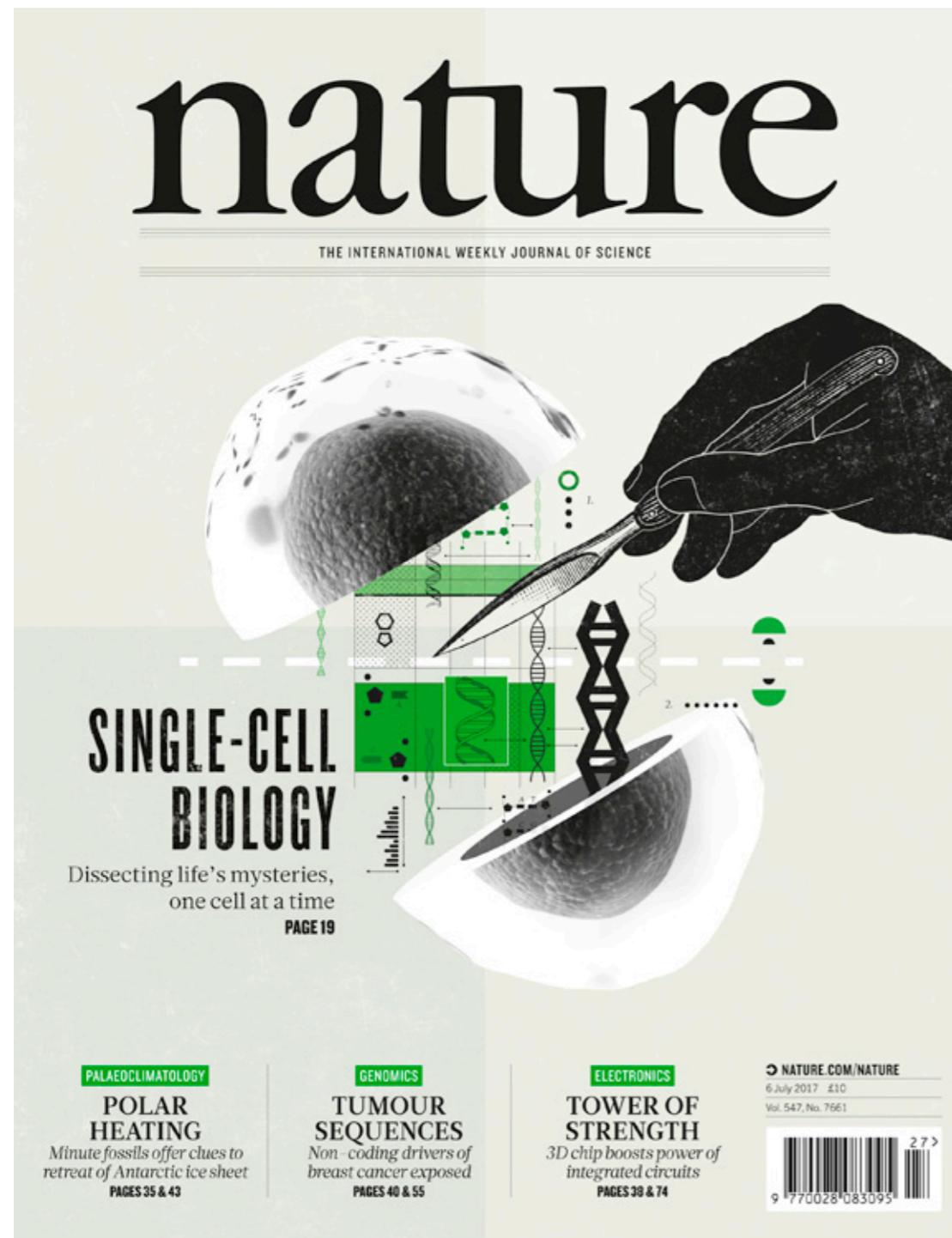
Brian Haas

Karthik Shekhar

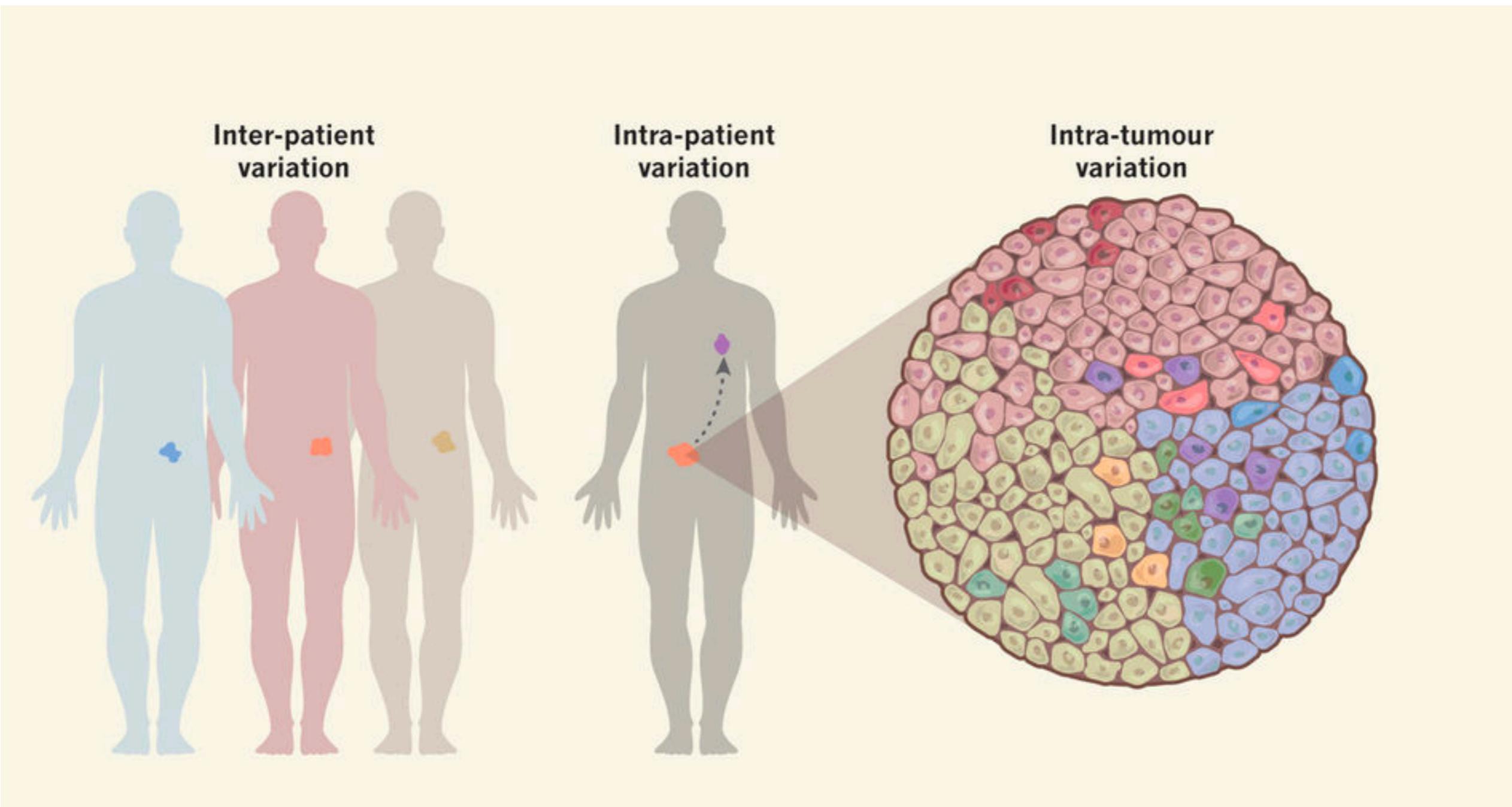
Timothy Tickle

Aviv Regev

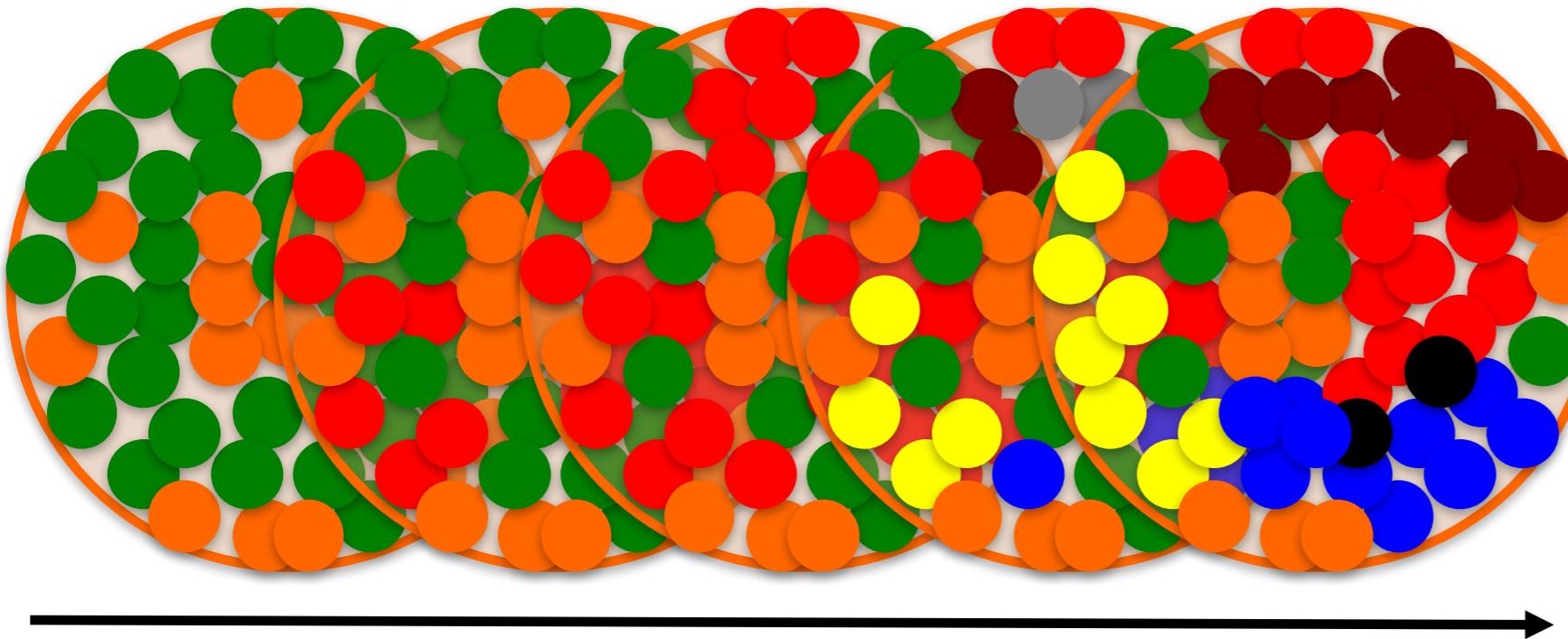
Part I of V Why should we study gene expression at the resolution of single-cells?



Tumors are heterogeneous cell populations that evolve over time



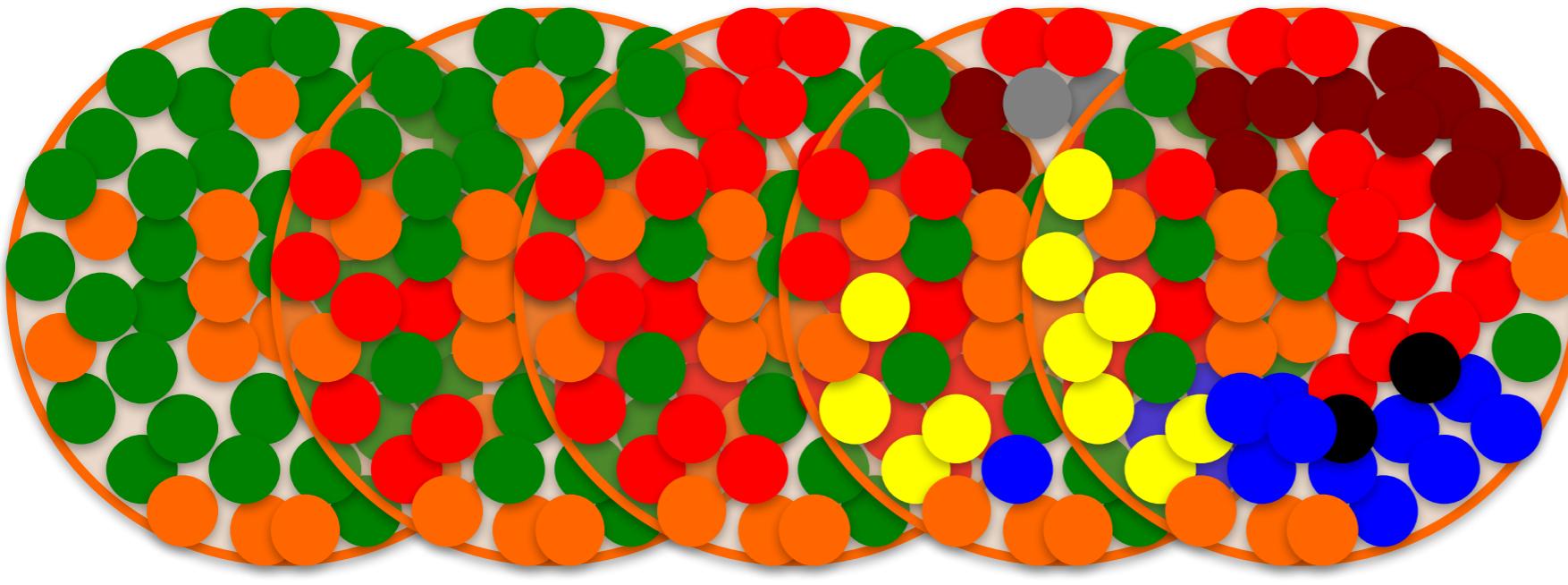
Rapid mutation and selection lead to tumor heterogeneity



A TUMOR TISSUE ACROSS TIME

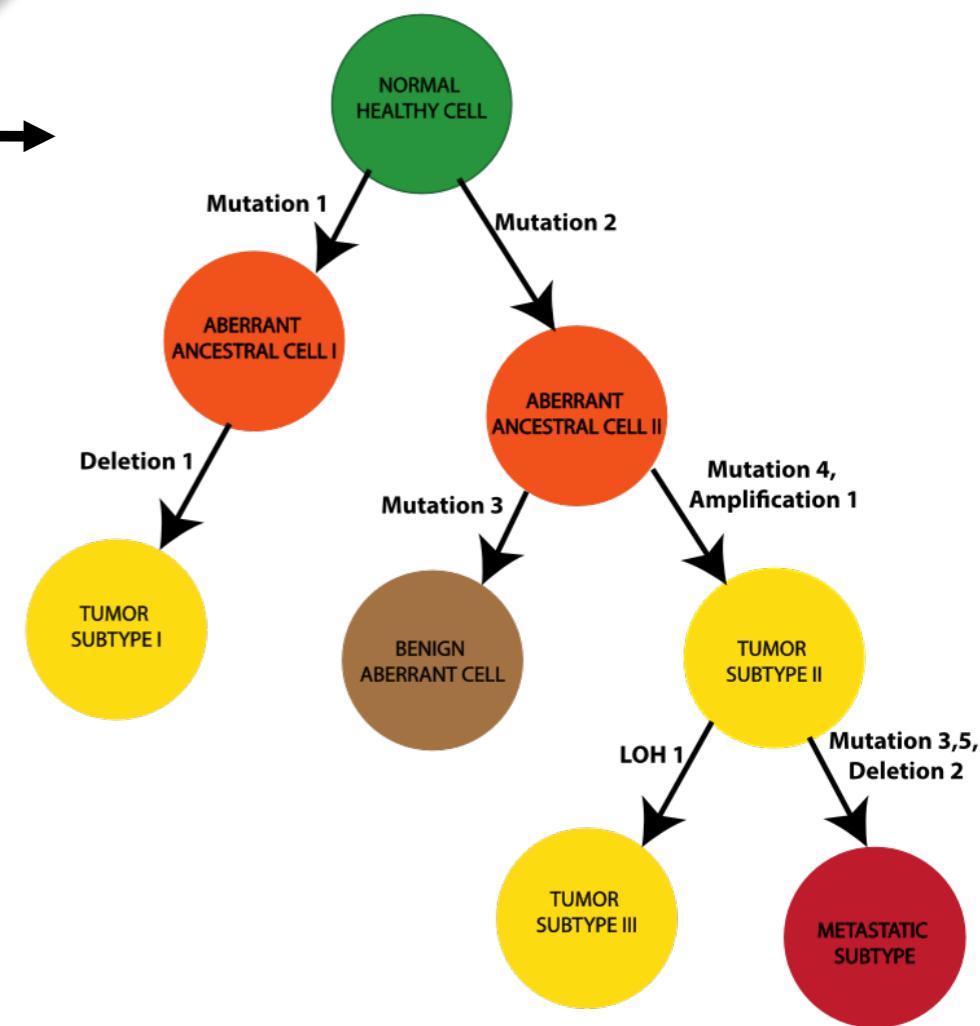
- Healthy Normal Cell from source tissue
- Stromal cells, blood cells
- Tumor cell 1
- Tumor cell 2
- Tumor cell 3
- Tumor cell 4
- Tumor cell 5
- Tumor cell 6

Rapid mutation and selection lead to tumor heterogeneity



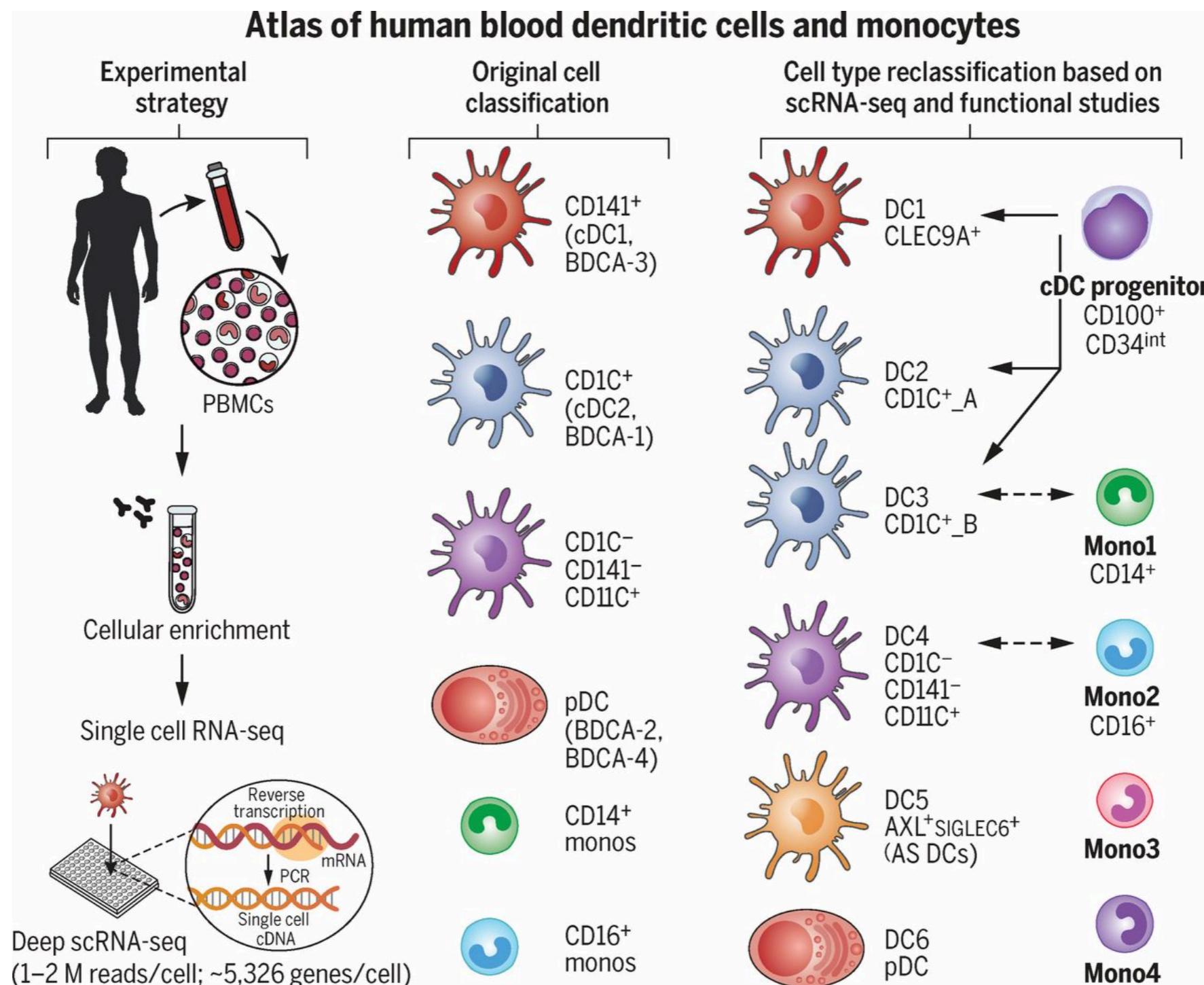
A TUMOR TISSUE ACROSS TIME

- Healthy Normal Cell from source tissue
- Stromal cells, blood cells
- Tumor cell 1
- Tumor cell 2
- Tumor cell 3
- Tumor cell 4
- Tumor cell 5
- Tumor cell 6



A tumor phylogeny

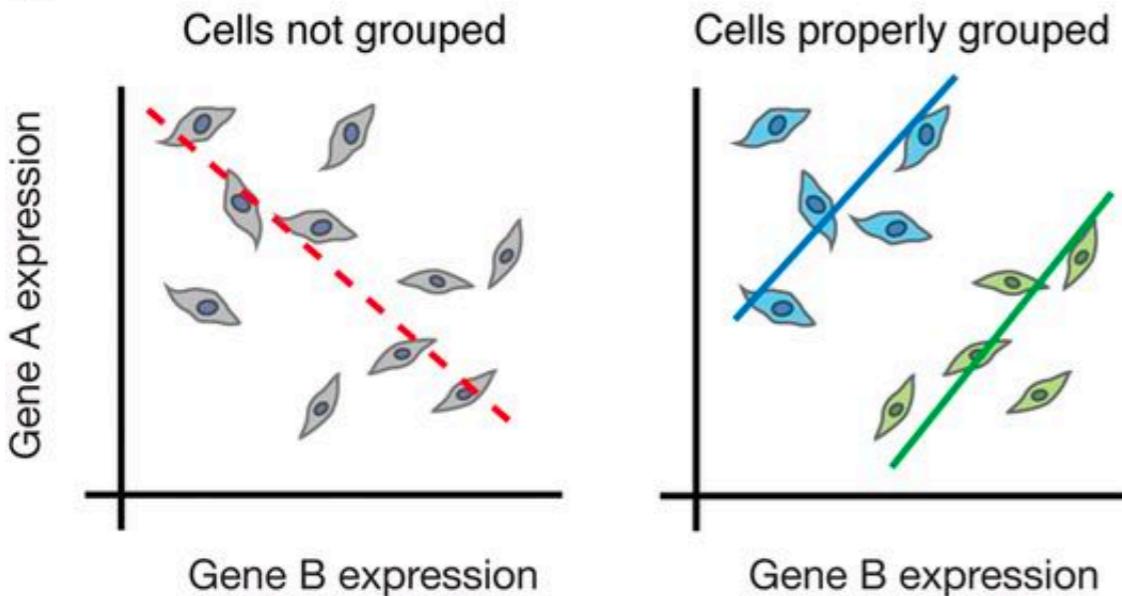
"Normal" tissue is also heterogenous



Bulk sequencing technologies measure average profiles from tissue samples

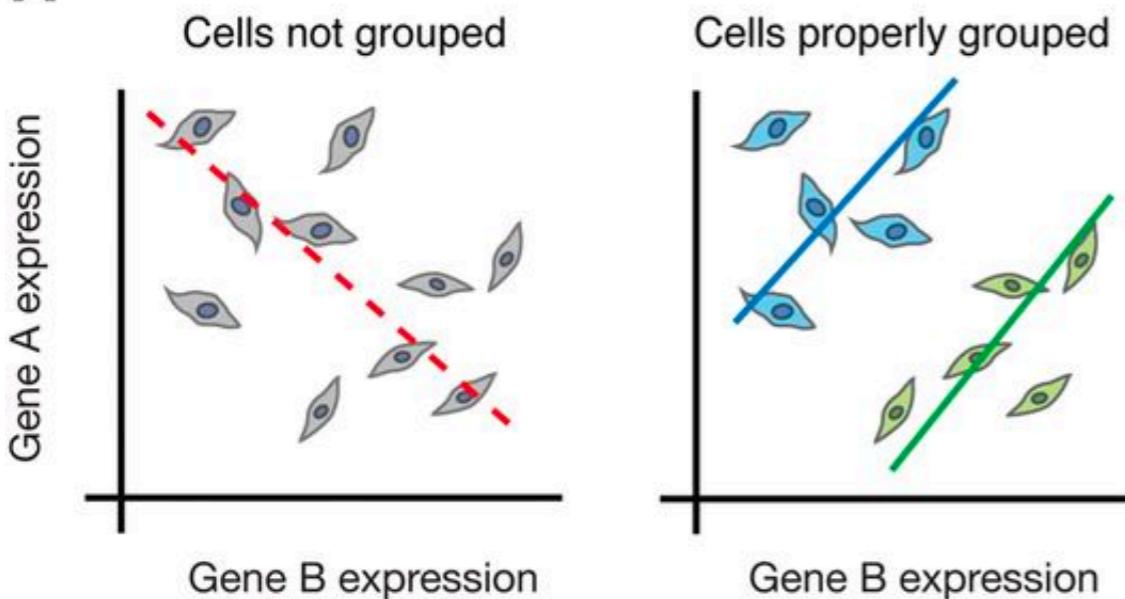
Bulk sequencing technologies measure average profiles from tissue samples

A

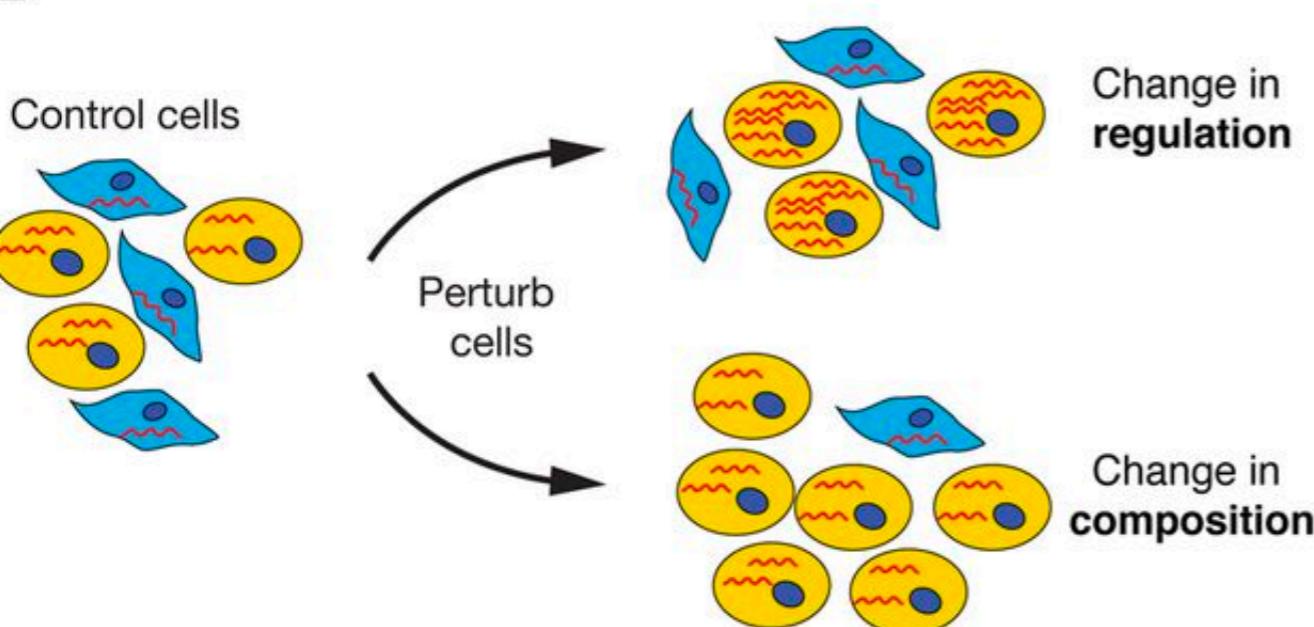


Bulk sequencing technologies measure average profiles from tissue samples

A

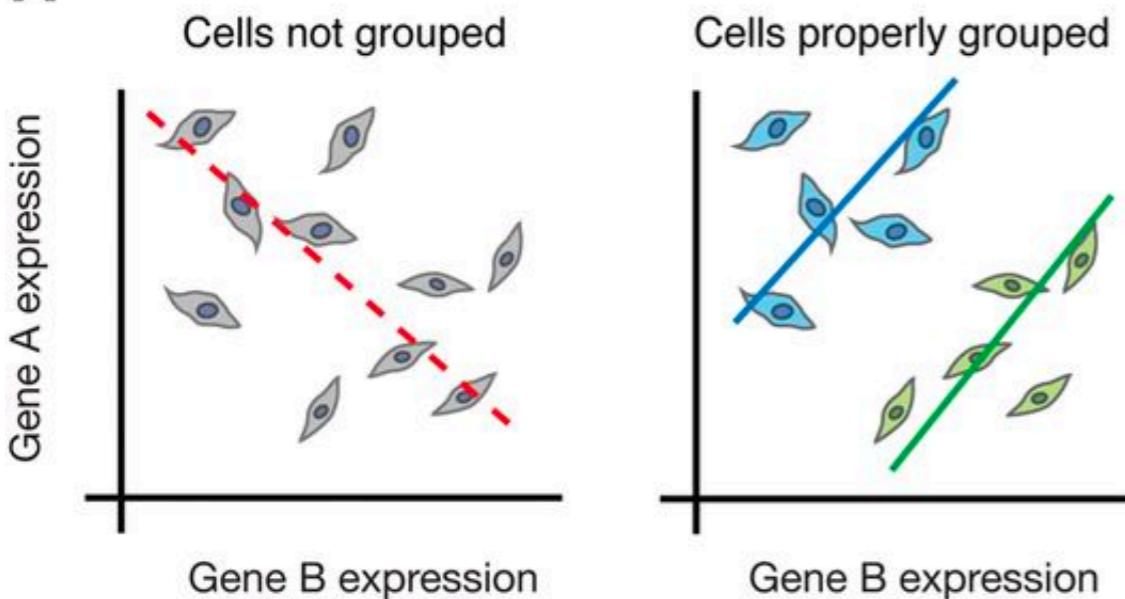


B

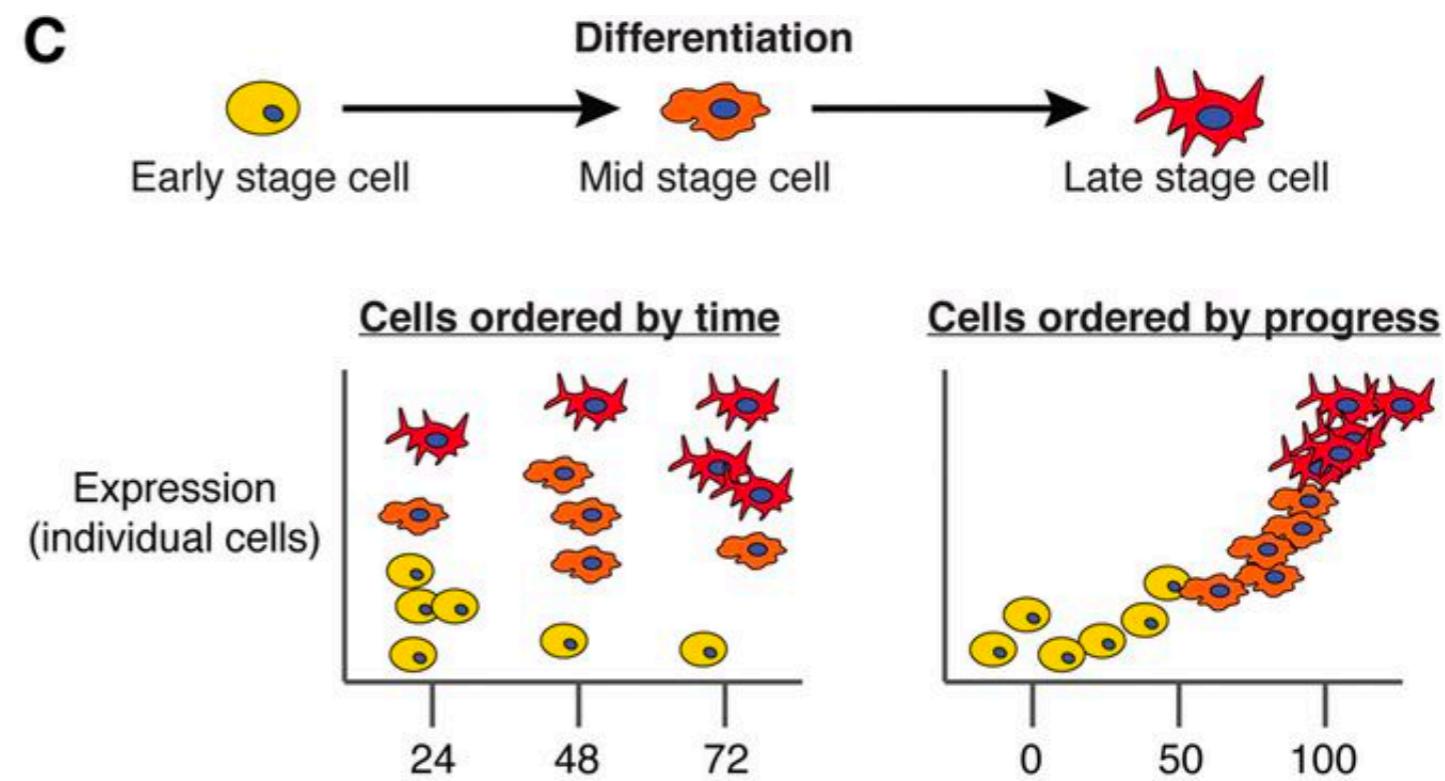


Bulk sequencing technologies measure average profiles from tissue samples

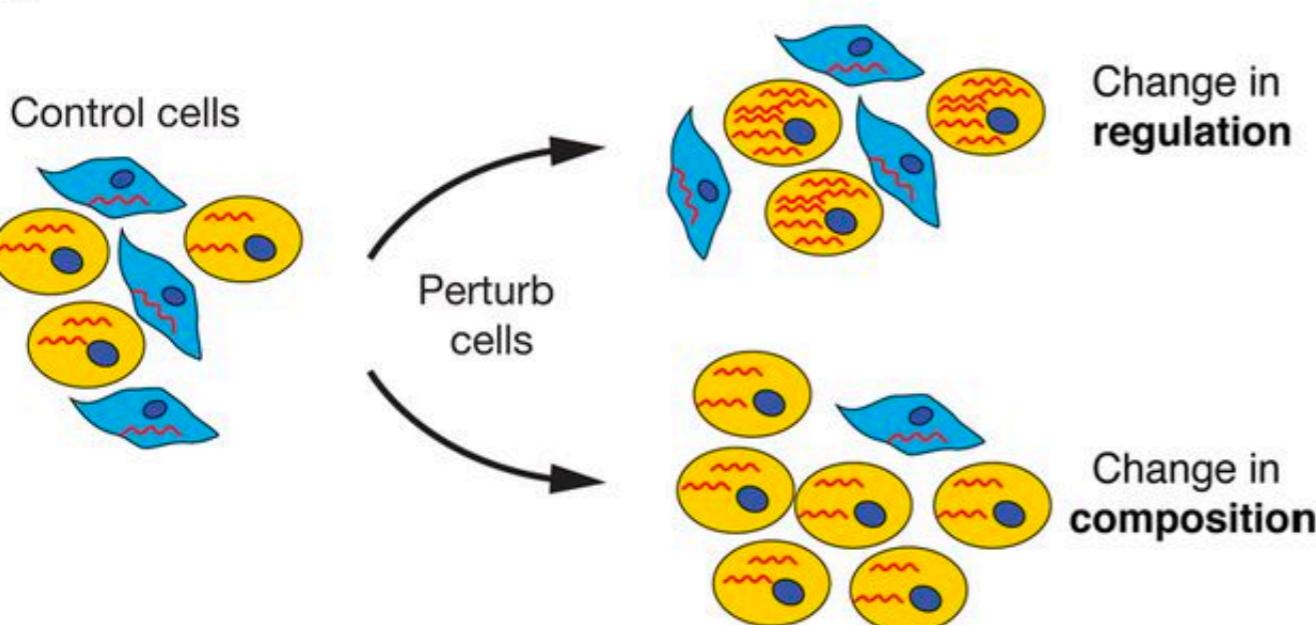
A



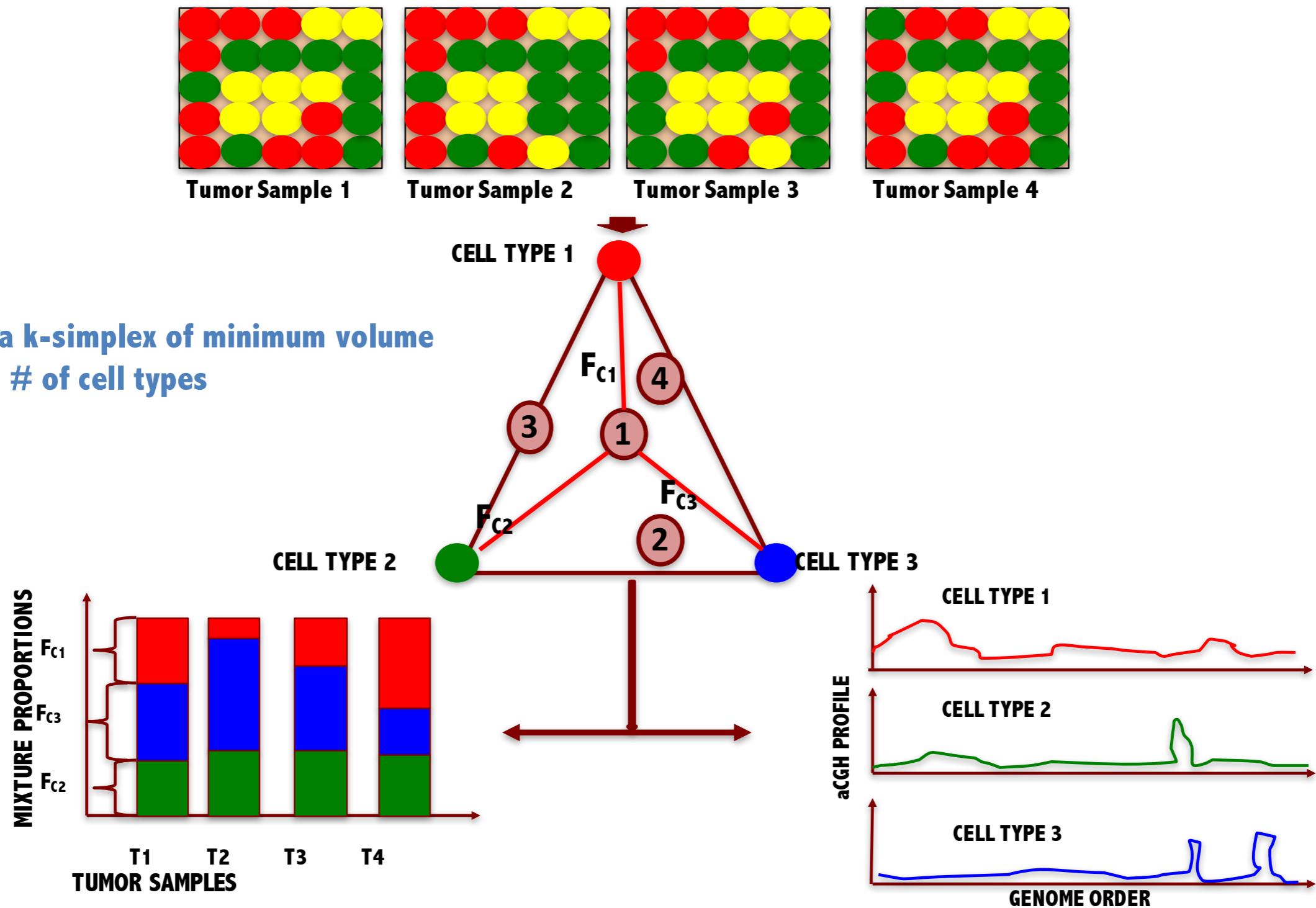
C



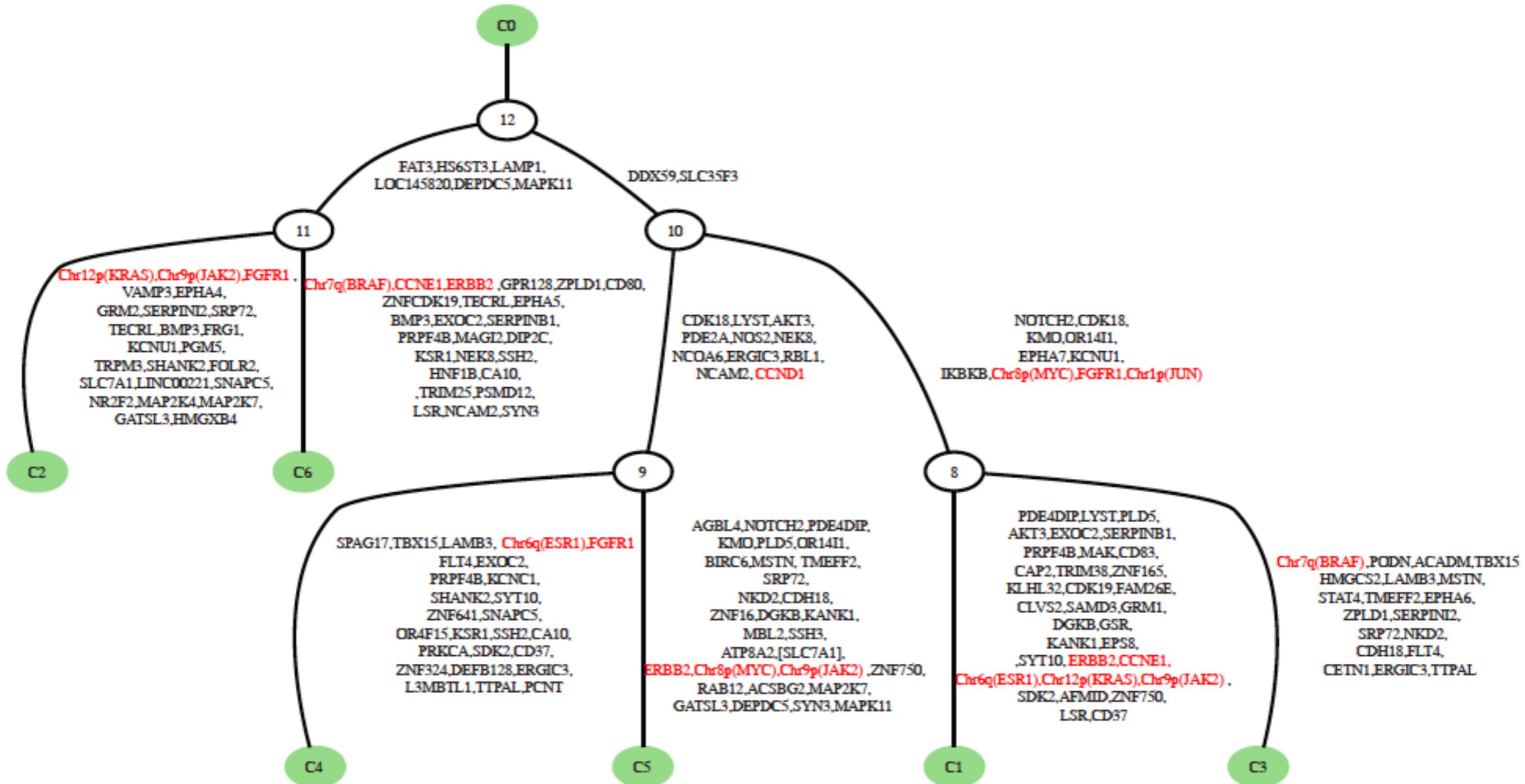
B



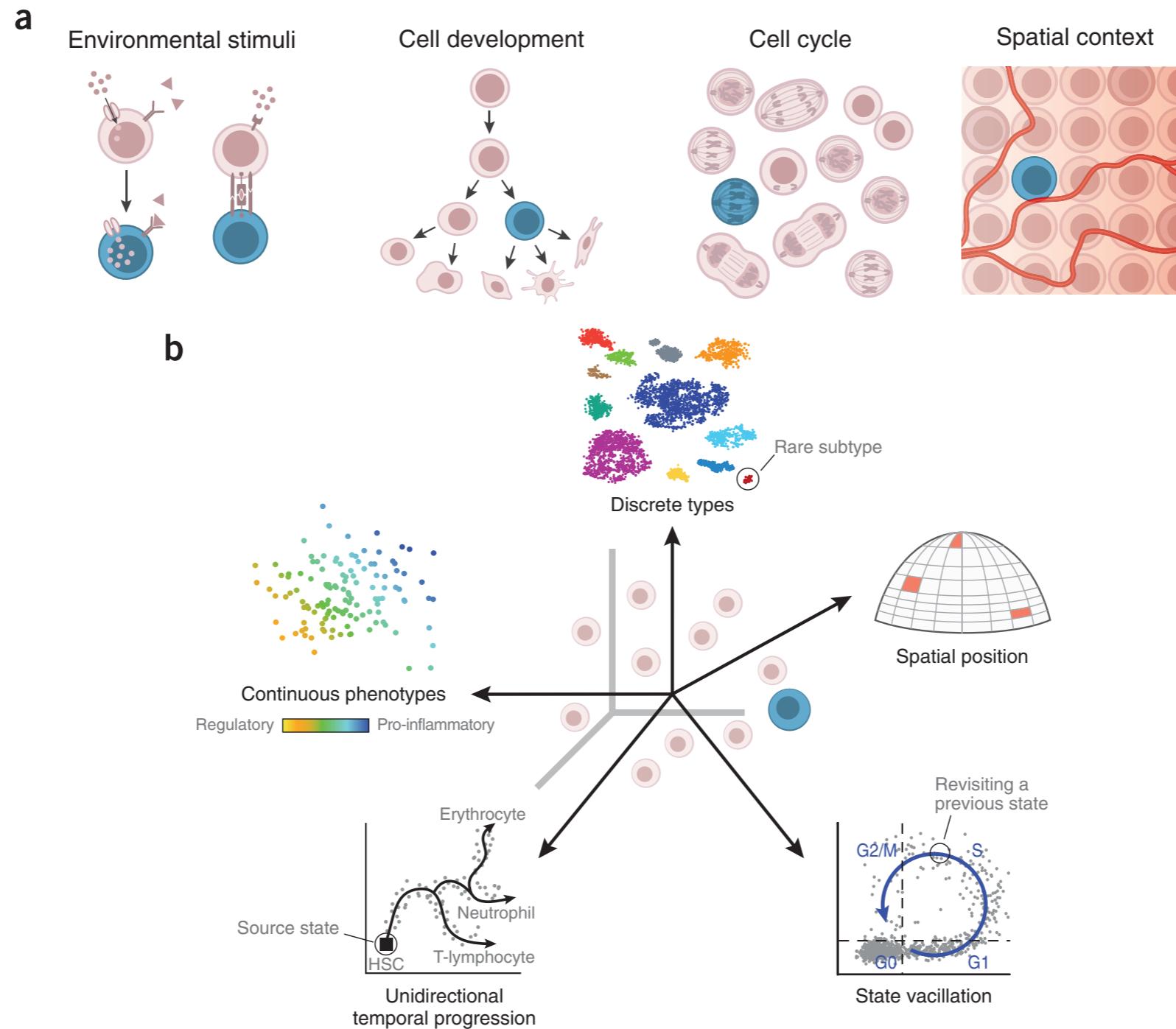
Computationally inferred cell states from bulk data need validation



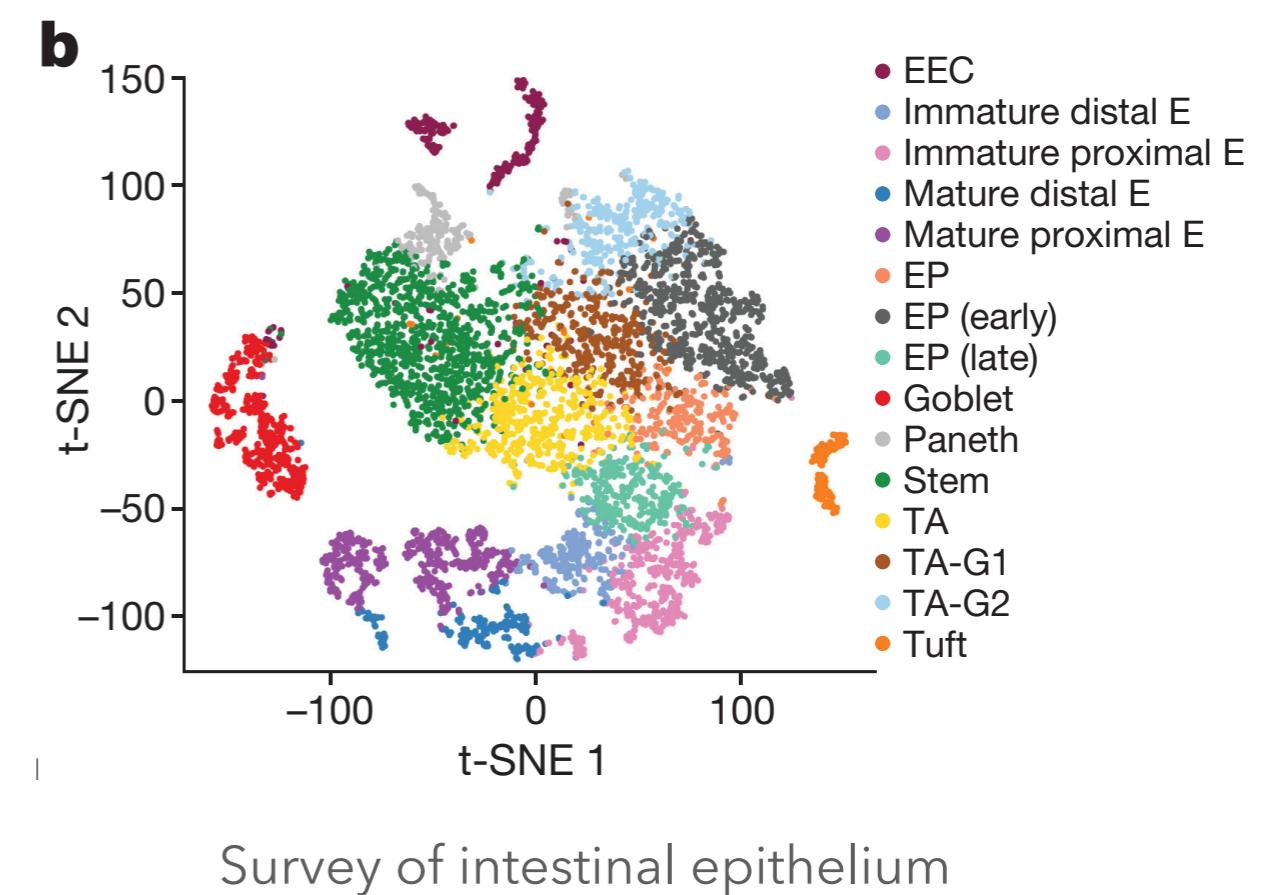
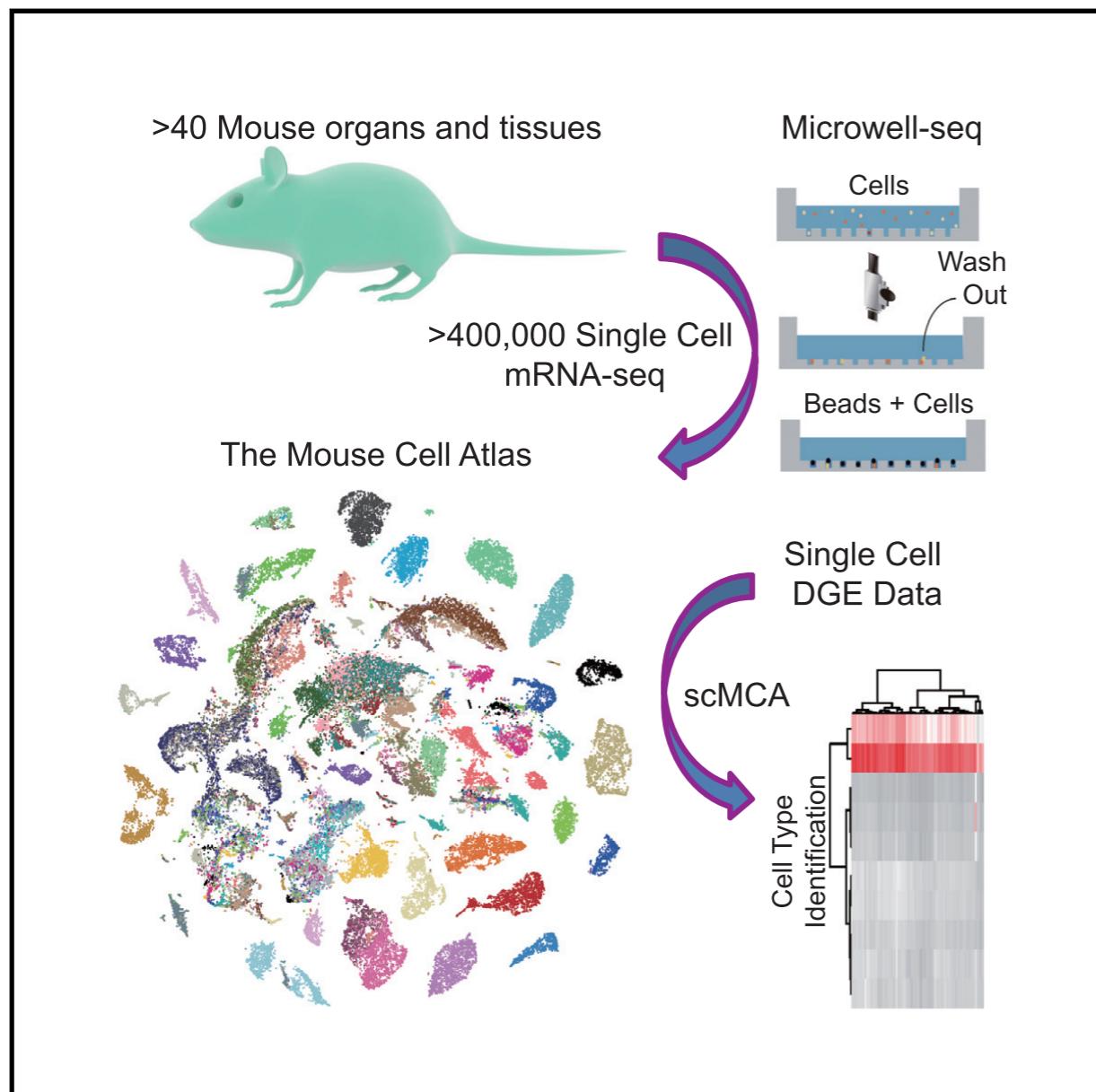
Inferring breast tumor evolution based on inferred cell states : cell states may again be mixtures



Part II of IV: What questions can we answer by studying single-cells?



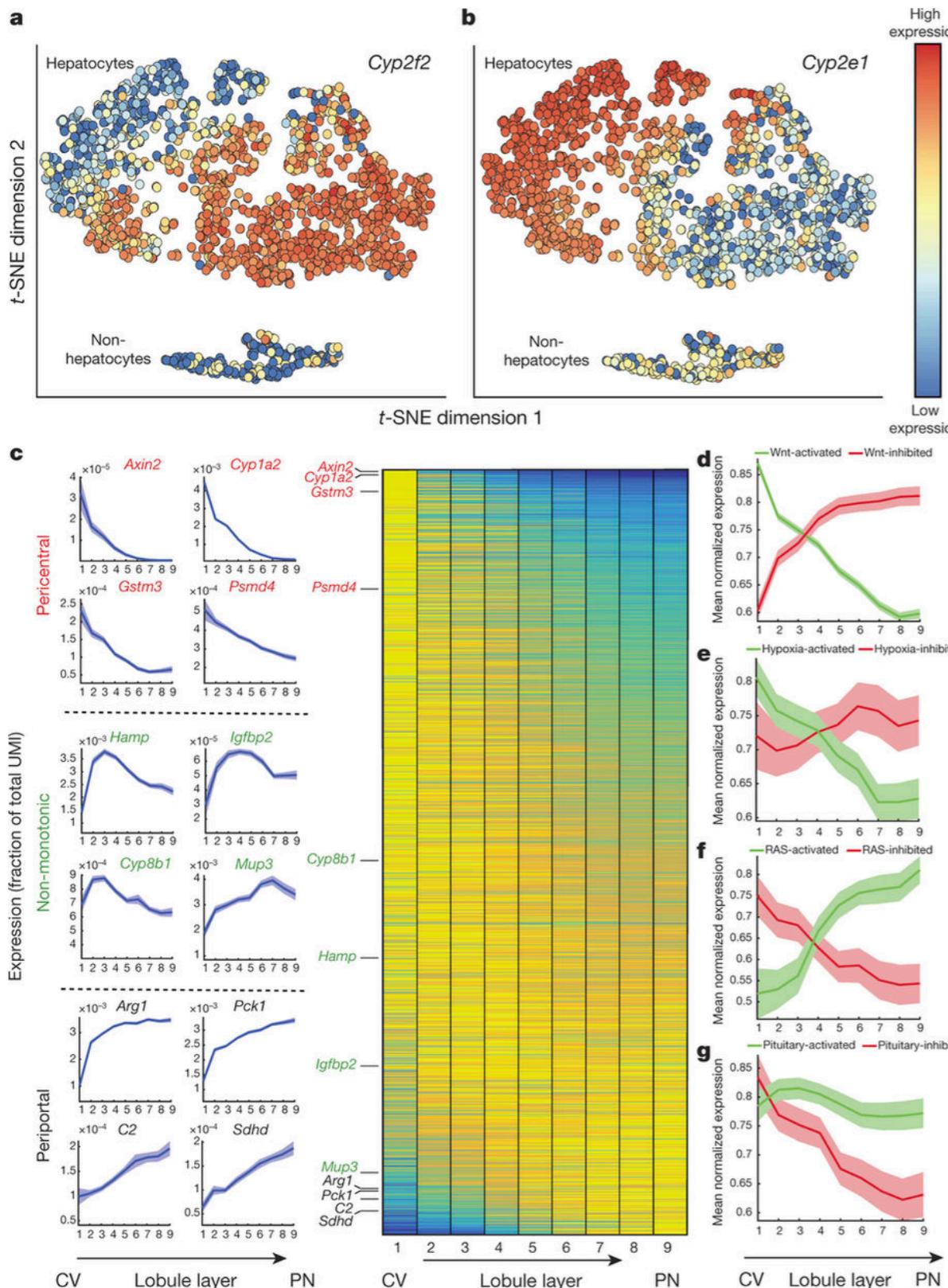
Taxonomy: Profile known and novel cell types and states in tissues



Han et al. Cell 2018

Haber, Moshe and Rogel et al. Nature 2017

Histology : Spatial orientation of cell states



Single-cell spatial reconstruction of liver

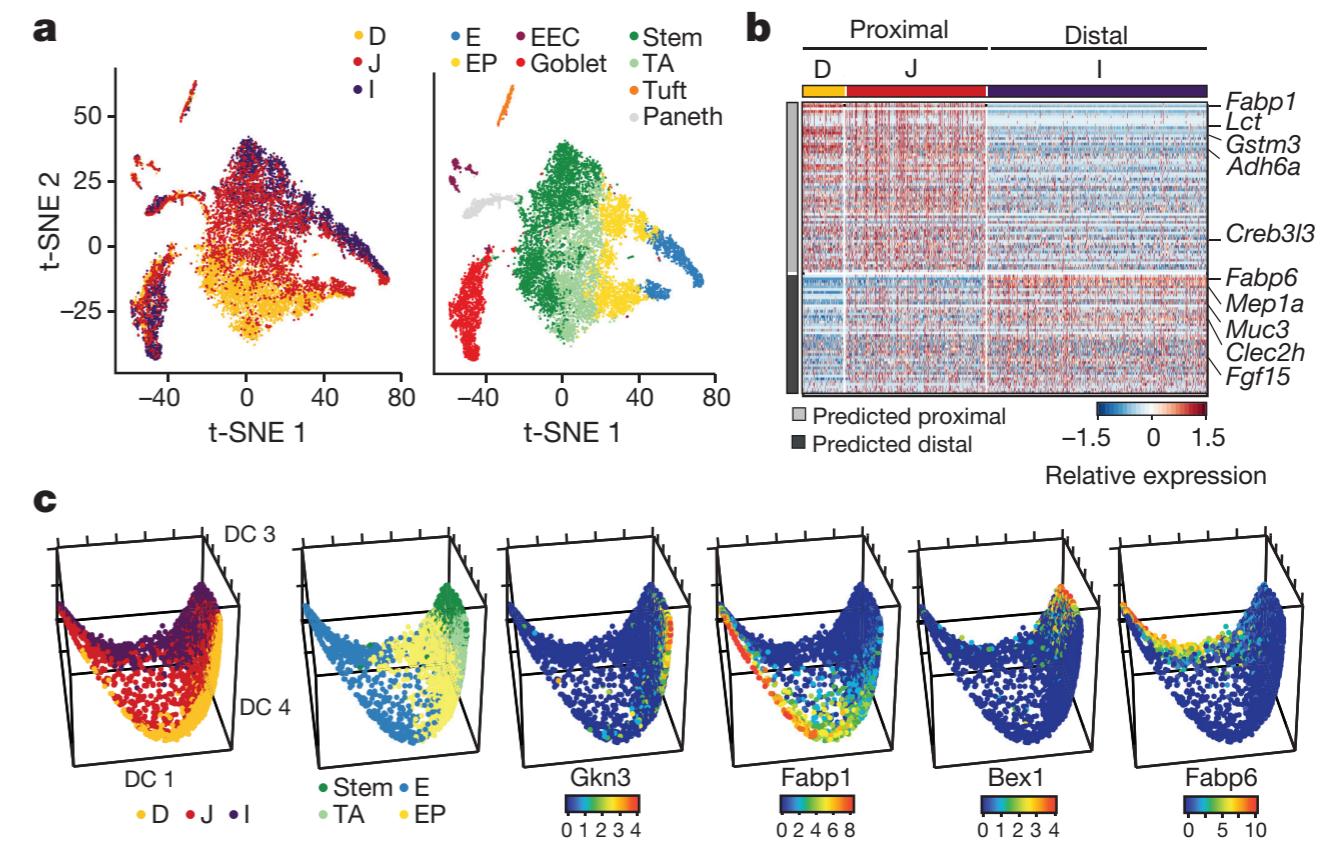
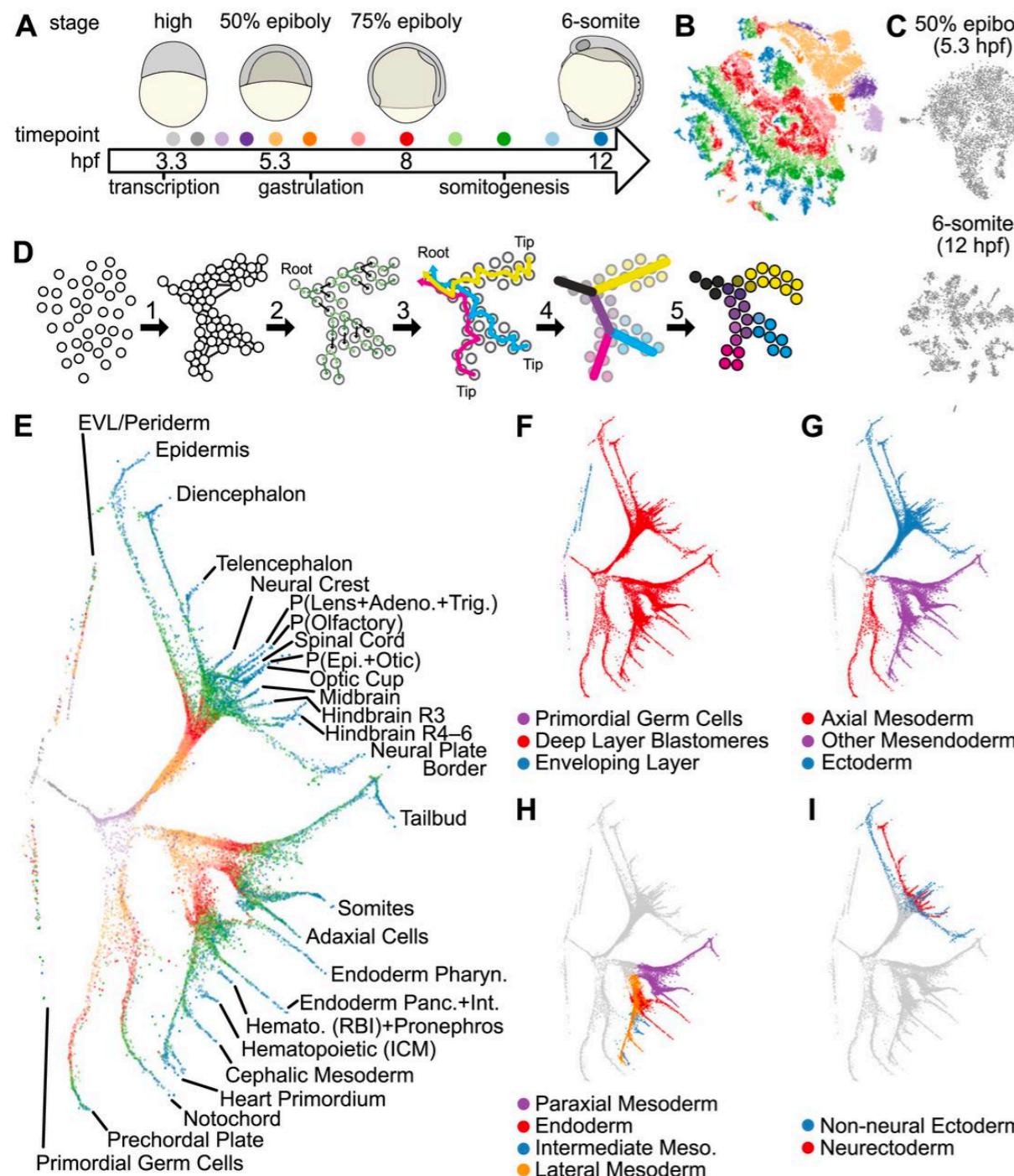


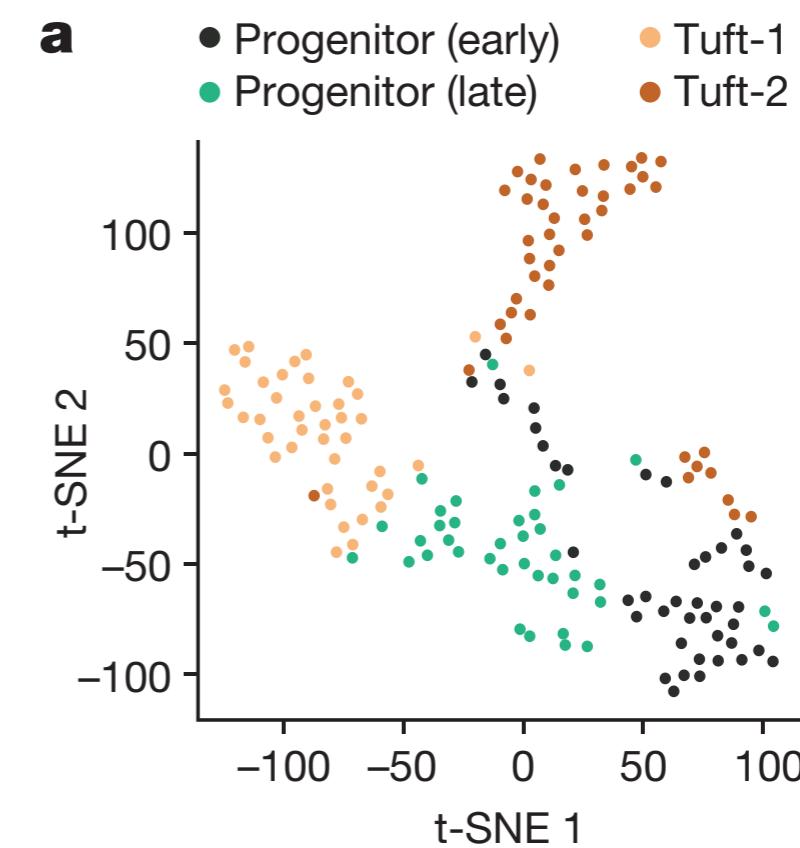
Figure 2 | Regional variation in cell type and differentiation. **a**, Regional surveys. t-SNE of 11,665 cells from the duodenum, jejunum and ileum

Halpern et al. Nature 2017
Haber, Moshe and Rogel et al. Nature 2017

Development or differentiation : Infer the trajectory of evolution of cell states

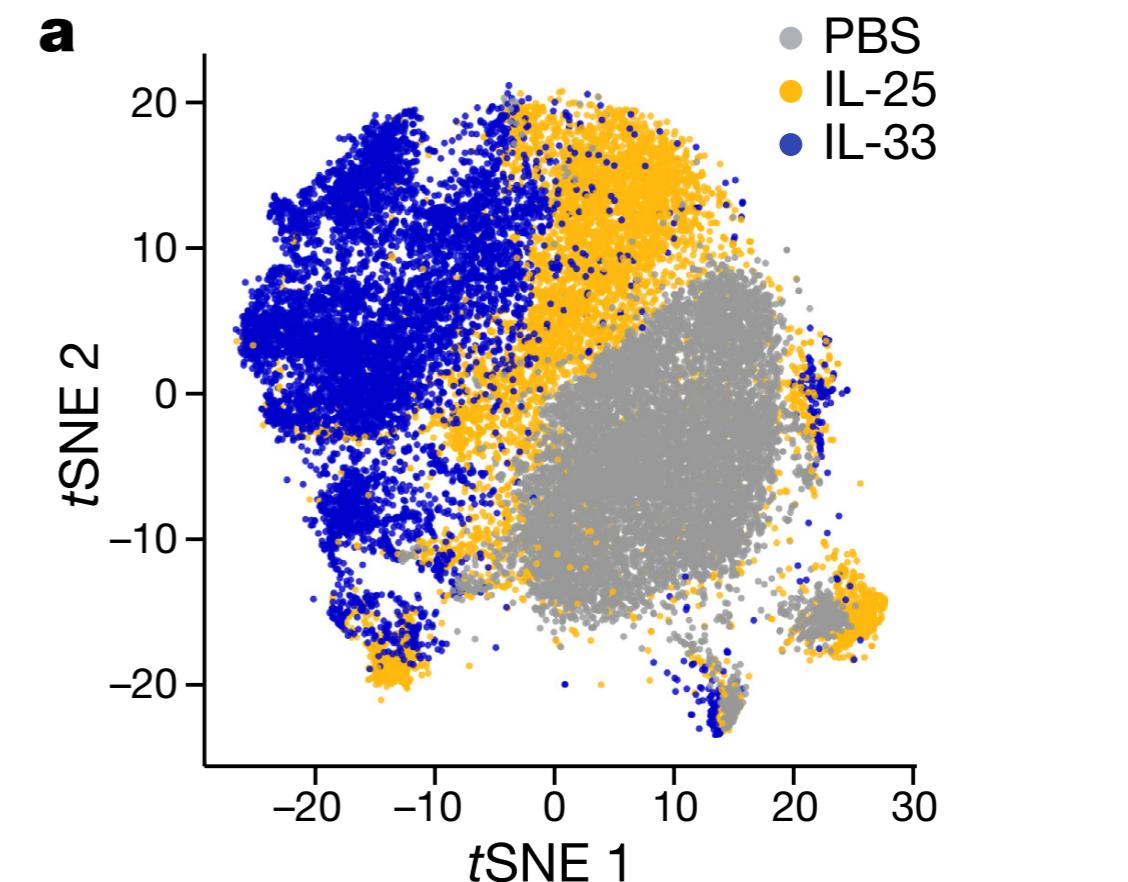
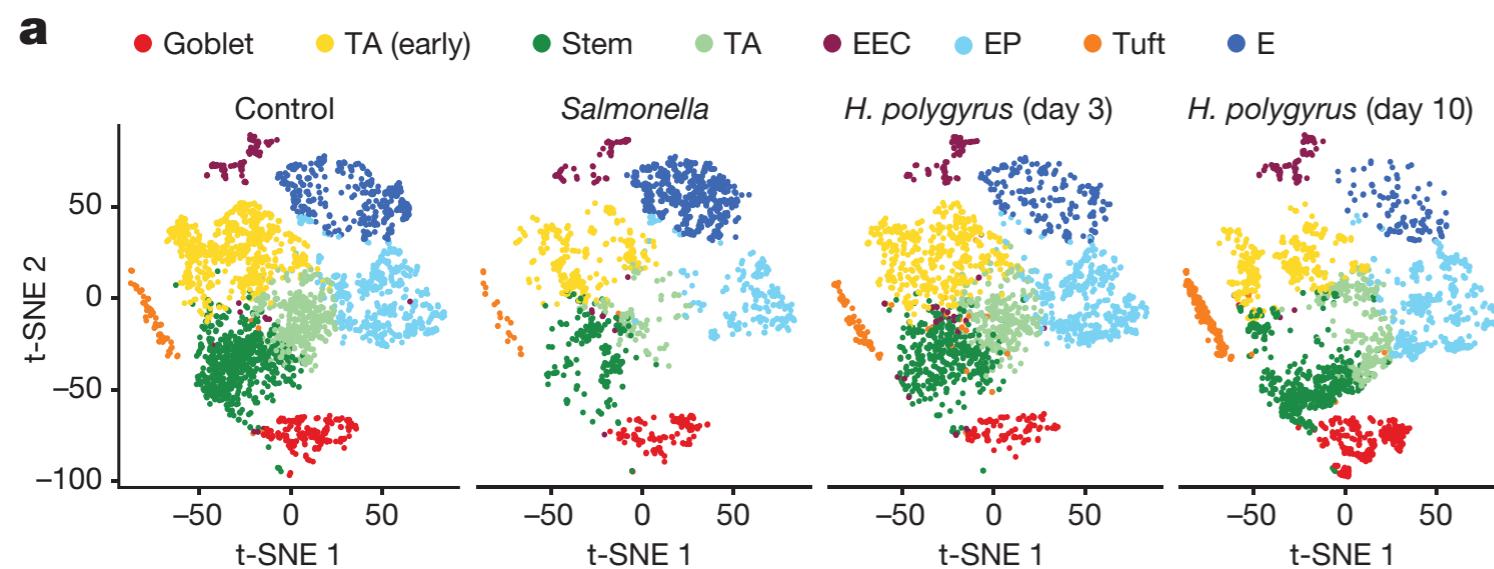


Model of early zebrafish embryogenesis

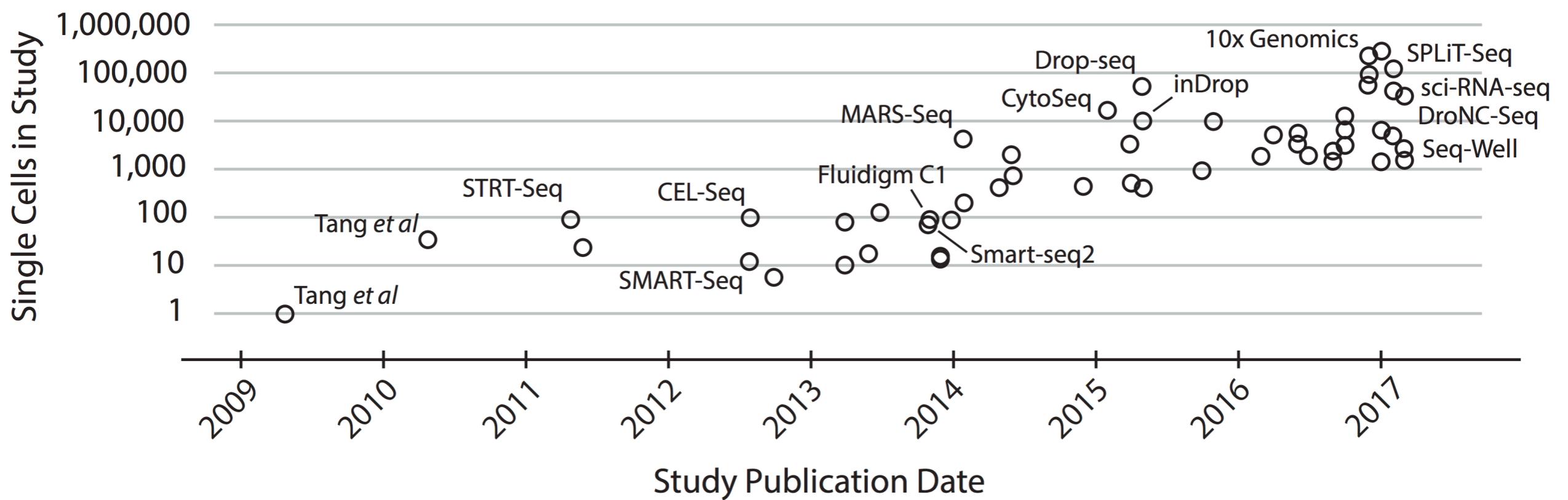


Haber, Moshe and Rogel et al. Nature 2017
Farrell and Wang et al. Science 2018

Physiology : Different states of cells in response to environmental cues



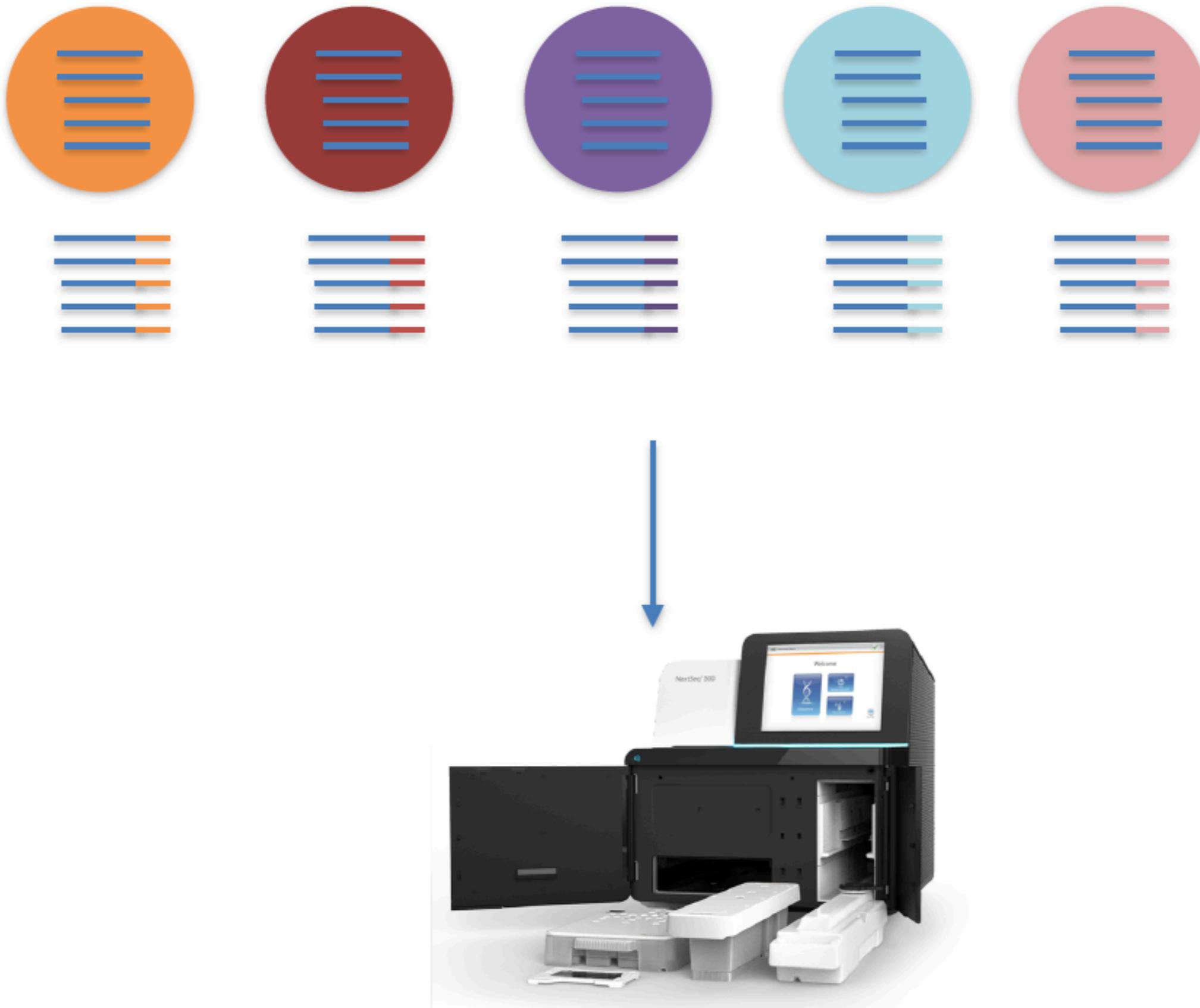
Part III of IV: How do we measure gene expression at single-cell resolution?



Tracking the cell-of-origin of individual transcripts



Tracking the cell-of-origin of individual transcripts



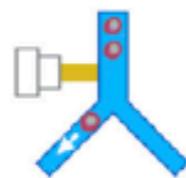
Single-cell RNA-seq pipeline



Single-cell RNA-seq pipeline



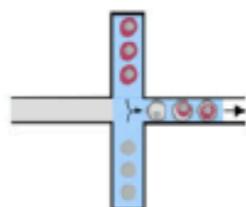
FACS



Micro-capture



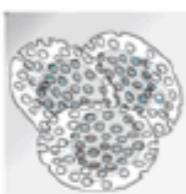
Droplet



Nanowell



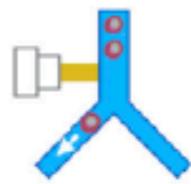
Split/pool
barcoding



Single-cell RNA-seq pipeline



FACS



PolyA vs random priming

3'/5' end tagging vs full-length

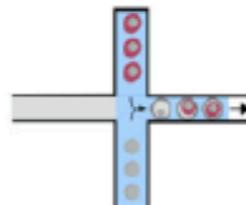
PCR vs *in vitro* transcription

...

Micro-capture



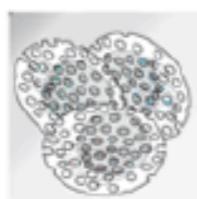
Droplet



Nanowell



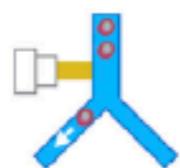
Split/pool
barcoding



Single-cell RNA-seq pipeline



FACS



PolyA vs random priming

3'/5' end tagging vs full-length

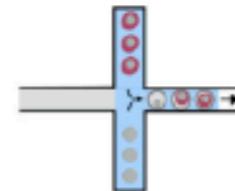
PCR vs *in vitro* transcription

...

Micro-capture



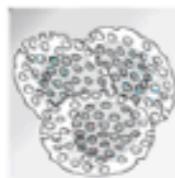
Droplet



Nanowell



Split/pool
barcoding



~ 20M reads total



Miseq

Nextseq

~ 500M reads total

HiSeq 4000

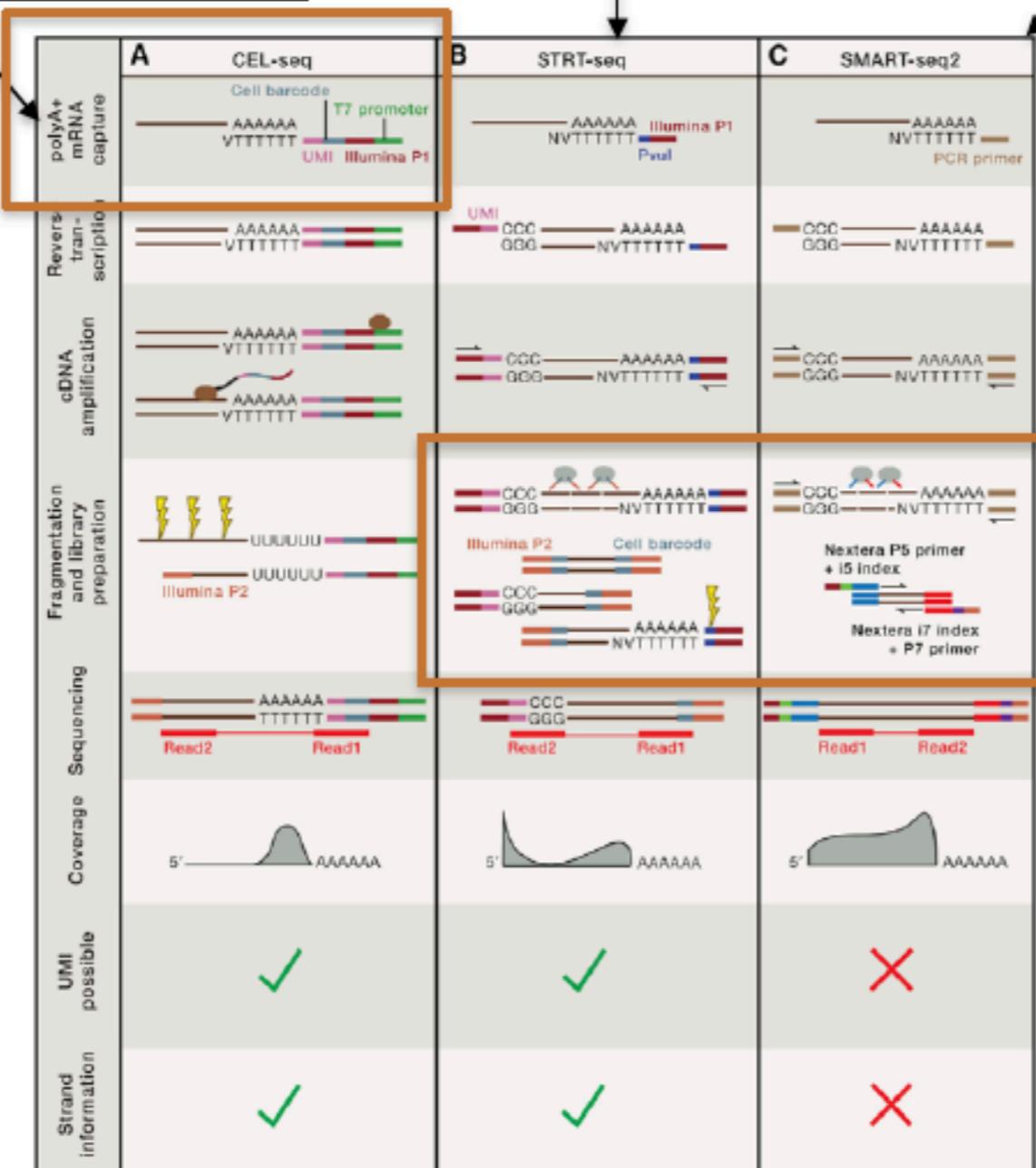
4 billion reads

Three commonly used scRNA-seq library prep methods

*in vitro transcription,
3' end tagging*

5' end tagging

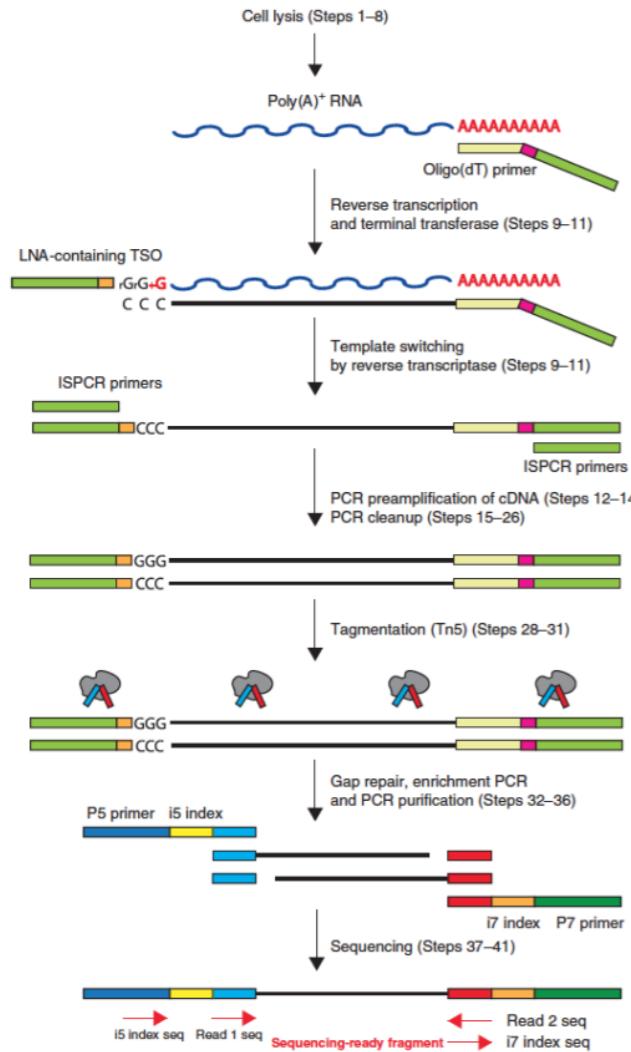
Full length coverage



Each protocol has **advantages** and **limitations**. What one ends up using is often dictated by multiple features - the **biological context, cost, objective** etc.

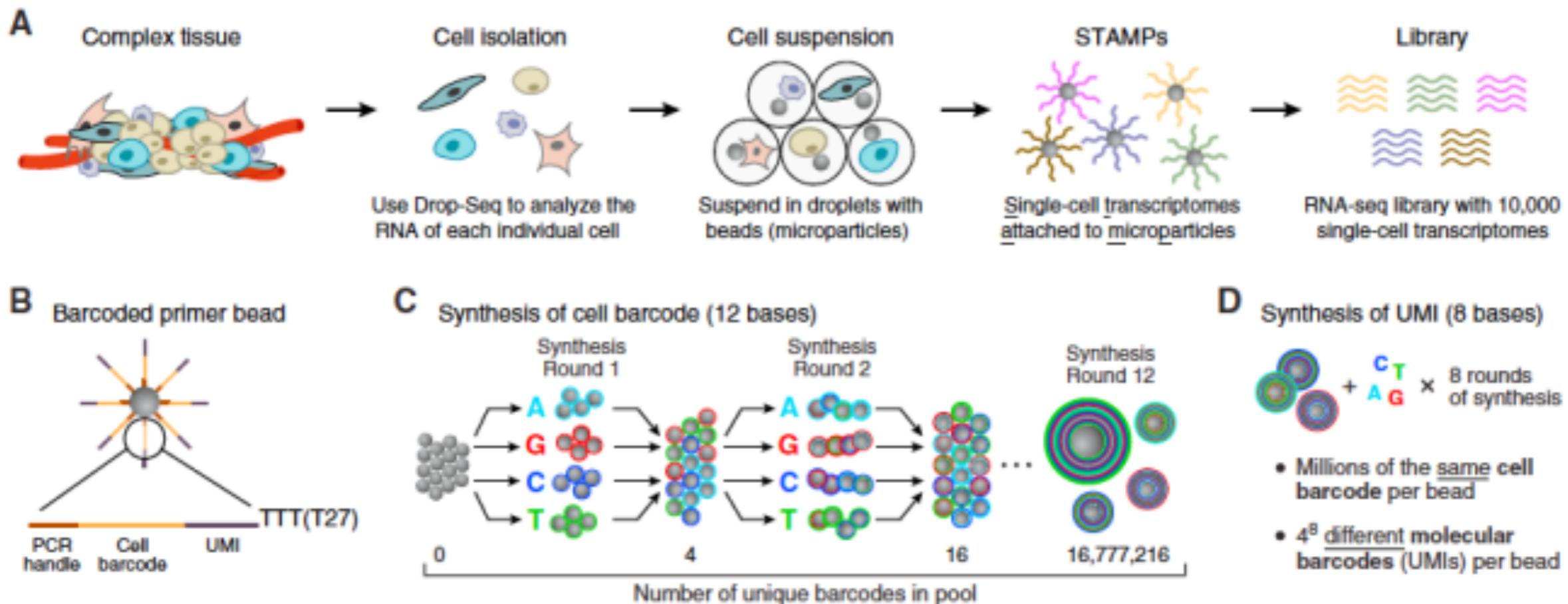
Grun and van oude naarden, Cell, 2015

Smart-seq2: Assay Overview

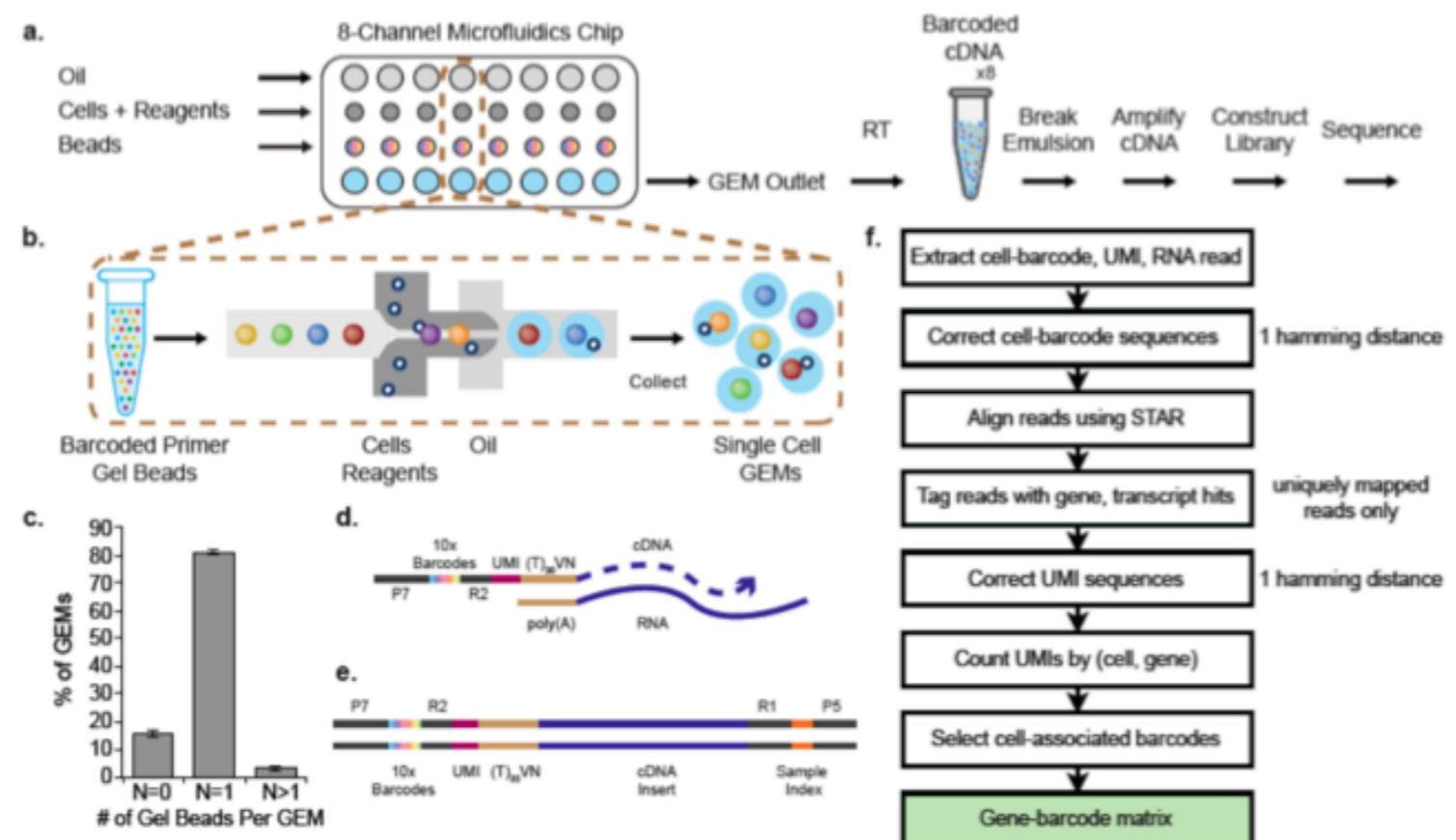


- Poly-A capture with 30nt polyT and 25nt 5' anchor sequence.
- RT adding untemplated C
- Template switching
- Locked Nucleic Acid binds to untemplated C
- RT switches template
- Preamplification / cleanup
- DNA fragmentation and adapter ligation together.
- Gap Repair, enrich, purify.

Dropseq: Capturing >1000s of cells using microfluidics



10x Droplet based technology for scRNAseq is getting widely adapted



Part IV of V: How do we analyze single-cell gene expression to draw biological insights?

Considerations for scRNA-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 :

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

More cells or more reads / cell

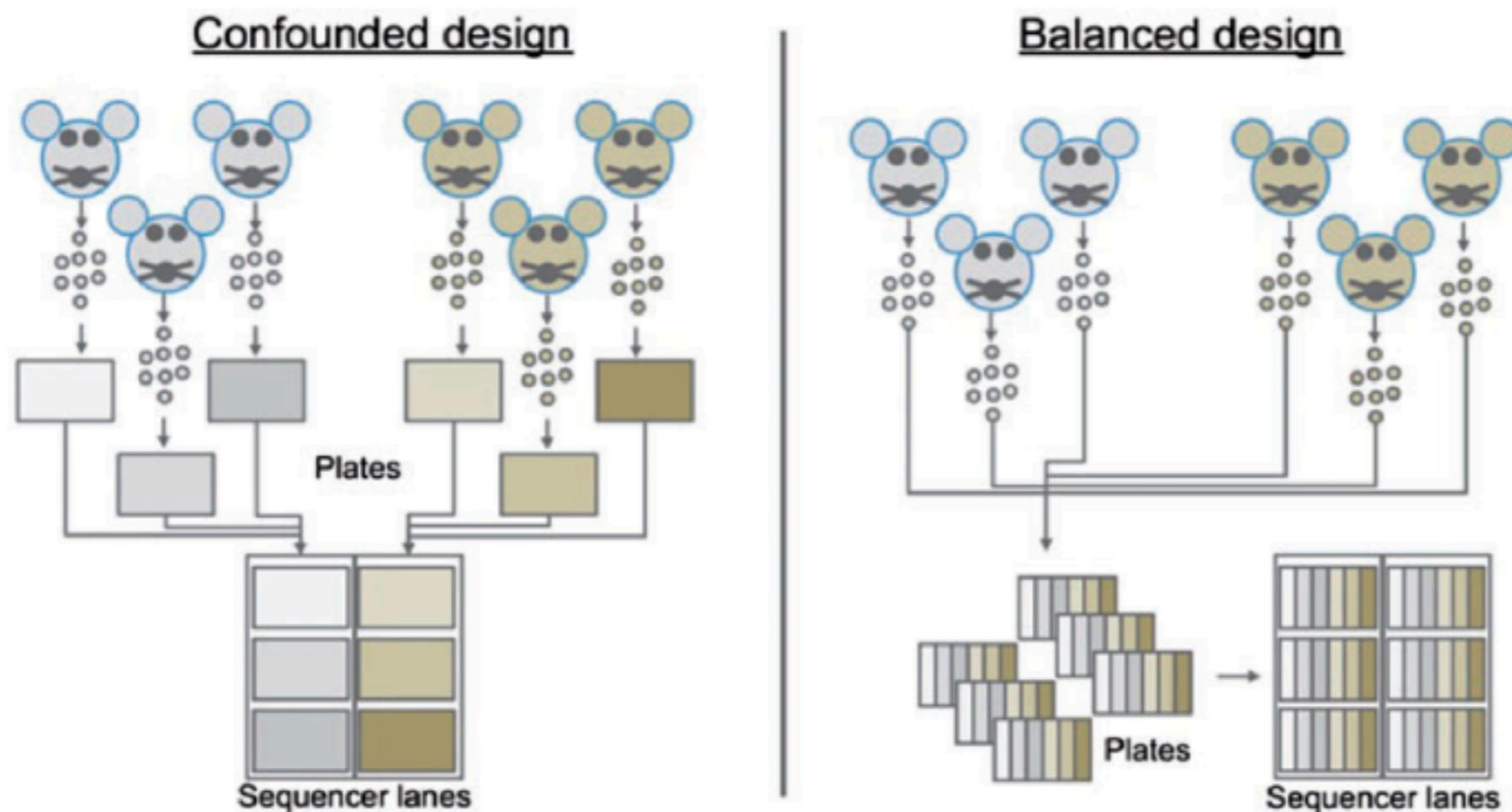
- Earlier scRNA-seq studies used to sequence each cell to > 10 million reads / cell - **this is now widely accepted as a ridiculous number!**
- **~50,000-100,000 reads/cell** is now widely regarded as sufficient for most applications. ~1M reads per cell effectively means saturation
- The modular nature of biology “guarantees” that key signals can be recovered at shallow sequencing depth*

Cleary et al., Cell, 2017

Experimental Design

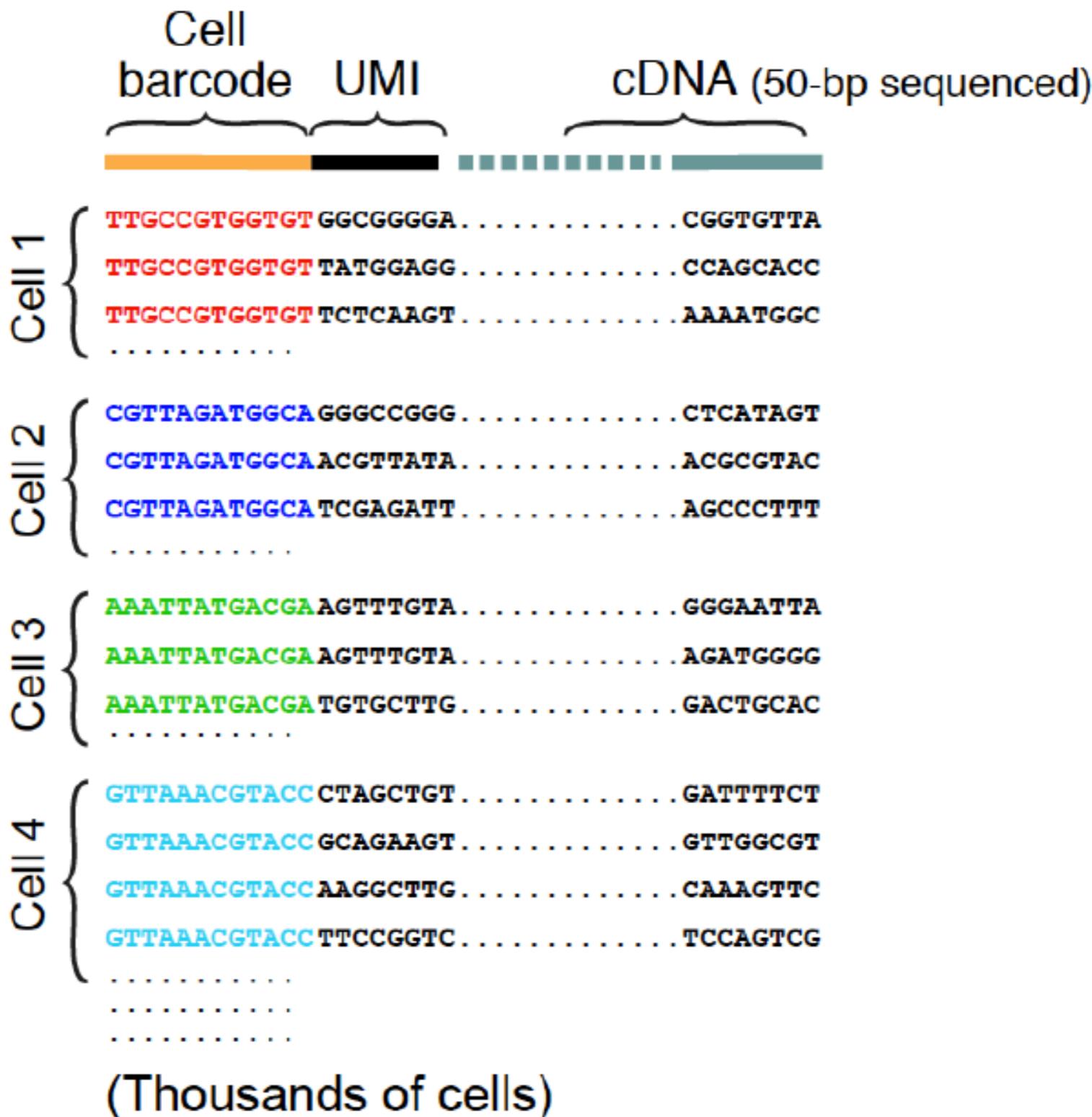
Sound experimental design : Replication, Randomization and Blocking

- R. A. Fisher, 1935



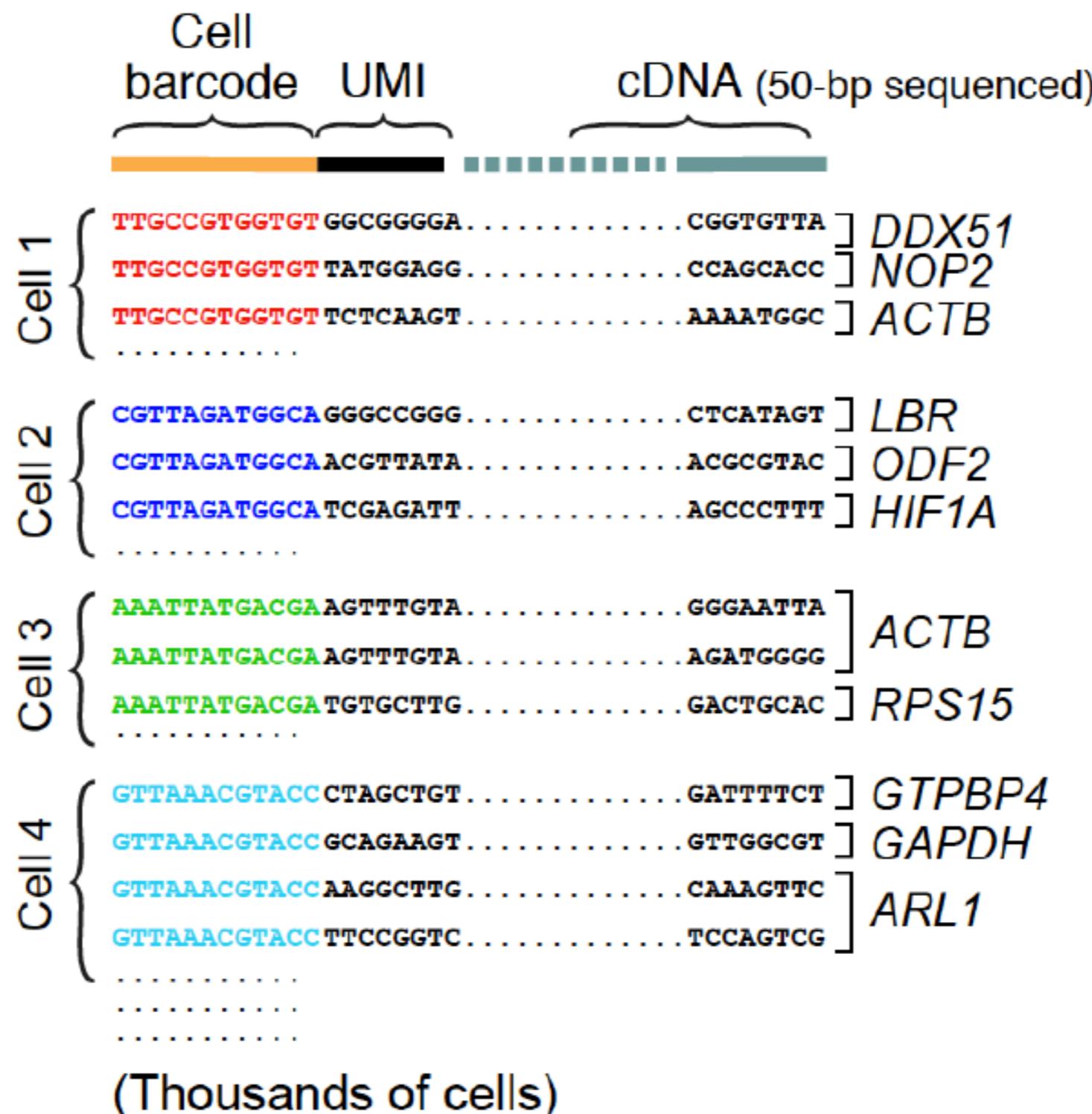
- For example, when analyzing tumor phenotypes in a patient process the tumor sample and a matched control on the same day, using the same reagents!
- Blocking is not always possible because of logistic limitations, in which case ensure that any biological conclusion is supported by multiple, independently collected samples

Group reads by cell barcode



Some platforms incorporate “error-correcting” barcodes which makes the pipeline robust to sequencing errors

Align reads to the genome using a splice-aware aligner



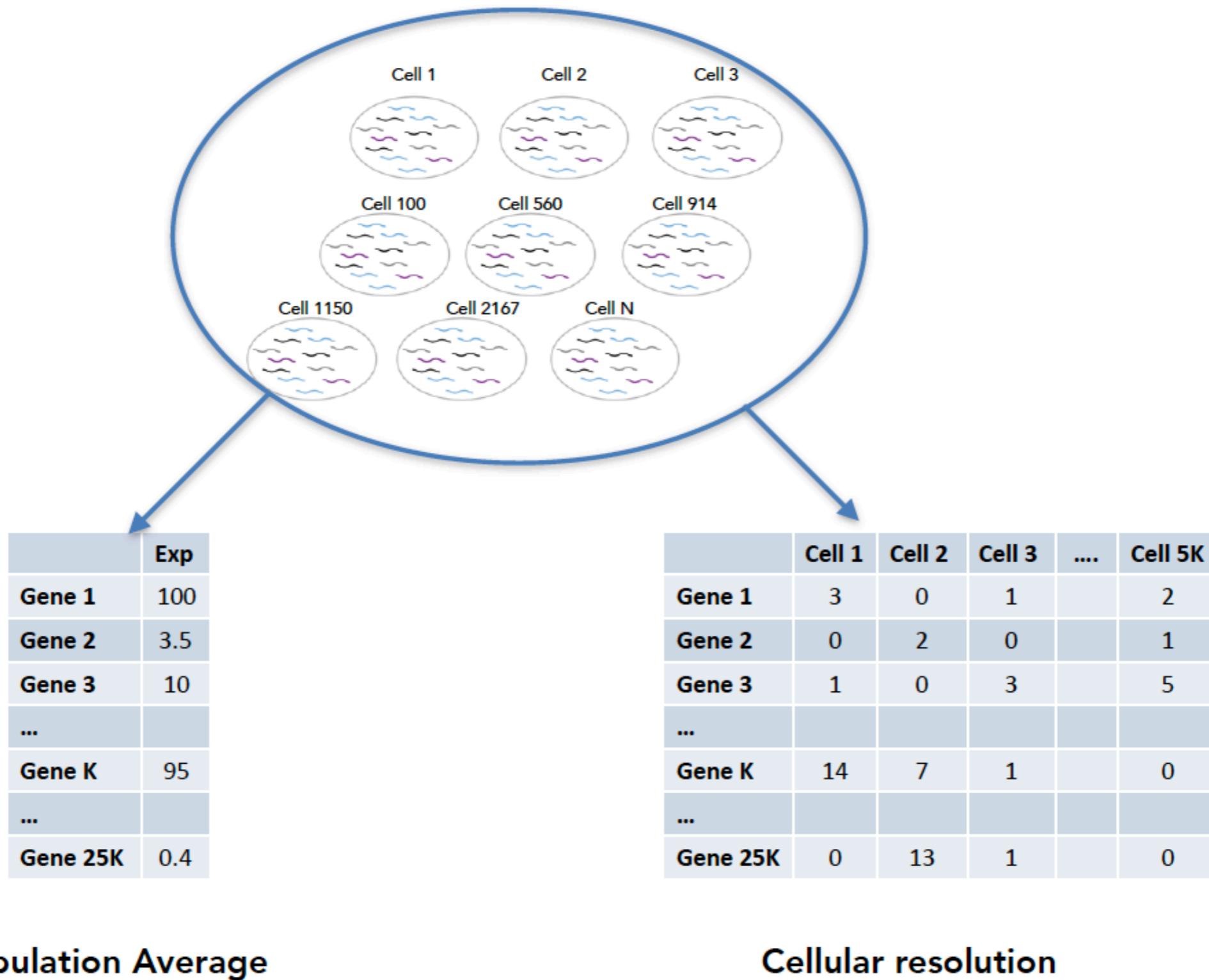
A lot of pipelines use the STAR aligner, which consumes a lot of memory,
but is EXTREMELY fast

Collapse UMIs to count transcripts

	Cell barcode	UMI	cDNA (50-bp sequenced)	
Cell 1	{	TTGCCGTGGTGT	GGCGGGGA.....CGGTGTTA] <i>DDX51</i>
		TATGGAGG.....CCAGCACC] <i>NOP2</i>
		TCTCAAGT.....AAAATGGC] <i>ACTB</i>
Cell 2	{	CGTTAGATGGCA	GGGCCGGG.....CTCATAGT] <i>LBR</i>
		ACGTTATA.....ACGGCTAC] <i>ODF2</i>
		TCGAGATT.....AGCCCTTT] <i>HIF1A</i>
Cell 3	{	AAATTATGACGA	AGTTTGTA.....GGGAATTA	<i>ACTB</i> → 2 reads, 1 molecule
		AGTTTGTA.....AGATGGGG		
		TGTGCTTG.....GACTGCAC] <i>RPS15</i>
Cell 4	{	GTTAACGTACC	CTAGCTGT.....GATTTCT] <i>GTPBP4</i>
		GCAGAAAGT.....GTTGGCGT] <i>GAPDH</i>
		AAGGCTTG.....CAAAGTTC] <i>ARL1</i> → 2 reads, 2 molecules
		TTCCGGTC.....TCCAGTCG		
			
			
			

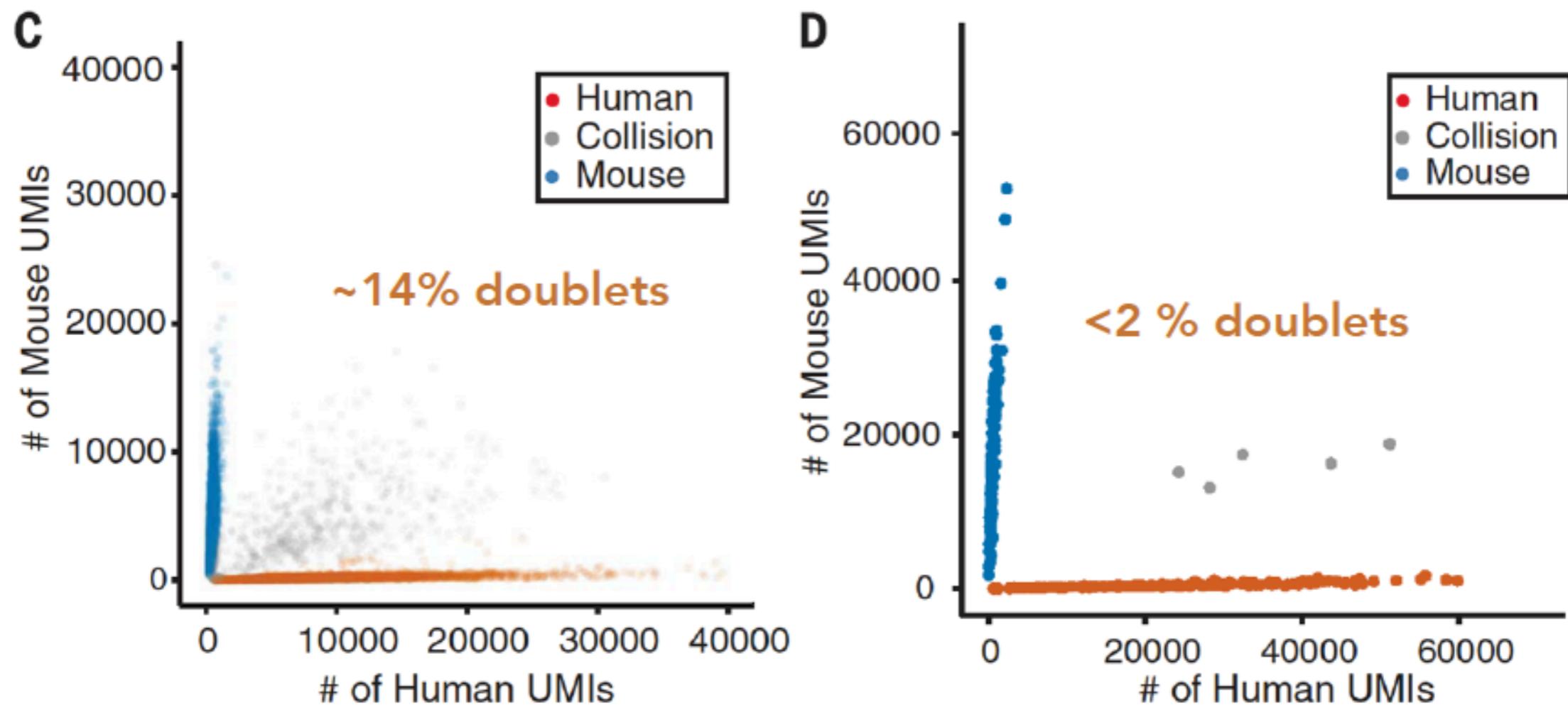
(Thousands of cells)

Bulk vs. single cell gene expression



Doublets

Because of the setup, it is possible that two or more cells can enter the same droplet. Studies estimate doublet frequency through a “mixed-species” experiment



The doublet frequency is +vely correlated with throughput

scRNA-seq by the numbers

2013, 18 cells

LETTER

doi:10.1038/nature12172

Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells

Alex K. Shalek^{1,2,*}, Rahul Satija^{2*}, Xian Adicono³, Rona S. Gerlinger³, Jellert T. Gaublomme³, Rakitma Raychowdhury², Schraga Schwartz², Nir Yosef², Christine Malboeuf³, Diana Lu³, John J. Trombetta³, Dave Gennert², Andreas Grütz², Alon Goren^{2,3}, Nir Hacohen^{2,3}, Joshua Z. Levin³, Hongkun Park^{2,3} & Aviv Regev^{2,3}

2014, 1700 cells

ARTICLE

doi:10.1038/nature13437

Single-cell RNA-seq reveals dynamic paracrine control of cellular variation

Alex K. Shalek^{1,2,3,*}, Rahul Satija^{2*}, Joe Shuga^{4*}, John J. Trombetta³, Diana Lu³, Peilin Chen⁴, Rona S. Gerlinger^{1,2}, Jellert T. Gaublomme^{1,2}, Nir Yosef², Schraga Schwartz², Brian Fowler⁴, Suzanne Weisner⁴, Jing Wang⁴, Xiaohui Wang⁴, Ruhua Ding^{1,2}, Rakitma Raychowdhury², Nir Friedman⁵, Nir Hacohen^{1,2}, Hongkun Park^{1,2,3}, Andrew P. May⁴ & Aviv Regev^{3,7}



2015, 45,000 cells

Resource

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

Evan Z. Macosko^{1,9,10}, Anindita Basu^{1,5}, Rahul Satija^{4,10}, James Nemesh^{1,2,3}, Karthik Shekhar⁴, Melissa Goldman^{1,3}, Ray Trivedi⁴, Allison R. Dallas³, Nolan Kamitaki^{1,11}, Emily M. Martensack³, John J. Trombetta³, David A. Weitz^{1,10}, Joshua R. Sanee⁷, Alex K. Shalek^{1,2,3,10}, Aviv Regev^{1,2,3,10} and Steven A. McCarroll^{1,10}



2016, 200,000 cells

Resource

Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens

Atay Dizit^{1,10}, Oren Parnas^{1,10}, Biyu Li¹, Jenny Gao^{1,10}, Charles P. Fulia^{1,10}, Urvat Jetty-Amon¹, Nemanja D. Marjanovic^{1,10}, Danielle Dionne⁷, Tyler Burks¹, Rakitma Raychowdhury², Britt Adamson⁹, Thomas M. Norman³, Eric S. Lander^{1,10}, Jonathan S. Weissman^{1,10}, Nir Friedman^{1,10} and Aviv Regev^{1,10}

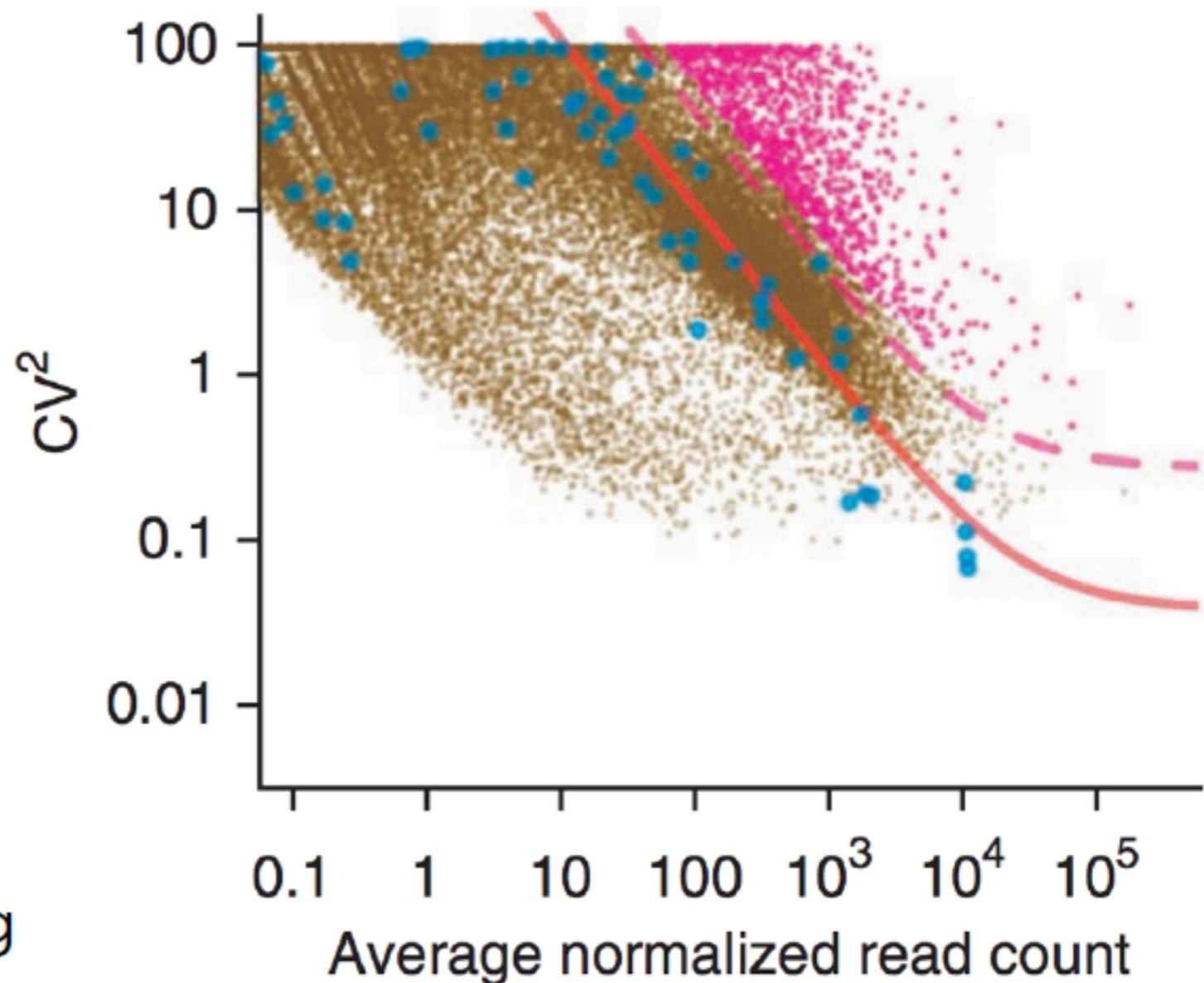
2017, 1.3 million cells (10X genomics)

1. Identifying relevant, “highly variable” features

First filter out,

- Lowly expressed genes
- “Housekeeping” genes

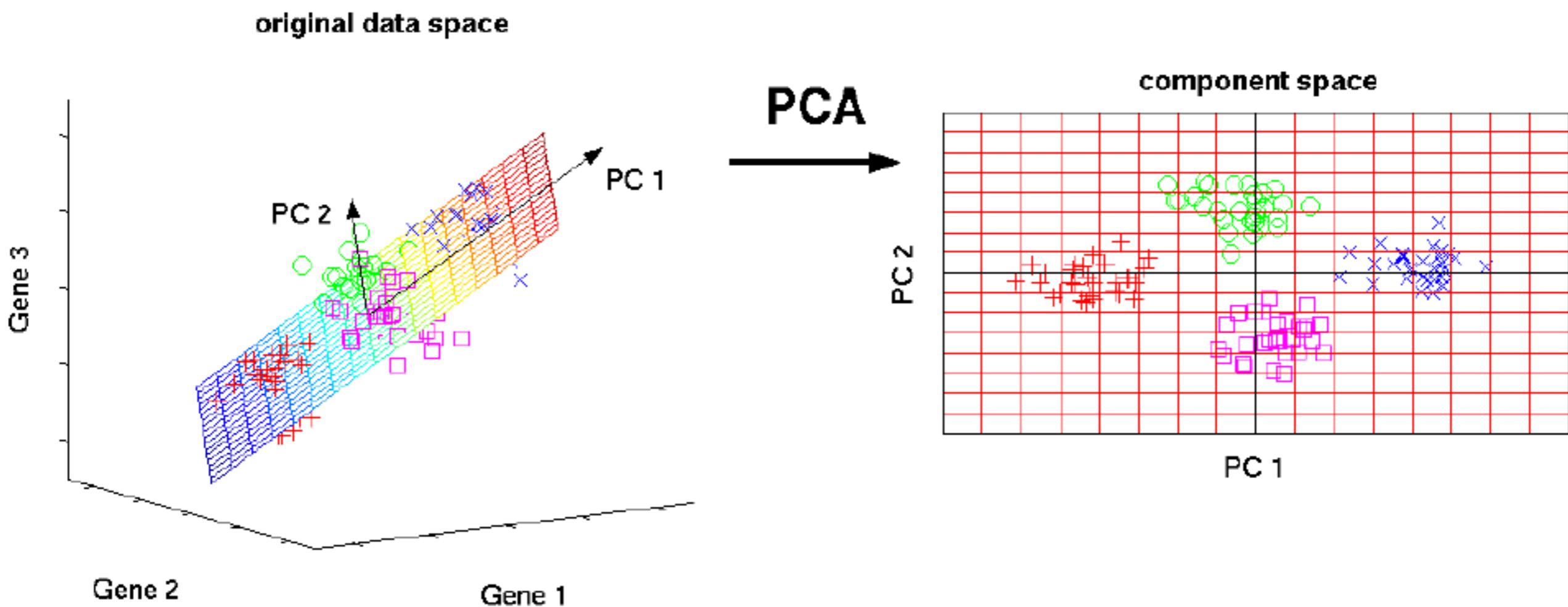
Typical practice to identify
“highly” variable genes is to
create a null model of statistical
variation based on housekeeping
or spike-in genes



Brennecke et al., *Nature Methods*, 2013

2. Dimensionality reduction

- Why? : Genes do not act independently, but as coregulatory “modules”. E.g. in a cell type, the activity of a handful of transcription factors might lead to the co-expression of hundreds of genes defining cell-identity
- Cells occupy a low dimensional manifold in gene-expression space defined by these modules



- Principal Component Analysis (PCA) is a **popular linear-method** to identify these modules

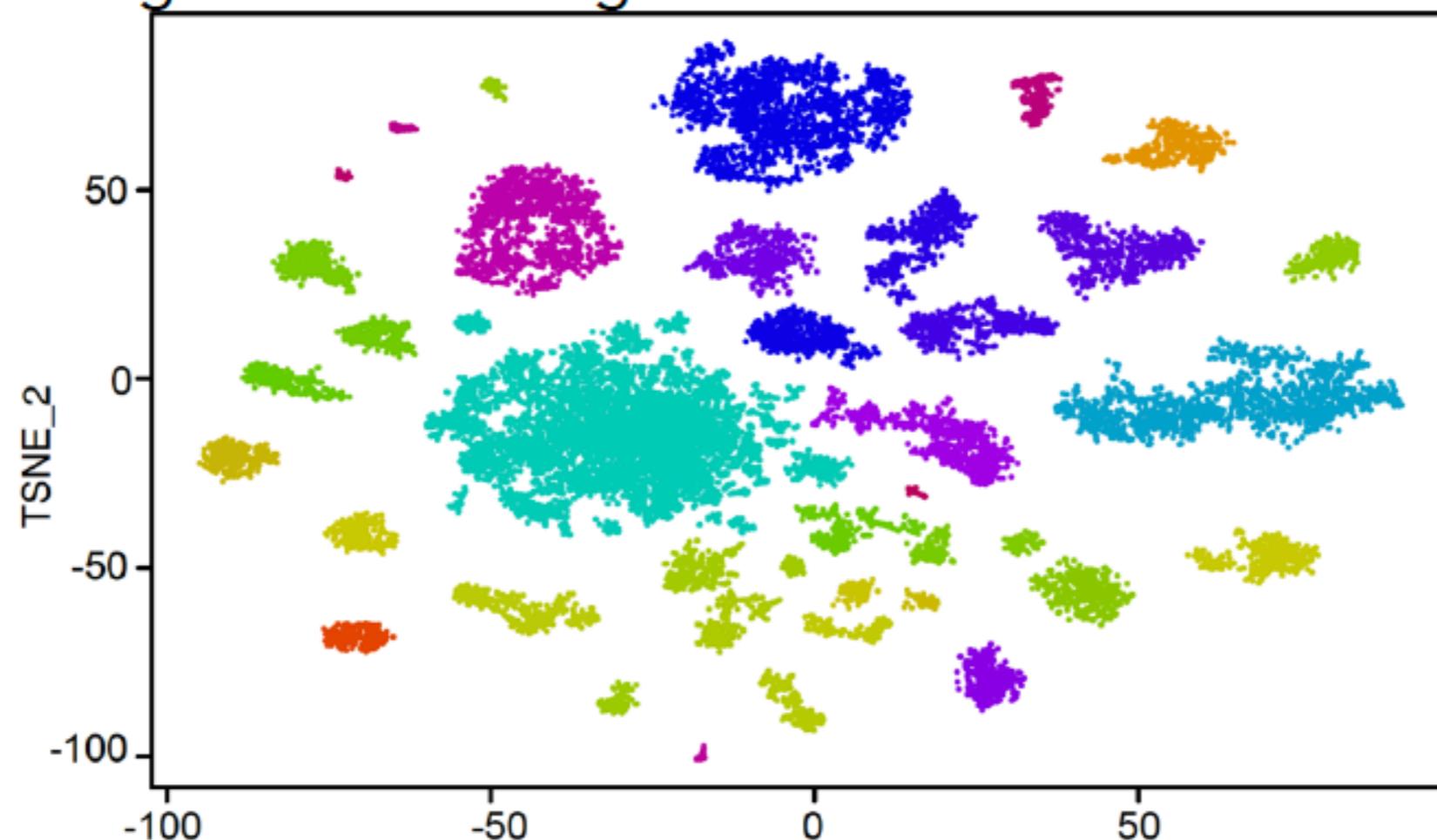
3. Visualization

t-distributed Stochastic Neighbor Embedding

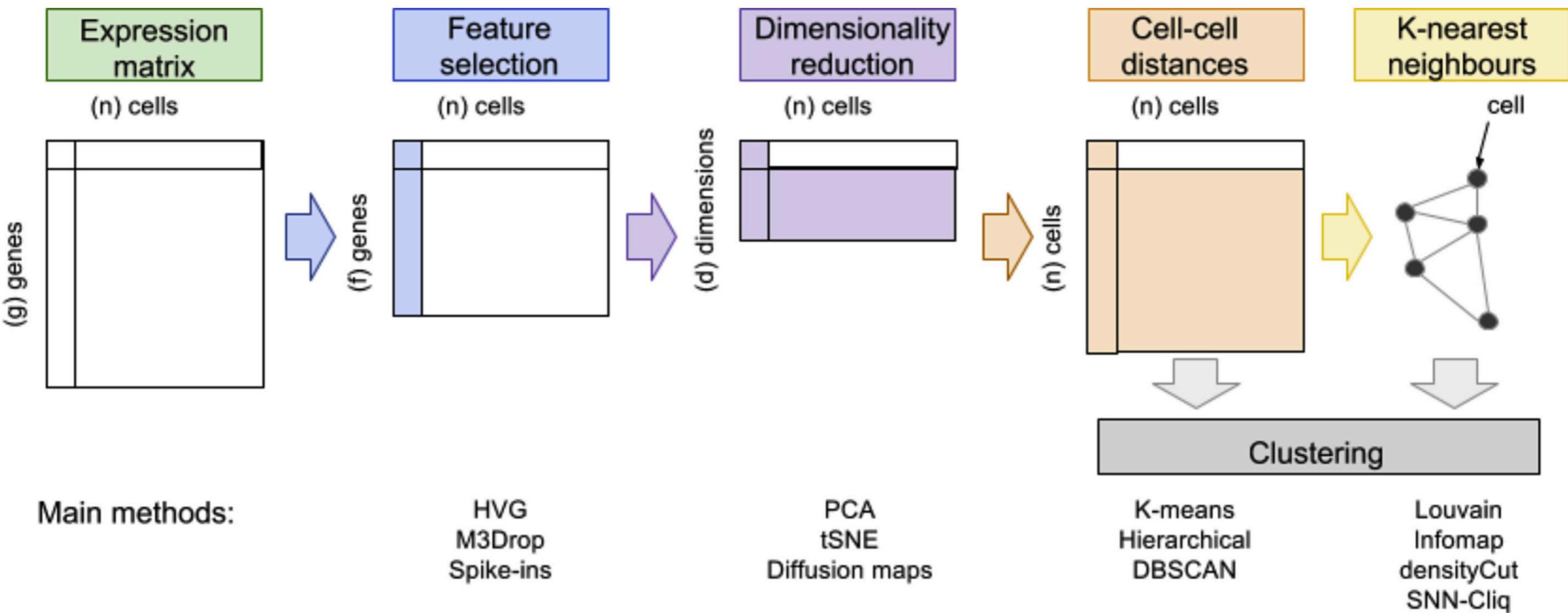
- A non-linear embedding method that preserves local distances between data-points in the low-dimensional space

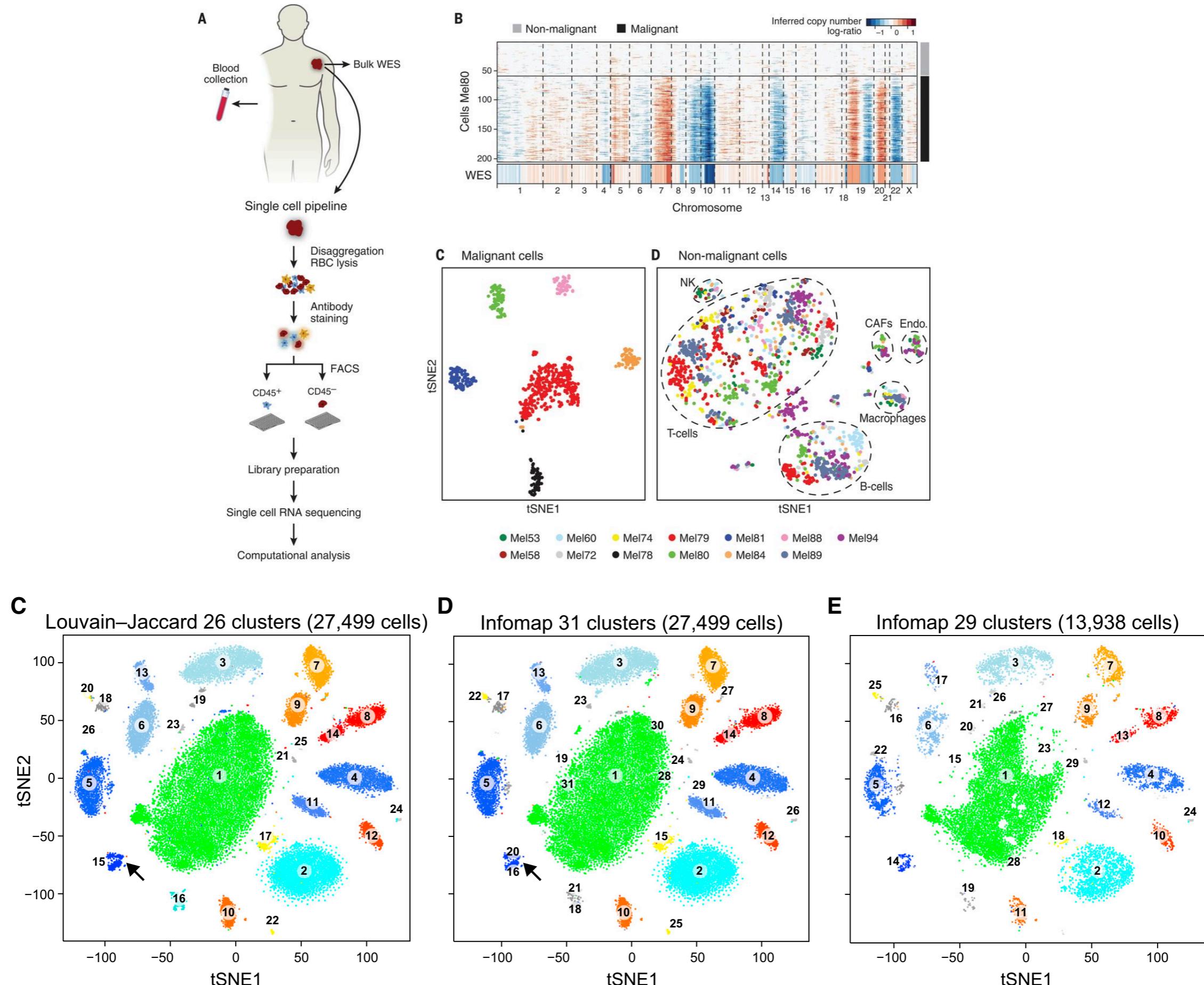
$$C = \sum_i KL(P_i || Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}},$$

- t-SNE embedding of 45,000 retinal neurons sequenced using Drop-seq and clustered using the DBScan algorithm

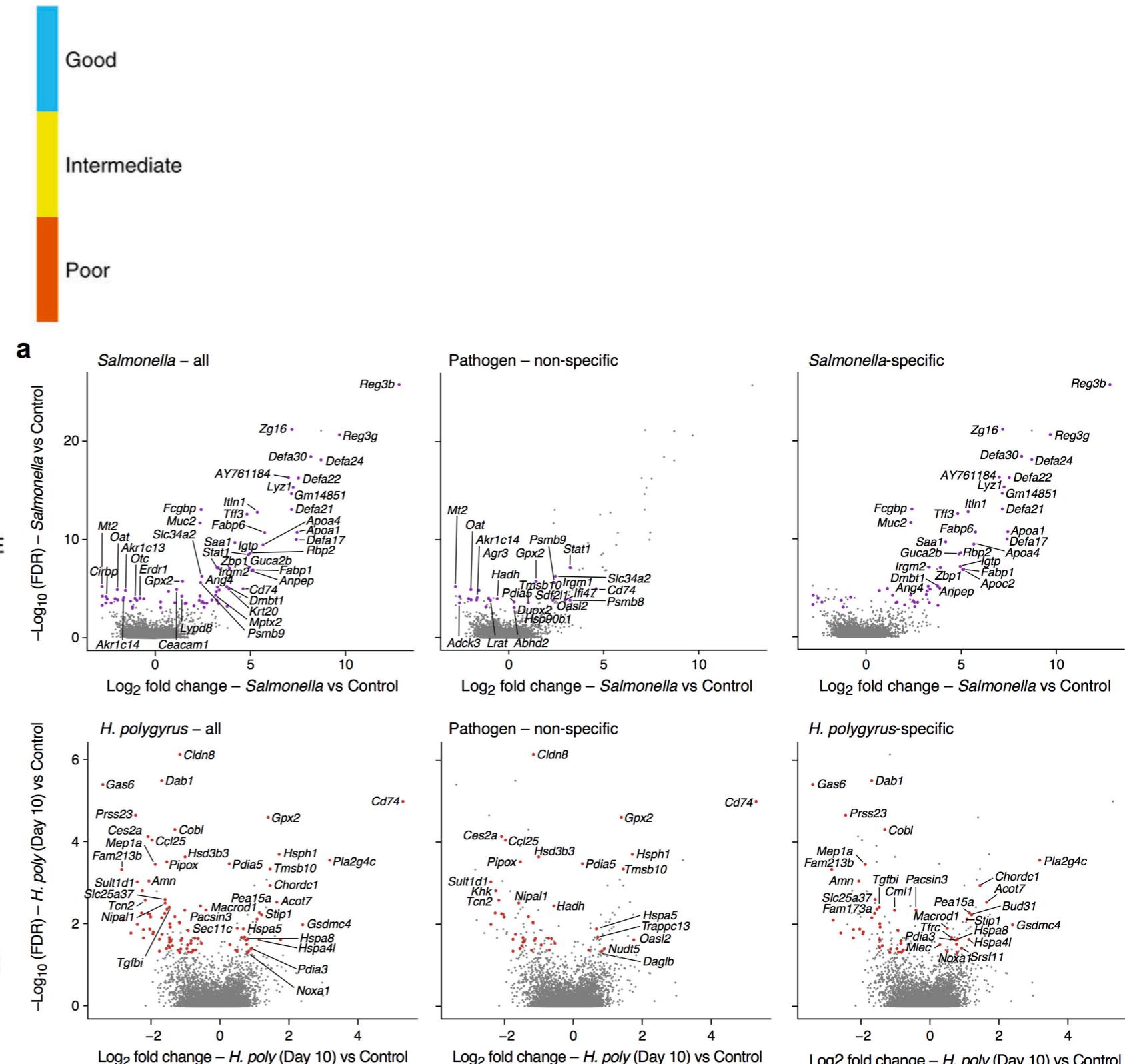
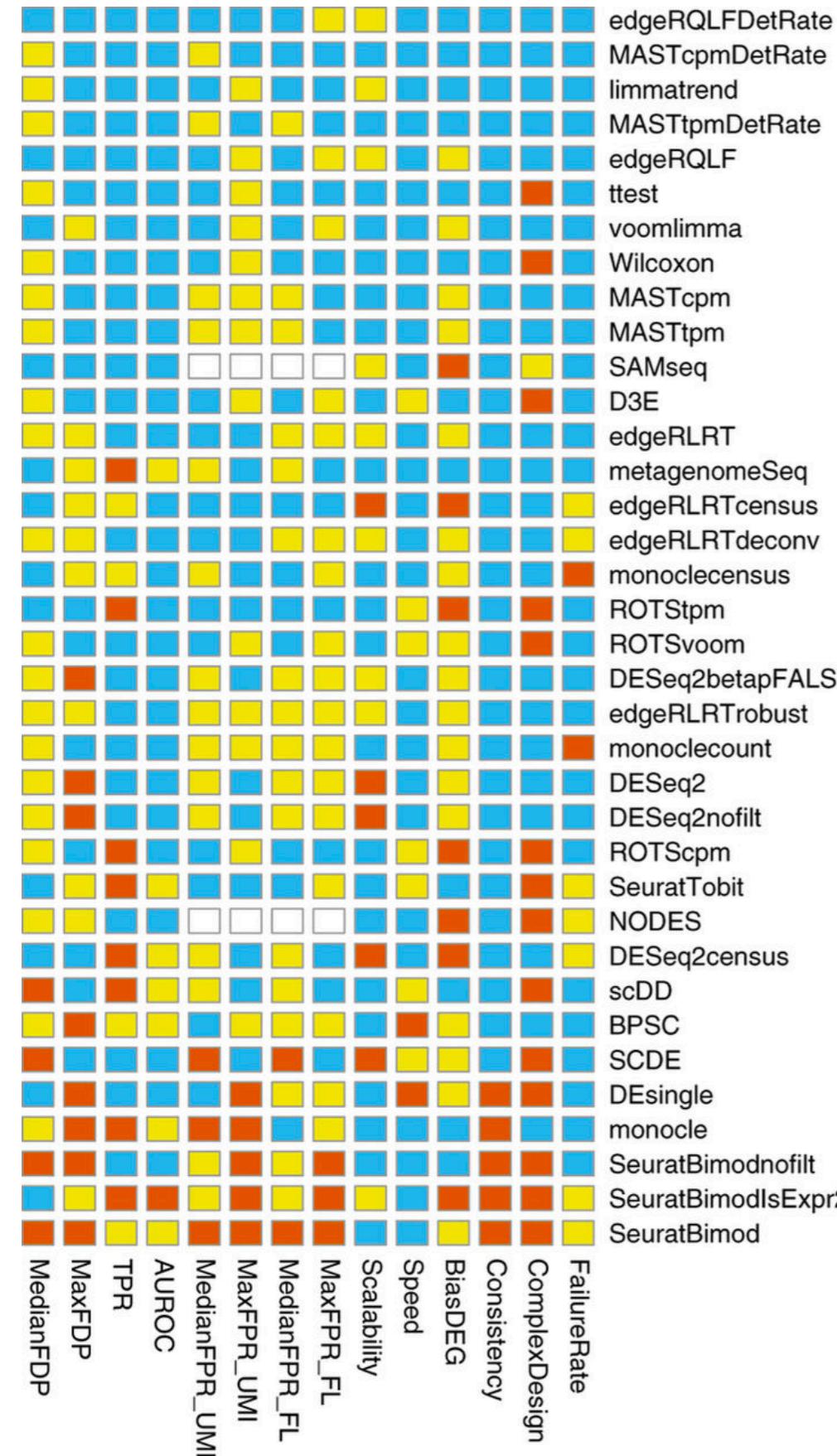


4. Clustering cells to identify cell-types

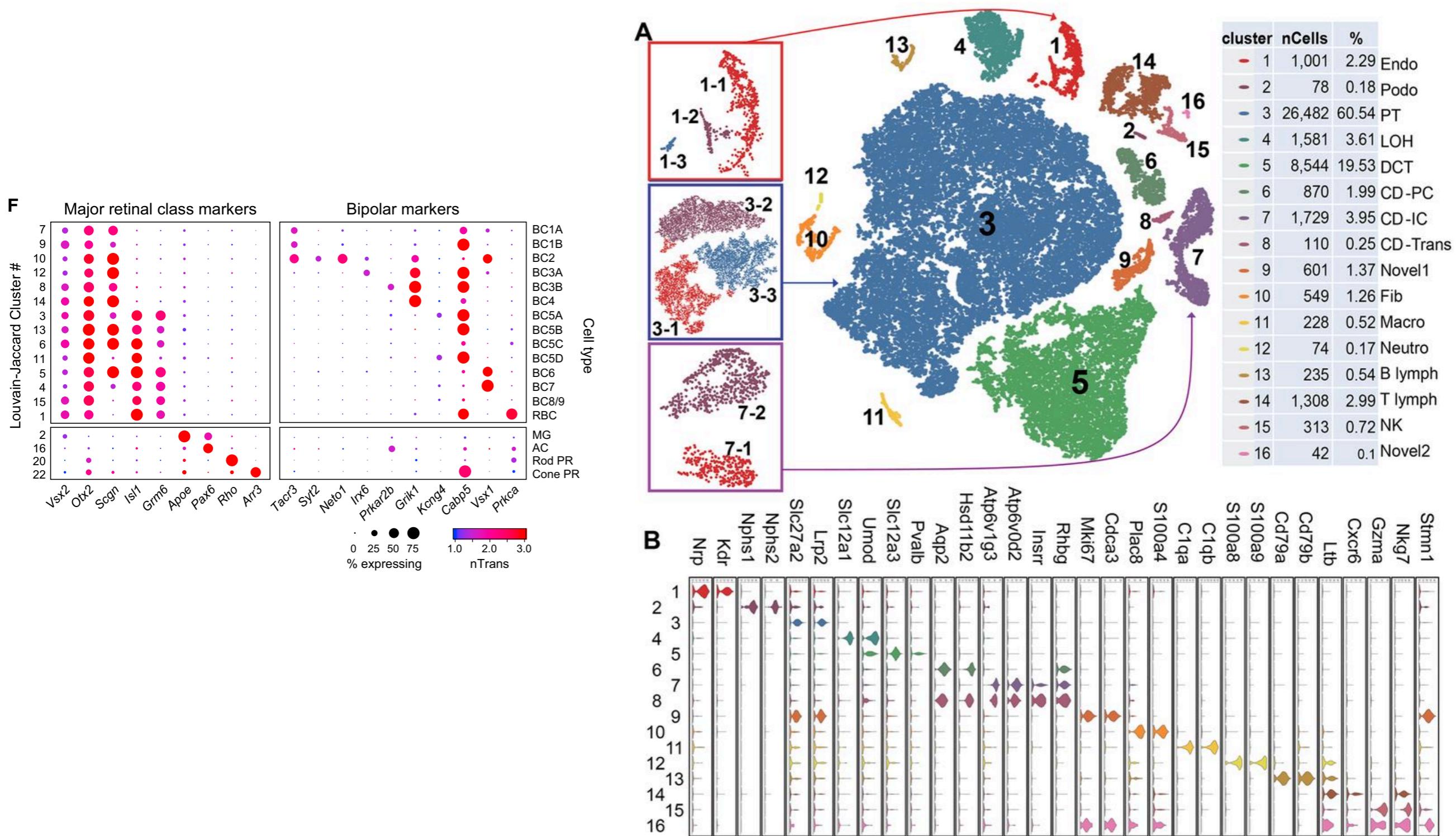




5. Assigning cell identity & comparing across conditions: Differential Expression Analysis



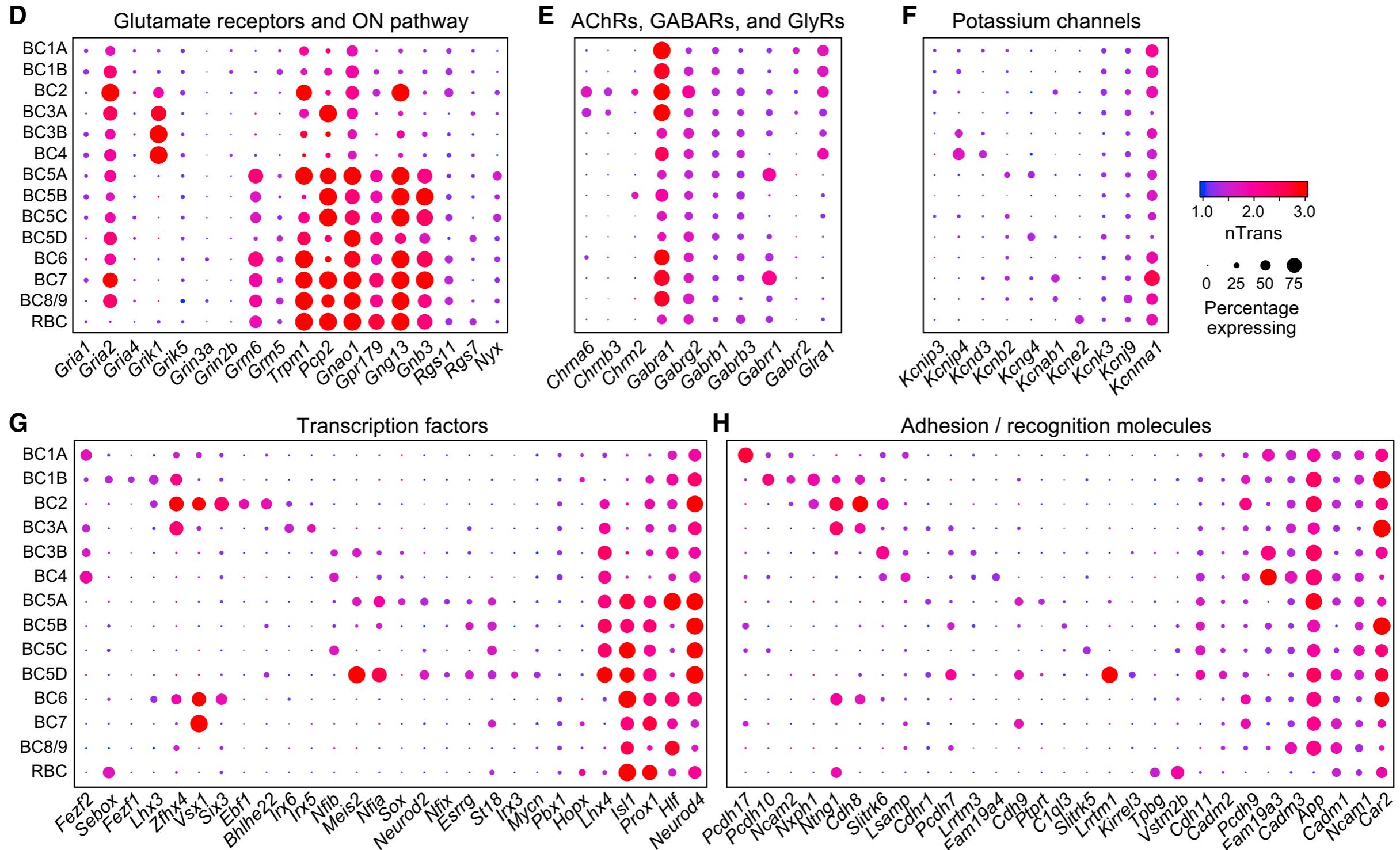
5. Assigning cell identity: Known marker genes



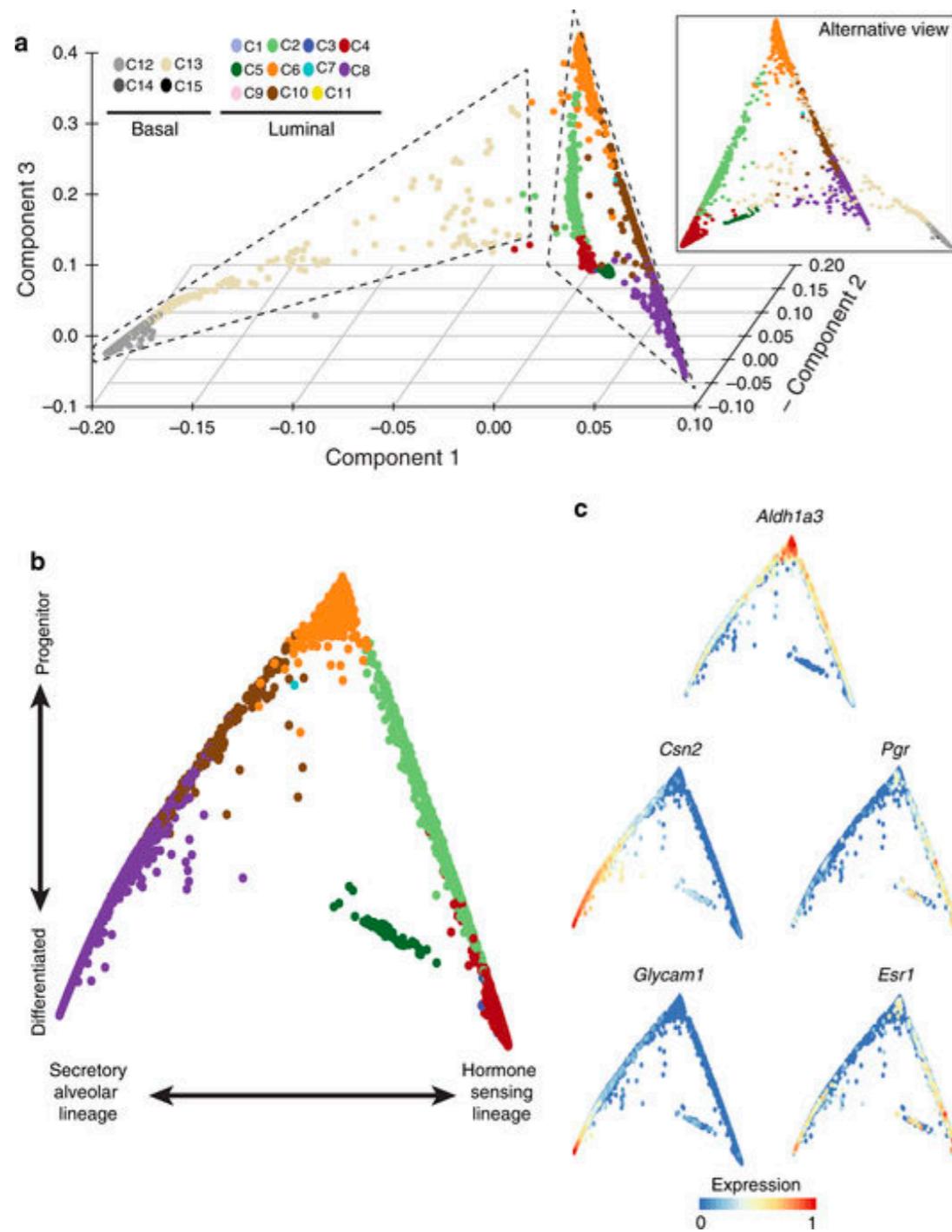
Shekhar et al. Cell 2016

Park and Shreshtha et al. Science 2018

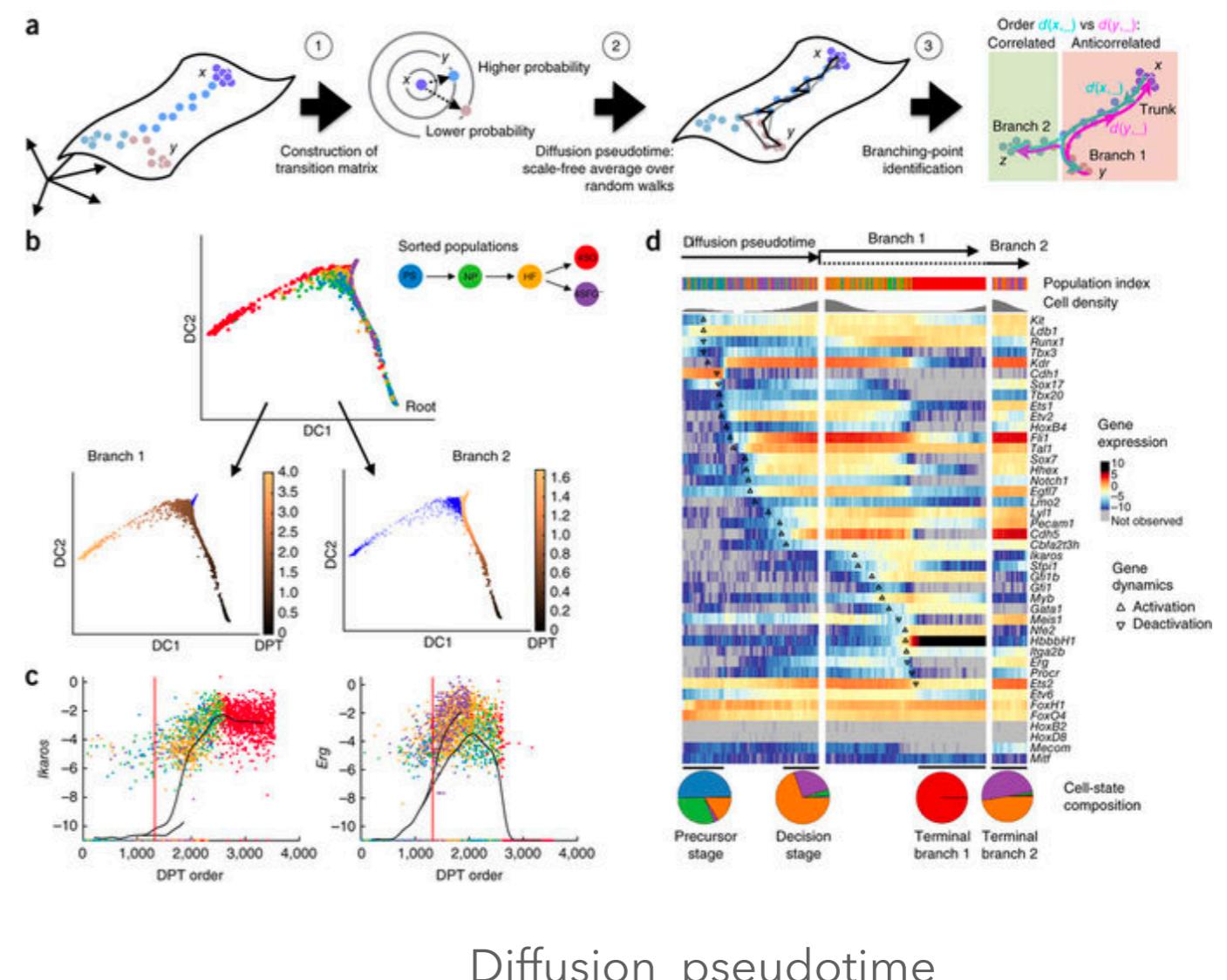
6. Functional annotation by pathway analysis and gene-set enrichment analysis



7. Trajectory inference



Diffusion Maps



Part V of V: Tools and resources



HUMAN
CELL
ATLAS

[Home](#) [HCA](#) [Areas of Impact](#) [News](#) [Publications](#) [Data Coordination](#) [Join HCA](#) [Contact](#)

MISSION

To create comprehensive reference maps of all human cells—the fundamental units of life—as a basis for both understanding human health and diagnosing, monitoring, and treating disease.



Human Cell Atlas Preview Datasets

The first single-cell sequencing datasets from the Human Cell Atlas are now available to the research community.

Census of Immune Cells¹

Profiling of immunocytes by single cell RNA-seq for understanding human health and disease.

Species	Homo sapiens
Organ	Umbilical cord blood and bone marrow
Method	10x
Cell count	~530,000 cells
File size	1.3 TB

DOWNLOADS

- [Raw Counts Matrix - Cord Blood](#)
- [Raw Counts Matrix - Bone Marrow](#)
- [Raw Counts Matrix - README](#)
- [Metadata Spreadsheet](#)
- [Primary Data Download Script](#)
Fastq and jsons. Additional instructions below.

Ischaemic Sensitivity of Human Tissue²

Assessment of ischaemic sensitivity of human spleen tissue by single cell RNA-seq.

Species	Homo sapiens
Organ	Spleen
Method	10x
Cell count	~2,000 cells
File size	14 GB

DOWNLOADS

- [Metadata Spreadsheet](#)
- [Primary Data Download Script](#)
Fastq and jsons. Additional instructions below.

Melanoma Infiltration of Stromal and Immune Cells³

Single cell RNA-seq of CD45+ and CD45- cells isolated from tumour and lymph nodes of a mouse model of melanoma.

Species	Mus musculus
Organ	Lymph node
Method	Smart-seq2
Cell count	6,639 cells
File size	380 GB

DOWNLOADS

- [Metadata Spreadsheet](#)
- [Primary Data Download Script](#)
Fastq and jsons. Additional instructions below.

All Preview Datasets include primary data (fastq and metadata). Additional analysis of the Preview Datasets will be made available here as donated by the data generators.

Broad Institute Single Cell Portal

Secure | https://portals.broadinstitute.org/single_cell

Single Cell Portal BETA

Help Sign In

Single Cell Portal BETA

Visualization portal for single cell RNA-seq data.

Now featuring 37 studies with 432,801 cells.

Browse Studies ?

Search Studies...

Most Recent Most Popular Reset Filters

Single nucleus RNA-seq of cell diversity in the adult mouse hippocampus (sNuc-Seq) ▾

View Study

Single nucleus RNA-seq of cell diversity in the adult mouse hippocampus. Habib N, Li Y, Heidenreich M, Swiech L, Avraham-David I, Trombetta J, Hession C, Zhang F, Regev A. Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science* 28 Jul 2016 DOI: 10.1126/science.aad7038 Contact: naomi@broadinstitute.org Single cell RNA-Seq provides rich information about cell types and states. However, it is difficult to capture rare dynamic processes, such as adult neurogenesis, because isolation of rare neurons from adult tissue is challenging and markers for each phase are limited. Here, we develop Div-Seq, which combines scalable single-nucleus RNA-Seq (sNuc-Seq) with pulse labeling of proliferating cells by EdU... (continued)

Retinal Bipolar Neuron Drop-seq ▾

View Study

Retinal Bipolar Neuron Drop-Seq Karthik Shekhar, Sylvain W. Lapan, Irene E. Whitney, Nicholas M. Tran, Evan Z. Macosko, Monika Kowalczyk, Xian Adiconis, Joshua Z. Levin, James Nemesh, Melissa Goldman, Steven A. McCarroll, Constance L. Cepko, Aviv Regev, Joshua R. Sanes. Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. *Cell*. Volume 166, Issue 5, p1308–1323.e30, 25 August 2016. DOI: http://dx.doi.org/10.1016/j.cell.2016.07.054 Contact: Karthik Shekhar at karthik@broadinstitute.org Patterns of gene expression can be used to characterize and classify neuronal types. It is challenging, however, to generate taxonomies that fulfill the essential criteria of being comprehensive, harmonizing with conventional classification schemes, and lacking superfluous subdivisions of genuine types. To address these challenges, we used massively parallel single-cell RNA profiling and optimized computational methods on a heterogeneous class of... (continued)

10X LucOS ▾

SATIJA LAB

HOME NEWS PEOPLE RESEARCH PUBLICATIONS SEURAT JOIN/CONTACT

SEURAT



R toolkit for single cell genomics

About

Install

Get Started

FAQ

Contact



Awesome Single-Cell

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

 README.md

awesome-single-cell

List of software packages (and the people developing these methods) for single-cell data analysis, including RNA-seq, ATAC-seq, etc. [Contributions welcome...](#)

Citation

DOI [10.5281/zenodo.1117763](https://doi.org/10.5281/zenodo.1117763)

Software packages

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.

More than ~80 software packages are available!

Data Repositories: JingleBells

jinglebells.bgu.ac.il

About Immune Non-immune



JingleBells

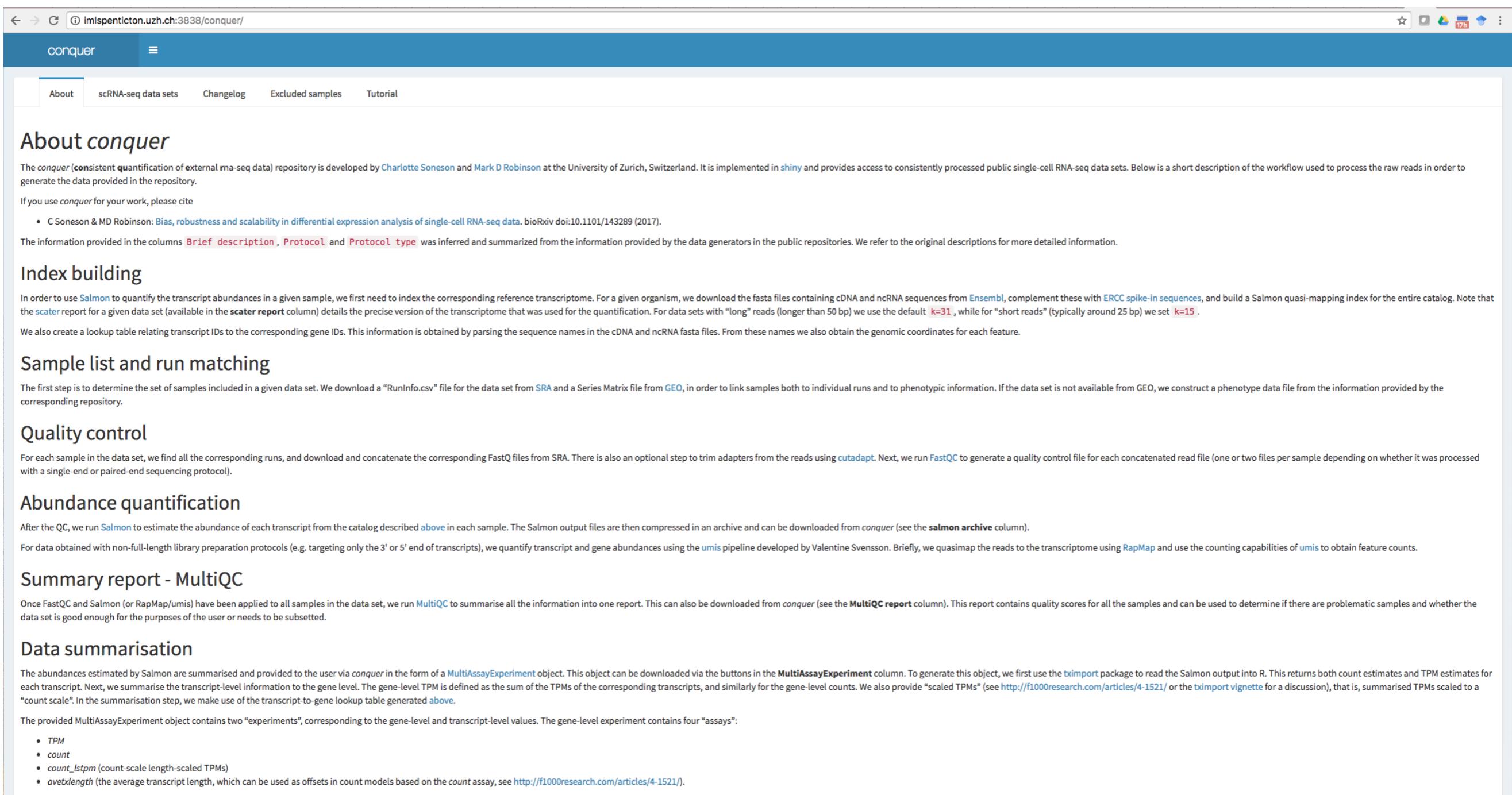
A repository of standardized single cell RNA-Seq datasets for analysis and visualization at the single cell level

Search

Immune Datasets (88 Entries)

Dataset ID	Dataset Description	Paper Title	Citation	Organism	Conditions (#)	Cell (#)	File Size (Gb)	BAM	BAI
GSE48968	Primary mouse dendritic cells (DCs) stimulated with three pathogenic components (LPS, PAM, PIC) at different time points after stimulation (0h, 1h, 2h, 4h, 6h)	Single-cell RNA-seq reveals dynamic paracrine control of cellular variation.	Shalek ... Regev Nature 2014	Mus musculus (mm10)	24	2412	1296		
GSE81682	Haematopoietic stem and progenitor cells	A single cell resolution map of mouse haematopoietic stem and progenitor cell differentiation.	Nestorowa ... Gottgens Blood 2016	Mus musculus (mm10)	3	1920	662.6		
E-MTAB-4388	transcriptional states of individual CD4+ T cells during blood-stage Plasmodium chabaudi infection in mice.	Single-cell RNA-seq and computational analysis using temporal mixture modeling resolves TH1/TFH fate bifurcation in malaria	Lonnberg ... Teichmann Sci Immunol. 2017	Mus musculus (mm10)	12	1598	566		
E-MTAB-4888	CD4+ T cells from young and old mice of CAST/EiJ and C57BL/6	Aging increases cell-to-cell transcriptional variability upon immune stimulation	Eling ... Odom Science 2017	Mus musculus (mm10)	11	1506	544.9		
GSE75478	Single-cell indexed RNA-Seq of human hematopoietic stem and progenitors	Human haematopoietic stem cell lineage commitment is a continuous process.	Velten ... Steinmetz Nat Cell Biol. 2017	Homo sapiens (hg38)	5	2414	377.9		

Data Repositories: Conquer



The screenshot shows a browser window with the URL imspenticton.uzh.ch:3838/conquer/. The page title is "conquer". The navigation bar includes links for "About", "scRNA-seq data sets", "Changelog", "Excluded samples", and "Tutorial". The main content area is titled "About conquer". It contains text about the repository's development by Charlotte Soneson and Mark D Robinson, its implementation in shiny, and access to public single-cell RNA-seq data sets. It also provides citation information and notes on data source inference.

About conquer

The *conquer* (**c**onsistent **q**uantification of **e**xternal rna-seq **d**ata) repository is developed by [Charlotte Soneson](#) and [Mark D Robinson](#) at the University of Zurich, Switzerland. It is implemented in [shiny](#) and provides access to consistently processed public single-cell RNA-seq data sets. Below is a short description of the workflow used to process the raw reads in order to generate the data provided in the repository.

If you use *conquer* for your work, please cite

- C Soneson & MD Robinson: [Bias, robustness and scalability in differential expression analysis of single-cell RNA-seq data](#). bioRxiv doi:10.1101/143289 (2017).

The information provided in the columns [Brief description](#), [Protocol](#) and [Protocol type](#) was inferred and summarized from the information provided by the data generators in the public repositories. We refer to the original descriptions for more detailed information.

Index building

In order to use [Salmon](#) to quantify the transcript abundances in a given sample, we first need to index the corresponding reference transcriptome. For a given organism, we download the fasta files containing cDNA and ncRNA sequences from [Ensembl](#), complement these with [ERCC spike-in sequences](#), and build a Salmon quasi-mapping index for the entire catalog. Note that the [scater report](#) for a given data set (available in the [scater report](#) column) details the precise version of the transcriptome that was used for the quantification. For data sets with "long" reads (longer than 50 bp) we use the default [k=31](#), while for "short reads" (typically around 25 bp) we set [k=15](#).

We also create a lookup table relating transcript IDs to the corresponding gene IDs. This information is obtained by parsing the sequence names in the cDNA and ncRNA fasta files. From these names we also obtain the genomic coordinates for each feature.

Sample list and run matching

The first step is to determine the set of samples included in a given data set. We download a "RunInfo.csv" file for the data set from [SRA](#) and a Series Matrix file from [GEO](#), in order to link samples both to individual runs and to phenotypic information. If the data set is not available from [GEO](#), we construct a phenotype data file from the information provided by the corresponding repository.

Quality control

For each sample in the data set, we find all the corresponding runs, and download and concatenate the corresponding FastQ files from [SRA](#). There is also an optional step to trim adapters from the reads using [cutadapt](#). Next, we run [FastQC](#) to generate a quality control file for each concatenated read file (one or two files per sample depending on whether it was processed with a single-end or paired-end sequencing protocol).

Abundance quantification

After the QC, we run [Salmon](#) to estimate the abundance of each transcript from the catalog described [above](#) in each sample. The [Salmon](#) output files are then compressed in an archive and can be downloaded from *conquer* (see the [salmon archive](#) column).

For data obtained with non-full-length library preparation protocols (e.g. targeting only the 3' or 5' end of transcripts), we quantify transcript and gene abundances using the [umis](#) pipeline developed by Valentine Svensson. Briefly, we quasimap the reads to the transcriptome using [RapMap](#) and use the counting capabilities of [umis](#) to obtain feature counts.

Summary report - MultiQC

Once [FastQC](#) and [Salmon](#) (or [RapMap/umis](#)) have been applied to all samples in the data set, we run [MultiQC](#) to summarise all the information into one report. This can also be downloaded from *conquer* (see the [MultiQC report](#) column). This report contains quality scores for all the samples and can be used to determine if there are problematic samples and whether the data set is good enough for the purposes of the user or needs to be subsetted.

Data summarisation

The abundances estimated by [Salmon](#) are summarised and provided to the user via *conquer* in the form of a [MultiAssayExperiment](#) object. This object can be downloaded via the buttons in the [MultiAssayExperiment](#) column. To generate this object, we first use the [tximport](#) package to read the [Salmon](#) output into R. This returns both count estimates and TPM estimates for each transcript. Next, we summarise the transcript-level information to the gene level. The gene-level TPM is defined as the sum of the TPMs of the corresponding transcripts, and similarly for the gene-level counts. We also provide "scaled TPMs" (see <http://f1000research.com/articles/4-1521/> or the [tximport vignette](#) for a discussion), that is, summarised TPMs scaled to a "count scale". In the summarisation step, we make use of the transcript-to-gene lookup table generated [above](#).

The provided [MultiAssayExperiment](#) object contains two "experiments", corresponding to the gene-level and transcript-level values. The gene-level experiment contains four "assays":

- TPM
- count
- count_lstpm (count-scale length-scaled TPMs)
- avetxlength (the average transcript length, which can be used as offsets in count models based on the count assay, see <http://f1000research.com/articles/4-1521/>).