UC DAVIS HEALTH

An Overview of Single Cell Sequencing, its Applications and Related Campus Resources

CLINICAL AND TRANSLATIONAL SCIENCE CENTER

Gerald Quon, Ph.D.

Assistant Professor

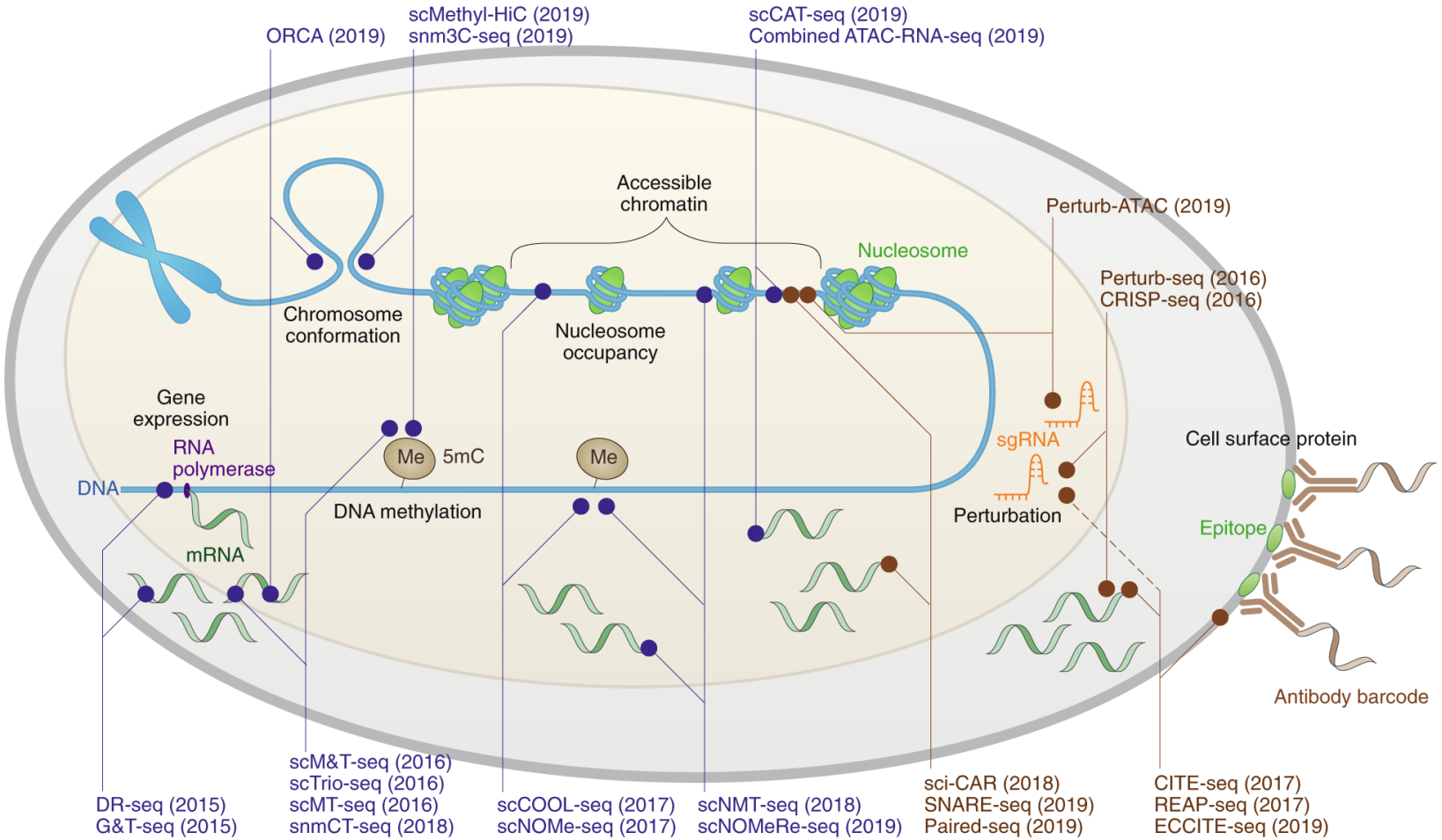
Department of Molecular and Cellular Biology, Genome Center

UC Davis

The UC Davis CTSC receives support from the NIH National Center for Advancing Translational Sciences (award TR001860).

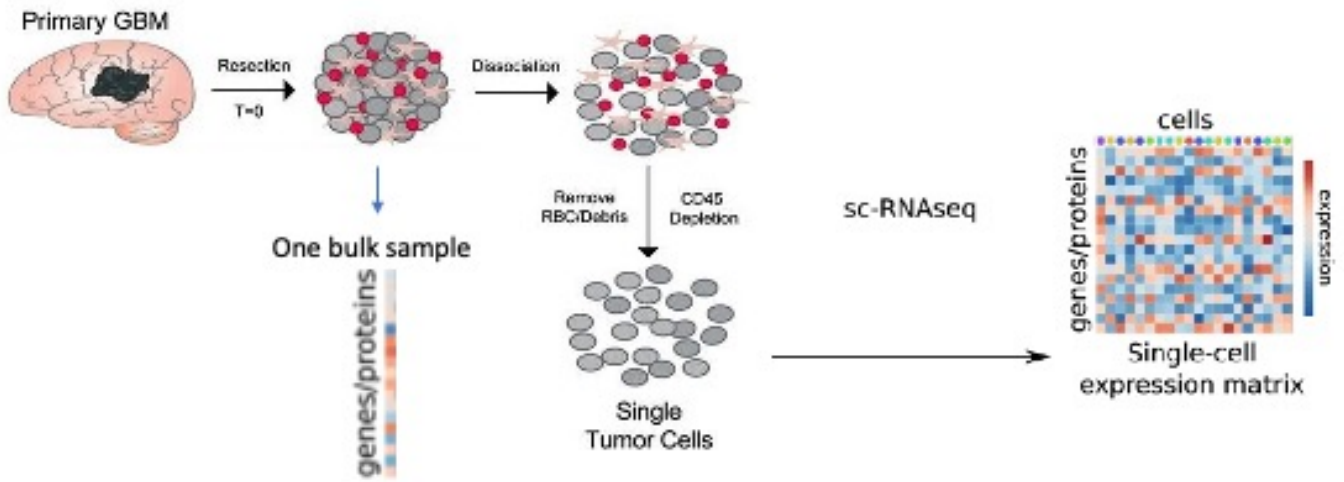
Overview

Measuring biology at the single cell level



Zhu, etc., *Nature Methods*, 2020, https://www.nature.com/articles/s41592-019-0691-5

Bulk versus single cell sequencing



- "Bulk" assays are named because you accumulate measurements over a large number of cells (millions)
- Measurements are an average (gene expression, TF binding, methylation, etc.) of many cells.
- The bulk measurement ignores the inter--cellular heterogeneities:
  - Different cell types.
  - Biological variation within the same cell type

Cell identity and function are fluid concepts

REVIEW

nature biotechnology

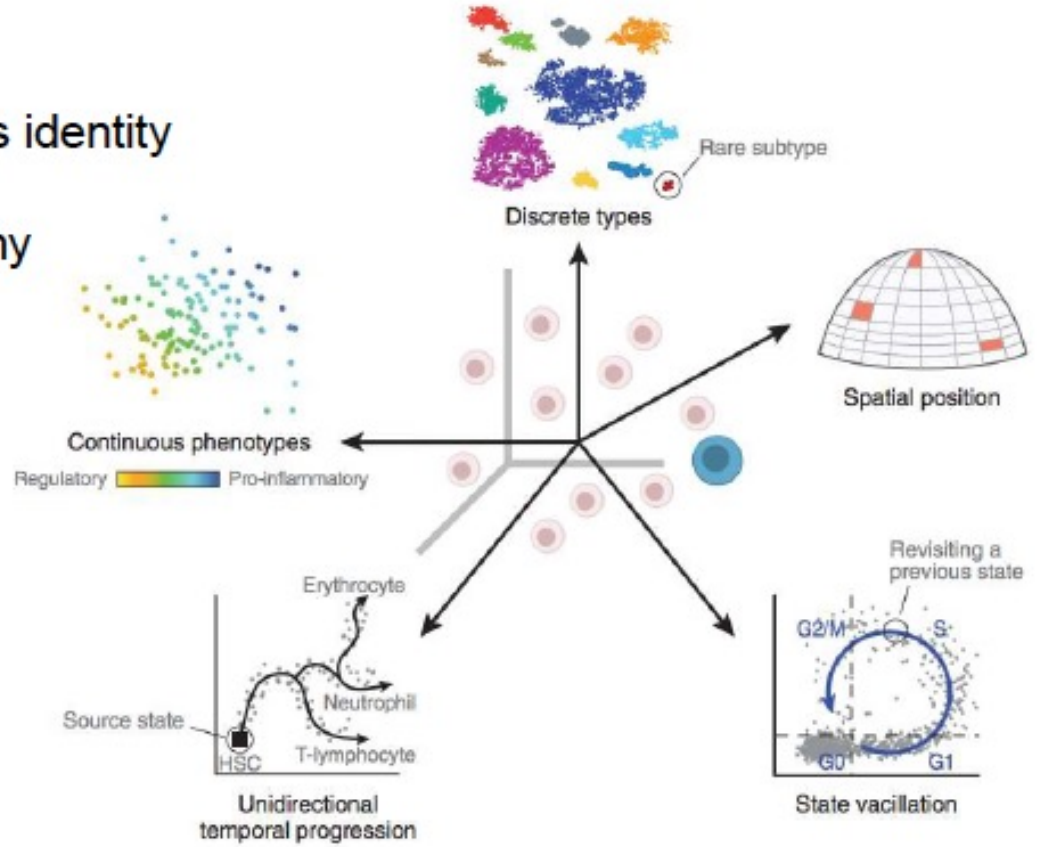
Revealing the vectors of cellular identity with single-cell genomics

Allon Wagner[1], Aviv Regev[2,3,5] & Nir Yosef[1,4,5]

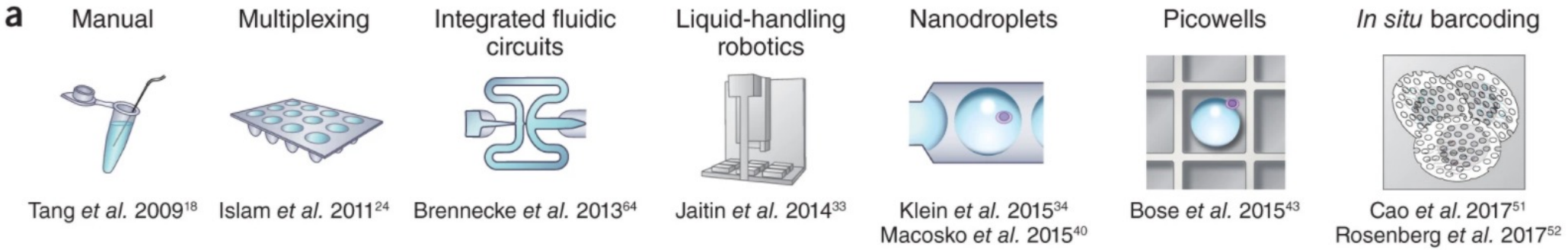
Multiple factors shape a cell's identity

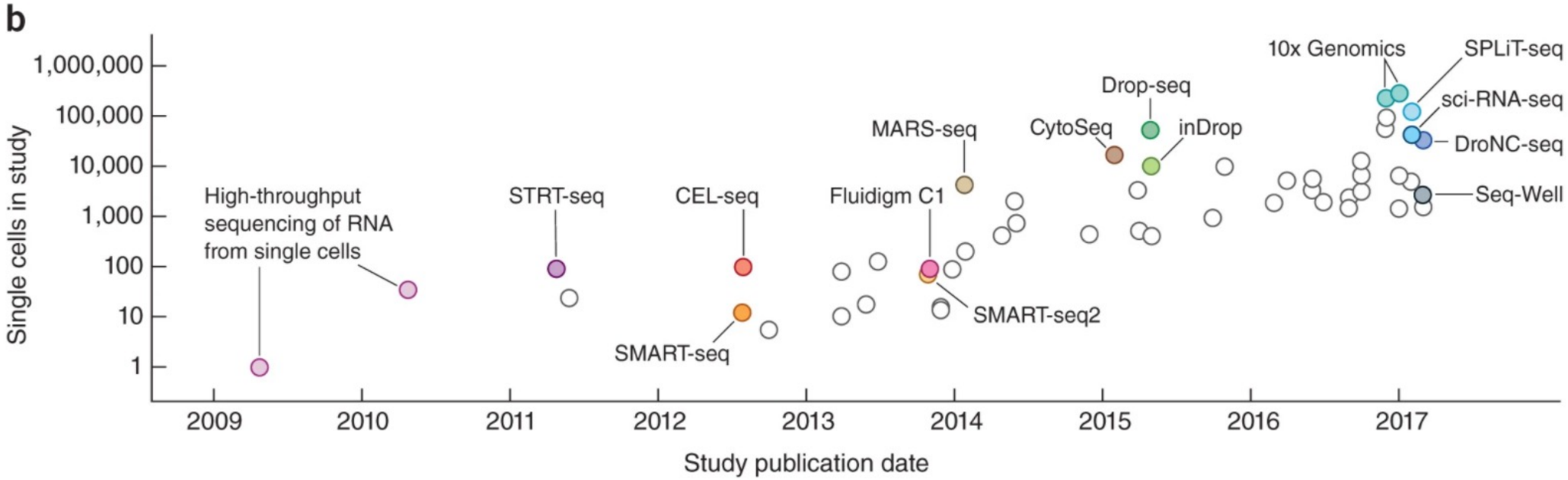
- Membership in a taxonomy of cell types
- Simultaneous time-dependent processes
- Response to the environment
- Spatial positioning

A cell participates in multiple cell contexts.



Single cell technologies are still under active development





Svensson, etc., 2018, Nature Protocols https://www.nature.com/articles/nprot.2017.149

Many technologies for single cell sequencing

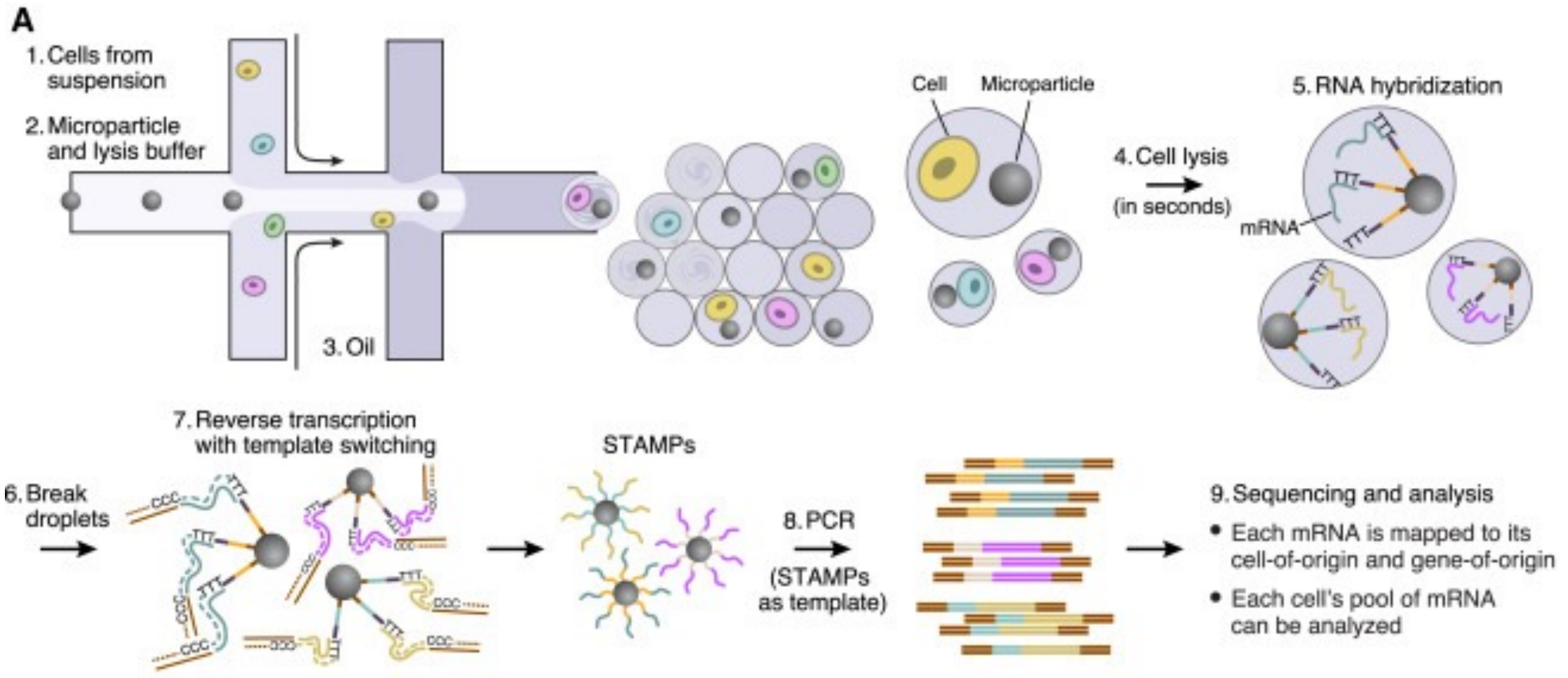
- Main distinction here: **Full length transcripts** (reads come from all regions of the transcript) versus **3' tagging** (sequence only 3' end of transcripts)
- Comparison: GTEx (one of the biggest producers of genotype and RNA samples from humans) provides 100's of samples at most per tissue

Table 1 Brief overview of scRNA-seq approaches

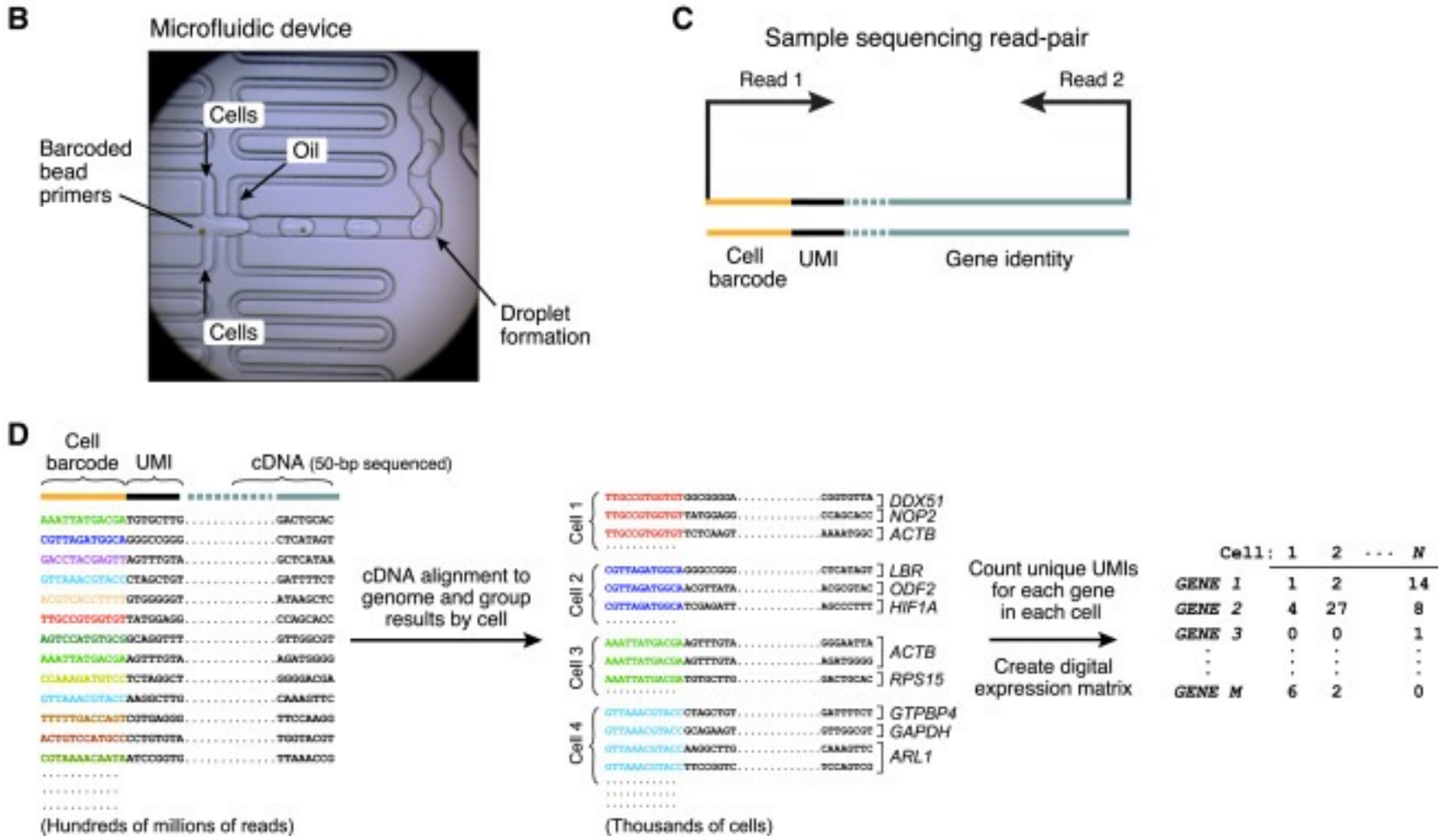
Protocol example	*C1 (SMARTer)*	*Smart-seq2*	*MATQ-seq*	*MARS-seq*	*CEL-seq*	*Drop-seq*	*InDrop*	*Chromium*	*SEQ-well*	*SPLIT-seq*
Transcript data	Full length	Full length	Full length	3'-end counting	3'-end counting	3'-end counting	3'-end counting	3'-end counting	3'-end counting	3'-end counting
Platform	Microfluidics	Plate-based	Plate-based	Plate-based	Plate-based	Droplet	Droplet	Droplet	Nanowell array	Plate-based
Throughput (number of cells)	10^2–10^3	10^2–10^3	10^2–10^3	10^2–10^3	10^2–10^3	10^3–10^4	10^3–10^4	10^3–10^4	10^3–10^4	10^3–10^5
Typical read depth (per cell)	10^6	10^6	10^6	10^4–10^5	10^4–10^5	10^4–10^5	10^4–10^5	10^4–10^5	10^4–10^5	10^4
Reaction volume	Nanoliter	Microliter	Microliter	Microliter	Nanoliter	Nanoliter	Nanoliter	Nanoliter	Nanoliter	Microliter

https://genomemedicine.biomedcentral.com/track/pdf/10.1186/s13073-017-0467-4

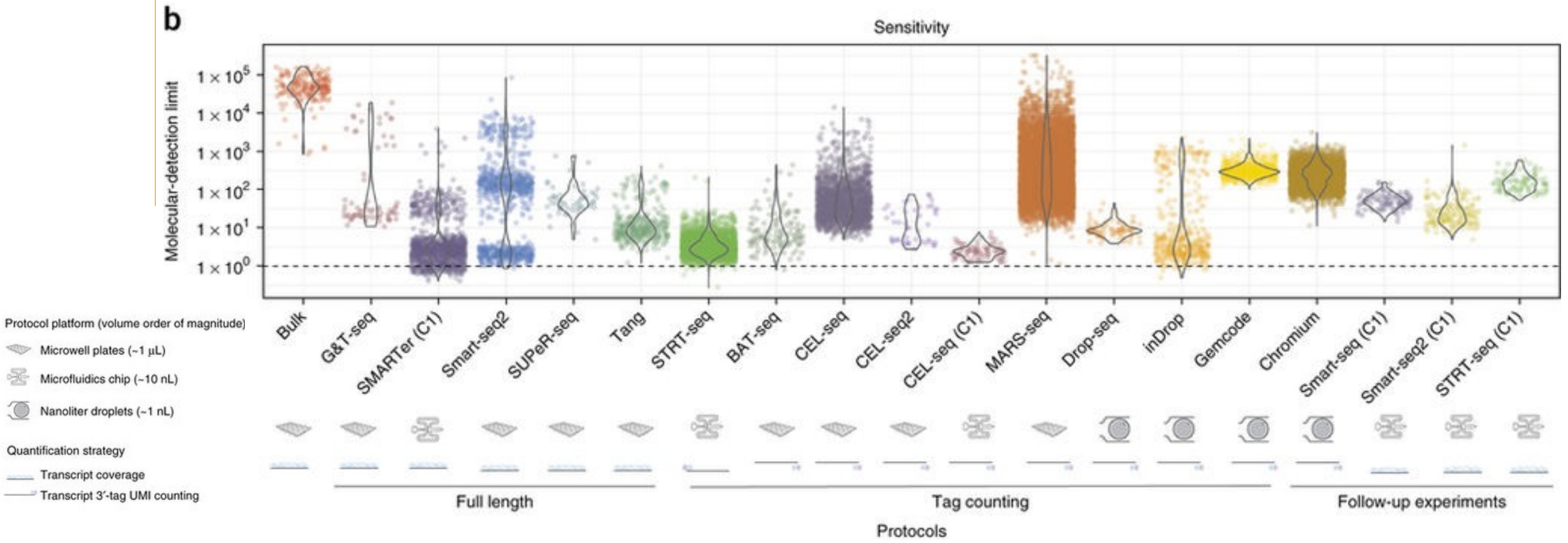
10x Chromium



10x Chromium

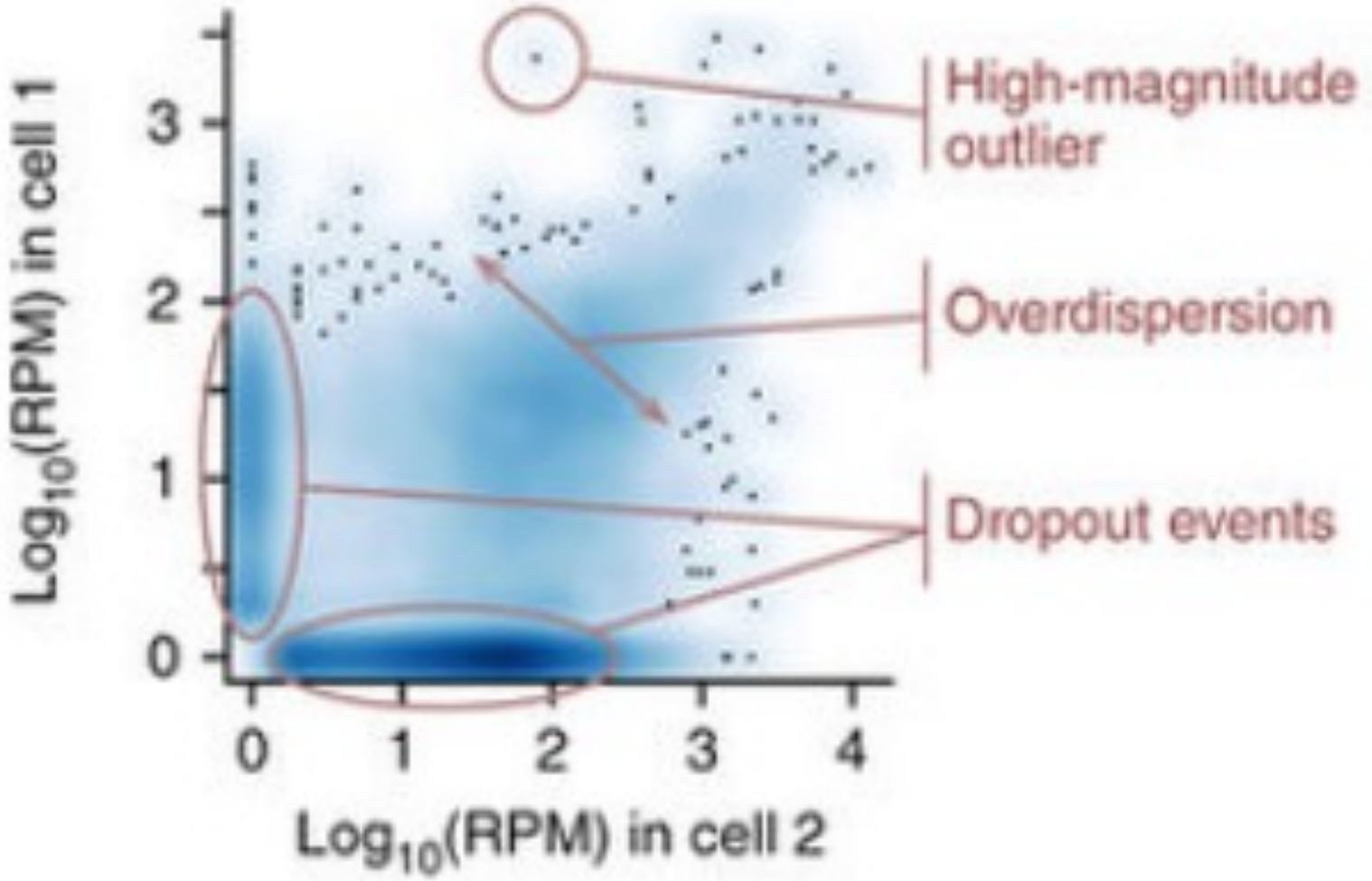


Not all technologies are created equal

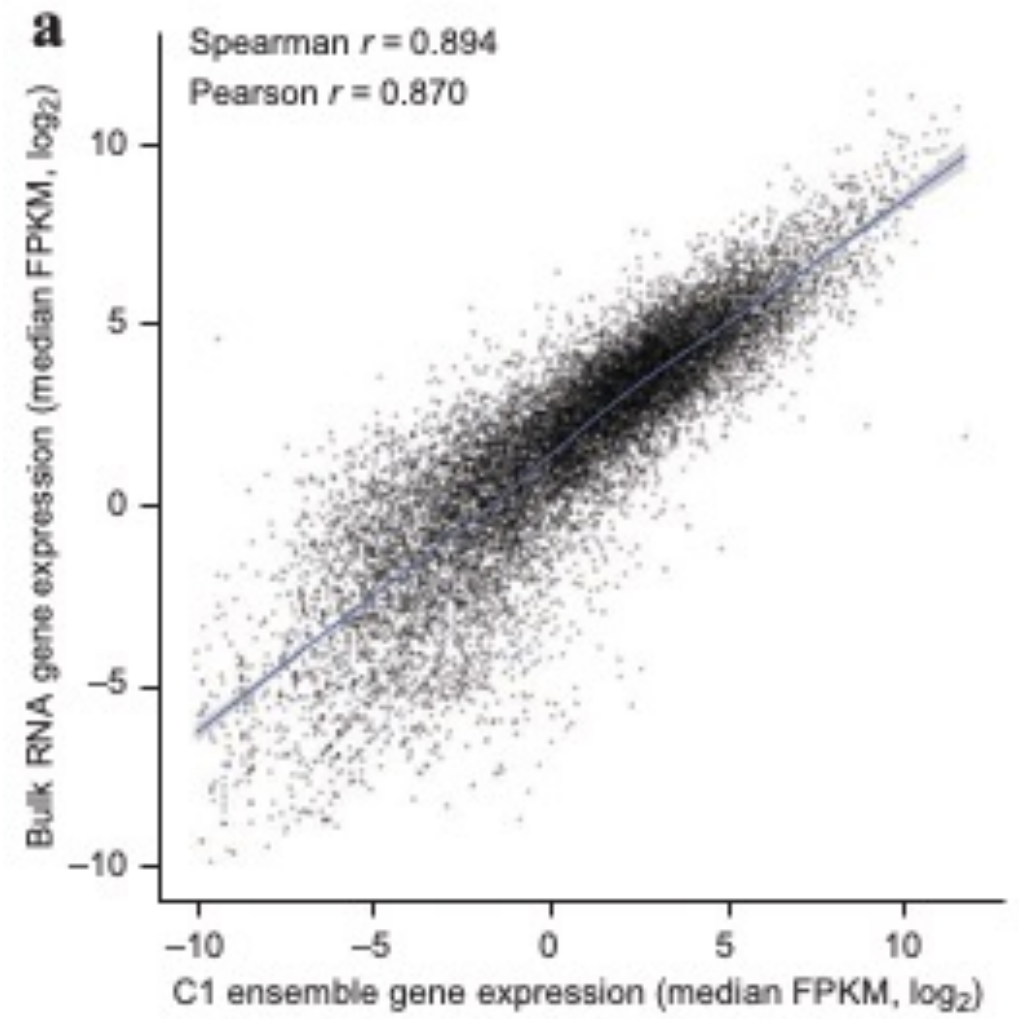


https://www.nature.com/articles/nmeth.4220/figures/4

scRNA-seq data has much more technical noise than bulk RNA-seq



Still, scRNA-seq and bulk RNA-seq broadly agree



There are many sources of biologically-relevant noise as well

Technical variation

- Batch effect
- Library quality
- Cell-specific capture efficiency
- Amplification bias

Allele-intrinsic variation

- Bursts of transcription
  - Stochastic initiation
  - Stochastic duration
- Varying rates of RNA processing

Allele-extrinsic variation (cell types and states)

- Fixed cell identity
  - Discrete
  - Continuous
- Temporal progression/oscillation
- Spatial location
  - Niche environments

Basic applications

Single-cell analyses define a continuum of cell state and composition changes in the malignant transformation of polyps to colorectal cancer

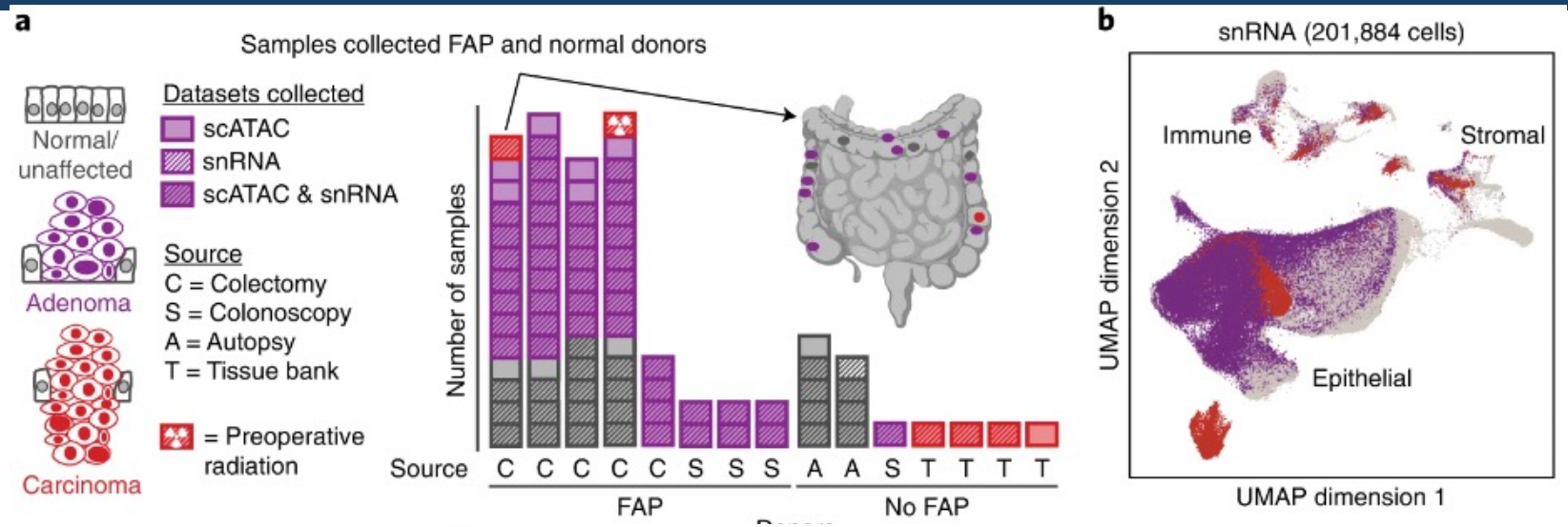
Winston R. Becker, Stephanie A. Nevins, Derek C. Chen, Roxanne Chiu, Aaron M. Horning, Tuhin K. Guha, Rozelle Laquindanum, Meredith Mills, Hassan Chaib, Uri Ladabaum, Teri Longacre, Jeanne Shen, Edward D. Esplin, Anshul Kundaje, James M. Ford, Christina Curtis, Michael P. Snyder ✉ & William J. Greenleaf ✉

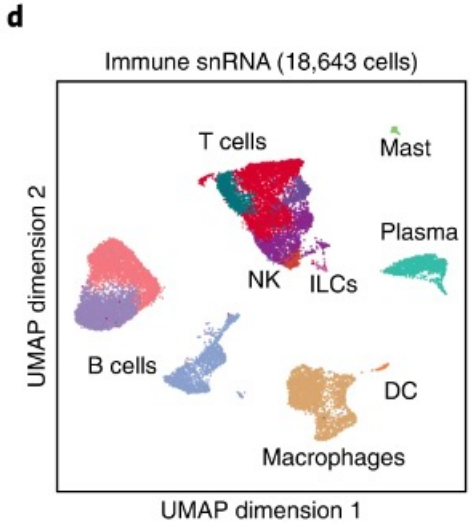
Nature Genetics **54**, 985–995 (2022) | Cite this article

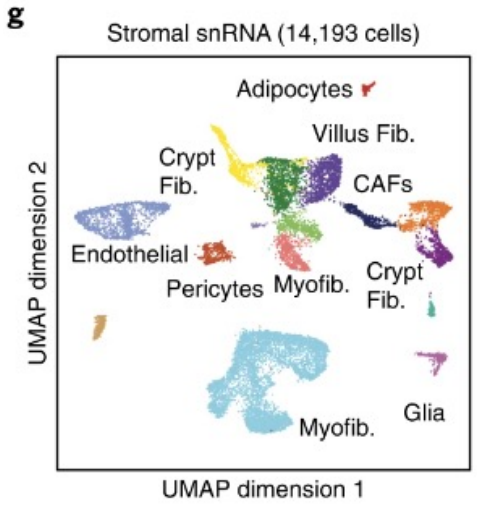
https://www.nature.com/articles/s41588-022-01088-x

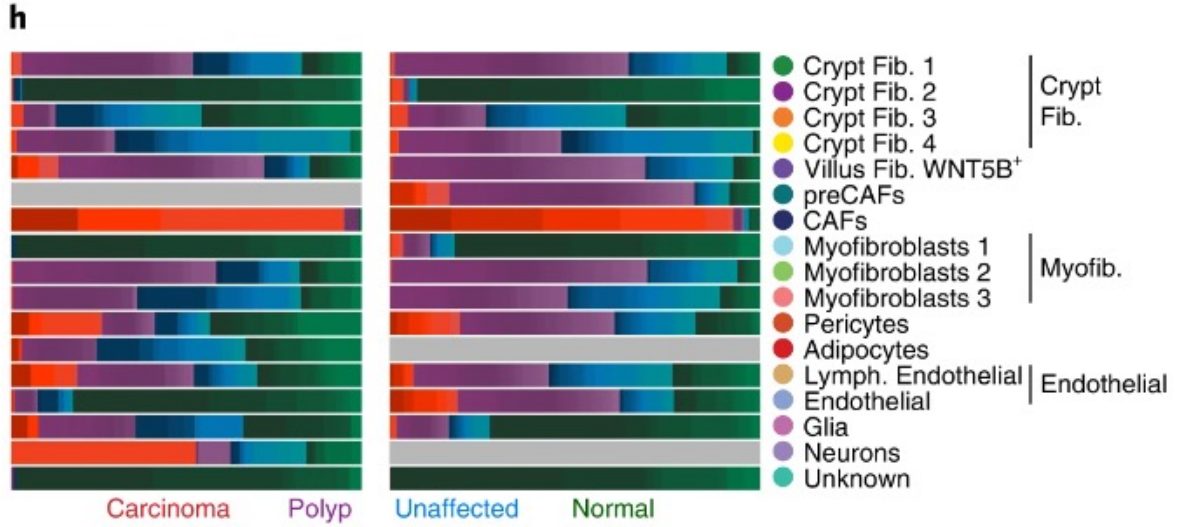
Cell census (cell type identification, clustering)

- Sequenced 48 polyps, 27 normal, 6 CRCs
- 1-10k cells/sample

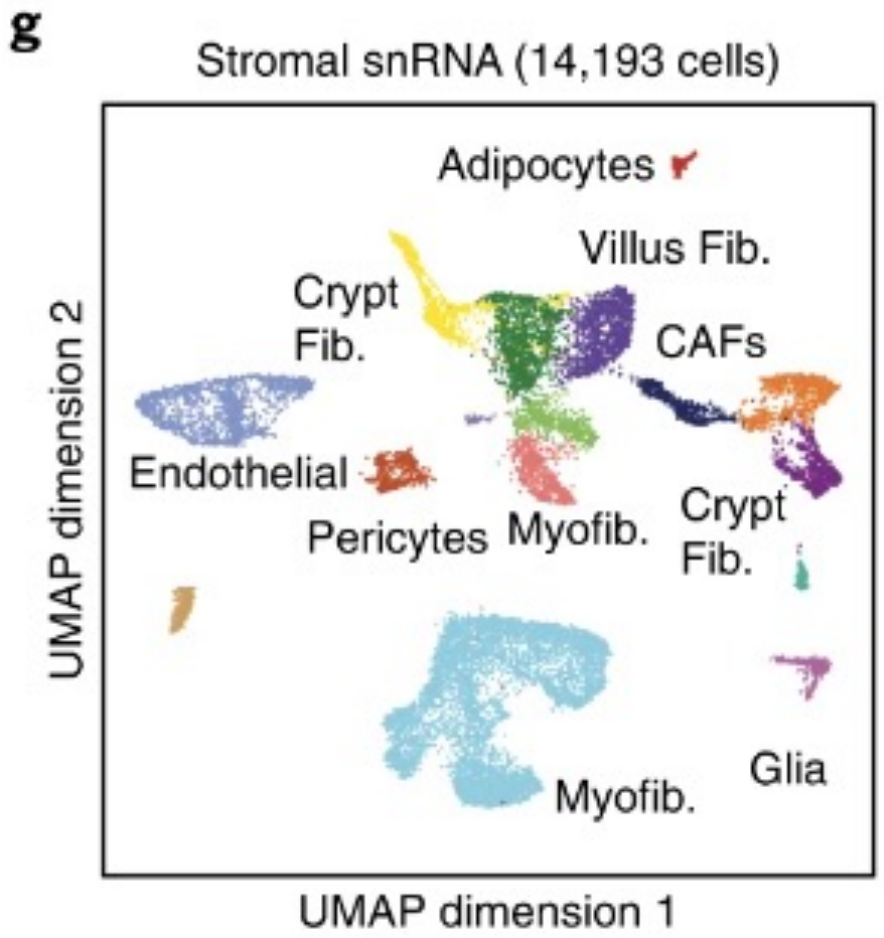


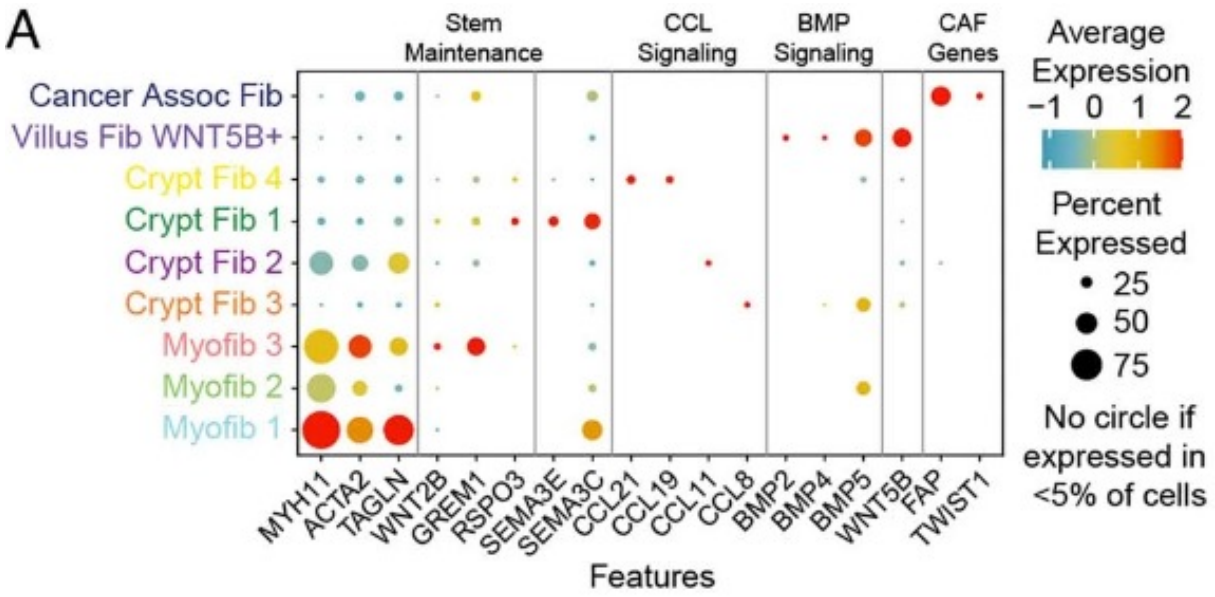




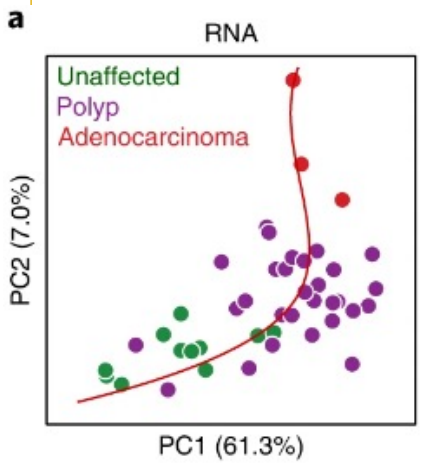


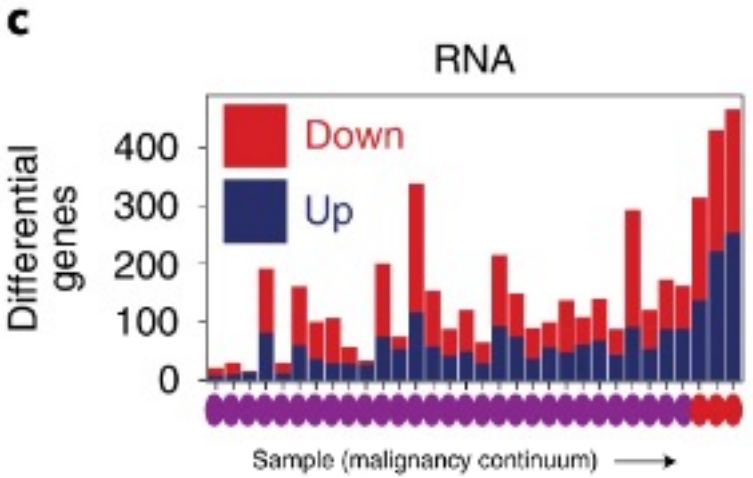
Gene marker identification

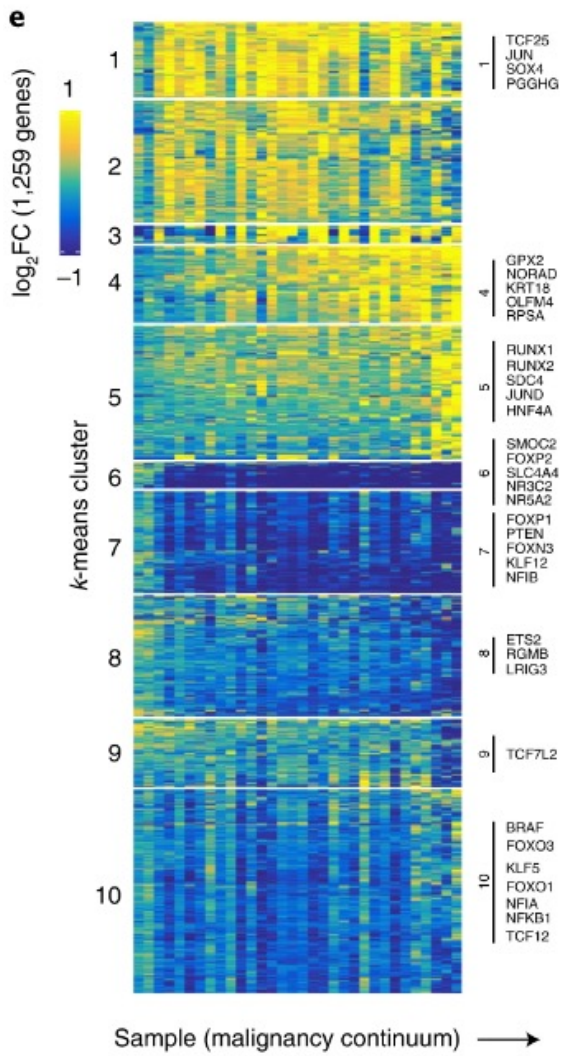


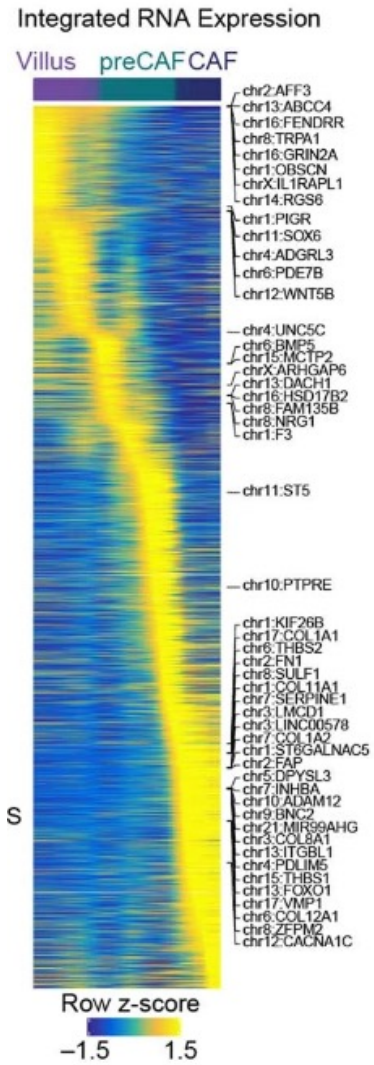


Trajectory inference

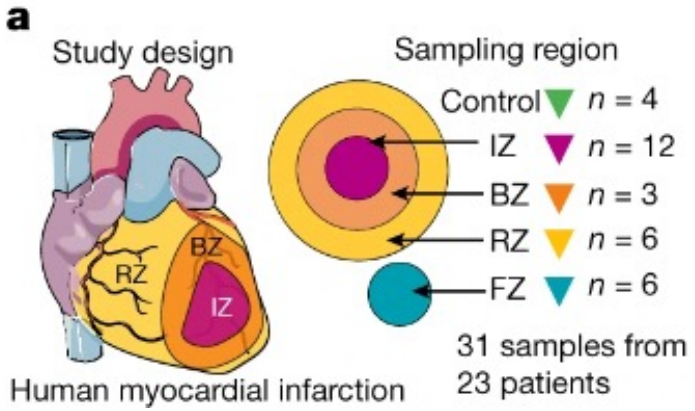


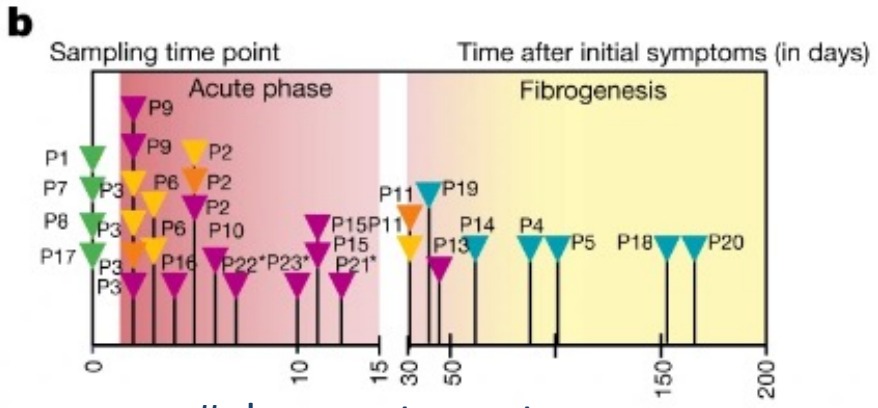




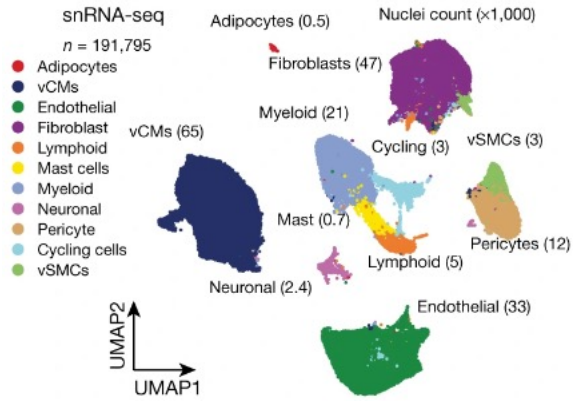


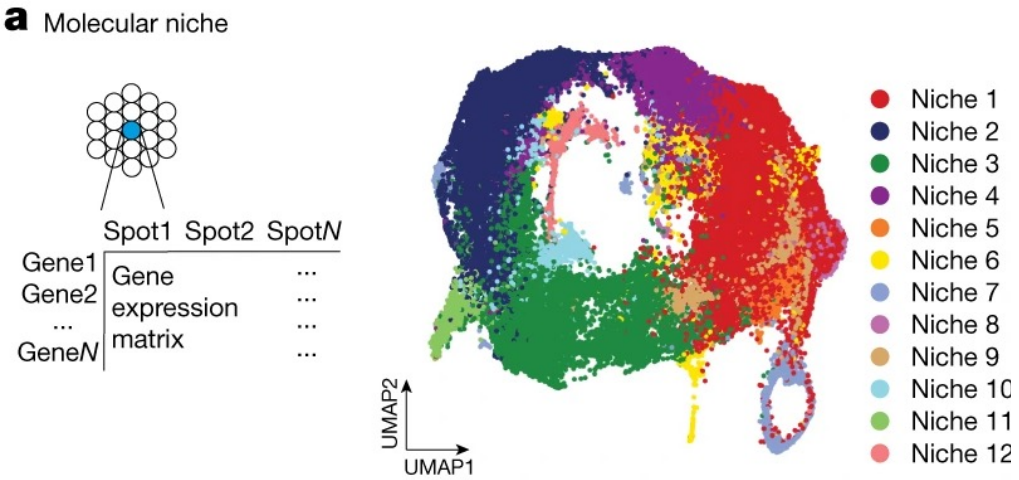
Cardiac remodelling after myocardial infarction

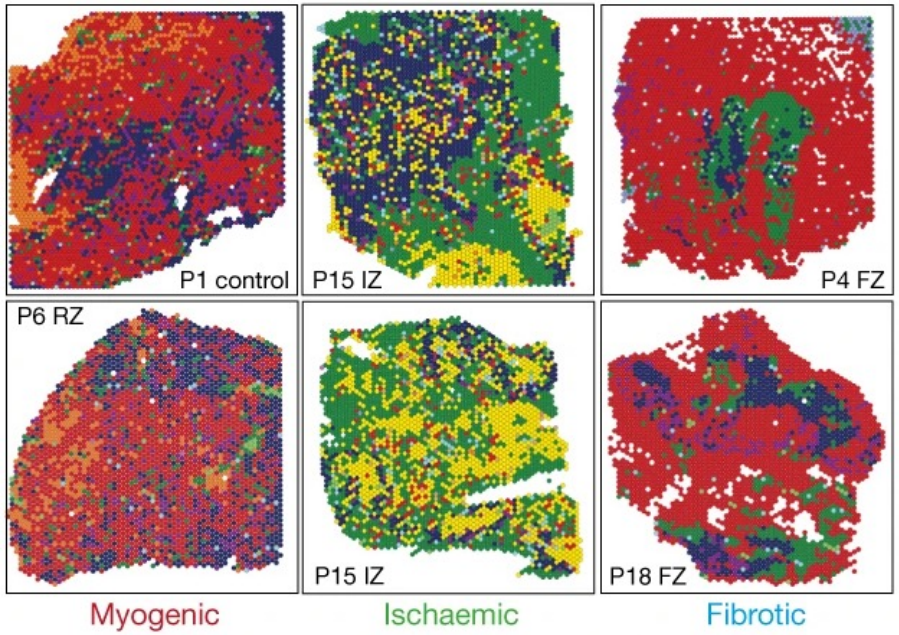




days post-onset



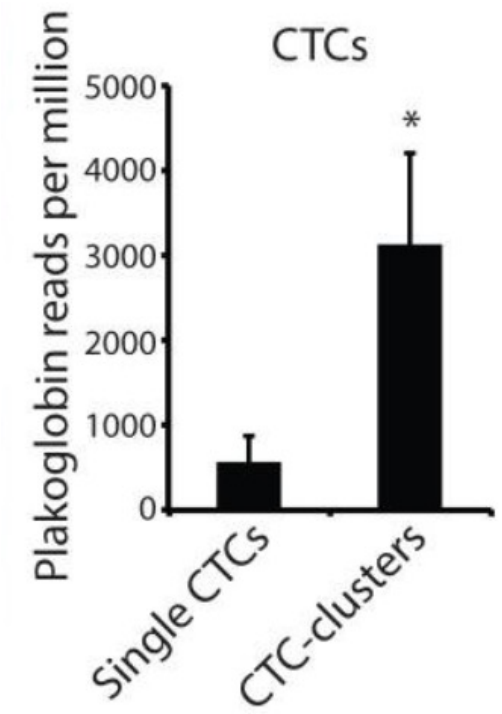


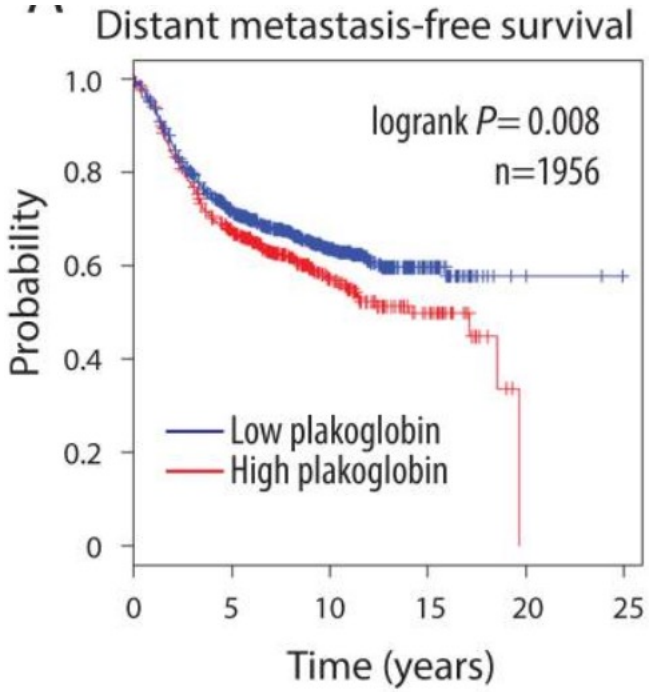


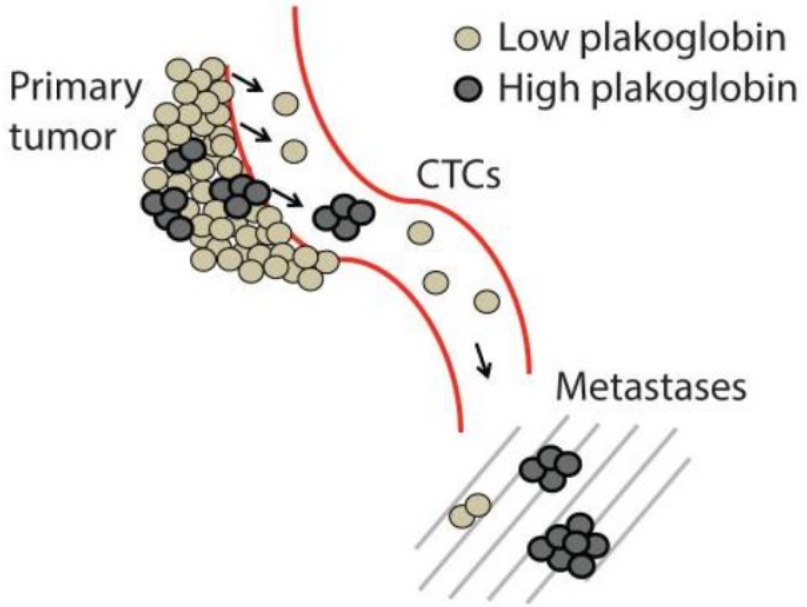
https://www.nature.com/articles/s41586-022-05060-x

Biomarker identification, potential metastasis mechanism

- CTC clusters originate from the primary tumor, exhibit increased metastatic propensity compared to single CTCs
- Abundance of CTC clusters in patients correlated with adverse outcome
- scRNA-seq revealed plakoglobin mediates CTC cluster formation, enhancing metastatic spread
- knockdown of plakoglobin abrogates CTC cluster formation and suppresses lung metastases

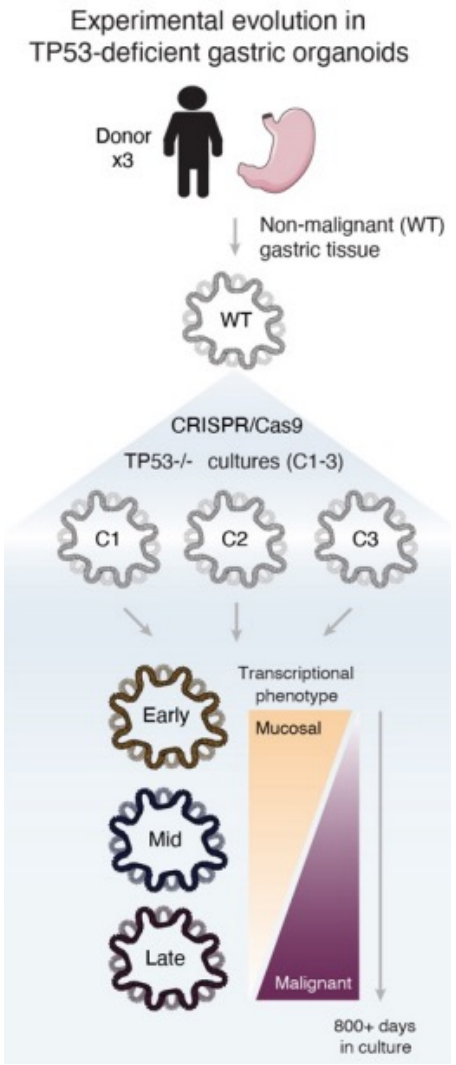




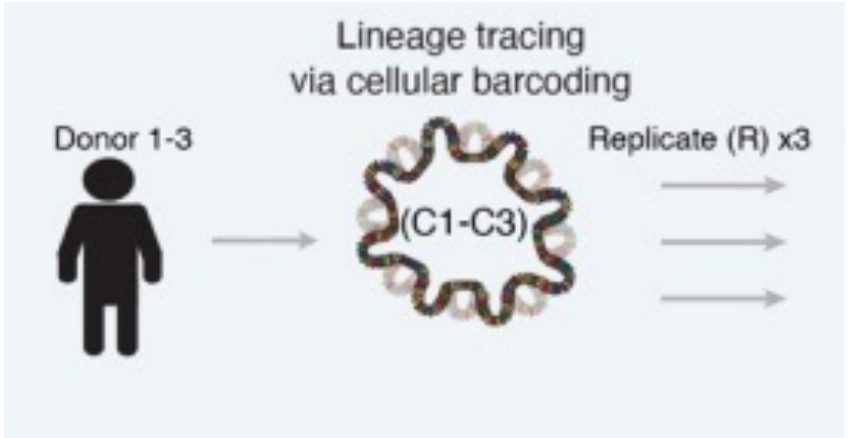


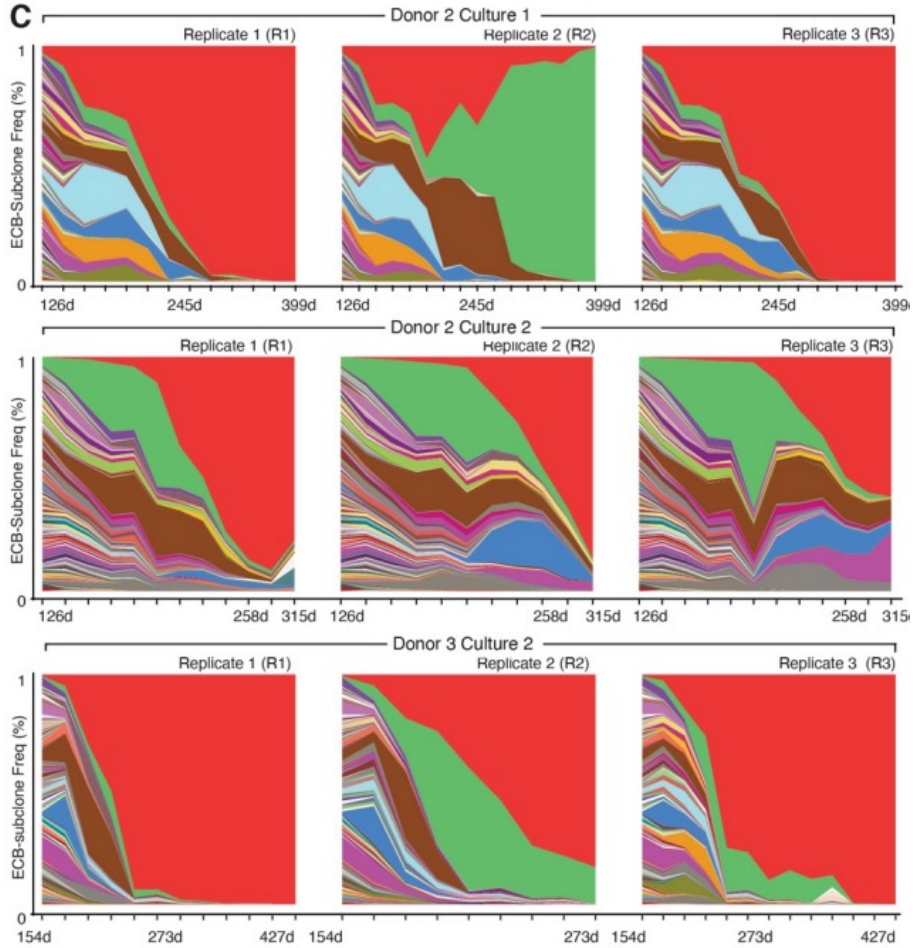
https://linkinghub.elsevier.com/retrieve/pii/S0092-8674(14)00927-1

Tumor clonal evolution



- Engineered TP53 deficiency into human gastric organoids, observe genomic changes over 2 years
- Observed reproducible clonal dynamics (initially rare subpopulation attain clonal dominance)





Experimental design considerations

Experiment design

- Beginning with the question of interest (and working backwards)

- Traditional statistical considerations and basic principals of statistical design of experiments apply.
  - Control for effects of outside variables, avoid/consider possible biases, avoid confounding variables in sample preparation.

    e.g. diet in a study of genetic risk factors for gastric cancer
  - Randomization of samples, plots, etc.
  - Replication is essential

Adapted from Jie Li (GC)

How many cells should I sequence per sample?

- The number of cells to target can be estimated based on:
  - The expected heterogeneity of all cells in a sample
  - The minimum frequency expected of a particular cell type within the sample, and
  - The minimum number of cells of each type desired in the resulting data set.

- For example, if we sequence a mixture of ~10 cell types where the frequency of the rarest cell type is ~0.03, then we would need to sequence ~2250 cells to have a 90% chance of capturing at least 50 of those rare cells.

- clearly difficult to know if you are sequencing new samples
  - Check cell atlases (e.g. Human Cell Atlas, Tabula Sapiens)
    https://tabula-sapiens-portal.ds.czbiohub.org/

- www.satijalab.org/howmanycells

Adapted from Jie Li (GC)

Sequencing depth required for scRNA-seq

- “Depth" (per cell) ~= # of reads/cell
- Factors to consider are (per lane):
  - # of sequenced reads
  - # of cells pooled for sequencing (estimate)
  - Expected percentage of usable data (80% used below)

$$\frac{\#\ \text{reads}}{\text{cell}} = 0.8 \times \frac{\#\ \text{sequenced reads}}{\#\ \text{cells pooled}}$$

- Read length, or single versus paired end, does not factor into depth above.
- Should expect minimum ~20k+ reads/sample (Chromium v3)

Adapted from Jie Li (GC)

Sequencing depth required for scRNA-seq

Other factors to consider for depth:

- Complexity of sample: the higher the complexity (more cell types, skewed composition), the higher the depth (number of cells needed to be sequenced).
- Interest in detecting genes expressed at low levels: the lower the level, the higher the depth (dropout).
- Some cell types yield higher duplication rate than others
- The fold change you want to be able to detect (smaller fold change requires more replicates and higher depth).
- Detection of novel transcripts, or quantification of isoforms (full-length libraries)

The sequencing depth needed is determined by the goals of the experiment and the nature of the sample.

Adapted from Jie Li (GC)

Getting it done

- Cell Isolation
- Library preparation (Per sample/pool)
- Sequencing (Number of lanes)
- Bioinformatics
  - General rule is to estimate the same dollar amount as data generation, i.e. double your budget

“Cost per cell”

Multiplexing cost calculator

Sample ‘multiplexing’, i.e. pooling cells from different samples together and running a single experiment, has significant potential benefits for single cell experiments. The 'demuxlet' algorithm (Ye lab, UCSF), leverages genetic polymorphisms to demultiplex pooled cells from different genetic backgrounds, while the 'Cell hashing' approach (Satija and Technology/Innovation labs, NYGC), accomplishes similar goals with barcoded antibodies. Both approaches also enable robust detection of cross-sample doublets, as they will exhibit multiple sample barcodes.
By identifying and discarding doublets, multiplexing enables the ‘super-loading’ of commercial droplet-based single cell platforms, which can greatly reduce costs. We provide a multiplexing cost calculator below, which models the costs of library prep and sequencing for different experimental designs.

Load preset 1 We aim to recover 20,000 single cells. By multiplexing 8 samples together, running one 10x lane yields a non-identifiable multiplet rate of 2.9% and a total cost of ~$4,700

Load preset 2 We aim to recover 20,000 single cells, without multiplexing. To achieve a similar non-identifiable multiplet rate, we need to spread the cells across 6 10x runs, with a total cost of ~$14,000.

Number of cells desired: 20000

Number of 10x lanes: 1

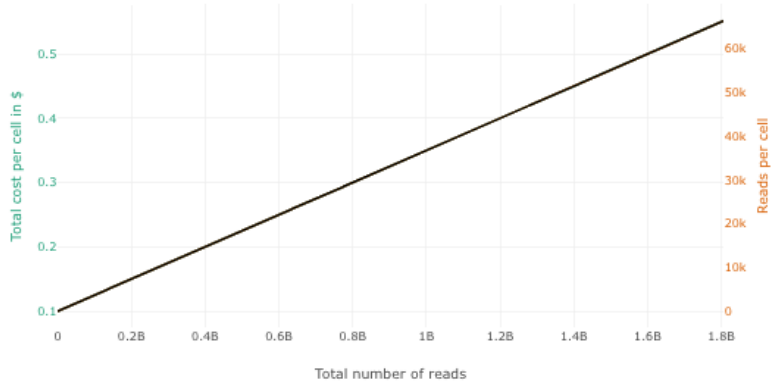
Number of multiplexed samples: 8

Show advanced settings

Results

More detail

Total cells needed for loading: 42,146
Multiplet rate: 2.92%
Total cost for library prep: $2,000
Sequencing cost @ 20,000 reads/cell: $2,726
Overall cost per cell @ 20,000 reads/cell: $0.24



https://satijalab.org/costpercell/

Who can help me construct the libraries and sequence?

DNA TECH

DNA Technologies & Expression Analysis
Core Laboratory

UCDAVIS GENOME CENTER

- Single Cell Gene Expression (3' GEX V3.1)
- **Fixed** Single Cell Gene Expression (human and mouse samples)
- Single Cell Immune Profiling (5' GEX V1.1 and V2 + V(D)J)
- Single Cell ATAC-seq
- **HT** Gene Expression and Immune Profiling
- Visium Library Preparation (slide preparation *not* included)
- Single Cell Multiome (combined ATAC + GEX)

See website for current costs (always changing) – expect $1000's for scRNA-seq, $100's for RNA-seq per sample

https://dnatech.genomecenter.ucdavis.edu

Who can help me do (some types of) analysis?

UC Davis Bioinformatics Core

RNA-Seq (Per Project)

RNA-Seq differential expression analysis for model organism (human, mouse), including QC of raw data, alignment to genome, and generation of tables of differentially expressed genes for typical two-factor experiment. For projects with 25 samples or more.

UC Campus	$2055
Non-UC Academic	$3227

Data Analysis (Per Hour)

Custom bioinformatics analysis from experimental design to publication.

UC Campus	$100
Non-UC Academic	$158
Private Enterprise	$195

Biostatistics (Per Hour)

Statistical analysis of bioinformatics data.

UC Campus	$160
Non-UC Academic	$251
Private Enterprise	$310

As of Sept 2022

Can I have someone in my lab trained to do the analysis?

- The Bioinformatics Core offers training workshops for scRNA-seq analysis:
  - https://bioinformatics.ucdavis.edu/training

Jul 21, 2022	Advanced Topics in Single Cell RNA-Seq Analysis: Multiomics ATAC-Seq & RNA-Seq
Jul 18 - Jul 20, 2022	Single Cell RNA-Seq Analysis

- MIND Institute Investigators: The Biostatistics, Bioinformatics, & Research Design Core
  - https://health.ucdavis.edu/mindinstitute/centers/intellectual-developmental-disabilities-research/cores/bbrd.html

- Software packages to perform single cell analyses
  - https://github.com/seandavi/awesome-single-cell

- Online tutorials:
  - https://www.singlecellcourse.org/
  - https://bioconductor.org/books/release/OSCA/

Questions?