

A Hitchhiker's Guide to Single-cell Data Analysis

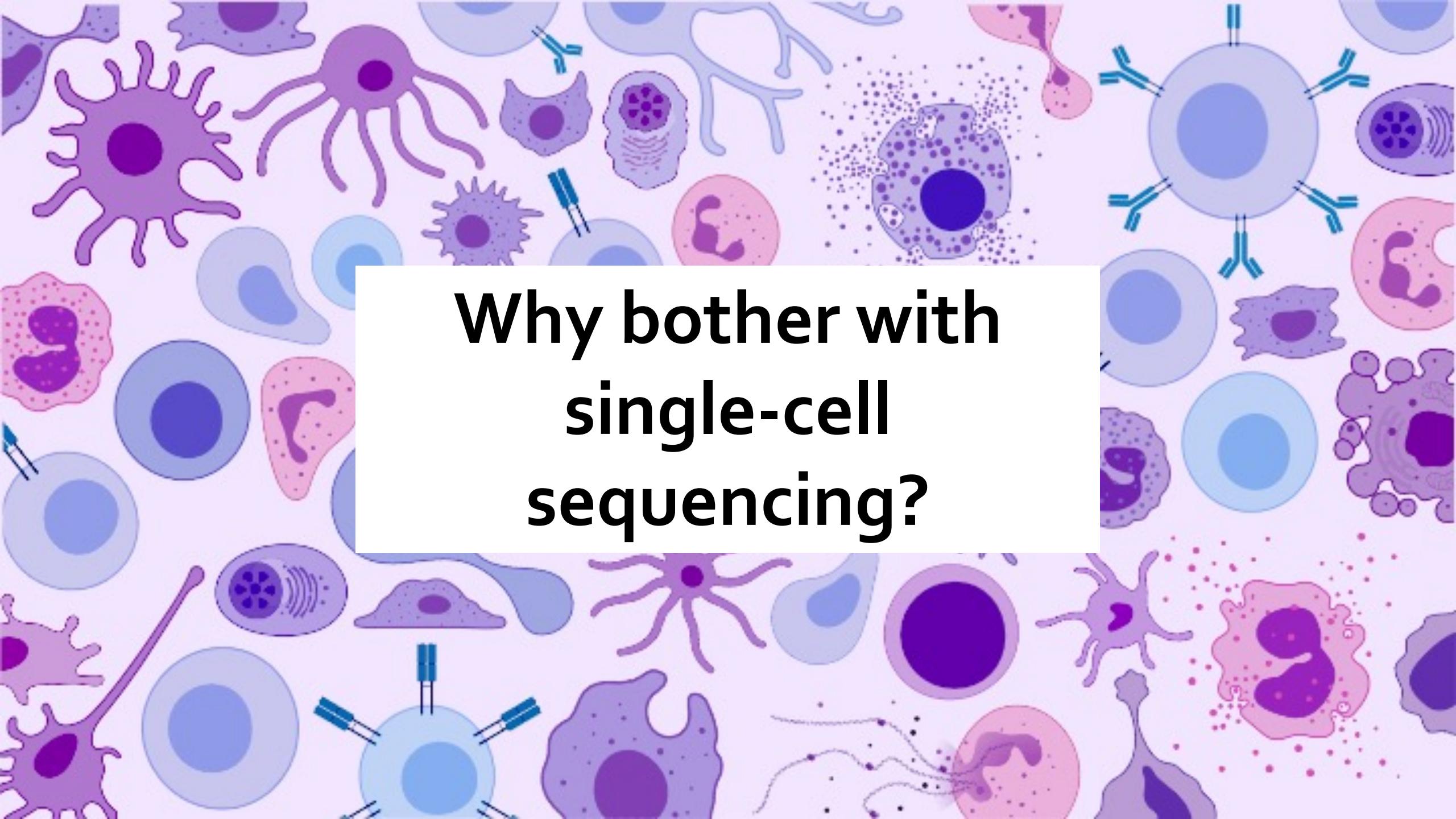
Workshop outline

	Monday	Tuesday	Wednesday	Thursday	Friday
9h00	Welcome	Good morning	Good morning	Good morning	Good morning
9h30	Get to know each other	Integration António Sousa	Adaptive immune receptor Lisa Dratva	Velocity Zhisong He	Spatial Transcriptomics Anna Maguza
10h40	Coffee break	Coffee break	Coffee break	Coffee break	Coffee break
11h00	Fundamentals of scRNA-seq Tomás Gomes	Integration António Sousa	Adaptive immune receptor Lisa Dratva	Velocity Zhisong He	Spatial Transcriptomics Anna Maguza
12h30	Lunch break	Lunch break	Lunch break	Lunch break	Lunch break
13h30	Fundamentals of scRNA-seq Tomás Gomes	Seminar Yang Li HZI Braunschweig	Seminar Nuno Morais iMM	Seminar Zhisheng He ETH Zurich	Spatial Transcriptomics Anna Maguza
14h00		Integration António Sousa	Adaptive immune receptor Lisa Dratva	Velocity Zhisheng He	
15h30	Coffee break	Coffee break	Coffee break	Coffee break	Bonsai Lab Ana Borges
16h00			Enzfarma Diogo Silvério	Novogene Orial Alejo-Valle	Coffee break
16h30	Fundamentals of scRNA-seq Tomás Gomes	Integration António Sousa	Adaptive immune receptor Lisa Dratva	Velocity Zhisheng He	Closing session
17h00					
19h00		Art Gallery	Dinner with speakers	Dinner with speakers	
20h00		All group dinner			
21h00					

More info:
scworkshop2024@medicina.ulisboa.pt

Today's outline

- Introductions
- Why bother with single-cell sequencing?
- Technical support – preparing your work environment
- Obtaining and pre-processing scRNA-seq data
- scRNA-seq data analysis
 - In R
 - In python
- scRNA-seq data exploration

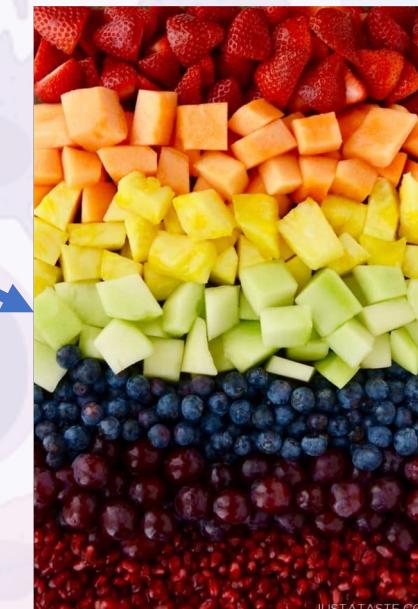


**Why bother with
single-cell
sequencing?**

Why transcriptomics?

- Able to capture **all transcripts** with little to no bias
- **Good proxy** for understanding the phenotype of a cell/tissue
- Easy to **access**
- Easy to **amplify the signal**

Bulk vs single-cell sequencing



JUSTATASTE.COM

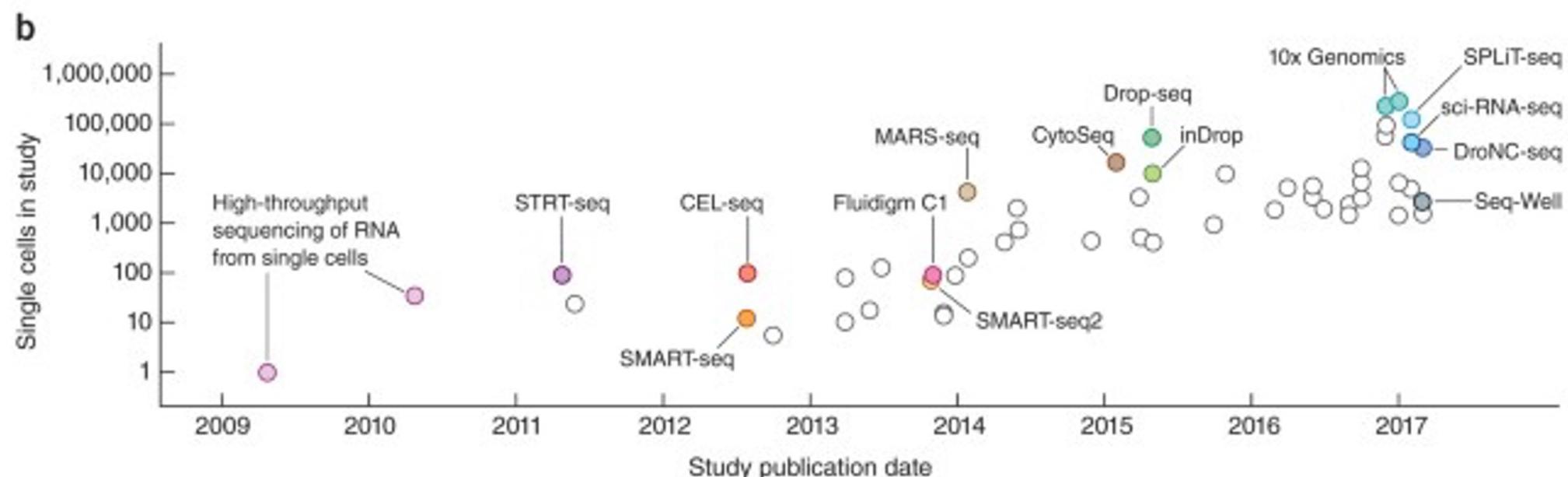
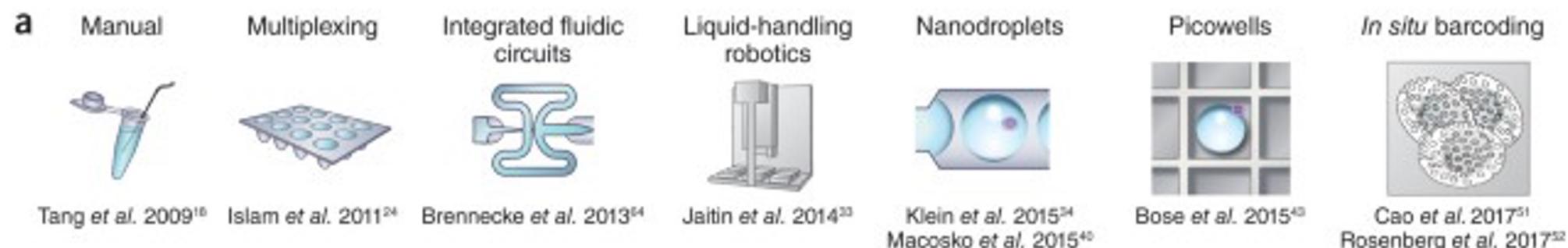
Bulk vs single-cell sequencing

Bulk sequencing	Single-cell sequencing
cheaper	More expensive
Simpler protocol	More steps/more sensitive
Lower resolution	Higher resolution



This implies a cell-centric reasoning, i.e.
at life's unit level.

The single-cell sequencing decade



The first single-cell RNA-seq paper

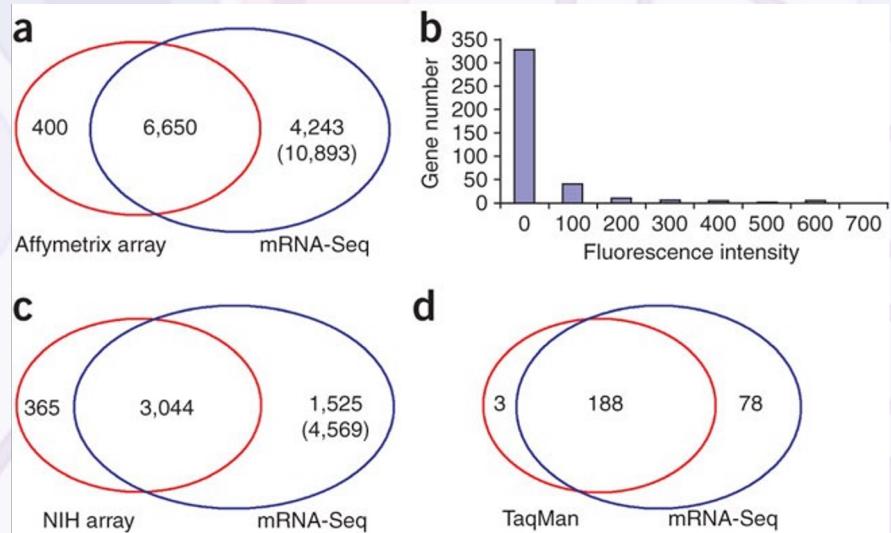
Published: 06 April 2009

mRNA-Seq whole-transcriptome analysis of a single cell

Fuchou Tang, Catalin Barbacioru, Yangzhou Wang, Ellen Nordman, Clarence Lee, Nanlan Xu, Xiaohui Wang, John Bodeau, Brian B Tuch, Asim Siddiqui, Kaiqin Lao✉ & M Azim Surani✉

Nature Methods **6**, 377–382 (2009) | Cite this article

14k Accesses | 1308 Citations | 89 Altmetric | Metrics



1 year after the first RNA-seq paper

Goal was to quantify RNA in a very low input system (but with many different genes expressed)

Setting the stage for the single-cell decade

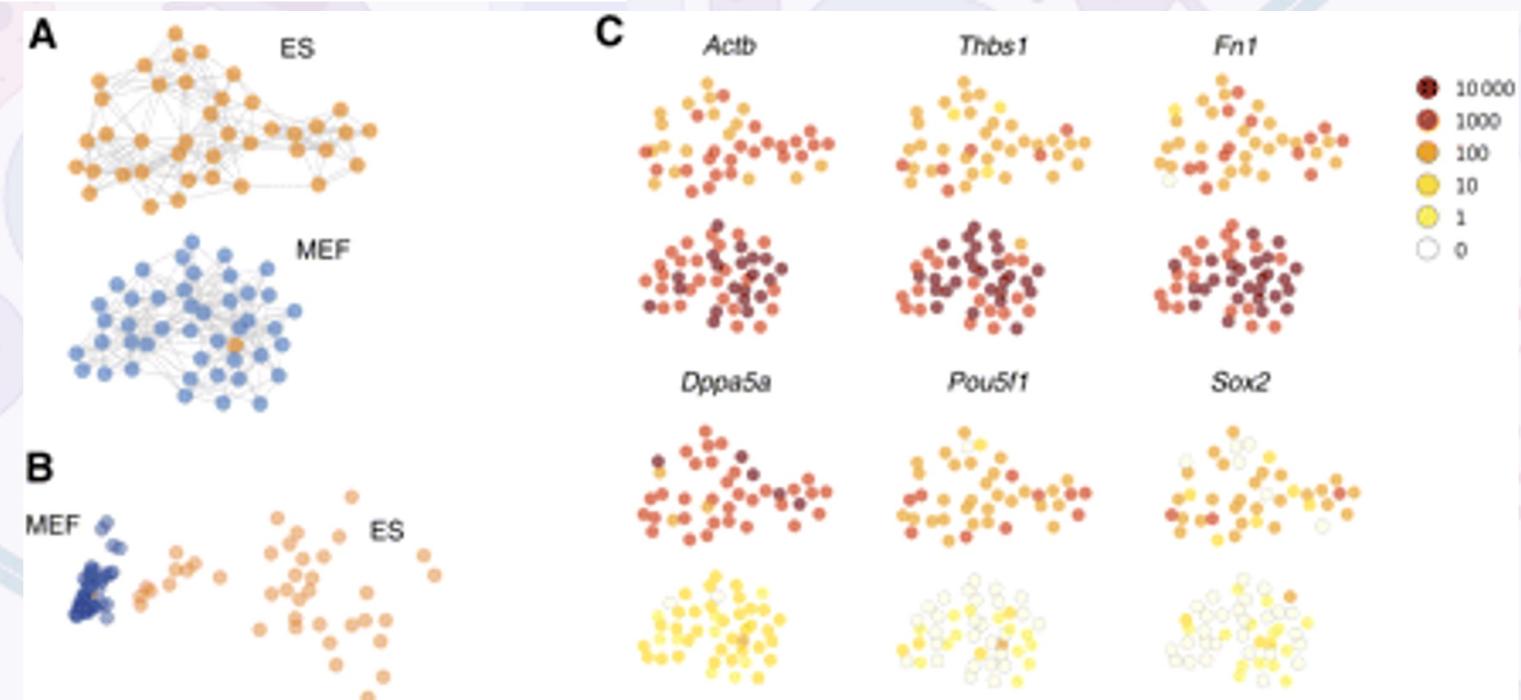
Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq

Saiful Islam^{1,4}, Una Kjällquist^{1,4}, Annalena Moliner², Paweł Zajac¹, Jian-Bing Fan³, Peter Lönnerberg¹ and Sten Linnarsson^{1,5}

¹Laboratory for Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-171 77 Stockholm, Sweden;

²Department of Neuroscience, Karolinska Institutet, SE-171 77 Stockholm, Sweden;

³Illumina Inc., San Diego, California 92121, USA



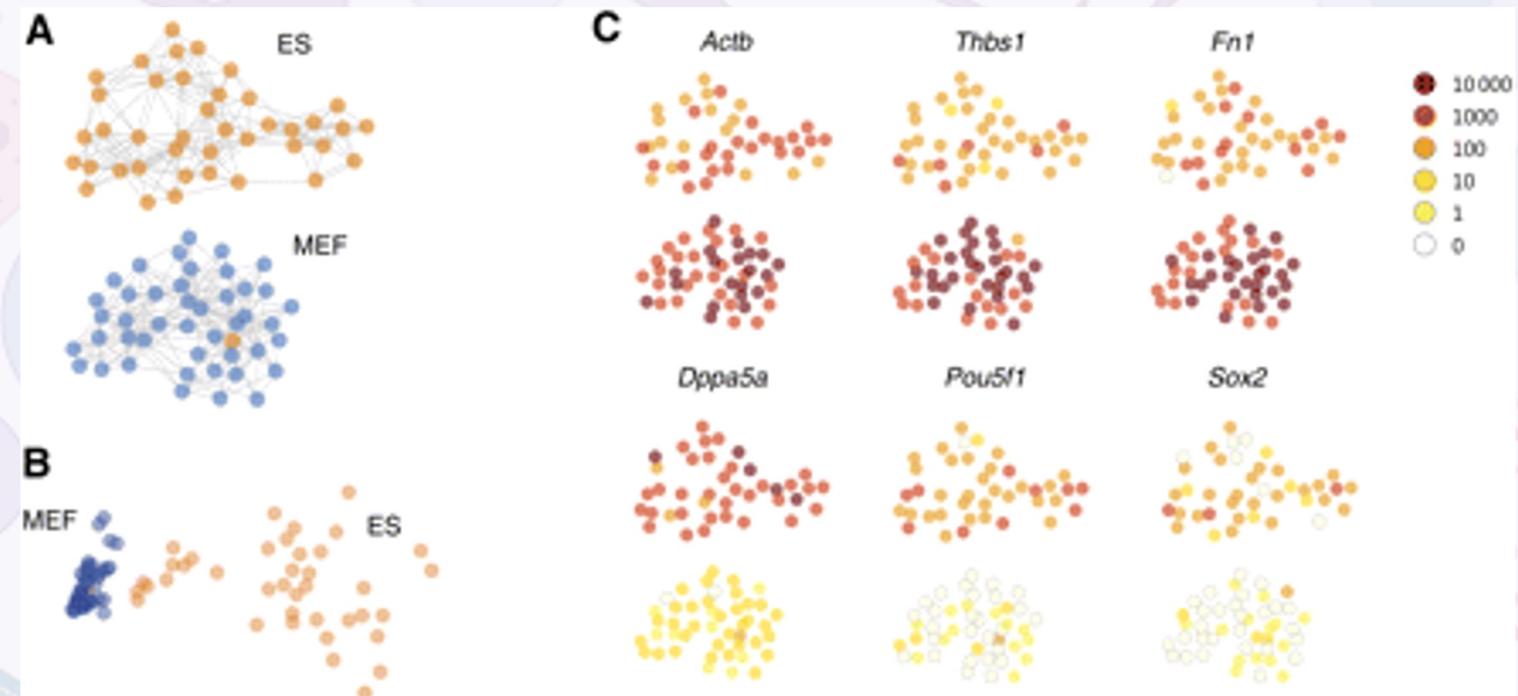
First paper (2011) doing multiplexing of single-cells in a 96 well plate

Use of control RNA molecules (ERCC, included at specific concentrations) to distinguish technical and biological variance

Setting the stage for the single-cell decade

We envisage the future use of very large-scale single-cell transcriptional profiling to build a detailed map of naturally occurring cell types, which would give unprecedented access to the genetic machinery active in each type of cell at each stage of development.

Force-directed graph
PCA

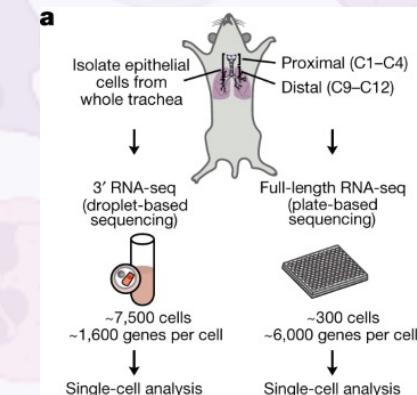
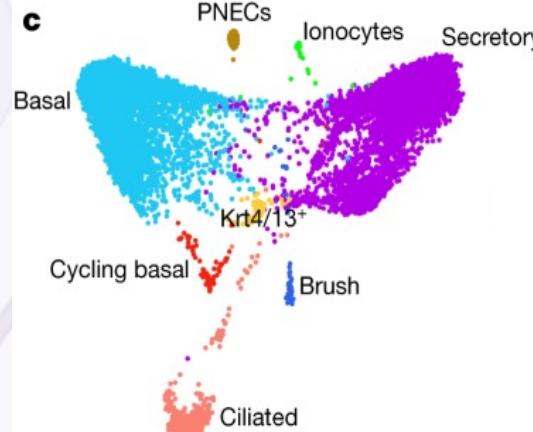
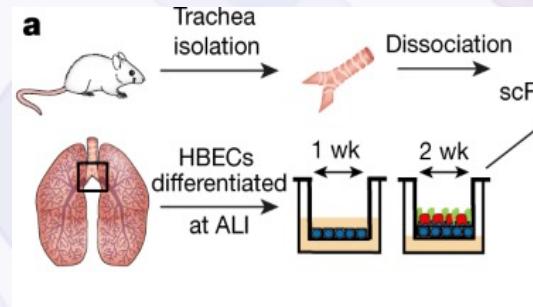


Uses of scRNA-seq – cell type discovery

Letter | Published: 01 August 2018

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

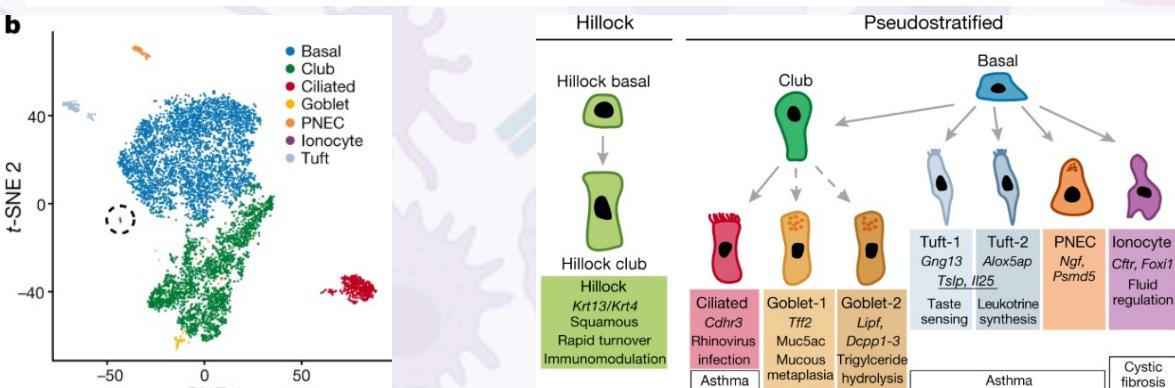
Lindsey W. Plasschaert, Rapolas Žilionis, Rayman Choo-Wing, Virginia Savova, Judith Knehr, Guglielmo Roma, Alon M. Klein & Aron B. Jaffe



Article | Published: 01 August 2018

A revised airway epithelial hierarchy includes CFTR-expressing ionocytes

Daniel T. Montoro, Adam L. Haber, Moshe Biton, Vladimir Vinarsky, Brian Lin, Susan E. Birket, Feng Yuan, Sijia Chen, Hui Min Leung, Jorge Villoria, Noga Rogel, Grace Burgin, Alexander M. Tsankov, Avinash Waghray, Michal Slyper, Julia Waldman, Lan Nguyen, Danielle Dionne, Orit Rozenblatt-Rosen, Purushothama Rao Tata, Hongmei Mou, Manjunatha Shivaraju, Hermann Bihler, Martin Mense, Guillermo J. Tearney, Steven M. Rowe, John F. Engelhardt, Aviv Regev & Jayaraj Rajagopal - Show fewer authors



Description of ionocytes in mouse (and human) airways

Specified by transcription factor Foxl1

Involved in cystic fibrosis

<https://www.nature.com/articles/s41586-018-0394-6>

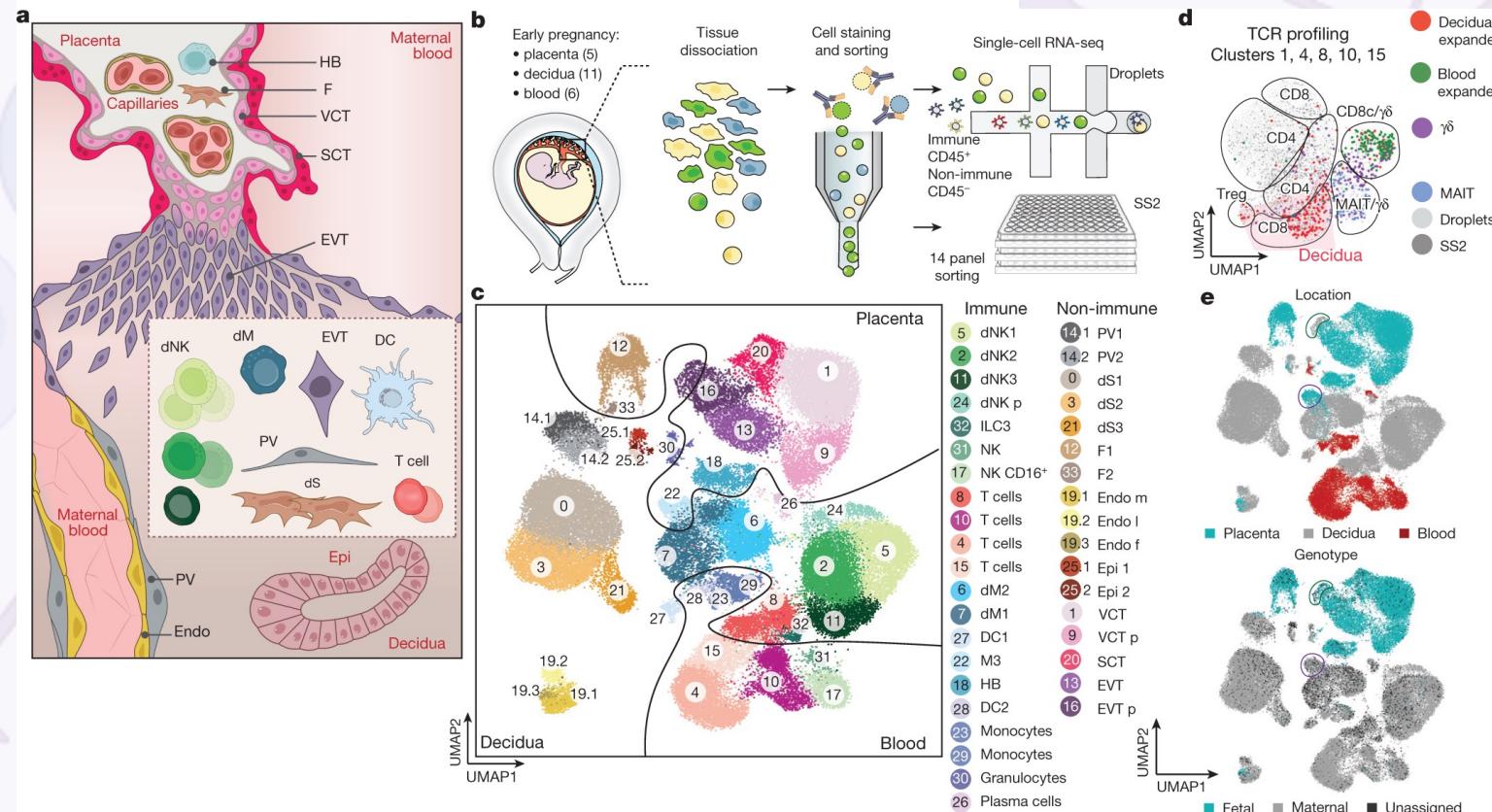
<https://www.nature.com/articles/s41586-018-0393-7>

Uses of scRNA-seq – tissue composition

Article | Published: 14 November 2018

Single-cell reconstruction of the early maternal–fetal interface in humans

Roser Vento-Tormo, Mirjana Efremova, Rachel A. Botting, Margherita Y. Turco, Miquel Vento-Tormo, Kerstin B. Meyer, Jong-Eun Park, Emily Stephenson, Krzysztof Polański, Angela Goncalves, Lucy Gardner, Staffan Holmqvist, Johan Henriksson, Angela Zou, Andrew M. Sharkey, Ben Millar, Barbara Innes, Laura Wood, Anna Wilbrey-Clark, Rebecca P. Payne, Martin A. Ivarsson, Steve Liso, Andrew Filby, David H. Rowitch, Judith N. Bulmer, Gavin J. Wright, Michael J. T. Stubbington, Muzlifah Haniffa✉, Ashley Moffett✉ & Sarah A. Teichmann✉ - Show fewer authors



Identification of all cell populations at the maternal-fetal interface

Discovering cellular interactions in immune and non-immune compartment

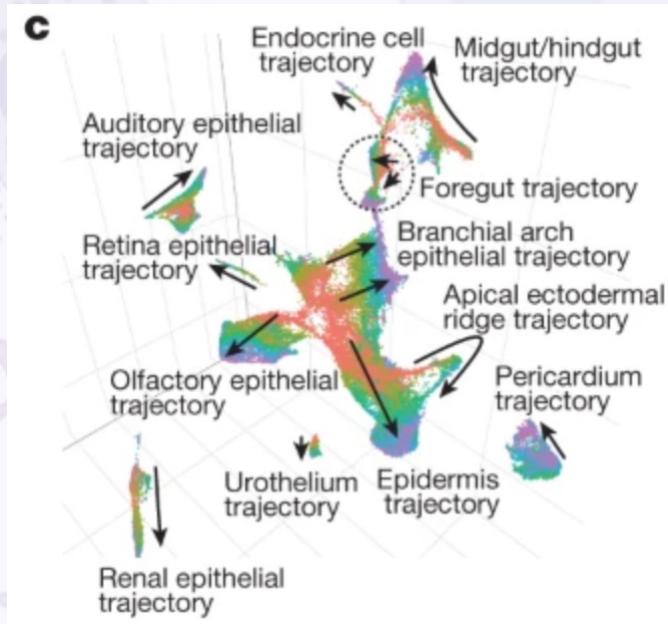
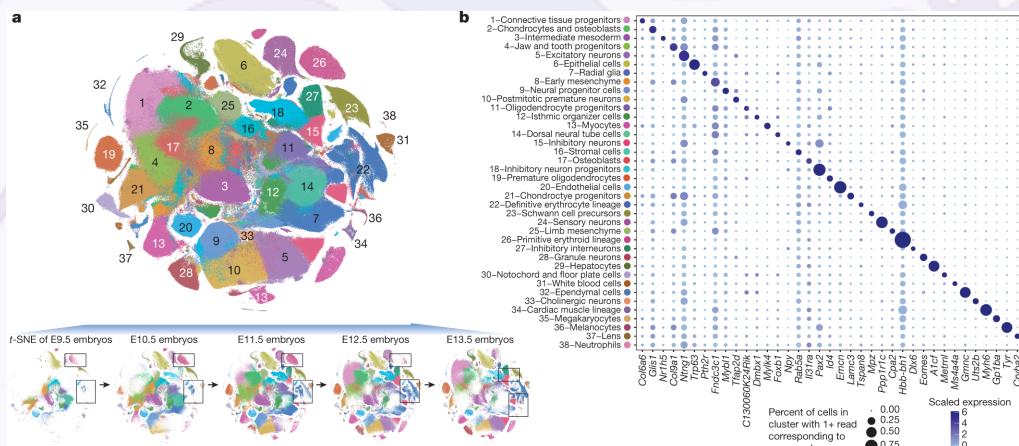
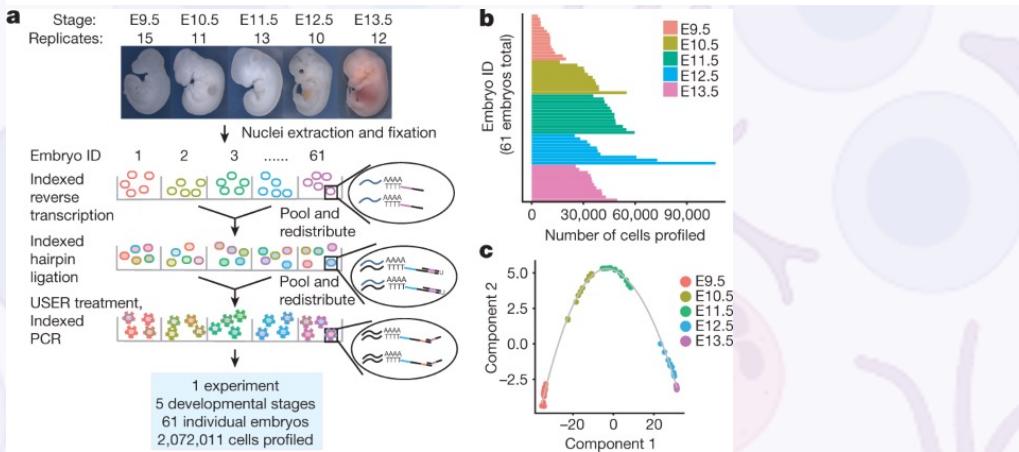
Unravel cell origin (maternal or fetal?) by combining variant calling with genome data

Uses of scRNA-seq – development

Article | Published: 20 February 2019

The single-cell transcriptional landscape of mammalian organogenesis

Junyue Cao, Malte Spielmann, Xiaojie Qiu, Xingfan Huang, Daniel M. Ibrahim, Andrew J. Hill, Fan Zhang, Stefan Mundlos, Lena Christiansen, Frank J. Steemers, Cole Trapnell & Jay Shendure



Profiling about 2M cells during mouse organogenesis

Defining the major developmental trajectories for cell types based on transcriptional changes

Uses of scRNA-seq – building gene networks

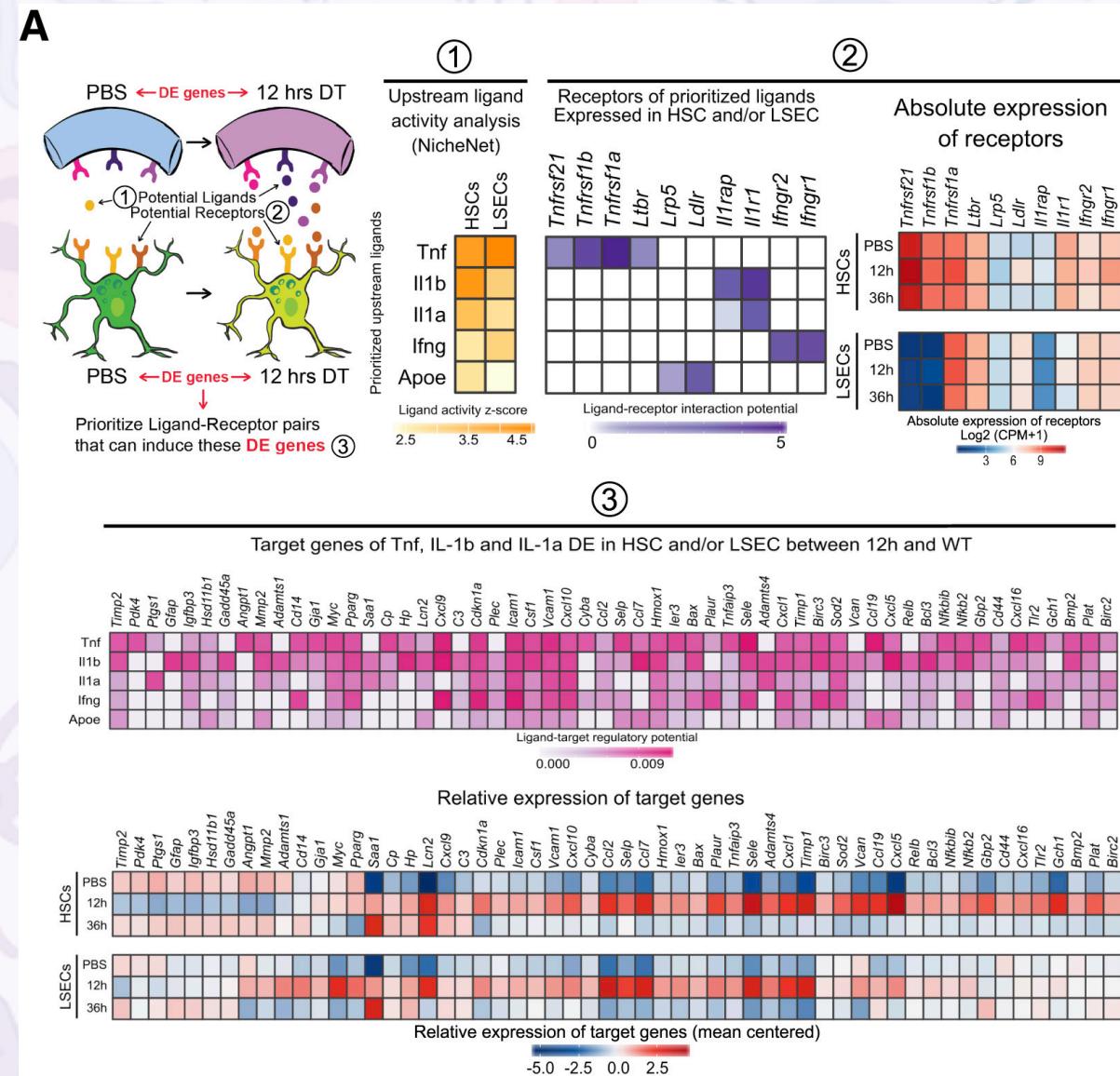
Stellate Cells, Hepatocytes, and Endothelial Cells
Imprint the Kupffer Cell Identity on Monocytes
Colonizing the Liver Macrophage Niche

Johnny Bonnardel ^{1, 2, 16}  , Wouter T'Jonc ^{1, 2, 16}, Djoere Gaublomme ^{3, 4}, Robin Browaeys ^{5, 6}, Charlotte L. Scott ^{1, 2, 7}, Liesbet Martens ^{2, 5}, Bavo Vanneste ^{1, 2}, Sofie De Prijck ^{1, 2}, Sergei A. Nedospasov ^{8, 9}, Anna Kremer ^{2, 10}, Evelien Van Hamme ^{2, 10}, Peter Borghgraef ^{2, 10}, Wendy Toussaint ^{4, 11}, Pieter De Blieser ^{2, 5}, Inge Mannaerts ¹², Alain Beschin ^{13, 14}, Leo A. van Grunsven ¹², Bart N. Lambrecht ^{4, 11}, Tom Taghon ¹⁵, Saskia Lippens ^{2, 10}, Dirk Elewaut ^{3, 4}, Yvan Saeyns ^{5, 6, 17}, Martin Guillemins ^{1, 2, 17, 18}  

What intercellular signaling drives activation of Hepatic Stellate cells?

Predicts ligands (TNF) with more active targets in receiving cell types; verified by blocking in vivo

Other tools focus on identifying GRN or CCN specifically

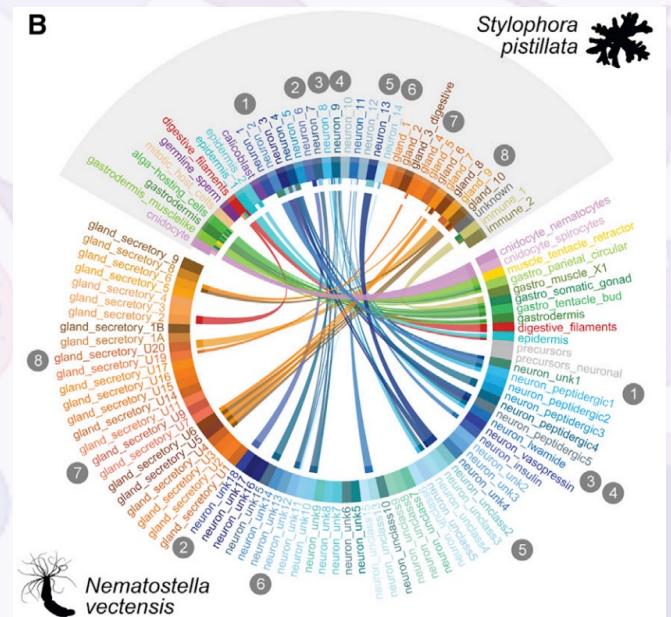
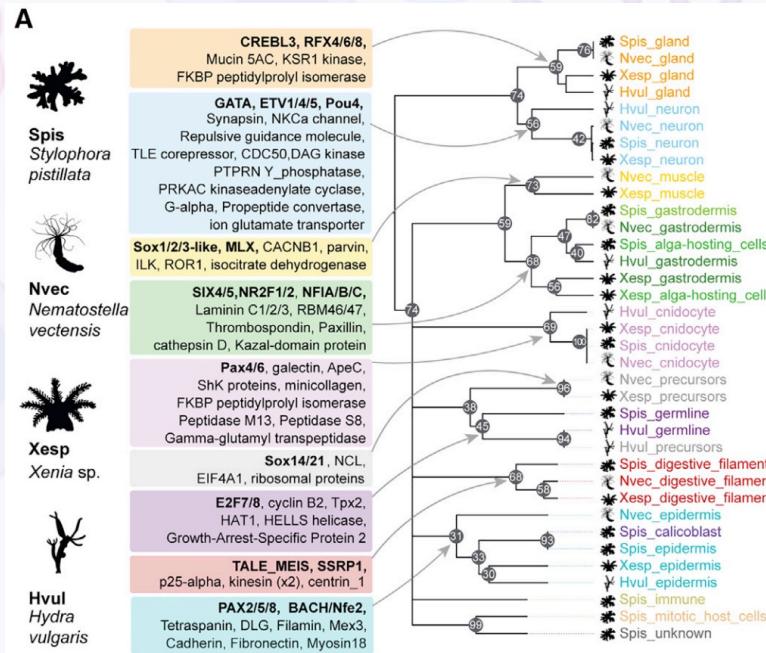
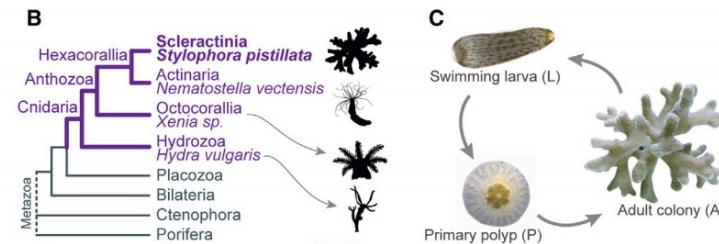


Uses of scRNA-seq – evolutionary comparisons

A stony coral cell atlas illuminates the molecular and cellular basis of coral symbiosis, calcification, and immunity

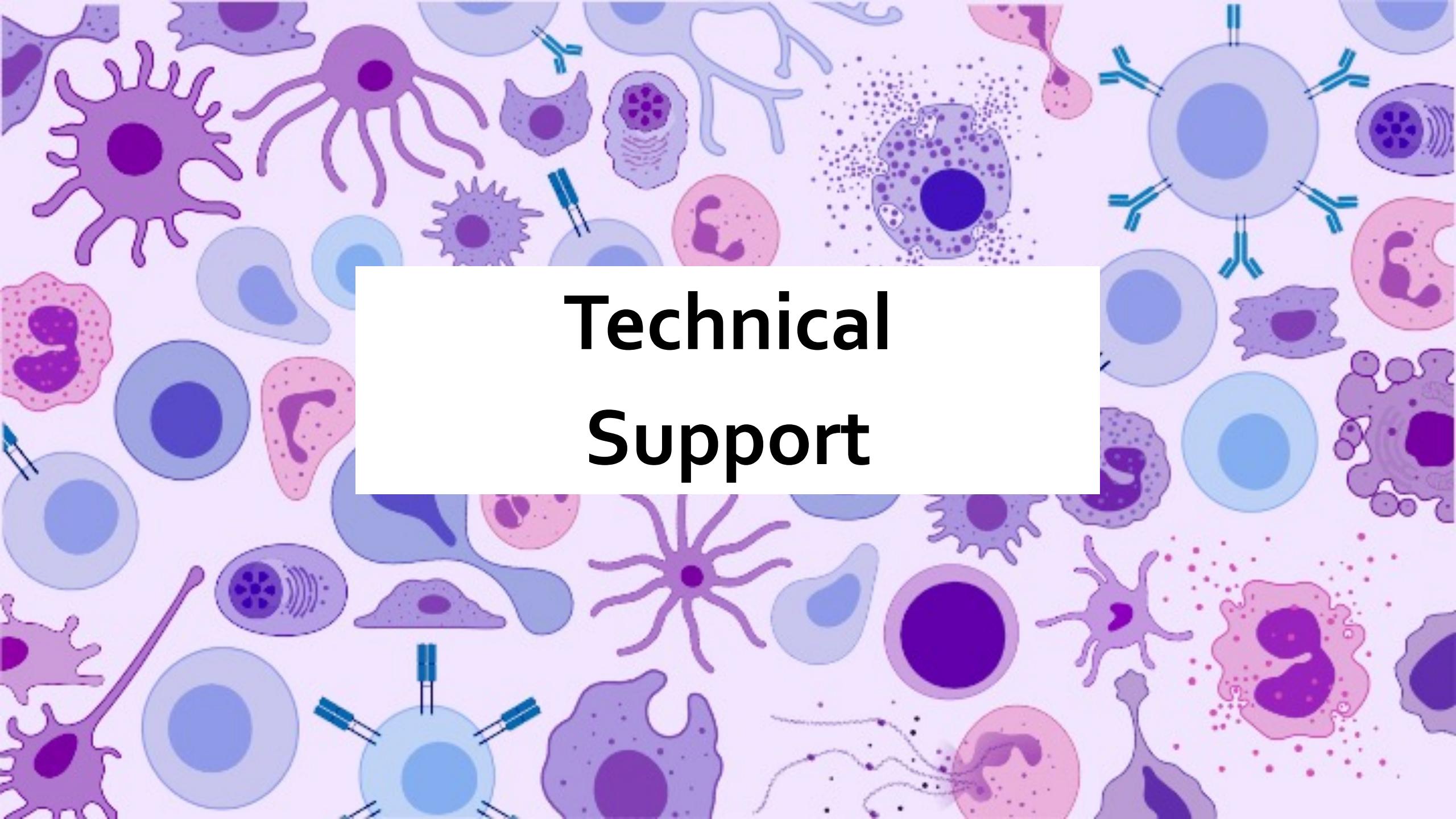
Shani Levy ⁶ • Anamaria Elek ⁶ • Xavier Grau-Bové [•] ... Amos Tanay [•] Tali Mass [•]

Arnaud Sebé-Pedrós [•] • Show all authors • Show footnotes



Comparison of Cnidaria cell type lineages

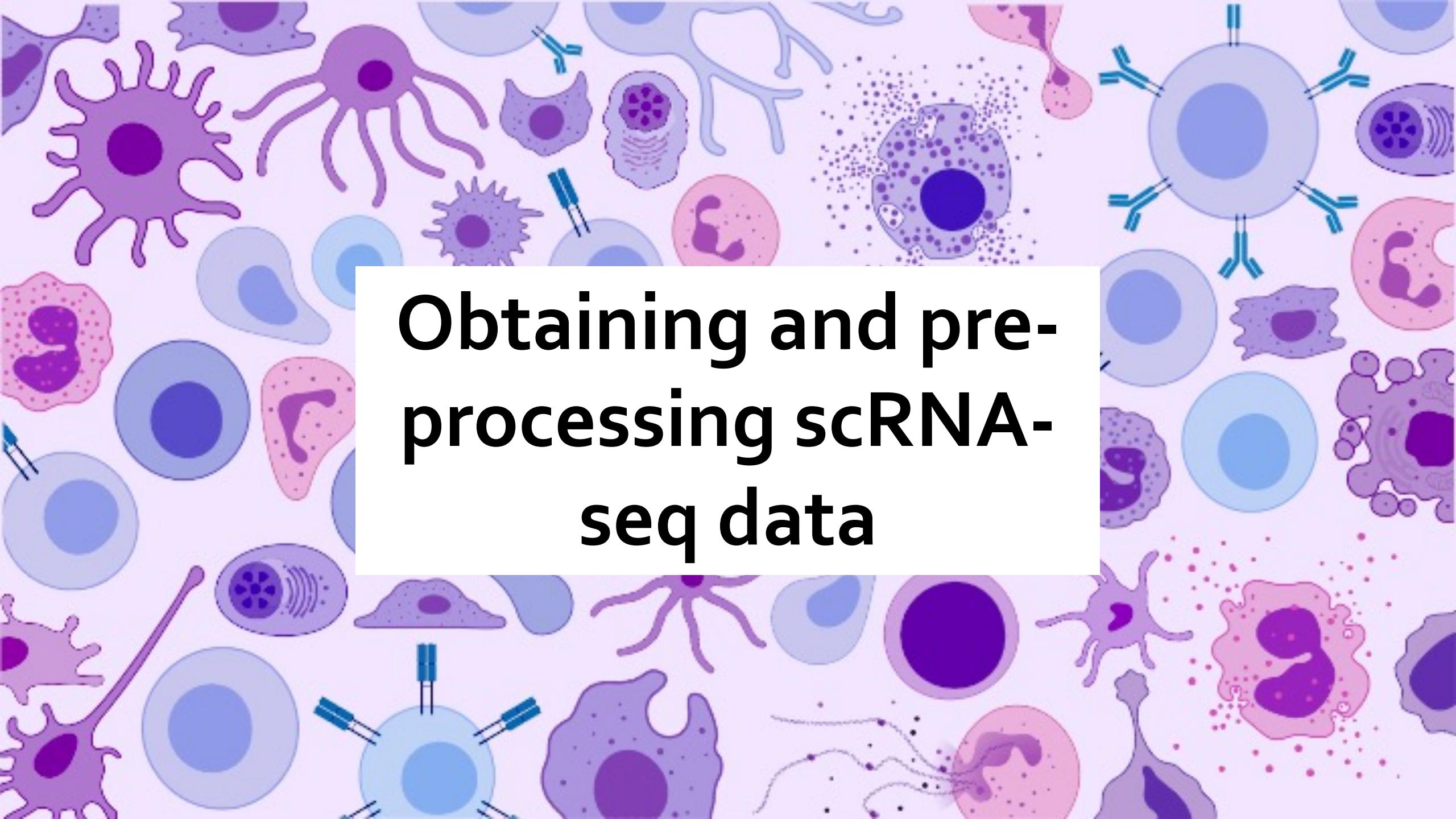
Finds origins of immune and calcification gene programmes



Technical Support

Basic Installations

1. R and RStudio: <https://posit.co/download/rstudio-desktop/>
2. Loupe Browser: <https://www.10xgenomics.com/support/software/loupe-browser/downloads>
3. Conda: <https://conda.io/projects/conda/en/latest/user-guide/install/index.html>



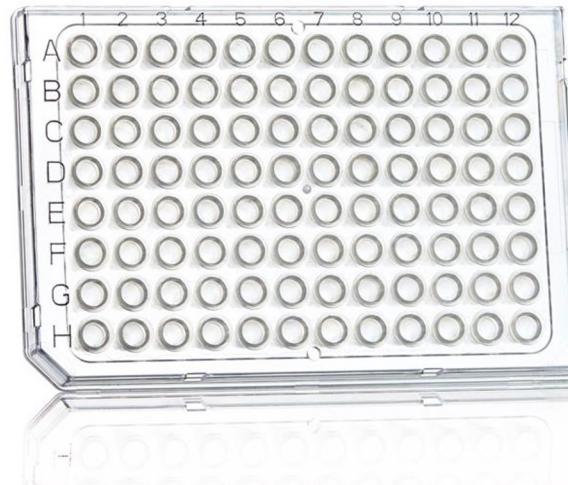
Obtaining and pre-processing scRNA-seq data

Common steps in a scRNA-seq experiment

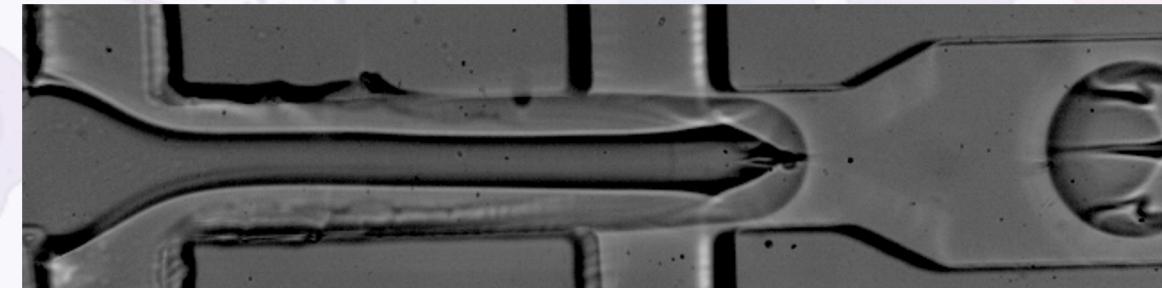
scRNA-seq experiment	<u>Bench work</u>	Cell isolation
		Library construction
		Sequencing
		Quantification
	<u>Computational work</u>	Quality control
		Normalisation
		Dimensionality reduction
		Clustering/Trajectory analysis
		Differential expression

Isolation - how do we get single-cells?

Plate-based
methods



Microdroplet-based
methods



Goal is to be able to perform RNA extraction and library construction for each cell in isolation

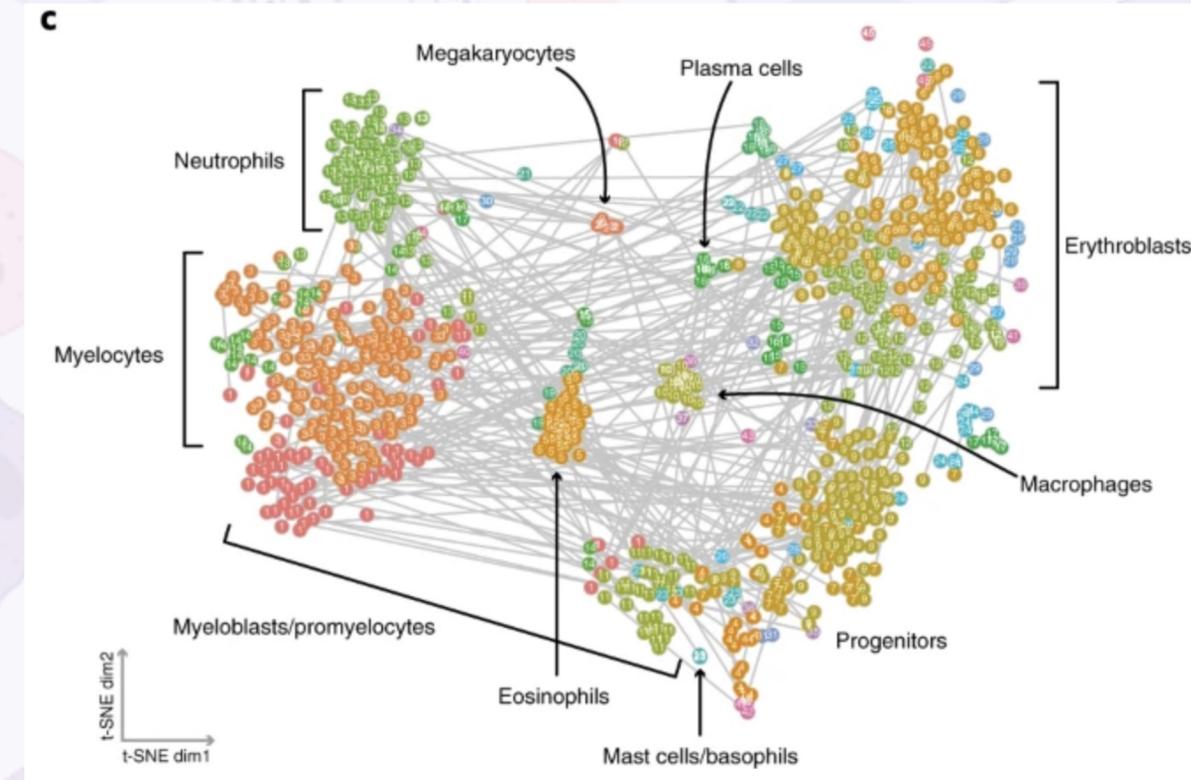
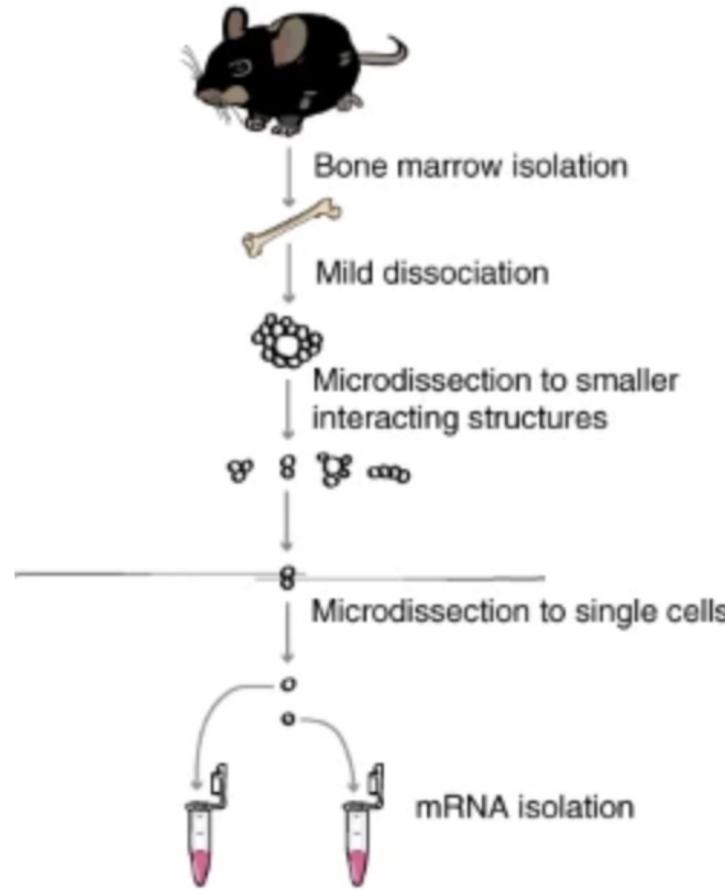
Microdissection: uncommon but useful

Article | Published: 21 May 2018

Mapping the physical network of cellular interactions

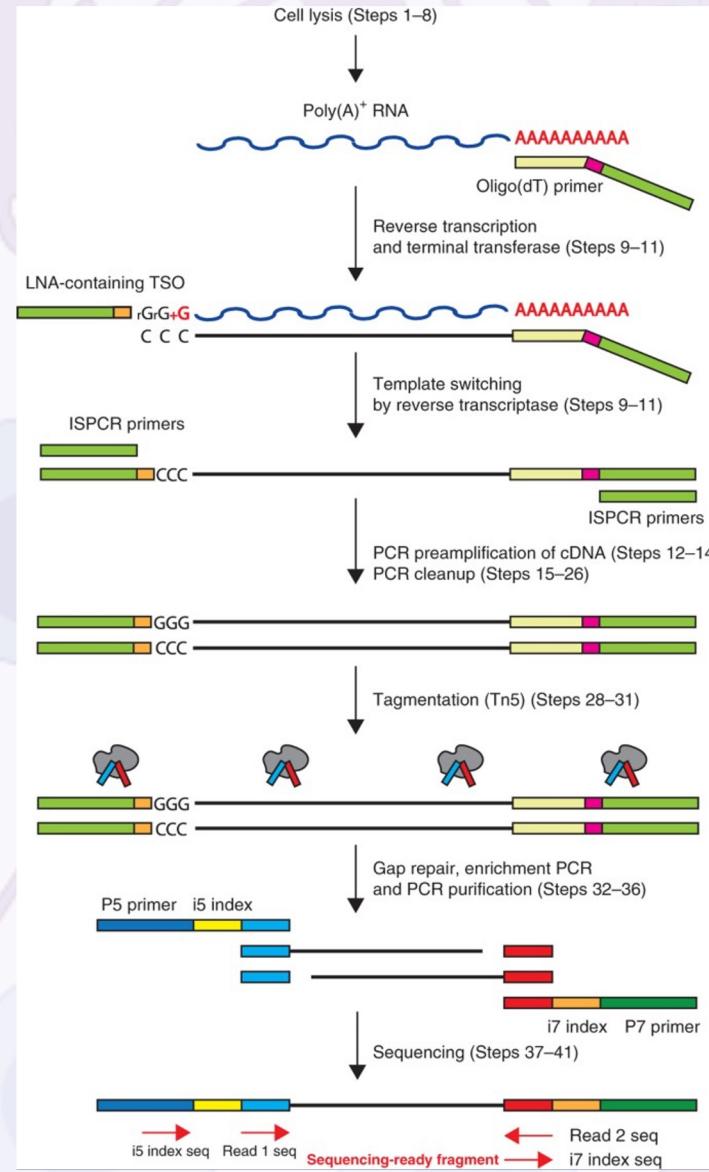
Jean-Charles Boisset, Judith Vivié, Dominic Grün, Mauro J. Muraro, Anna Lyubimova & Alexander van Oudenaarden [✉](#)

Nature Methods 15, 547–553 (2018) | [Cite this article](#)



Microdissection of interacting cell pairs to identify preferentially interacting cell types

Getting RNA from single-cells - Smart-seq2

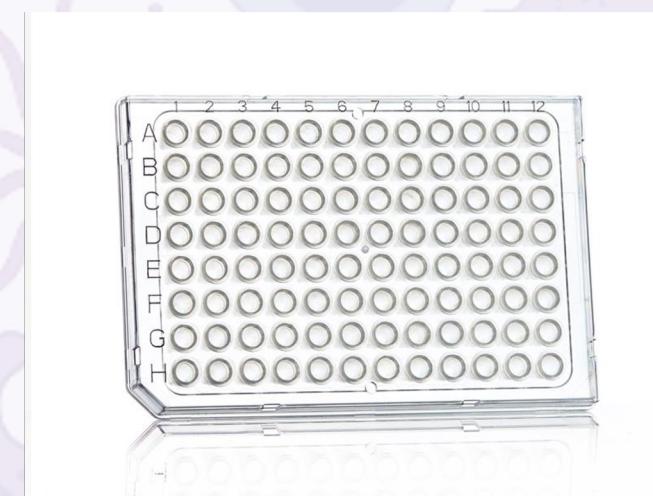


Smart-seq (current verison V4, the most used/well know is V2) is a protocol used to sequence the full-length transcriptome of individual cells

Requires extensive liquid handling, usual throughput is ~1000 cells

Since fewer cells are sequenced, sequencing depth tends to be higher and thus more genes are captured

Another (commercial) plate-based option is SORTseq



Getting RNA from single-cells - Chromium

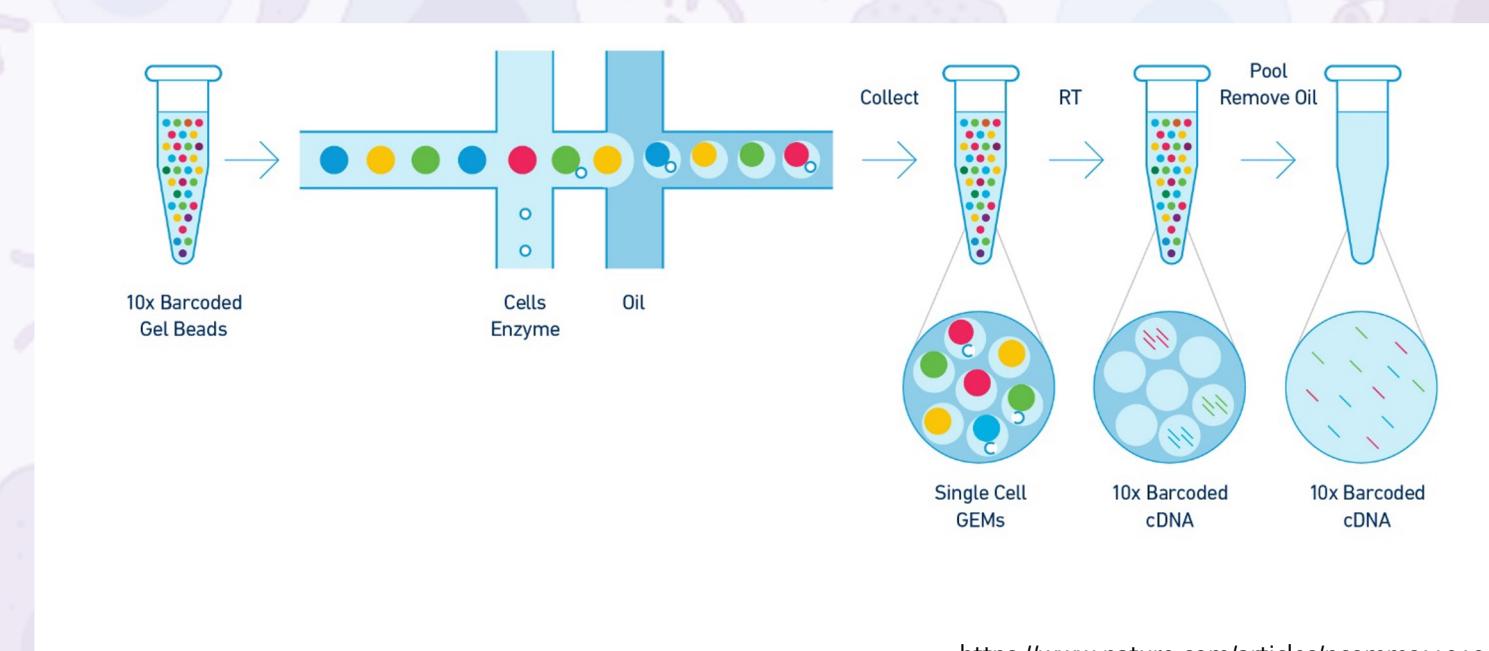
Open Access | Published: 16 January 2017

Massively parallel digital transcriptional profiling of single cells

Grace X. Y. Zheng, Jessica M. Terry, Phillip Belgrader, Paul Ryvkin, Zachary W. Bent, Ryan Wilson, Solongo B. Ziraldo, Tobias D. Wheeler, Geoff P. McDermott, Junjie Zhu, Mark T. Gregory, Joe Shuga, Luz Montesclaros, Jason G. Underwood, Donald A. Masquelier, Stefanie Y. Nishimura, Michael Schnall-Levin, Paul W. Wyatt, Christopher M. Hindson, Rajiv Bharadwaj, Alexander Wong, Kevin D. Ness, Lan W. Beppu, H. Joachim Deeg, Christopher McFarland, Keith R. Loeb, William J. Valente, Nolan G. Ericson, Emily A. Stevens, Jerald P. Radich, Tarjei S. Mikkelsen, Benjamin J. Hindson & Jason H. Bielas - Show fewer authors

Nature Communications 8, Article number: 14049 (2017) | Cite this article

67k Accesses | 1210 Citations | 201 Altmetric | Metrics



10x Genomics Chromium platform

Uses microdroplet encapsulation for individual lysis and amplification

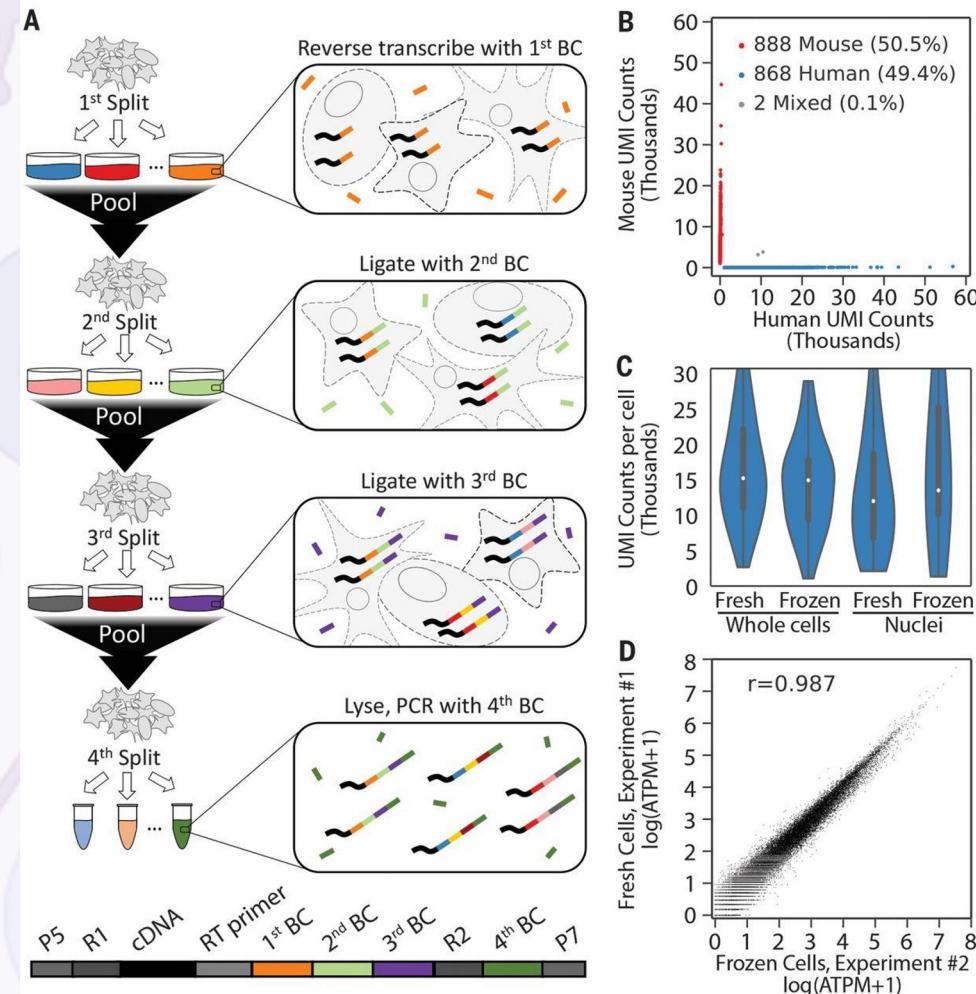
Current version is v3.1; has 3' and 5' sequencing variants

High throughput (5-10k cells/well)

The return of plate-based methods?

Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding

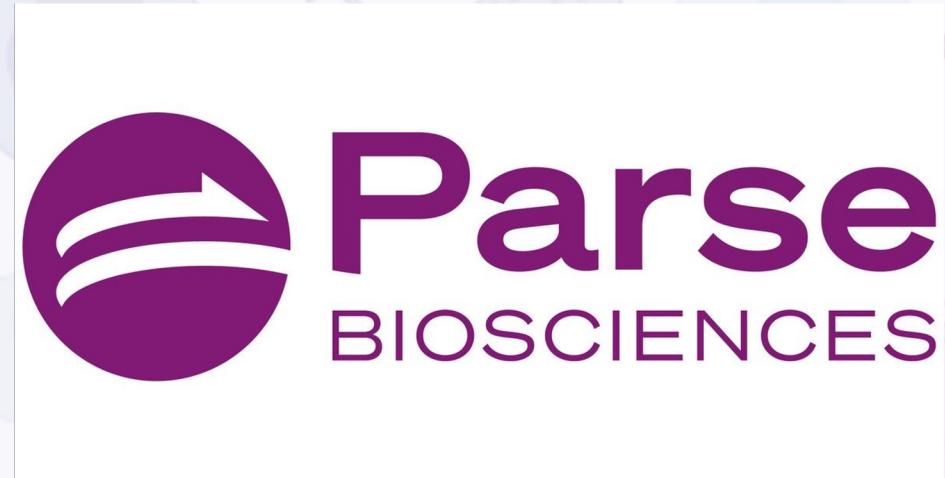
^{1,6} Alexander B. Rosenberg^{1,*†}, ^{1,6} Charles M. Roco^{2,*}, ^{1,6} Richard A. Muscat¹, ^{1,6} Anna Kuchina¹, ^{1,6} Paul Sample¹, ^{1,6} Zizhen Yao³, ^{1,6} Lucas T. Graybuck³, ^{1,6} David J. Peeler², ^{1,6} Sumit Mukherjee¹, ^{1,6} Wei Chen⁴, ^{1,6} Suzie H. Pun², ^{1,6} Drew L. Sellers^{2,5}, ^{1,6} Bosiljka Tasic³, ^{1,6} Georg Seelig^{1,4,6,†}



Pooling and splitting cells, and adding consecutive barcodes, will give a very close to unique cell barcoding

This allows for greater throughput with lower costs, as well as no specialised instrument required (just a 24/48/96 well plate)

Methods: SPLiT-seq, sci-RNA-seq

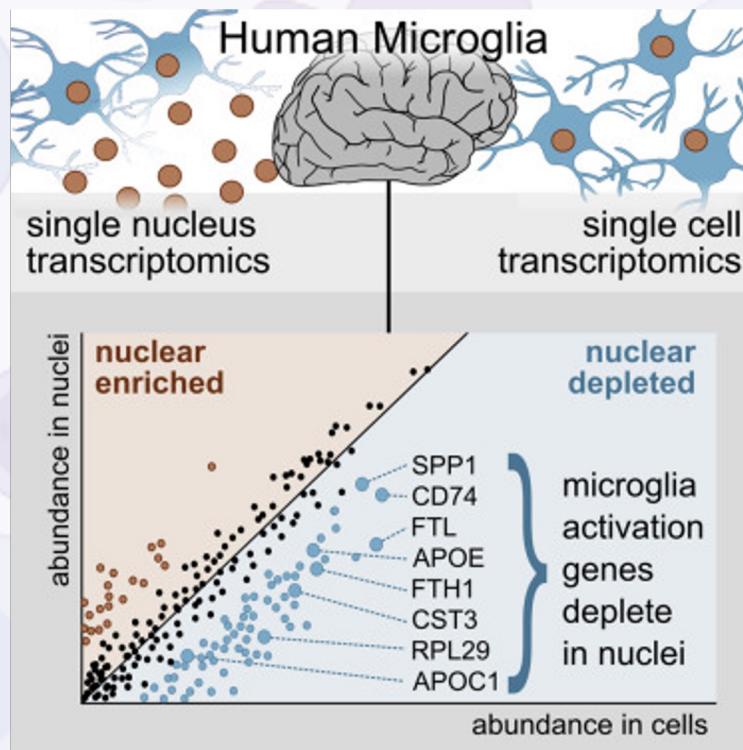


Single-cell vs single-nuclei

Report

Single-Nucleus RNA-Seq Is Not Suitable for Detection of Microglial Activation Genes in Humans

Nicola Thrupp ^{1, 2}, Carlo Sala Frigerio ^{1, 2, 3}, Leen Wolfs ^{1, 2}, Nathan G. Skene ⁴, Nicola Fattorelli ^{1, 2}, Suresh Poovathingal ^{1, 2}, Yannick Fourne ^{1, 2}, Paul M. Matthews ⁴, Tom Theys ⁵, Renzo Mancuso ^{1, 2}, Bart de Strooper ^{1, 2, 3, 6}                                                                   



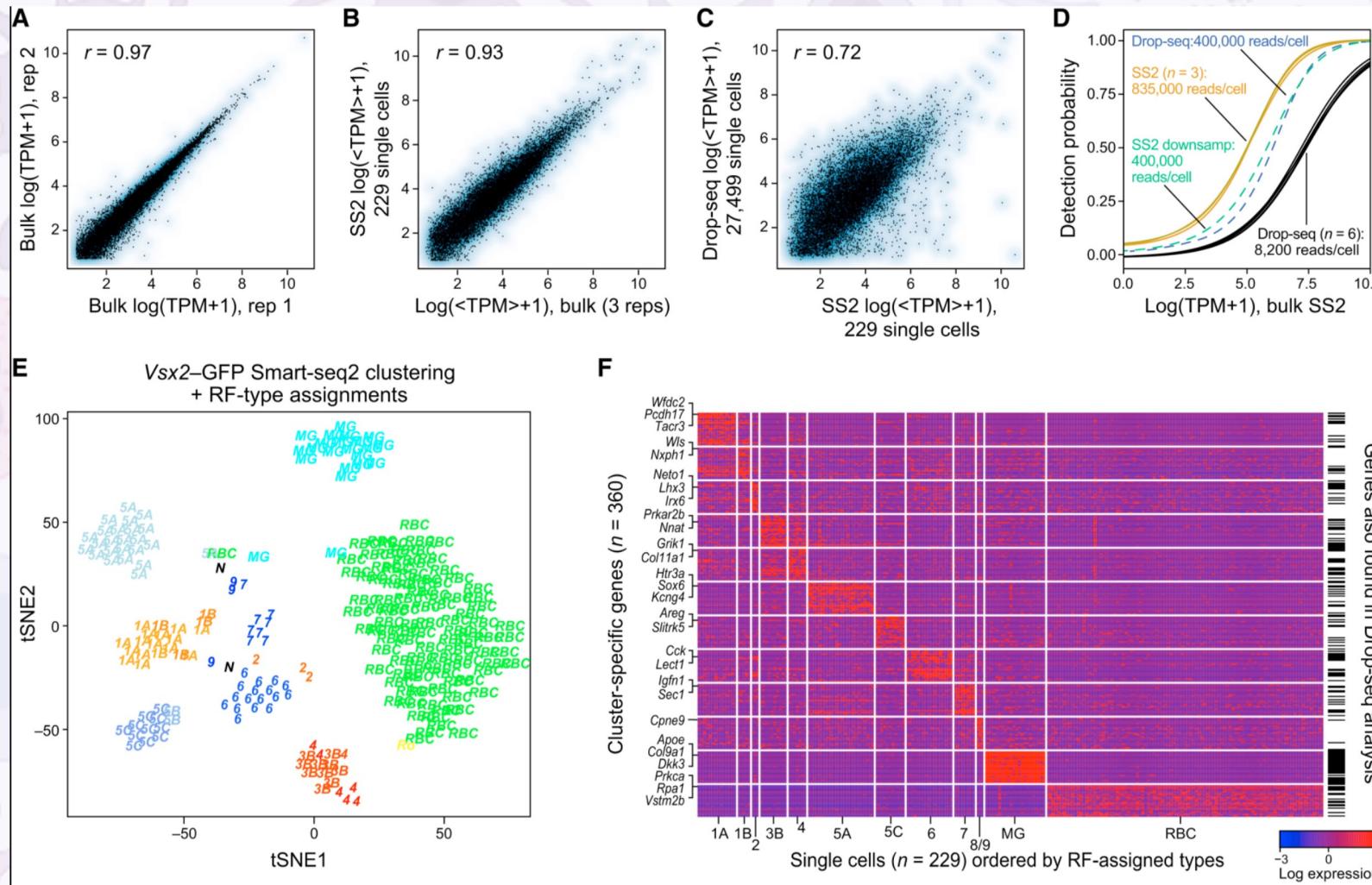
In some situations, it might be necessary to sequence single-nuclei instead of single-cells

Applications: frozen samples, oddly-shaped cells (e.g. neurons)

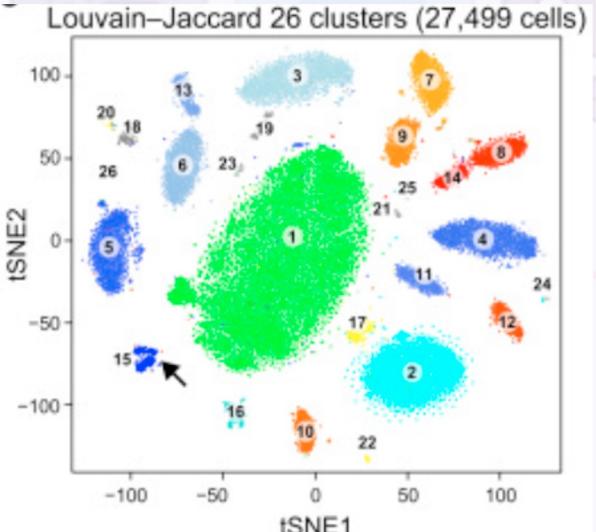
Usually RNA content and gene diversity is lower than whole cell, but still possible to identify cell populations

Transcripts in the nucleus are not exactly the same as those in the cytoplasm

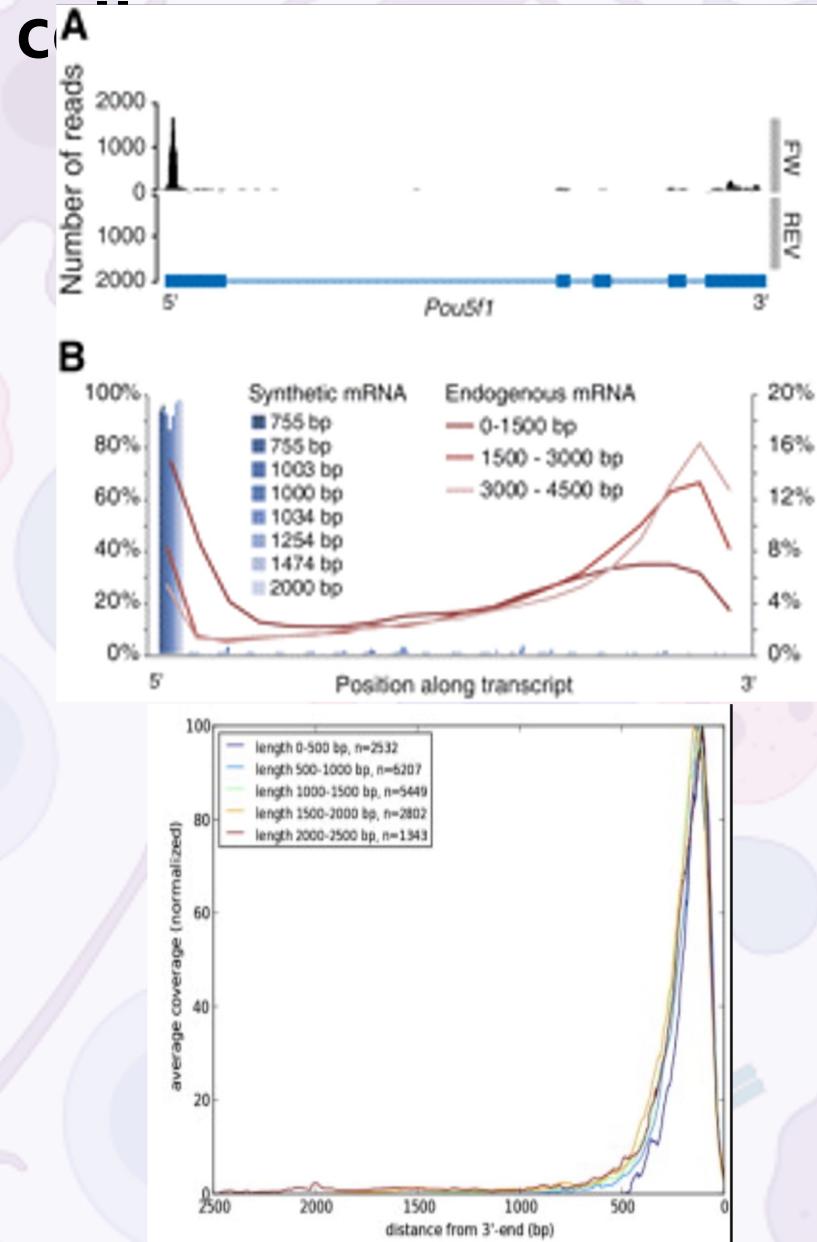
Identifying cell types from scRNA-seq



- **SS2 (plate)** experiments tend to be sequenced deeper per cell – **more genes detected**
- **Droplet** experiments usually have lower depth per cell, but **capture more cells**
- **To identify cell types, more cells is better than more genes**

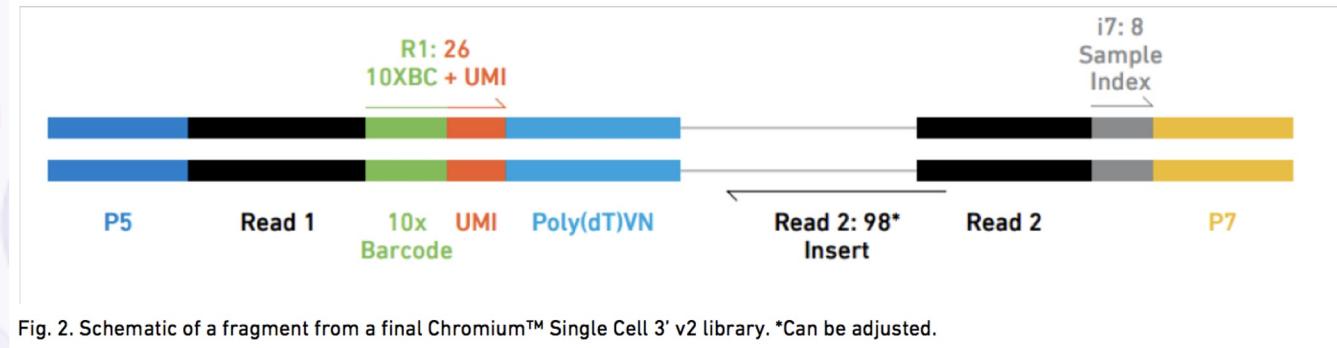


Isolating and amplifying RNA from single-



- 5' end capture (e.g. STRT-seq)
- 3' end capture (e.g. CEL-seq)
- Whole transcript capture (e.g. Smart-seq2)

Sequencing more than just mRNA



For highly multiplexed experiments, transcripts from each cell should be tagged with a cell barcode to allow sequencing in bulk and further deconvolution

To avoid counting the same transcript twice (usually caused by PCR duplicates), a Unique Molecular Identifier (UMI) is also added to the constructed libraries

This also means that counting UMI duplication can be a QC metric for the sample/sequencing

Common steps in a scRNA-seq experiment

scRNA-seq experiment	Bench work	Cell isolation
		RNA capture
	Computational work	Library construction
		Sequencing
		Quantification
		Quality control
		Normalisation
		Dimensionality reduction
		Clustering/Trajectory analysis
		Differential expression
		...

File formats

FASTQ – (decoded) sequencing output

Diagram illustrating the structure of FASTQ sequencing output:

- Label:** @FORJUSP02AJWD1
- Sequence:** CCGTCAATTCACTTAAGTTAACCTTGCAGCCGTACTCCCCAGGGGT
- Q scores (as ASCII chars):** +AAAAAAA:::99@:::?:?@:::FFAAAAACCAA:::BB@@?A?
- Base=T, Q=':'=25**

fastq header format (version > 1.8)

Sequence Header	+	Sequence ID								
a	b	c	d	e	f	g	h	i	j	k
@HWI-ST486:166:C06K9ACXX:7:1101:1443:1995	1:N:0:ACAGTG									

a. unique instrument name
b. run id
c. flowcell id
d. flowcell lane
e. tile number within the flowcell lane
f. x-coordinate of the cluster within the tile
g. y-coordinate of the cluster within the tile

h. the member of a pair, 1 or 2 (paired-end or mate-pair reads only)
i. Y if the read fails filter (read is bad), N otherwise
j. 0 when no control bits are on
k. index sequence

Count matrix

Cells

0	0	1	2	1	0	0	0
0	1	3	5	2	0	1	0
0	0	0	1	1	0	0	1
3	2	0	0	1	0	0	0
2	2	0	0	0	0	0	0
1	1	1	0	0	3	1	2
1	2	0	0	0	1	1	1
1	2	0	0	0	2	2	2
2	1	0	0	0	4	4	3
0	0	0	0	0	2	1	1
1	0	0	0	0	2	3	1
0	0	0	0	0	2	1	1

Genes

File formats

FASTQ – (decoded) sequencing output

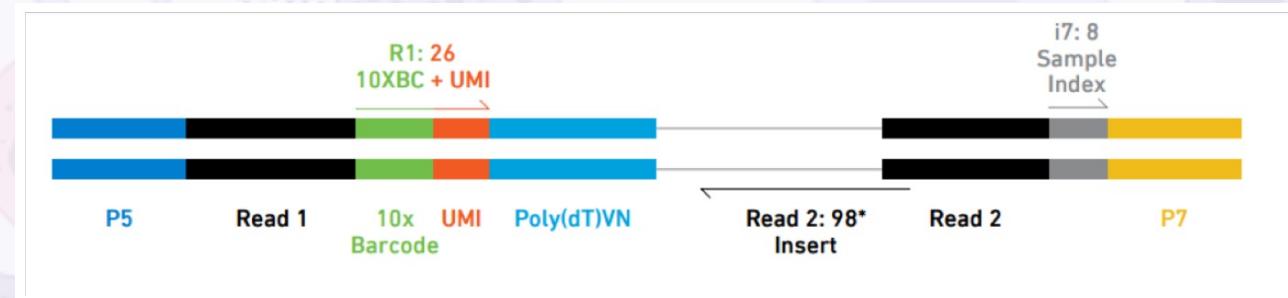
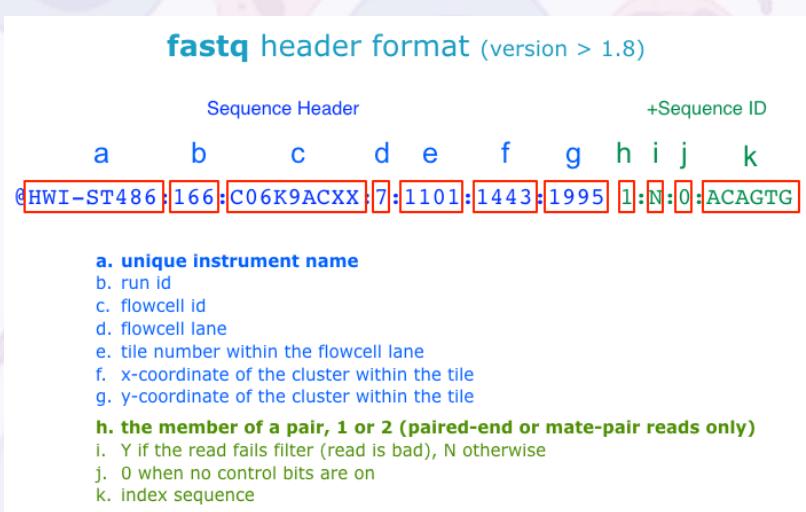
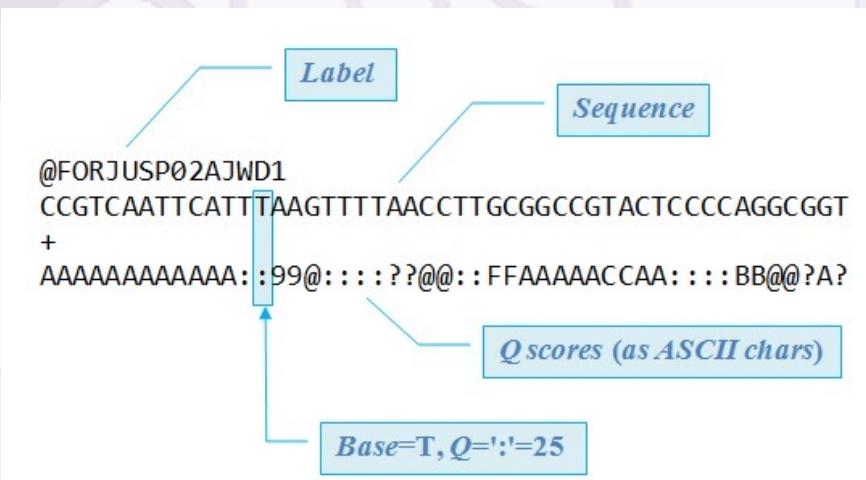


Fig. 2. Schematic of a fragment from a final Chromium™ Single Cell 3' v2 library. *Can be adjusted.

Number of fastq files received depends on method

In library construction, sample barcodes are added to then separate them computationally (usually as part of the sequencing service)

- For plate-based methods, each sample is a well, and therefore a cell
- For droplet based methods, a sample is a mix of cells

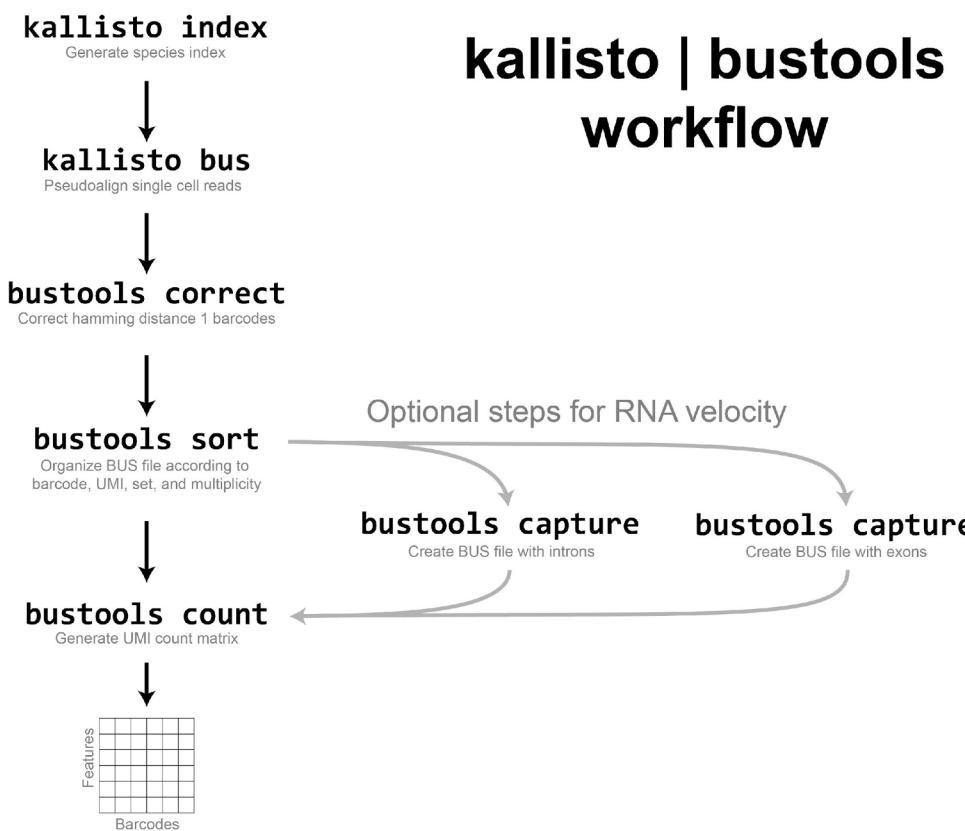
Cells are deconvolved from their cell barcode (as part of the pipeline)

Quantification and demultiplexing – kallisto-bustools

Letter | Published: 01 April 2021

Modular, efficient and constant-memory single-cell RNA-seq preprocessing

Páll Melsted, A. Sina Booeshaghi, Lauren Liu, Fan Gao, Lambda Lu, Kyung Hoi (Joseph) Min, Eduardo da Veiga Beltrame, Kristján Eldjárn Hjörleifsson, Jase Gehring & Lior Pachter 



Kallisto does pseudoalignment of reads to the transcriptome

This is more efficient – faster, low memory and CPU footprint (should work on a standard laptop!)

Kallisto individually is sufficient for plate-based scRNA-seq

BUSTools processes/deconvolutes scRNA-seq data from microdroplet methods

Salmon-Alevin are similar methods

Quantification and demultiplexing - CellRanger

Open Access | Published: 16 January 2017

Massively parallel digital transcriptional profiling of single cells

Grace X. Y. Zheng, Jessica M. Terry, Phillip Belgrader, Paul Ryvkin, Zachary W. Bent, Ryan Wilson, Solongo B. Ziraldo, Tobias D. Wheeler, Geoff P. McDermott, Junjie Zhu, Mark T. Gregory, Joe Shuga, Luz Montesclaros, Jason G. Underwood, Donald A. Masquelier, Stefanie Y. Nishimura, Michael Schnall-Levin, Paul W. Wyatt, Christopher M. Hindson, Rajiv Bharadwaj, Alexander Wong, Kevin D. Ness, Lan W. Beppu, H. Joachim Deeg, Christopher McFarland, Keith R. Loeb, William J. Valente, Nolan G. Ericson, Emily A. Stevens, Jerald P. Radich, Tarjei S. Mikkelsen, Benjamin J. Hindson & Jason H. Bielas - Show fewer authors

Nature Communications 8, Article number: 14049 (2017) | Cite this article

67k Accesses | 1210 Citations | 201 Altmetric | Metrics

A command-line software suite to quantify (and analyse) data from 10x Genomics scRNA-seq/snRNA-seq

Based on STAR for alignment

Parses Cell barcodes and UMIs

Detects real cells and empty droplets

Performs some default analysis



Products Research Areas Resources Support Company

Support > Single Cell Gene Expression > Software

SEARCH

Q&A CONTACT SUPPORT

SOFTWARE > PIPELINES

CELL RANGER

Introduction

- [What is Cell Ranger?](#)
- [What is Feature Barcode Data?](#)

What is Cell Ranger?

Cell Ranger is a set of analysis pipelines that process Chromium single-cell data to align reads, generate feature-barcode matrices, perform clustering and other secondary analysis, and more. Cell Ranger includes four pipelines relevant to the 3' Single Cell Gene Expression Solution and related products:

```
$ cellranger count --id=sample345 \
--transcriptome=/opt/refdata-gex-GRCh38-2020-A \
--fastqs=/home/jdoe/runs/HAWT7ADXX/outs/fastq_path \
--sample=mysample \
--expect-cells=1000 \
--localcores=8 \
--localmem=64
```

<https://www.nature.com/articles/ncomms14049>

<https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger>

Running CellRanger

```
$ cellranger count --id=sample345 \
--transcriptome=/opt/refdata-gex-GRCh38-2020-A \
--fastqs=/home/jdoe/runs/HAWT7ADXX/outs/fastq_path \
--indices=SI-3A-A1
```

Martian Runtime - v4.0.8

Running preflight checks (please wait)...

2016-01-01 10:23:52 [runtime] (ready) ID.sample345.CELLRANGER_CS.CELLRANGER.SETUP_CHUNKS ...

2012-01-01 12:10:09 [runtime] (join_complete)

ID.sample345.CELLRANGER_CS.CELLRANGER.SUMMARIZE_REPORTS

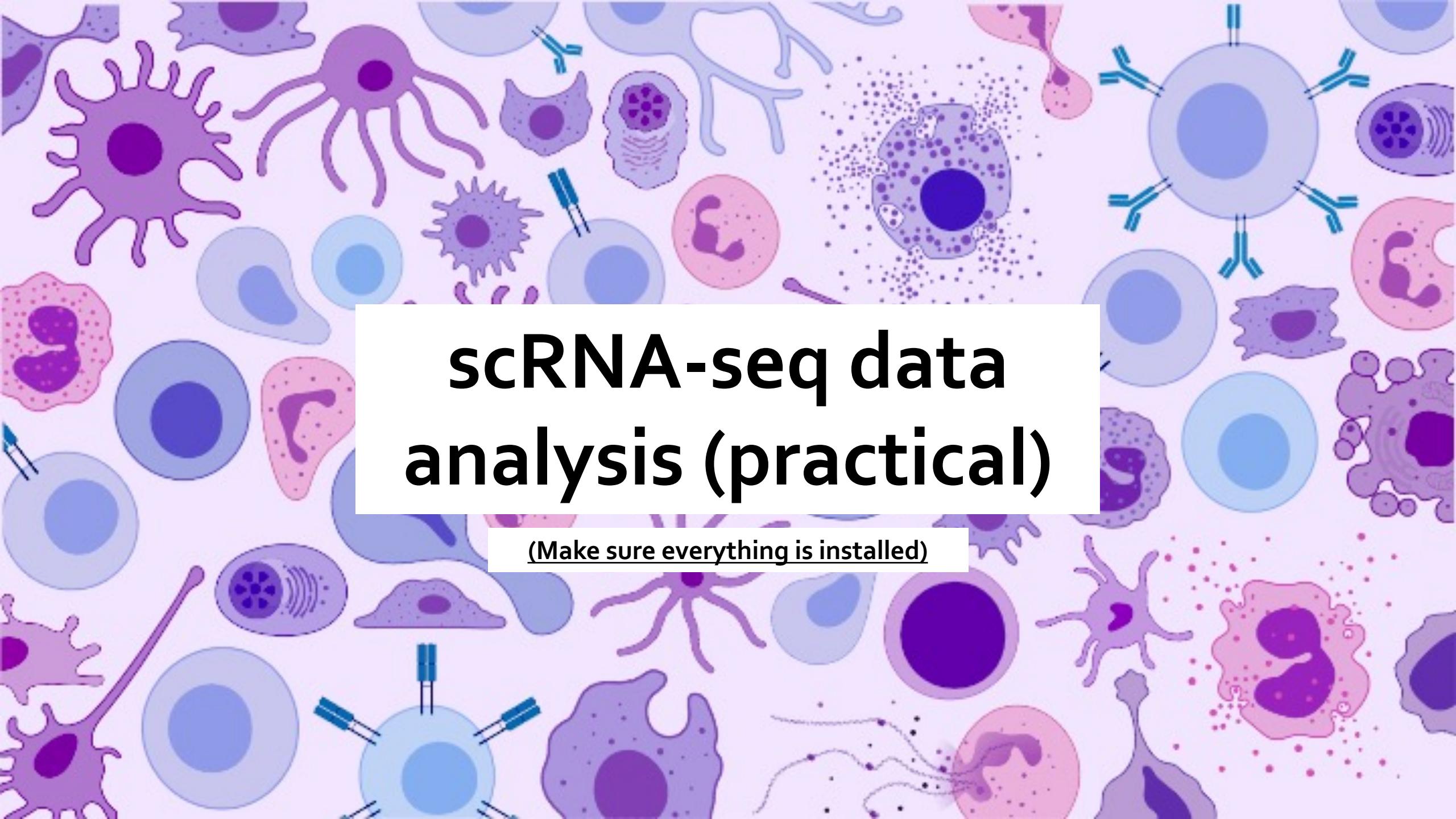
Outputs:

- Run summary HTML: /home/jdoe/runs/sample345/outs/web_summary.html
- Run summary CSV: /home/jdoe/runs/sample345/outs/metrics_summary.csv
- BAM: /home/jdoe/runs/sample345/outs/possorted_genome_bam.bam
- BAM index: /home/jdoe/runs/sample345/outs/possorted_genome_bam.bam.bai
- Filtered feature-barcode matrices MEX: /home/jdoe/runs/sample345/outs/filtered_feature_bc_matrix
- Filtered feature-barcode matrices HDF5: /home/jdoe/runs/sample345/outs/filtered_feature_bc_matrix.h5
- Unfiltered feature-barcode matrices MEX: /home/jdoe/runs/sample345/outs/raw_feature_bc_matrix
- Unfiltered feature-barcode matrices HDF5: /home/jdoe/runs/sample345/outs/raw_feature_bc_matrix_h5.h5
- Secondary analysis output CSV: /home/jdoe/runs/sample345/outs/analysis
- Per-molecule read information: /home/jdoe/runs/sample345/outs/molecule_info.h5
- Loupe Browser file: /home/jdoe/runs/sample345/outs/cloupe.cloupe

Pipestance completed successfully!

Activity – reading CellRanger outputs

1. Obtain the files from the following link:
<https://drive.google.com/drive/folders/1MaOHyFuKlvLtErJev4ctYb6LASA5CLvi?usp=sharing>
2. Explore the web_summary.html output file. Which are the most important metrics to determine whether your experiment was successful? Why?
3. Explore the clouf file (for this, Loupe Browser installation is necessary).
 1. What does each dot represent on the graph shown?
 2. What is the top marker for cluster 6?
 3. Which clusters express the gene CD3E?



scRNA-seq data analysis (practical)

(Make sure everything is installed)

Two major platforms for scRNA-seq analysis



```
library(Seurat)

# load data and create Seurat object
pbmc.data = Read10X(data.dir = "../data/pbmc3k/filtered_gene_bc_matrices/hg19/")
pbmc = CreateSeuratObject(counts = pbmc.data, min.cells = 3, min.features = 200)

# normalise data
pbmc = NormalizeData(pbmc, normalization.method = "LogNormalize", scale.factor =
10000)

# find highly variable genes
pbmc = FindVariableFeatures(pbmc, selection.method = "vst", nfeatures = 2000)

# scale data
pbmc = ScaleData(pbmc, features = rownames(pbmc))

# linear dimensionality reduction
pbmc = RunPCA(pbmc, features = VariableFeatures(object = pbmc))

# clustering
pbmc = FindNeighbors(pbmc, dims = 1:40)
pbmc = FindClusters(pbmc, resolution = 0.5)

# non-linear dimensionality reduction
pbmc = RunUMAP(pbmc, dims = 1:10)

# find marker genes
pbmc.mk = FindAllMarkers(pbmc, only.pos = TRUE, min.pct = 0.25, logfc.threshold =
0.25)
```

<https://satijalab.org/seurat/>



```
import scanpy as sc

# load data and create AnnData (scanpy) object
adata = sc.read_10x_mtx('data/filtered_gene_bc_matrices/hg19/',
var_names='gene_symbols')

# normalise data
sc.pp.normalize_total(adata, target_sum=1e4)
sc.pp.log1p(adata)

# find highly variable genes
sc.pp.highly_variable_genes(adata, min_mean=0.0125, max_mean=3,
min_disp=0.5)

# scale data
sc.pp.scale(adata, max_value=10)

# linear dimensionality reduction
sc.tl.pca(adata, svd_solver='arpack')

# clustering
sc.pp.neighbors(adata, n_neighbors=10, n_pcs=40)
sc.tl.leiden(adata)

# non-linear dimensionality reduction
sc.tl.umap(adata)

# find marker genes
sc.tl.rank_genes_groups(adata, 'leiden', method='t-test')
```

<https://scanpy.readthedocs.io/en/stable/> <https://www.sc-best-practices.org/preamble.html>

Activity – analyse scRNA-seq data in R

Obtain the scripts and data from the following link: https://drive.google.com/drive/folders/1io-klkw_m7QU1OOIAjNhQzuRXfHUyiXT?usp=sharing

Each of you will be assigned a dataset randomly. At the end of this session, you should be able to say:

- What cell types are in your dataset;
- What are the main markers for these cell types;
- What tissue do you think it is;
- What species you are working with

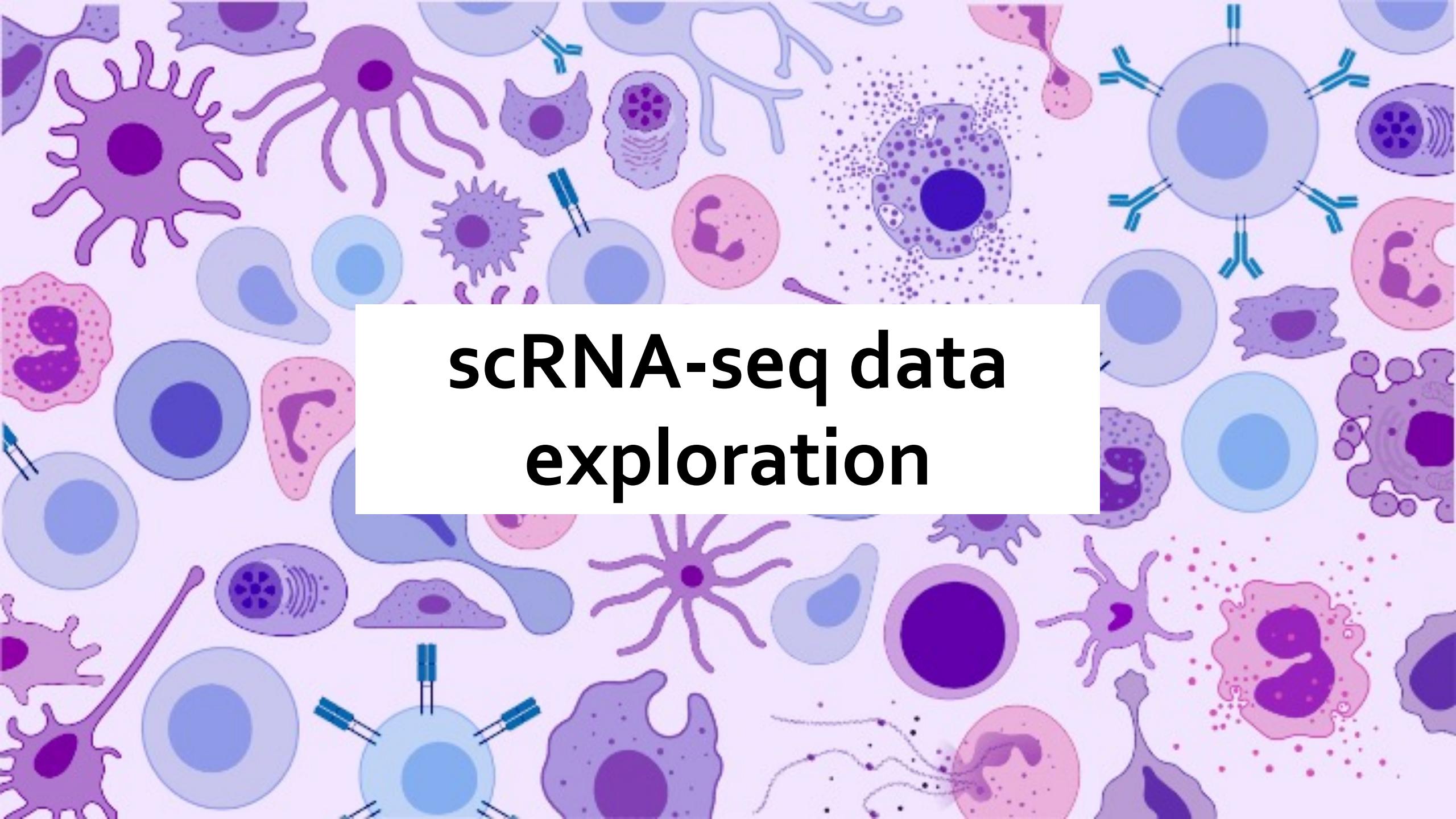
Activity – SOLUTION

dataset	species	tissue
1		
2		
3		

Activity – analyse scRNA-seq data in Python

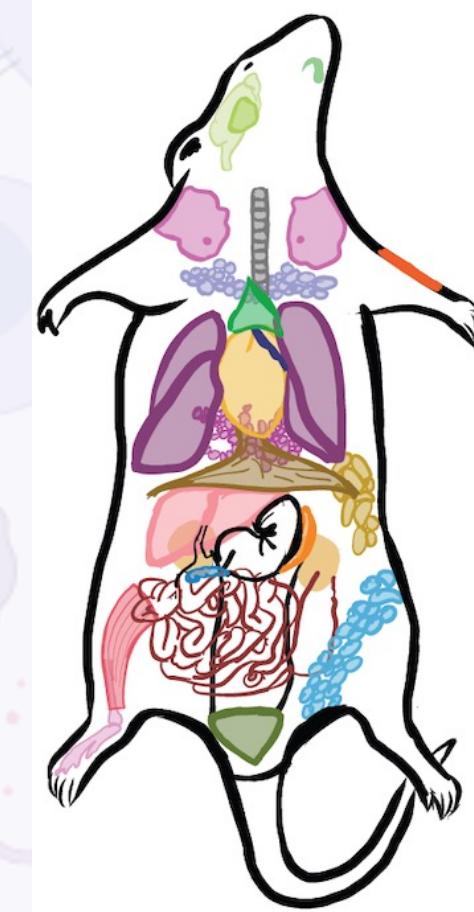
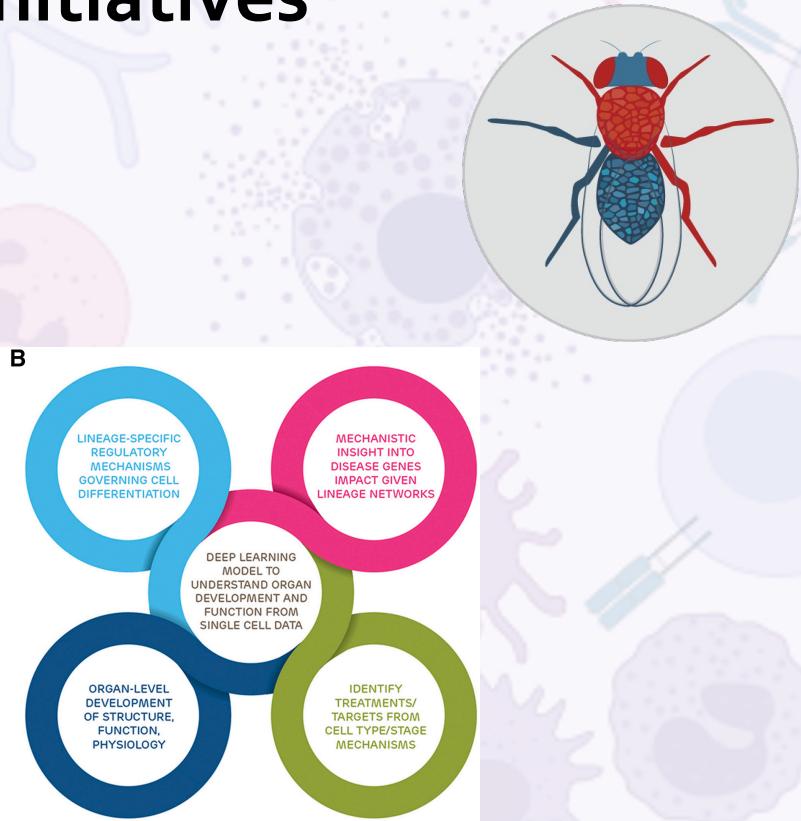
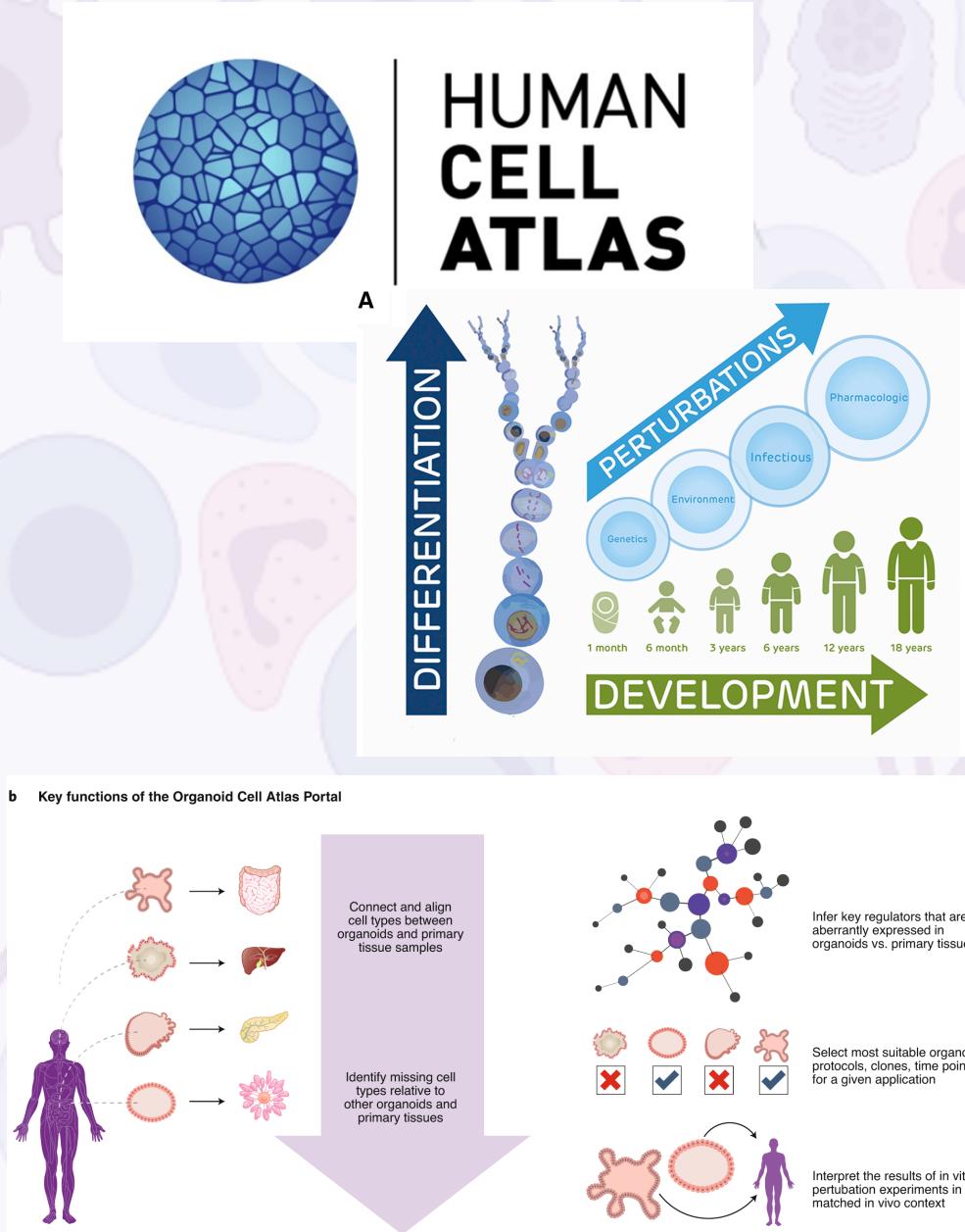
Access and copy the following colab notebook:

https://colab.research.google.com/drive/1_p6wLEW7XkoQnsuzsGSOaNqoVbyxxYwB?usp=sharing



scRNA-seq data exploration

Uses of scRNA-seq – atlasing initiatives



<https://www.sciencedirect.com/science/article/pii/S1534580719301789>
<https://cdn.elife sciences.org/articles/27041/elife-27041-v2.pdf>
<https://www.nature.com/articles/s41587-020-00762-x>

Activity – explore Human Cell Atlas data

1. Go to <https://data.humancellatlas.org/> and choose a dataset
 1. Can you colour the UMAP by cell type? And by disease?
 2. Can you colour the UMAP based on a gene of choice?
 3. Which cell types express your gene of choice?