מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

Introduction to Single-cell RNA-sequencing (scRNA-seq) Analysis

Dena Leshkowitz

An Introduction to Deep-Sequencing Data Analysis For Biologists 2019-2020

Agenda

- Introduction to single cell RNA-Seq (scRNA-Seq)
- Gil – Bioinformatics analysis of scRNA
- scRNA-Seq example: Single-cell molecular and cellular architecture of the mouse neurohypophysis

Bulk RNA-Seq Versus Single Cell RNA-Seq

Bulk RNA-Seq

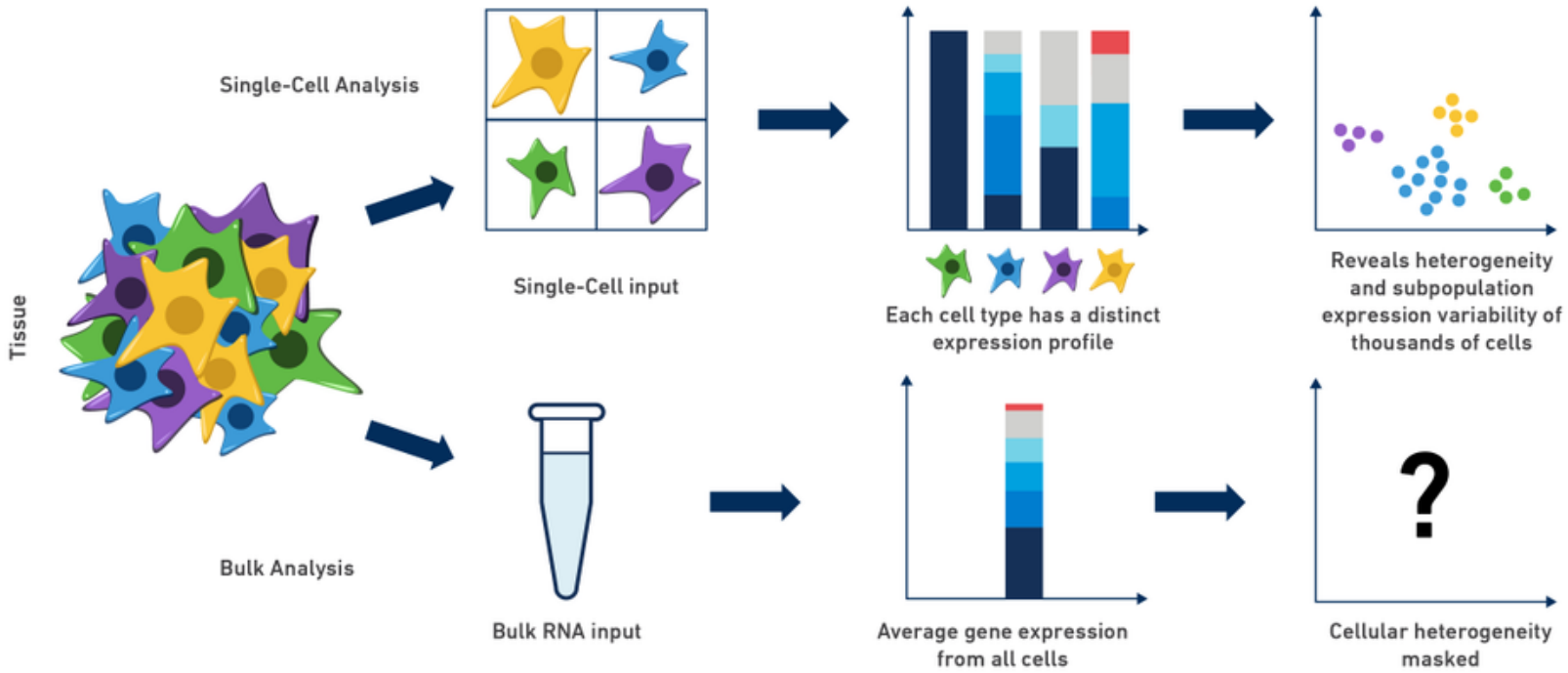
Taste is the average of all the fruit

scRNA-Seq

Taste each fruit individually

Categorize

Bulk RNA Versus scRNA-Seq



community.10xgenomics.com

Bulk RNA-Seq

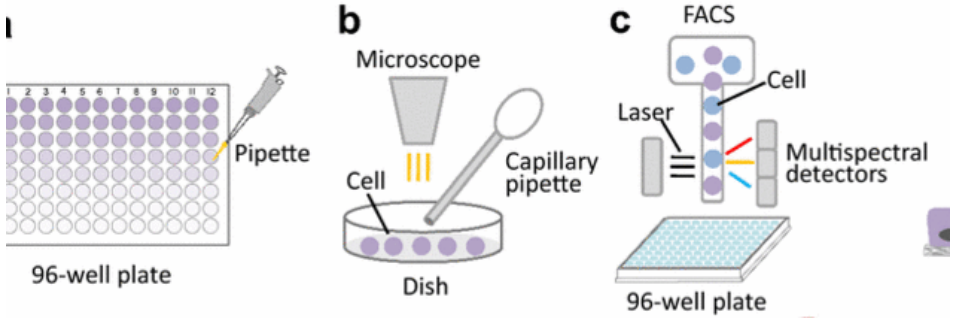
- Measures the average expression level for each gene across a large population of input cells
- Useful for comparative transcriptomics (different conditions, tissues, genotypes…)
- Insufficient for studying heterogeneous systems, such as complex tissues (brain) due to the difficulty of studying a specific population within the system
- Does not provide insights into the stochastic nature of gene expression

Methods for Cell Capture

Low-throughput

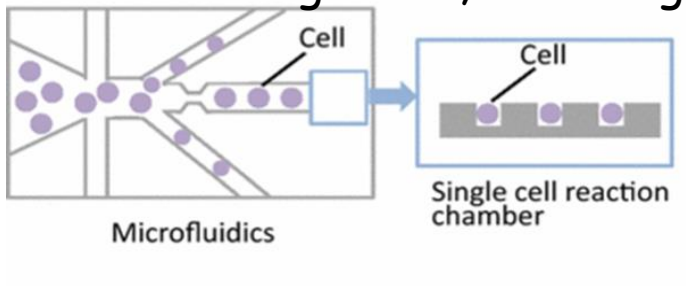
Well-based platform

Cells are isolated using for example pipette or laser capture or FACS and placed in microfluidic wells

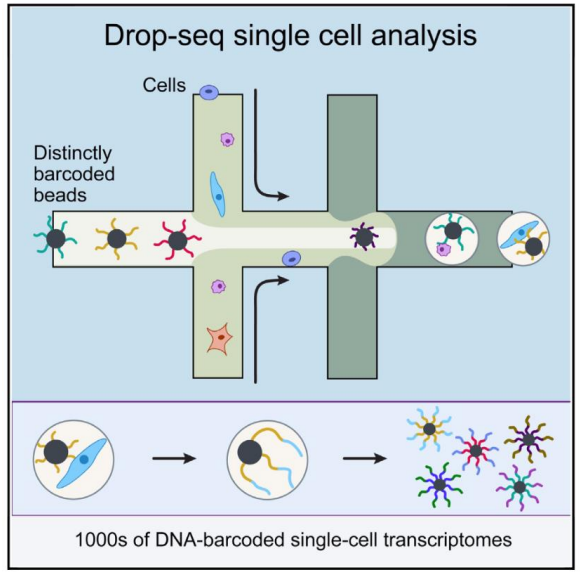


Microfluidic platforms

Such as Fluidigm's C1, advantage small volume

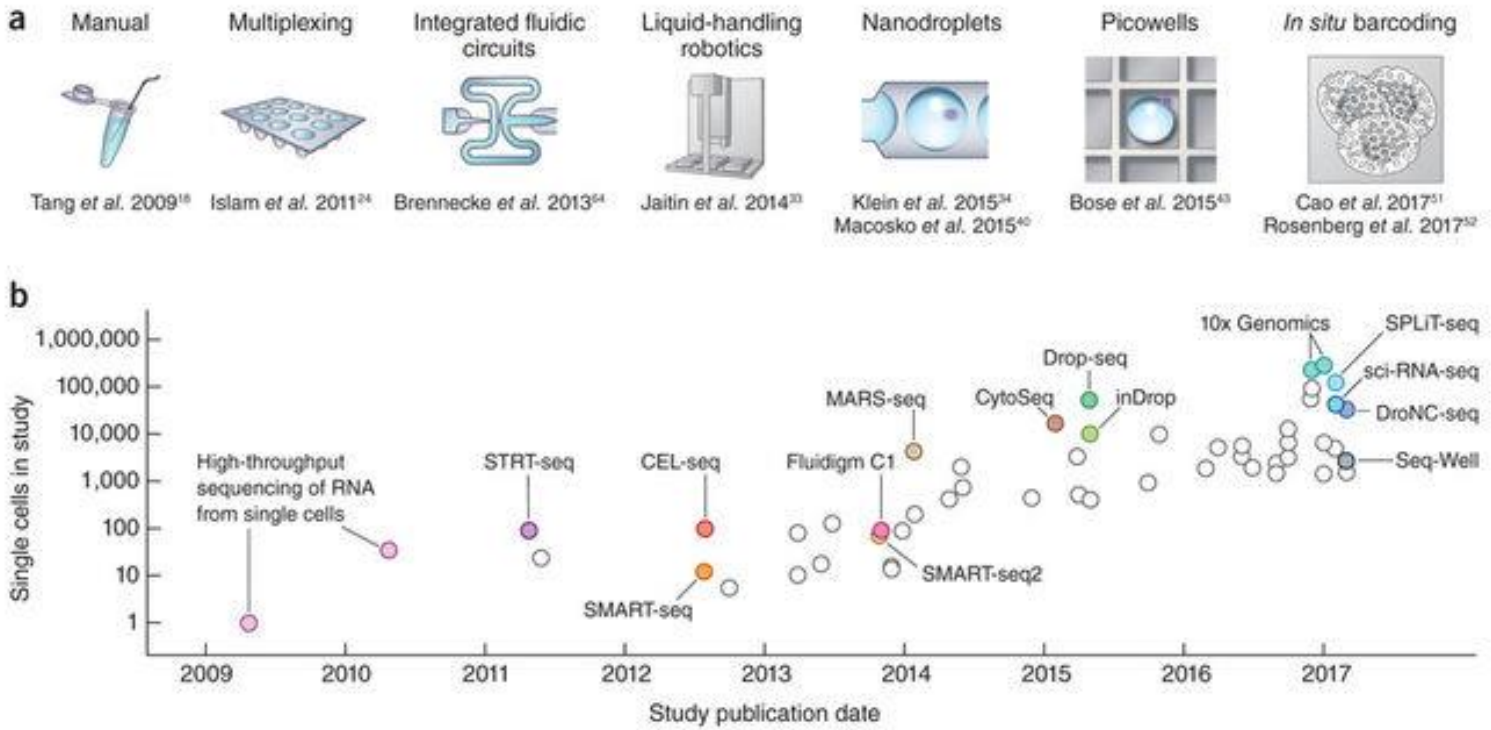


High-throughput Droplet based platform



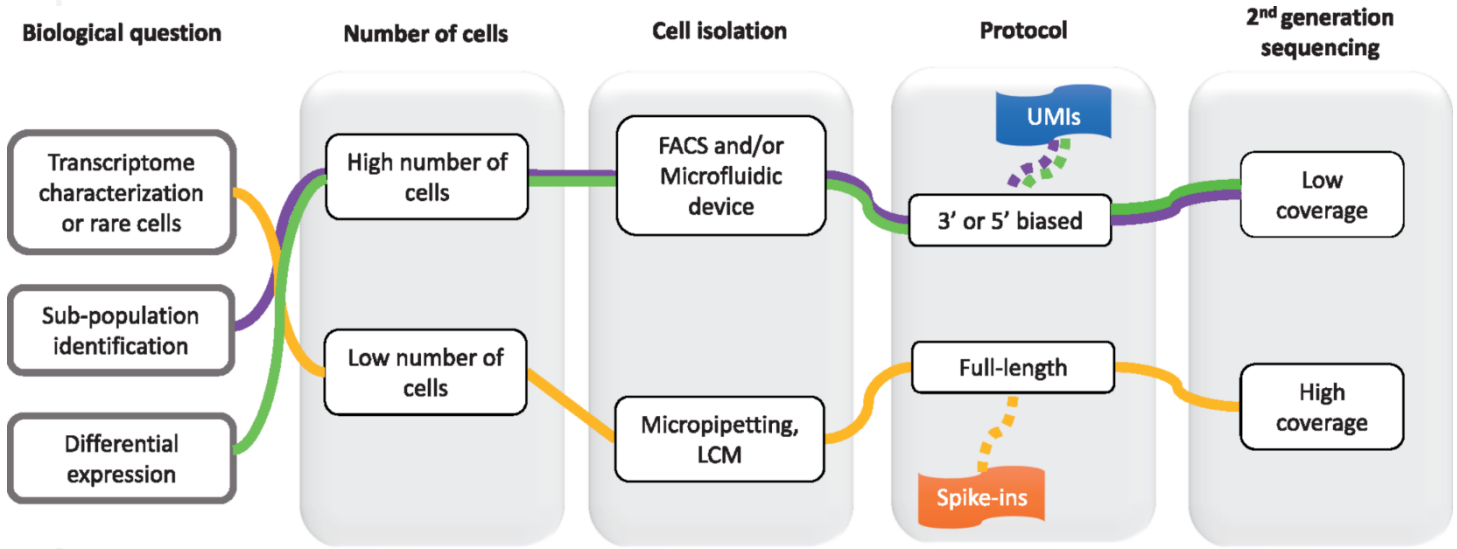
- Each individual cell is captured inside a nanoliter droplet together with a bead.
- The bead is loaded with the enzymes required to construct the library and unique barcode
- All of the droplets can be pooled and sequenced together

Exponential Growth of Scale in Number of Cells per scRNA-seq experiment



Svensson et al. Nature Protocols volume 13, pages 599–604 (2018)

Major Differences in Single Cell Platforms



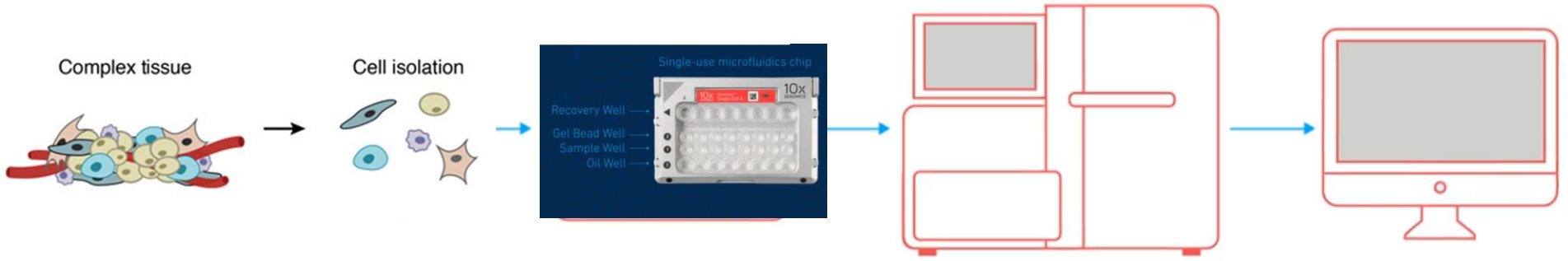
From: How to design a single-cell RNA-sequencing experiment: pitfalls, challenges and perspectives

Brief Bioinform. Published online January 31, 2018. doi:10.1093/bib/bby007

Spike-in : Extrinsic molecules added to the cell lysate. Used to filter empty wells & normalize

scRNA-Seq Workflow (Sandbox LSCF)
10X Genomics - Chromium™

Movie - https://www.10xgenomics.com/solutions/single-cell/



Microfluidic Chip

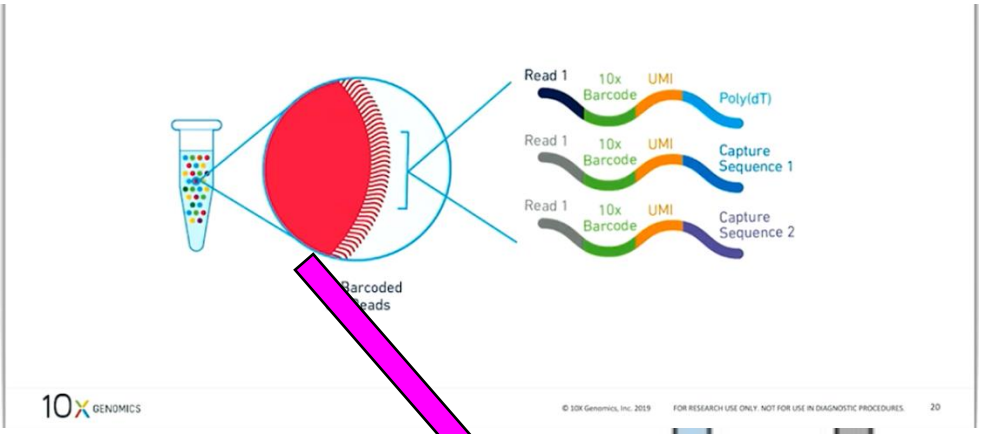
Library creation and sequencing

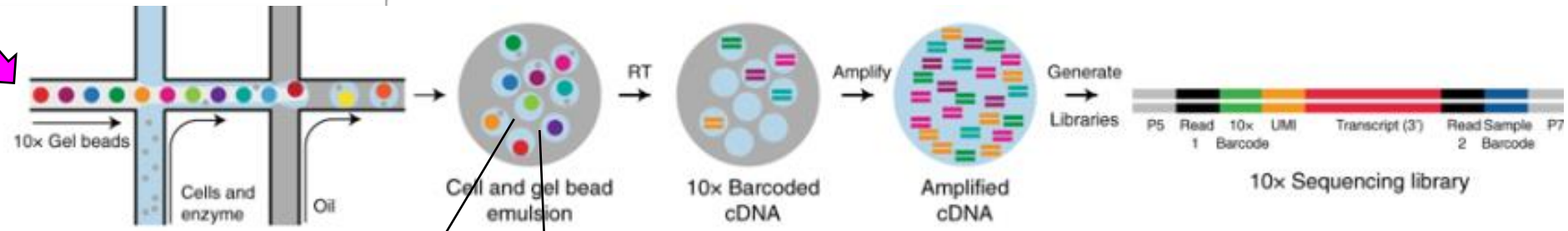
Bioinformatics analysis

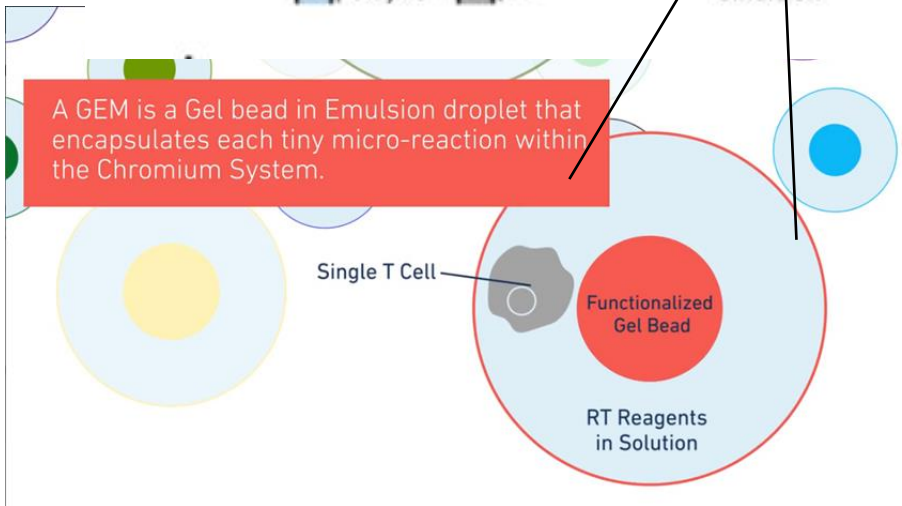
The microfluidic chip enables to:

- Capture single cells up to 10,000 per chamber in ~6 min
- Prepare cell-barcoded cDNA libraries to sequence with Illumina machines

From a Cell to Sequencing Library







From Library construction to Sequencing and Data analysis

After additional steps we have a library compatible with Illumina sequencing

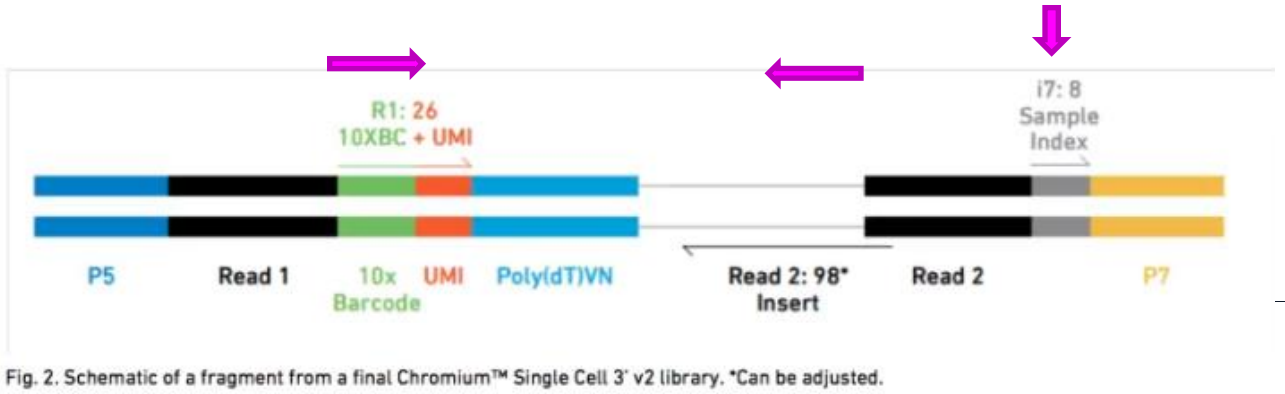


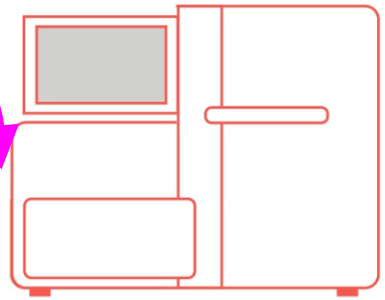
Fig. 2. Schematic of a fragment from a final Chromium™ Single Cell 3' v2 library. *Can be adjusted.

Read 1: 26 bases

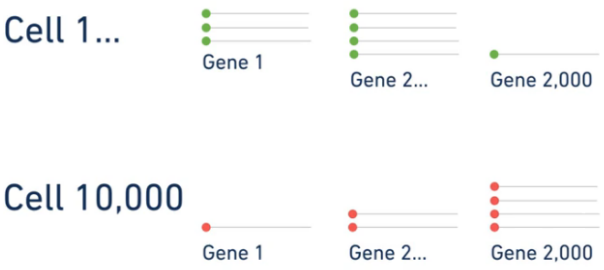
- 10 bases of cell barcode
- 16 bases of UMI

Read 2: contains the cDNA insert 3' end

After sequencing we separate the sequences per sample and cell



...enables massive transcription profiling of thousands of individual cells...



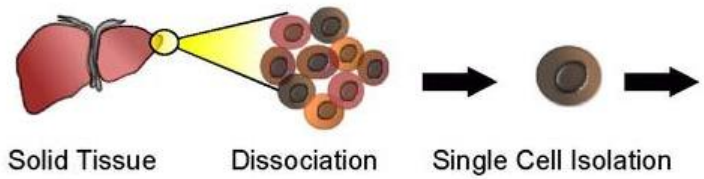
Challenges

Despite its power and high resolution, scRNA-seq has some open challenges related to the higher level of technical noise and data complexity with respect to bulk RNA-seq :

- Need for Amplification (up to 1 million fold) : The amount of RNA present in a single cell is limited, and ranges from 1–to 50 pg depending on cell type.
  - Solution - UMIs
- Gene 'dropouts' : High chance of missing nonzero gene expression values detected as zero.
  - Because of sampling 5-20% of transcripts.
- Doublets : some droplets/wells contain more than a single cell (two cells may be physically captured together)
  - 0.9% per 1000 cells; 7.6% per 10,000 cells (10X)
- Empty droplets: need to distinguish cells from empty droplets
  - In 10X technology majority of the droplets – we have ~3M barcodes

Why is their a problem to distinguish cells from empty droplets?

Experiment Design



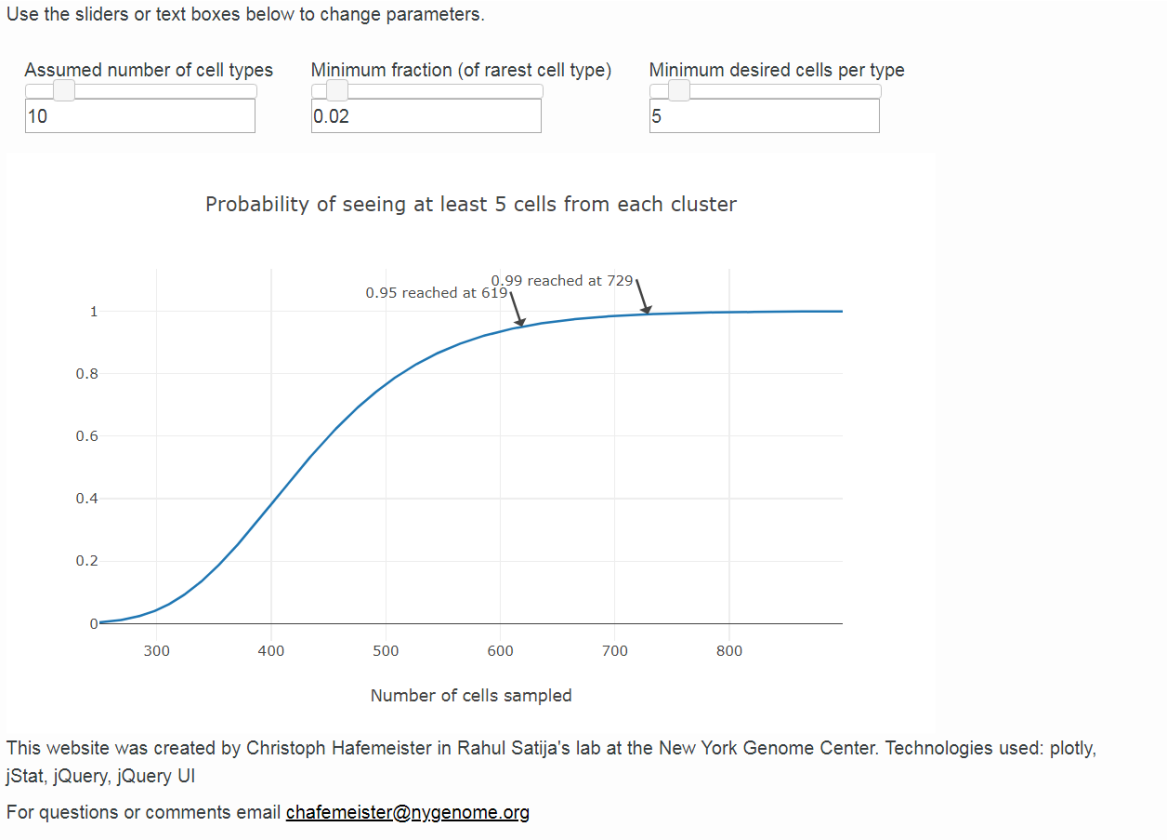
- Before conducting a single-cell experiment, trial single-cell separation methods and assess cell viability
- Cell isolation should be performed as fast as possible with all downstream work being carried out on ice
- There are cell-type-specific differences in recovery after FACS sorting, possibly because of cell size, with larger cells resulting in fewer cells recovered
- Avoid batch effects between experiments

Deciding on Appropriate Cell Number

The required number of cells can be estimated based on the expected heterogeneity of the population using a negative binomial distribution.

- Interactive tool

https://satijalab.org/howmanycells



Agenda

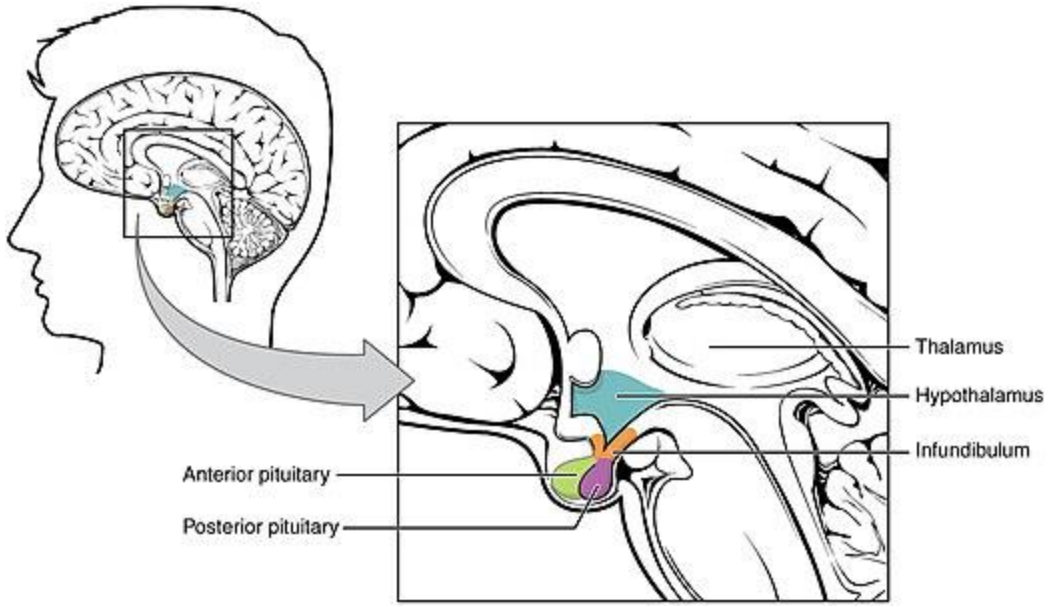
- Introduction to single cell RNA-Seq (scRNA-Seq)
- Gil – Bioinformatics analysis of scRNA
- scRNA-Seq example: Single-cell molecular and cellular architecture of the mouse neurohypophysis

Single-cell molecular and cellular architecture of the mouse neurohypophysis

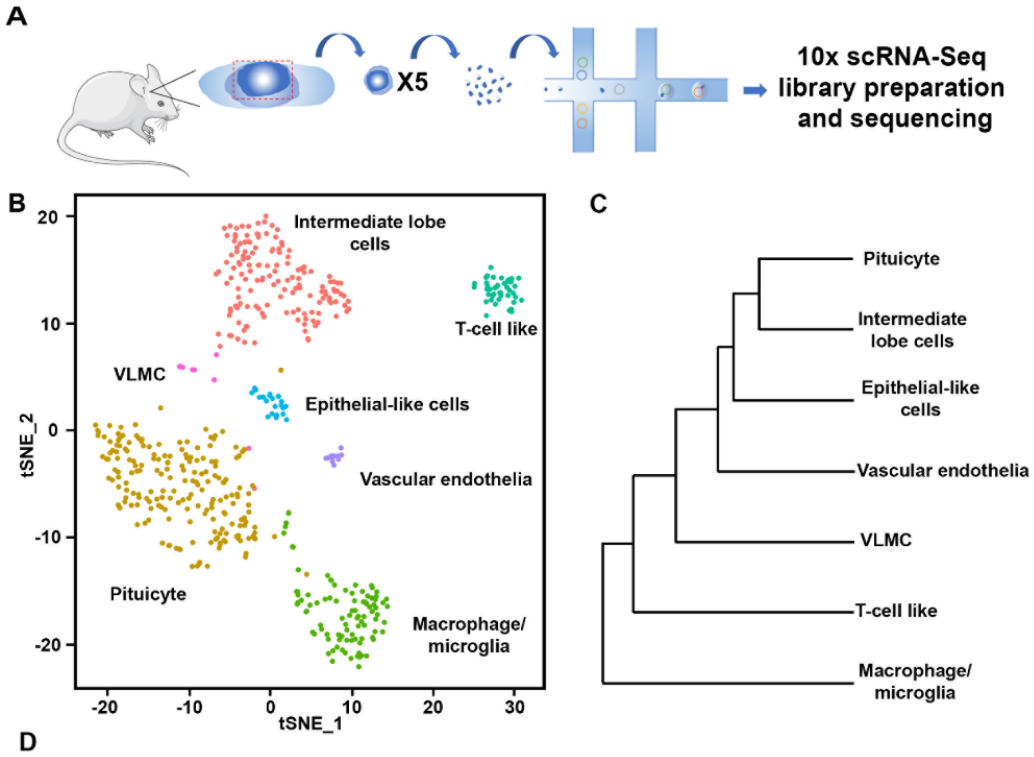
Qiyu Chen, Dena Leshkowitz, Janna Blechman, and Gil Levkowitz

Accepted to eNeuro

- Utilizing single cell RNA sequencing technology to identify cell populations in the neurohypophysis (NH) also named posterior pituitary



Single-cell RNA-Seq reveals seven cell types

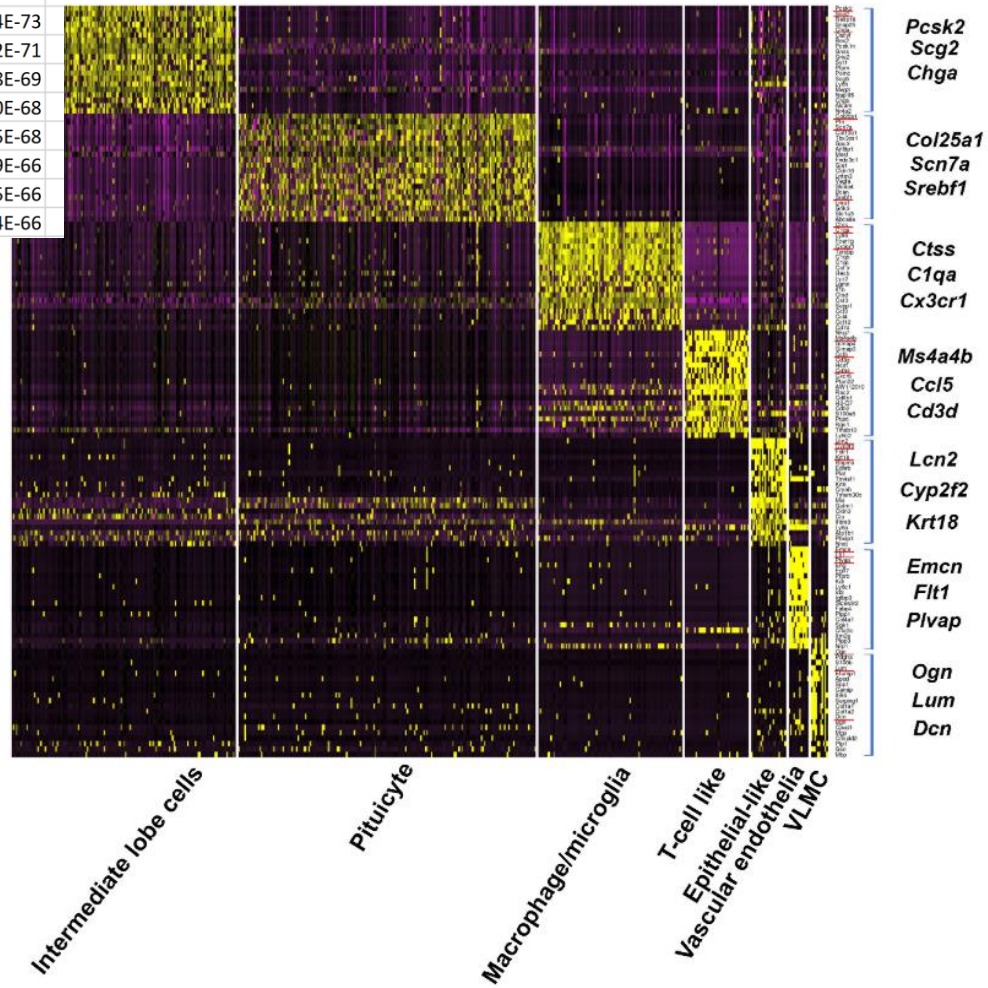


D

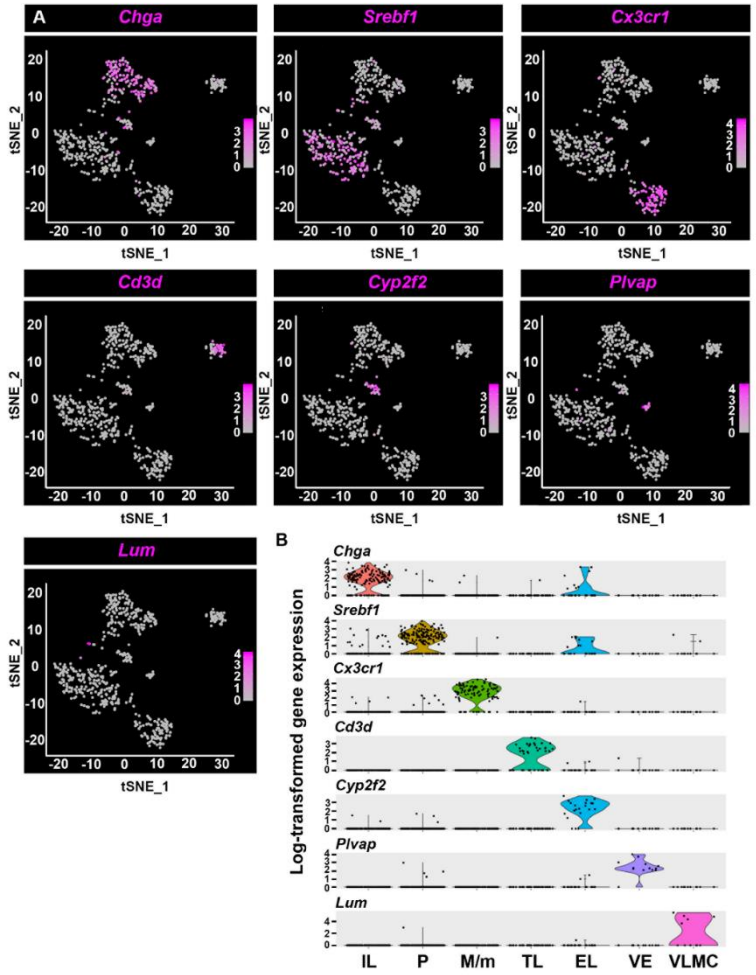
Cell type	No. of cells	Mean No. of genes	Mean No. of UMI
Pituicyte	196	1188	2496
Intermediate lobe cells	148	1524	3587
Macrophage/microglia	95	905	2069
T-cell like	42	797	2000
Epithelial-like cells	23	1608	3410
Vasculature endothelia	13	952	1833
VLMC	11	1046	2144

Heatmap of differentially expressed genes

	A	B	C	D	E	F	G
1	**cluster**	**gene**	**avg_logFC**	**p_val_adj**	**pct.1**	**pct.2**	**p_val**
2	intermediate lobe cells	Pcsk2	3.610686698	2.64E-83	0.878	0.037	1.75E-87
3	intermediate lobe cells	Scg2	1.963521122	1.73E-76	0.851	0.034	1.15E-80
4	intermediate lobe cells	Resp18	2.024051796	1.70E-71	0.831	0.045	1.13E-75
5	intermediate lobe cells	Snap25	1.871651061	1.57E-69	0.764	0.021	1.04E-73
6	intermediate lobe cells	Chga	1.905223518	2.74E-67	0.797	0.039	1.82E-71
7	intermediate lobe cells	Oacyl	2.207470522	4.33E-65	0.703	0.011	2.88E-69
8	intermediate lobe cells	Bex2	2.016276723	4.05E-64	0.878	0.134	2.70E-68
9	intermediate lobe cells	Pcsk1n	2.247596475	1.24E-63	0.946	0.271	8.25E-68
10	intermediate lobe cells	Gnas	1.9419483	7.80E-62	0.986	0.532	5.19E-66
11	intermediate lobe cells	Gria2	1.69800663	1.06E-61	0.676	0.011	7.05E-66
12	intermediate lobe cells	Syt1	1.65999712	1.15E-61	0.723	0.029	7.64E-66



Featured genes representing the landscape of the seven cell types



- EL, epithelial-like cells
- IL, intermediate lobe cells
- M/m, macrophage/microglia
- P, pituicyte
- TL, T-cell like
- VE, vascular endothelia
- VLMC, vascular and leptomeningeal cells

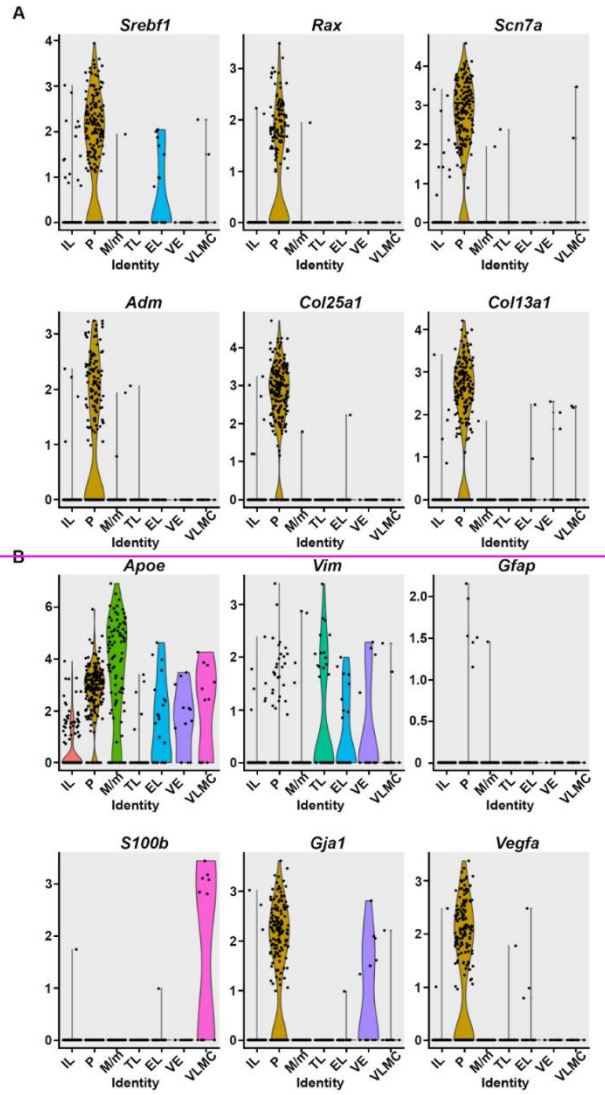
Chen et al. Figure. 3

Novel Pituicyte Markers

A. Novel markers

Show higher specificity a robustness compared to previously used markers

B. Previous markers



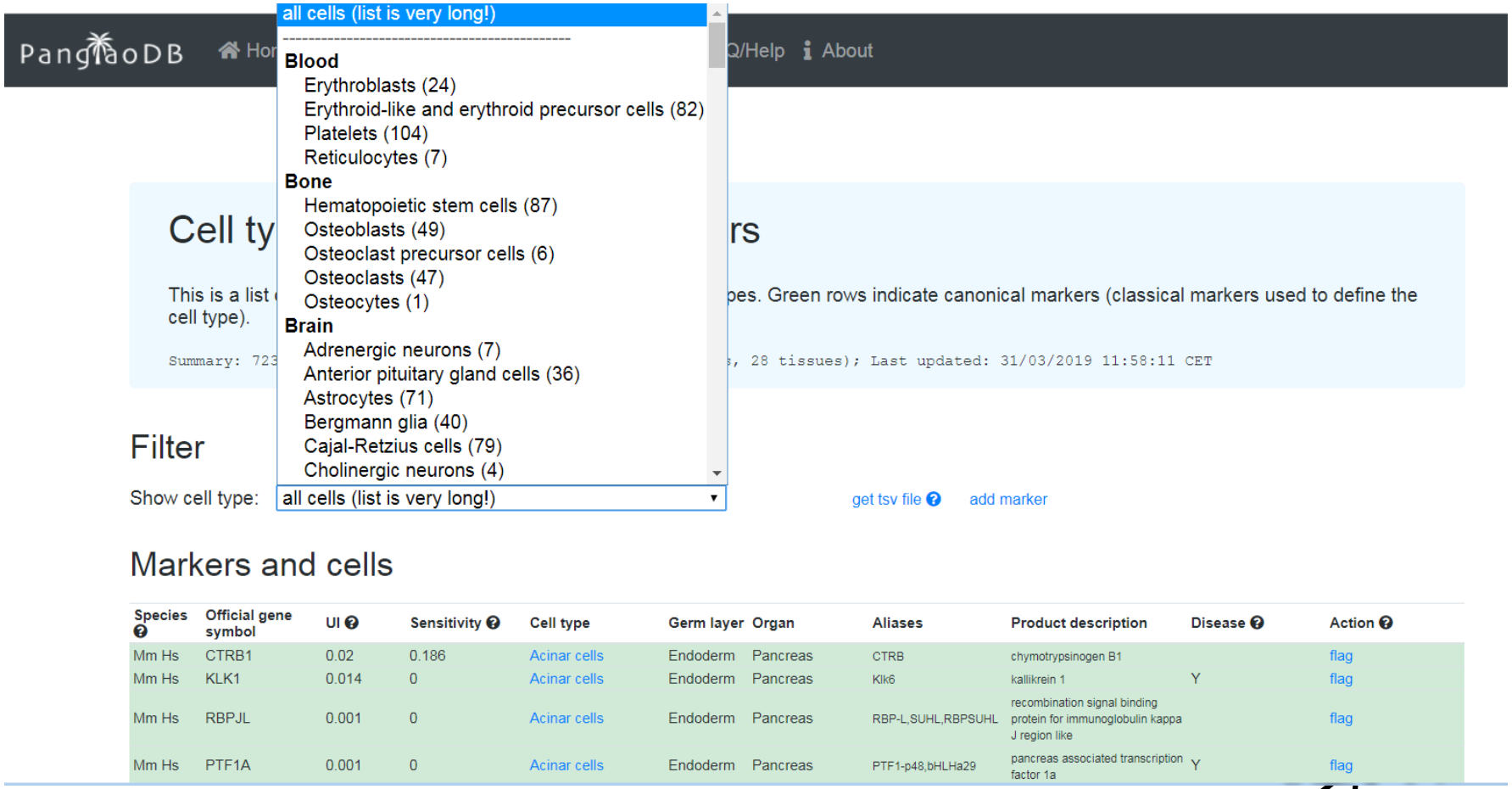
Chen et al. Figure. 4

PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data

Oscar Franzén ✉, Li-Ming Gan, Johan L M Björkegren

Database, Volume 2019, 2019, baz046, https://doi.org/10.1093/database/baz046

Published: 05 April 2019 **Article history** ▾



Over representation analysis (ORA) of Pituicytes DEG to PanglaoDB Cell Type Markers

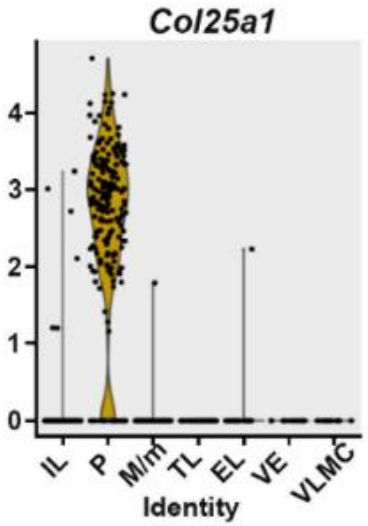
A	B	C	D	E
Ranking criteria: fdr smallest to largest				
cluster	**cell.type**	***p* -value**	**overlapping genes with PanglaoDB**	**FDR**
pituicytes	Tanycytes	5.84E-24	Rax; Igfbp5; Col23a1; Mest; Slc16a2; Lhx2; Rgcc; Adm; Ptn; Scn7a; Col25a1; Rgs7bp	1.20E-21
pituicytes	Astrocytes	1.53E-09	Gja1; Slc1a3; Sox9; Slc4a4; Slc16a2; Cldn10; Htra1	1.18E-07
pituicytes	Bergmann glia	9.14E-08	Gja1; Slc1a3; Slc4a4; A2m; Metrn	5.63E-06
pituicytes	Retinal progenitor cells	7.53E-07	Rax; Sox9; Hes1	3.57E-05
pituicytes	Müller cells	8.56E-06	Slc1a3; Dbi; Vegfa; Hes1	2.93E-04
pituicytes	Anterior pituitary gland cells	1.35E-05	Rax; Mest; Nkx2-1; Fgf10	4.06E-04

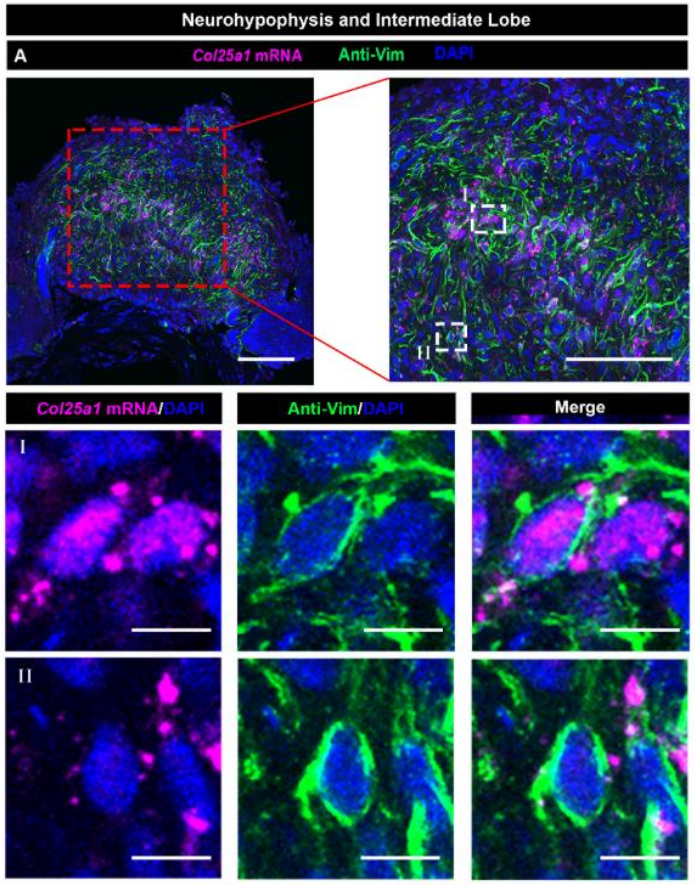
Our finding that Rax, Scn7a, Col25a1 and Adm are expressed in pituicyte is in line with the notion that tanycytes and pituicytes are of a common astrocytic lineage (Clasadonte and Prevot, 2018; Rodríguez et al., 2019; Wittkowski, 1998)

Validation of scRNA-Seq Results

- *in situ* hybridization with an antisense Col25a1 **fluorescent mRNA probe** followed by immunostaining with an antibody directed to the Vim protein and visualized by confocal microscopy.

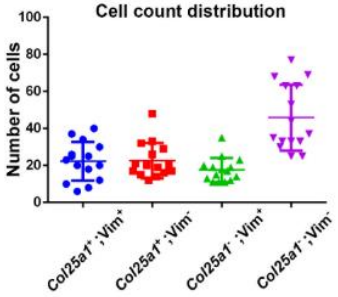
  This analysis shows that Vim immunoreactivity is detected in a subset of Col25a1-positive cells





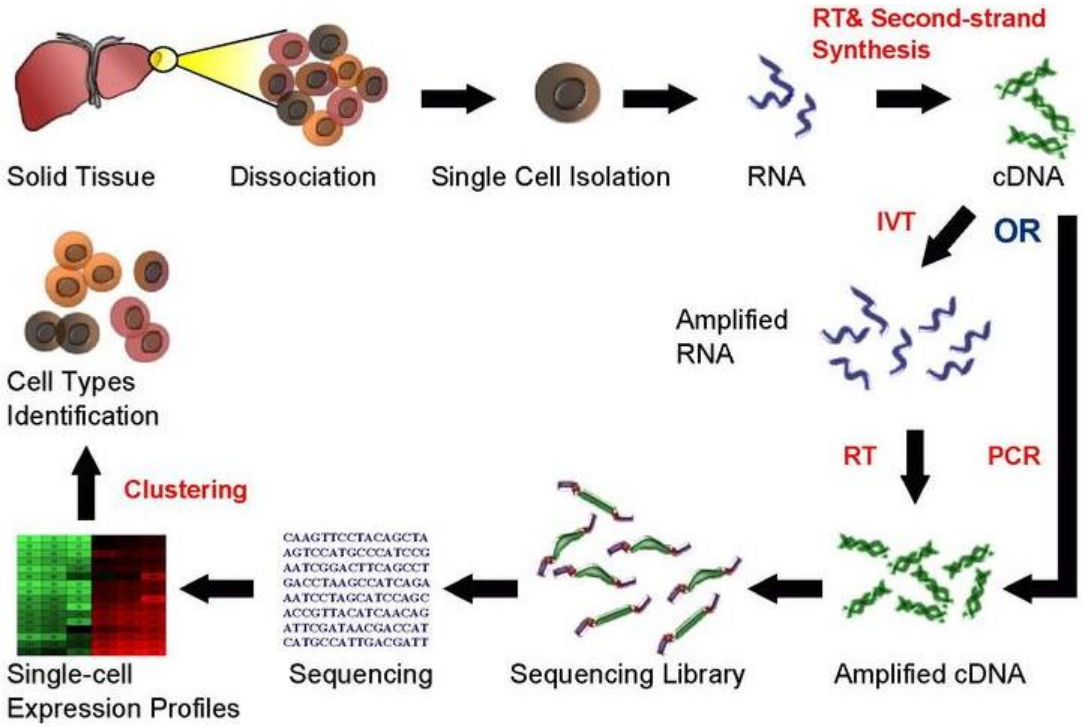
B

Cell population	Mean No. of Cells	Ratio
Col25a1$^+$; Vim$^+$	22	21%
Col25a1$^+$; Vim$^-$	23	21%
Col25a1$^-$; Vim$^+$	18	16%
Col25a1$^-$; Vim$^-$	46	43%



Chen et al. Figure. 5

Summarizing scRNA-Seq Workflow



https://hemberg-lab.github.io/scRNA.seq.course/introduction-to-single-cell-rna-seq.ht

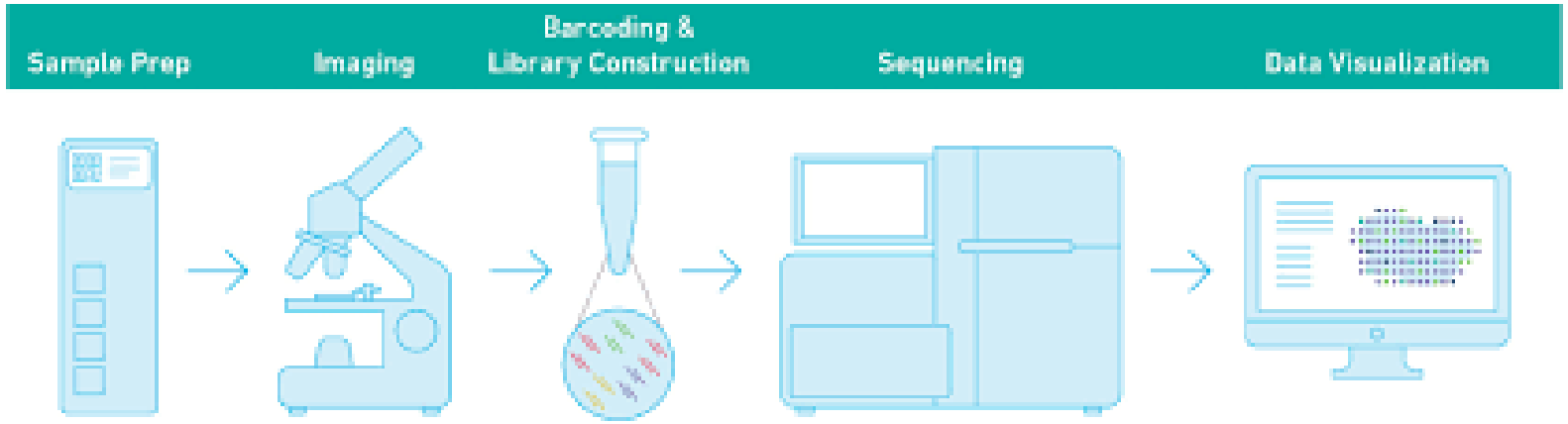
10X Genomics Products

SINGLE CELL PRODUCTS

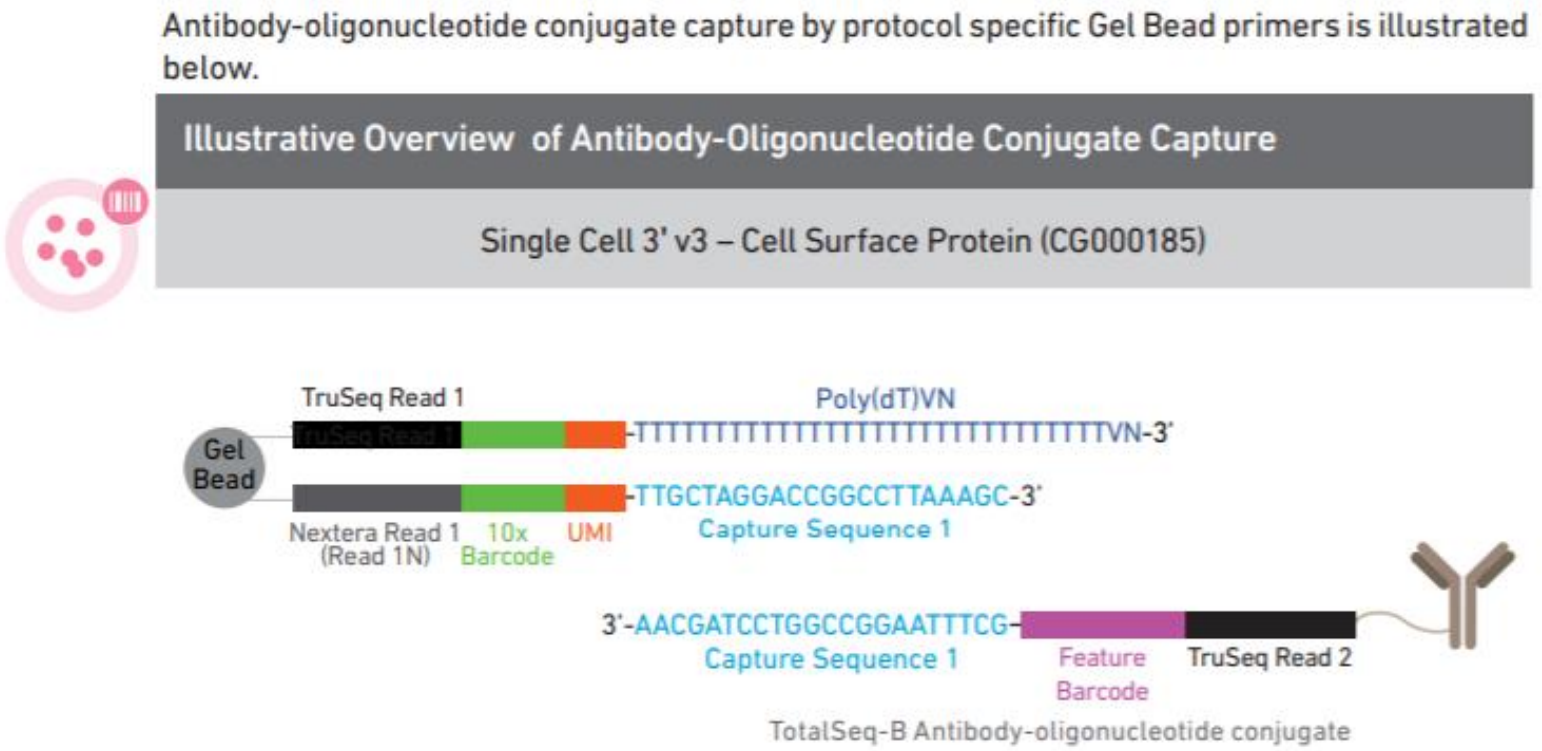
- Single Cell Gene Expression
  with CRISPR Screening & Cell Surface Protein
- Single Cell Immune Profiling
  with Antigen Specificity & Cell Surface Protein
- Single Cell CNV
- Single Cell ATAC

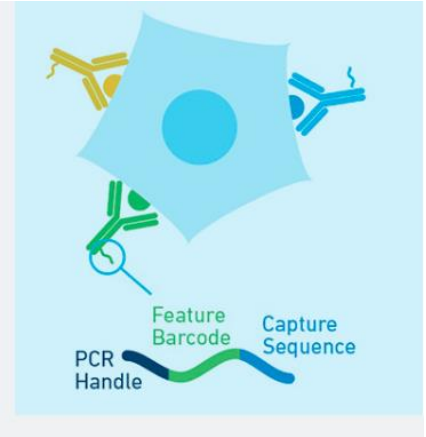
SPATIAL PRODUCTS

- Spatial Gene Expression



Simultaneously Measure Cell Gene Expression and Protein Expression





Simultaneously measure 5' gene expression and TCR, B cell Ig, cell surface protein expression in the same cells

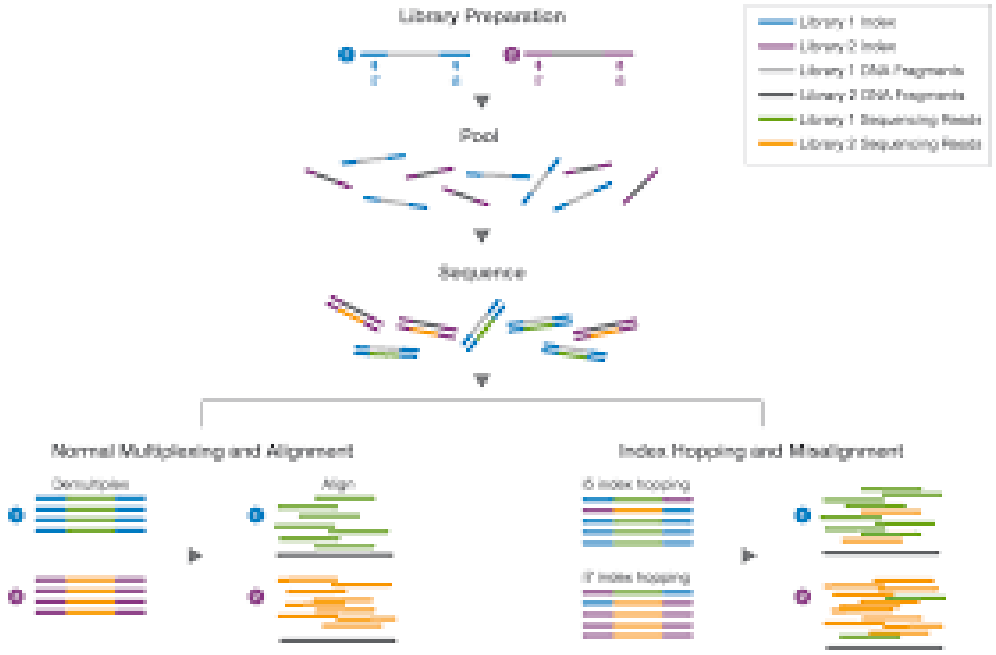
The END

THANKS FOR LISTENING
QUESTIONS?

In the Exercise we will review a cell ranger report and analyse clusters with Loupe

NovaSeq

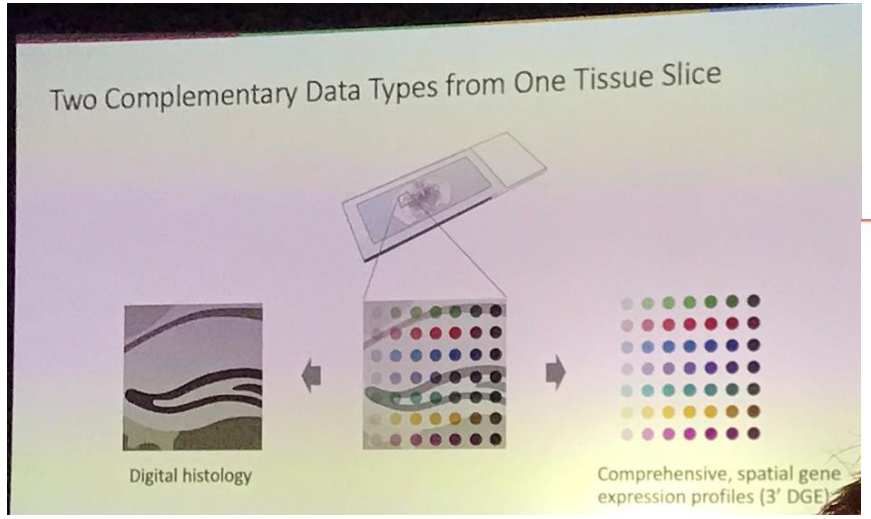
- Index hopping or index switching is a known phenomenon in NovaSeq. It causes incorrect assignment of libraries from the expected index to a different index (in the multiplexed pool).
- When sequencing with NovaSeq use unique **dual** indexing pooling combinations (unique i5 and i7 **indexes**).

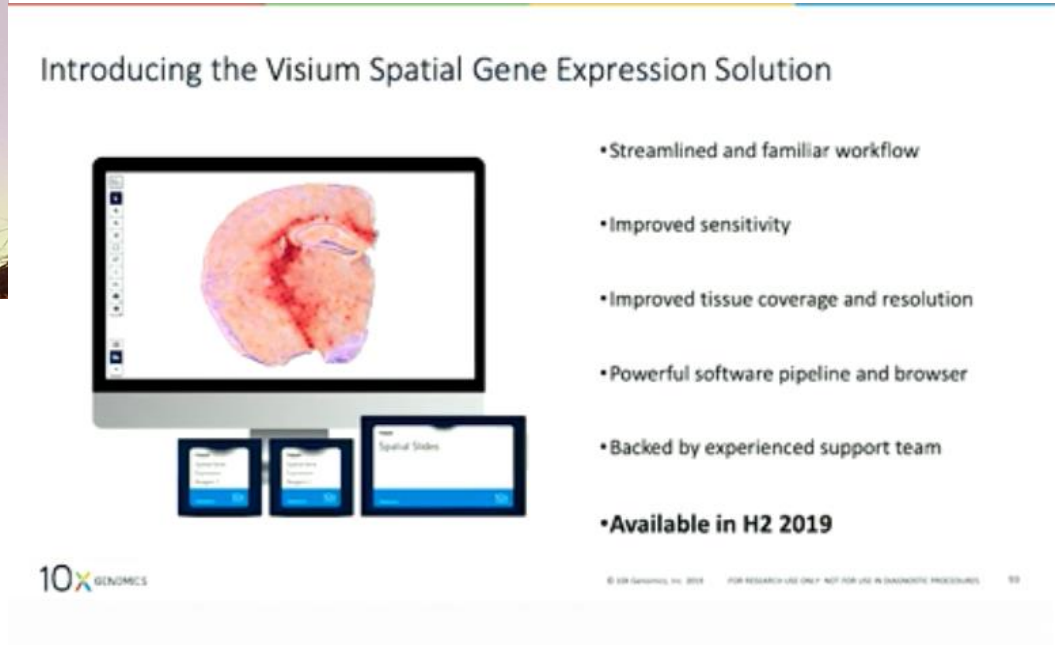


scRNA-Seq References

- https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/algorithms/overview#header
- https://hemberg-lab.github.io/scRNA.seq.course/introduction-to-single-cell-rna-seq.html
- How to design a single-cell RNA-sequencing experiment: pitfalls, challenges and perspectives; Molin & Camillo ; Briefings in Bioinformatics, 31 January 2018
- Bioinformatics workflows:
  - https://bioconductor.org/packages/release/workflows/html/simpleSingleCell.html Lun et al.
  - https://satijalab.org/seurat/ Seurat v2.0 Butler et al., Nature Biotechnology 2018.

Spatial Gene Expression Technology

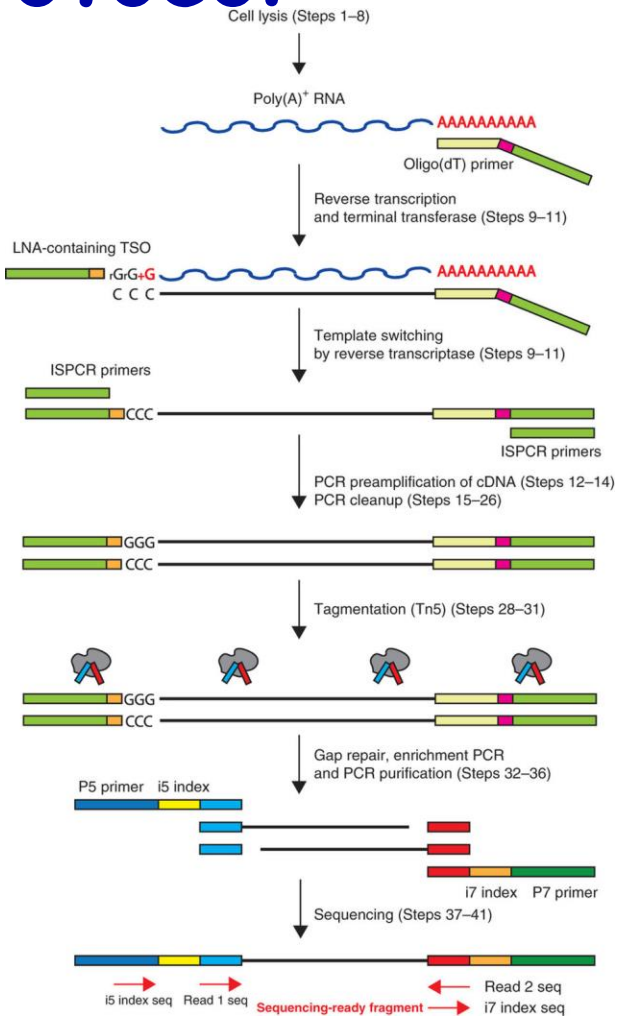




Smart-seq2 protocol

Full-length amplification of the transcripts :

- MMLV reverse transcriptase is able to add at the 5'-end of the RNA template, which corresponds to the 3'-end of the new cDNA strand, a few non-templated cytosines.
- These cytosines serve as an extended template for a helper oligonucleotide (called Template Switching Oligo - Locked Nucleic Acid) that allows the reverse transcriptase to 'switch' the template and synthesize the new cDNA strand



Platforms available at Weizmann

Read more - Experimental design for single-cell RNA sequencing
Baran-Gale et al. Briefings in Functional Genomics, Published: 08 November 2017

Protocol	Number of cells	Sequencing method	Number of reads per per cell	Sequencing machine	Advantages	Disadvantages	Usage
10x Genomics Chromium Includes Cell Isolation + library construction	Up to 100000, usually 100 to thousands	3' tag method	50,000 reads/cell for RNA-rich cells; 30,000 reads/cell for small primary cells	**LSCF Sandbox: NextSeq or NovaSeq**	**UMI** – ability to remove PCR duplicates	1. Doublets 2. Dropouts 3. Empty drops	Assess large numbers of cells
SmartSeq2 Nextera Construct Library - mRNA capture, RT and amplification	Up to 384	Full length	Minimum 1M reads/cell		Can add **spike-ins** more genes per cell	Batch effect; 3'- bias	Alleles, isoforms