

# Brainomics 2.0

Day 1 - Hands-on session

'Essential workflow  
for single-cell RNA-seq  
computational analyses'

Nicolo Caporale

Cristina Cheroni

Gaja Matassa

*November 18<sup>th</sup>, 2024*



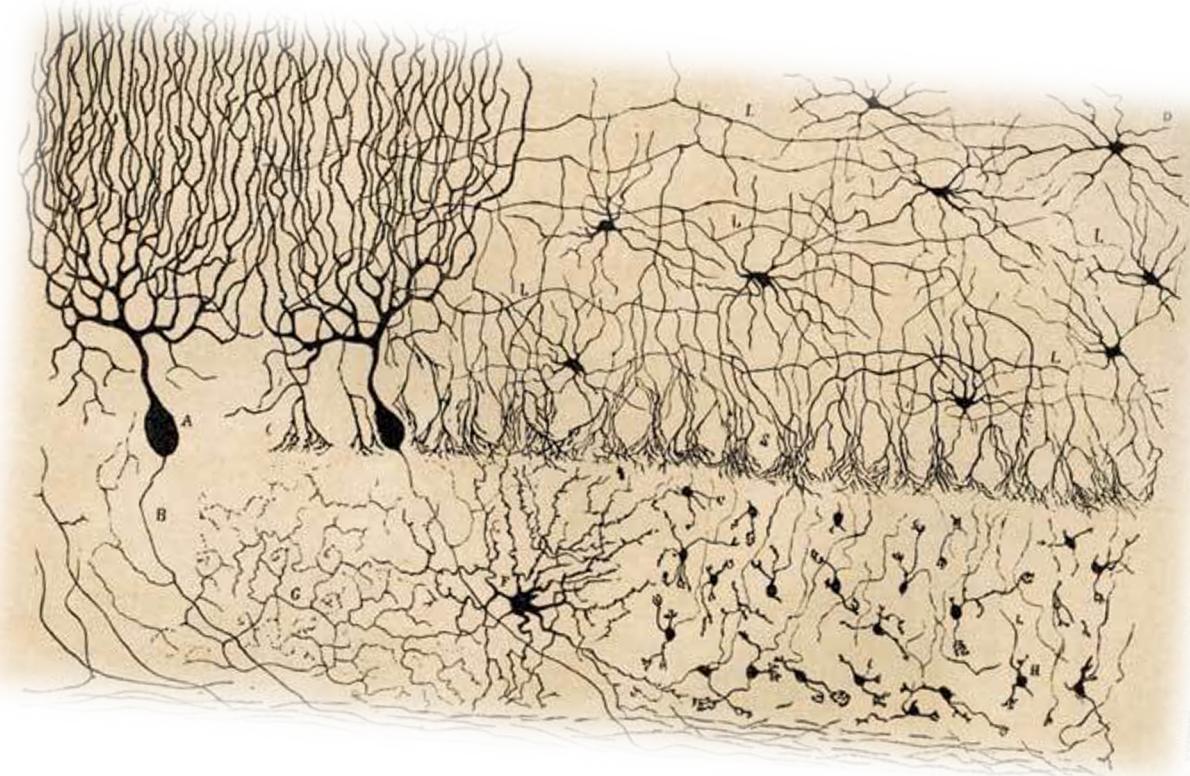
## Outline

- The challenges of studying the human brain: single cell omics approaches
- Single cell transcriptomics
  - From cells to count matrix
  - From count matrix to cell clusters – *more in the practical session*
- External dataset: human fetal cerebral cortex at mid-gestation
- Set up the analytical environment for the hands-on session

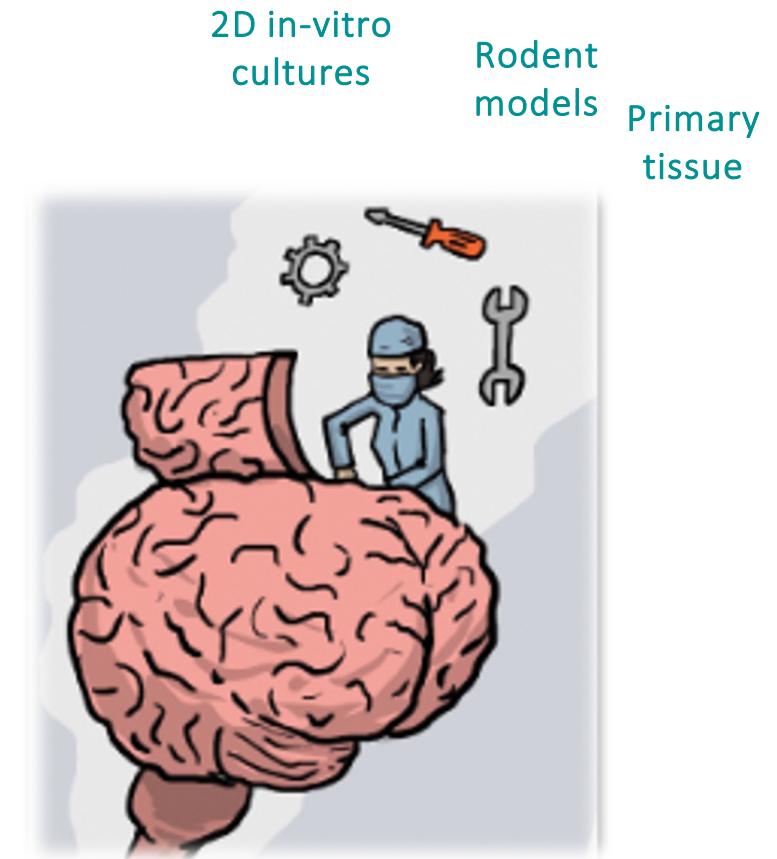
## Outline

- The challenges of studying the human brain: single cell omics approaches
- Single cell transcriptomics
  - From cells to count matrix
  - From count matrix to cell clusters – *more in the practical session*
- External dataset: human fetal cerebral cortex at mid-gestation
- Set up the analytical environment for the hands-on session

# The challenges of studying the human brain

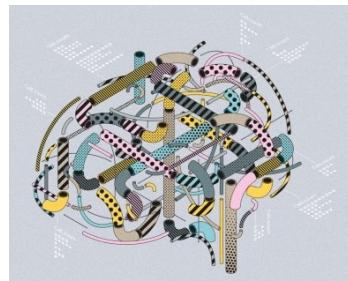


- Complexity (cell types, connectivity, ...)
- Human-specific features
- Accessibility



# The challenges of studying the human brain

Heterogeneity  
and complexity



scRNASeq

Single cell  
omics

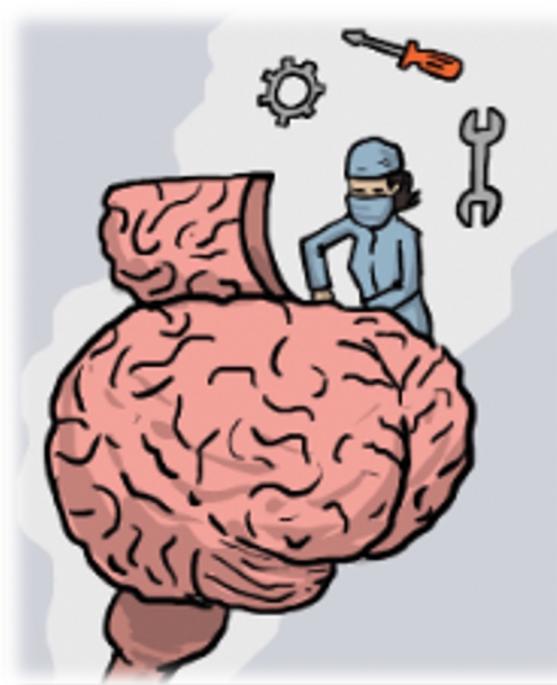
2D in-vitro  
cultures

Rodent  
models

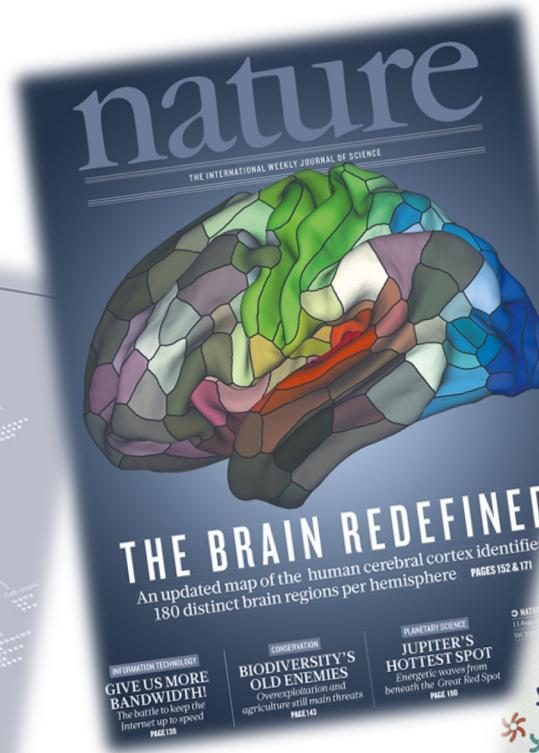
Primary  
tissue

Brain  
organoids

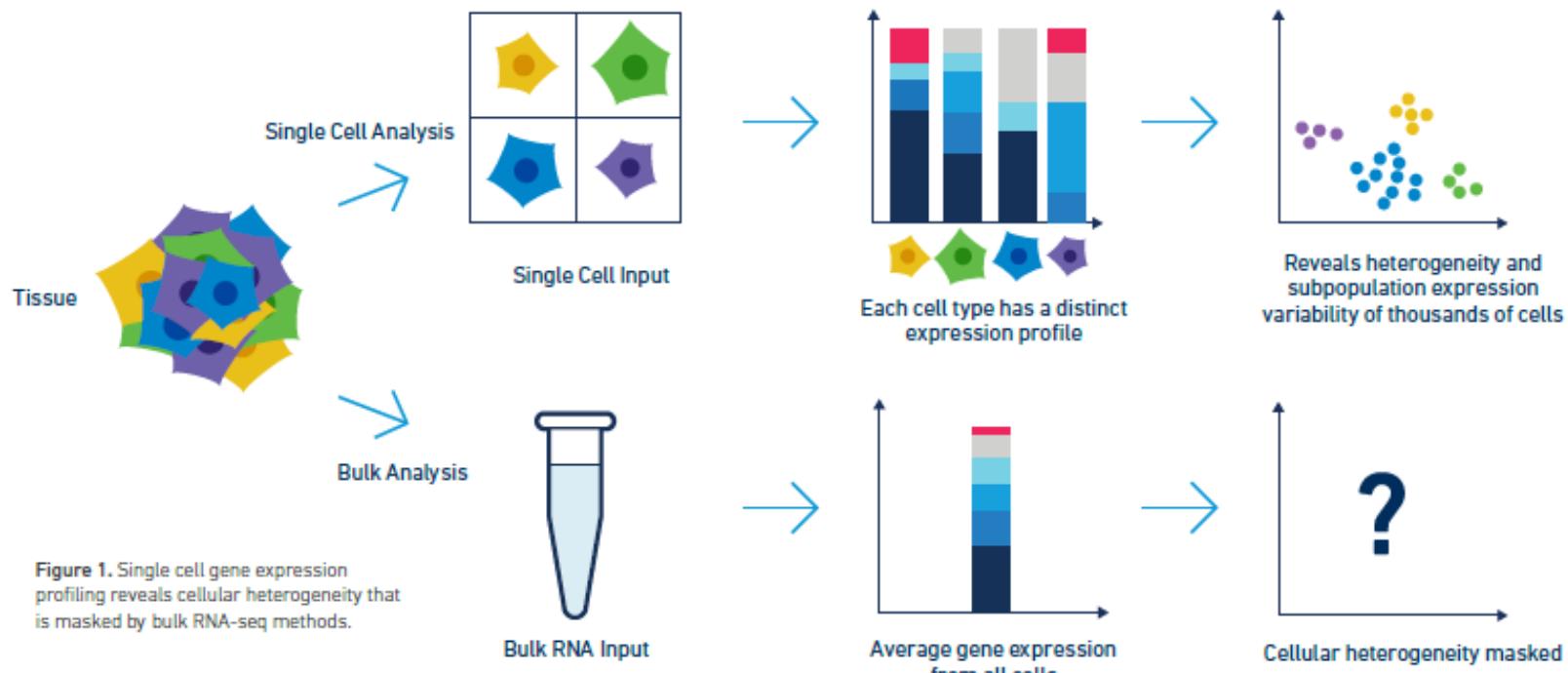
Accessibility



# Single cell omics of the human brain



# Why single cell?



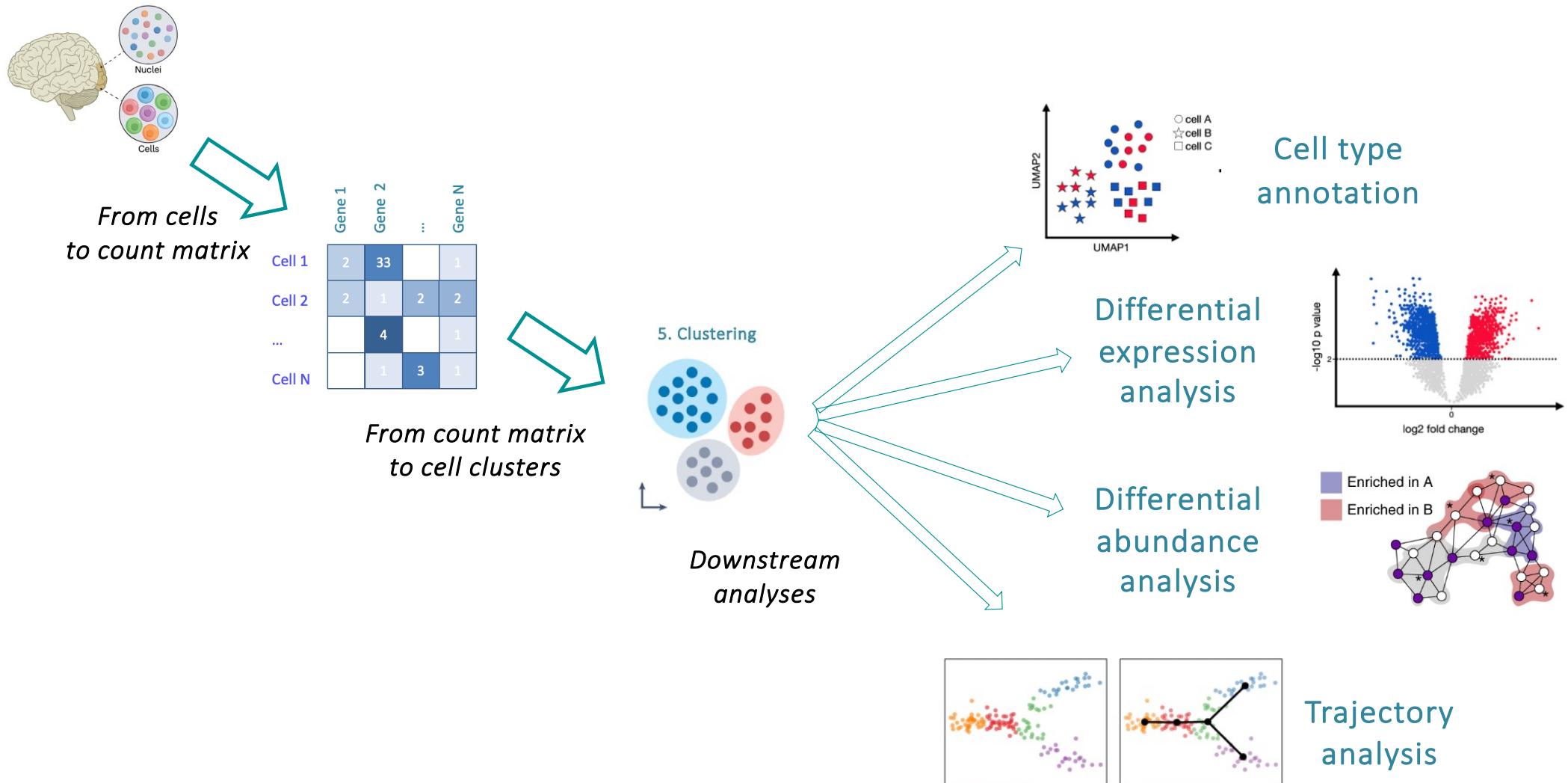
To address CNS complexity and heterogeneity

- Identify expression profiles of each specific cell types
- Identify rare cell types / states
- Understand how perturbations affect each cell type

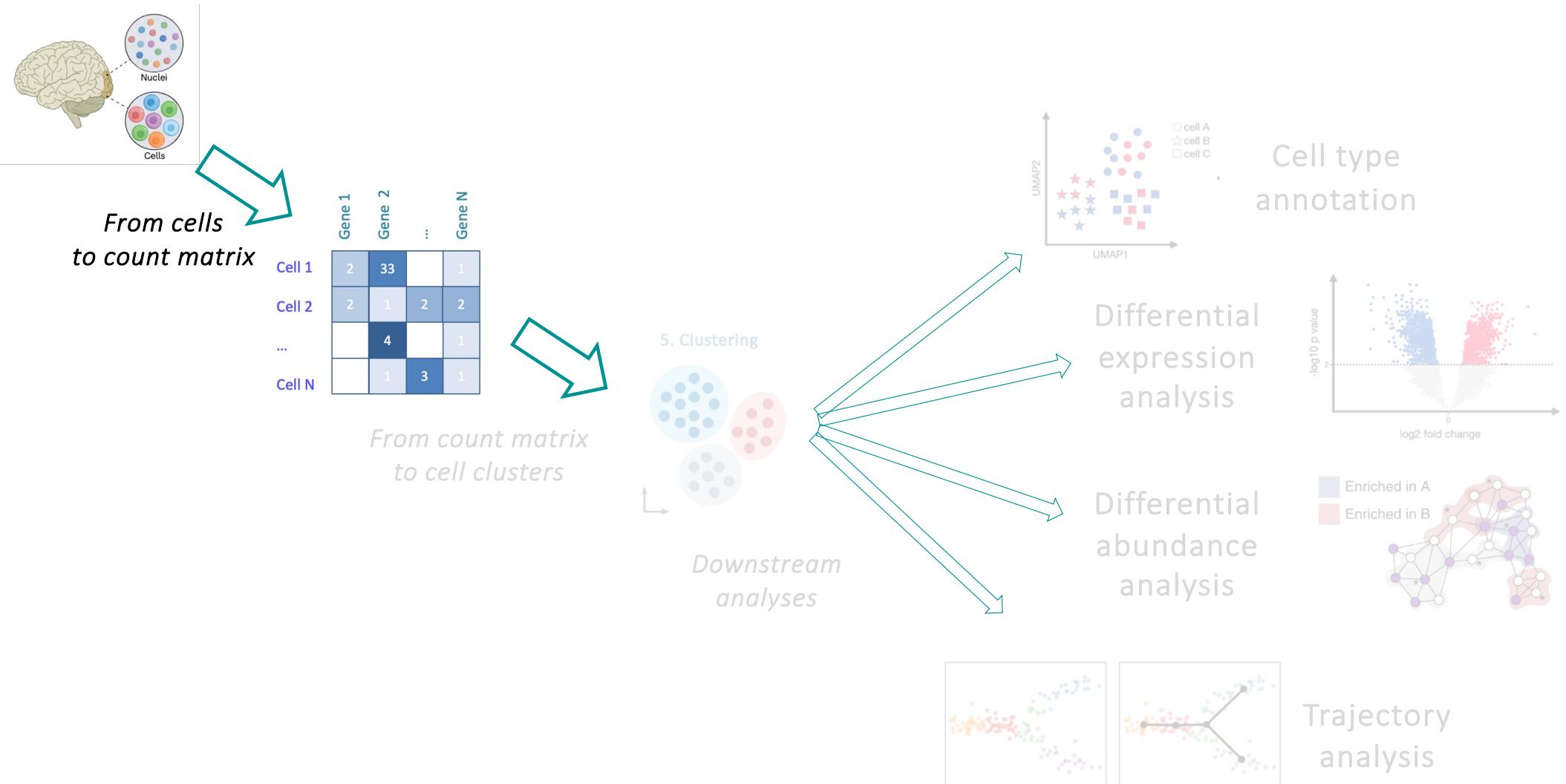
# Outline

- The challenges of studying the human brain: single cell omics approaches
- Single cell transcriptomics
  - **From cells to count matrix**
  - From count matrix to cell clusters – *more in the practical session*
- External dataset: human fetal cerebral cortex at mid-gestation
- Set up the analytical environment for the hands-on session

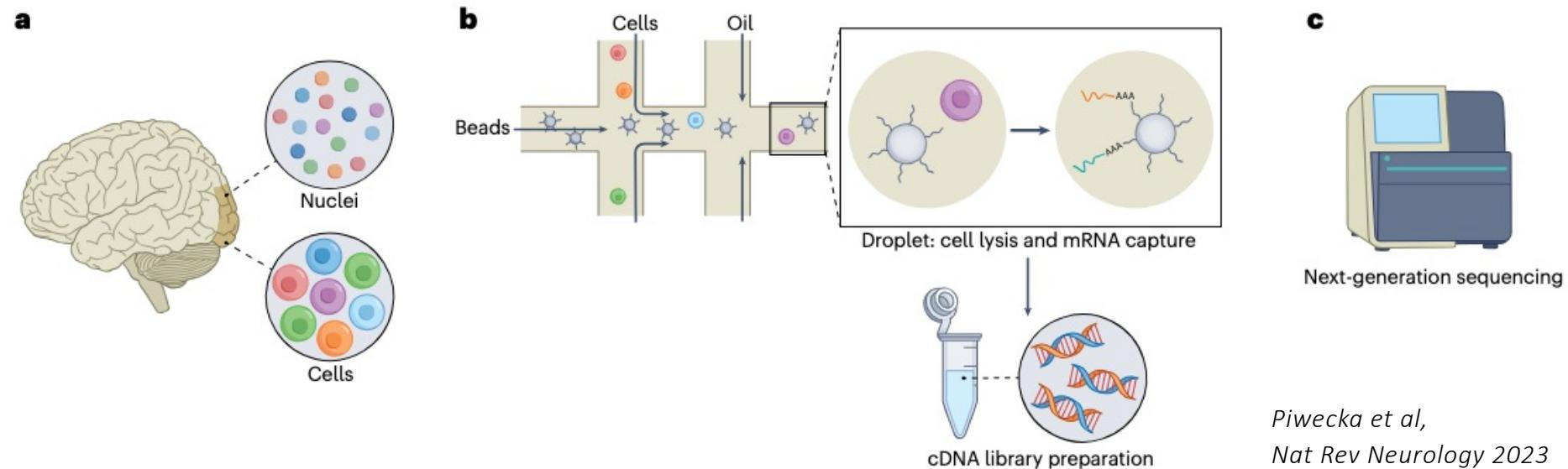
# Single cell transcriptomics: overview



# Single cell transcriptomics: overview



## Single cell transcriptomics: basic principles (10X Genomics)



Piwecka et al,  
Nat Rev Neurology 2023

- Single cell (or nuclei) suspension
- Microfluidic devices to generate droplets containing a barcoded bead and a single cell
- Cell lysis and capture of polyadenylated RNA
- Library preparation: cDNA containing barcode and UMI
- Sequencing

## Count matrix in single-cell experiments

	Gene 1	Gene 2	:	Gene N
Cell 1	2	33		1
Cell 2	2	1	2	2
...		4		1
Cell N		1	3	1

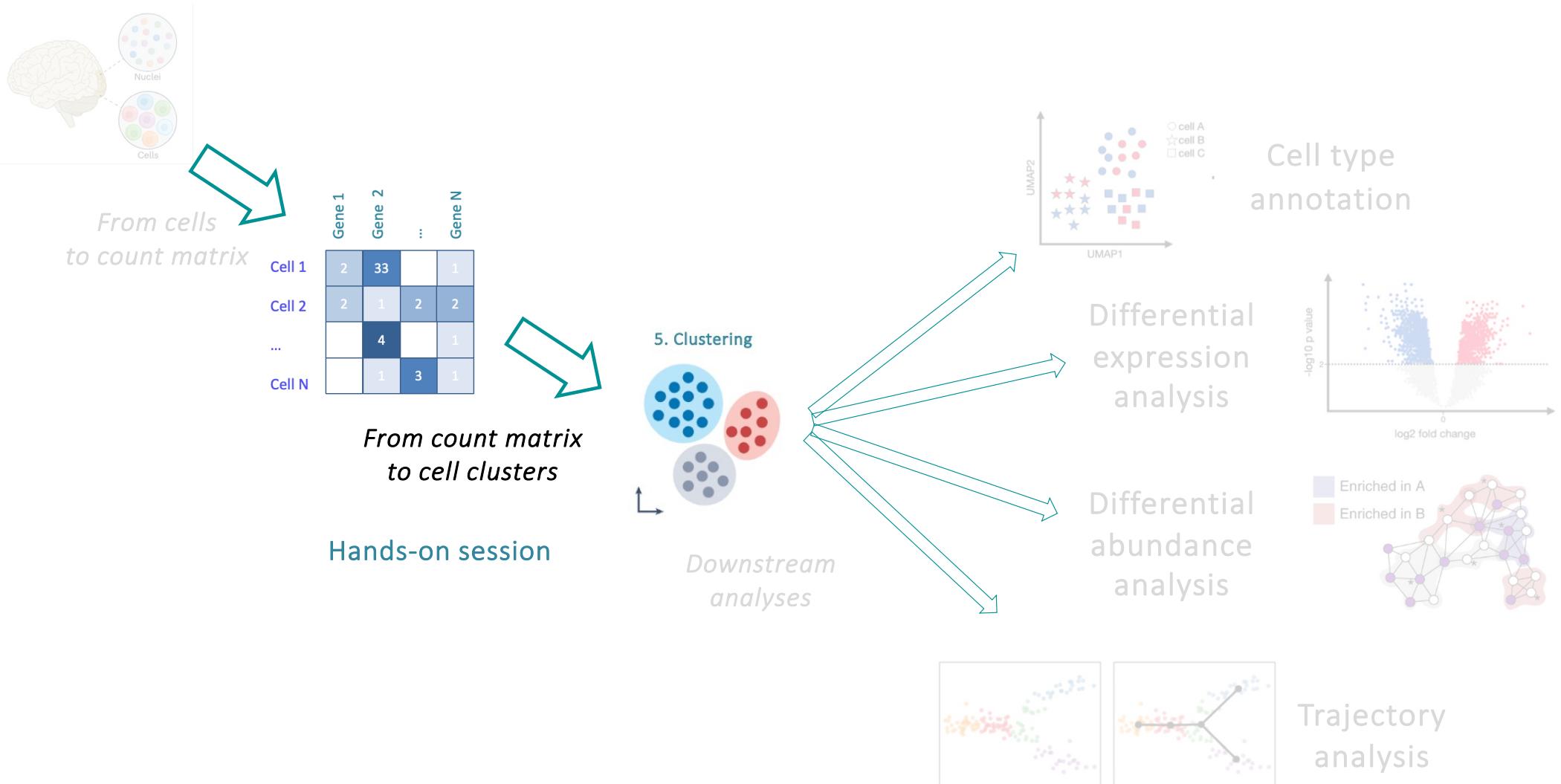
- **High dimensionality**  
e.g. 60,000 cells x 50,000 genes
- **High Sparsity**

Ad-hoc methodologies and procedures to process, visualize and analyse the data

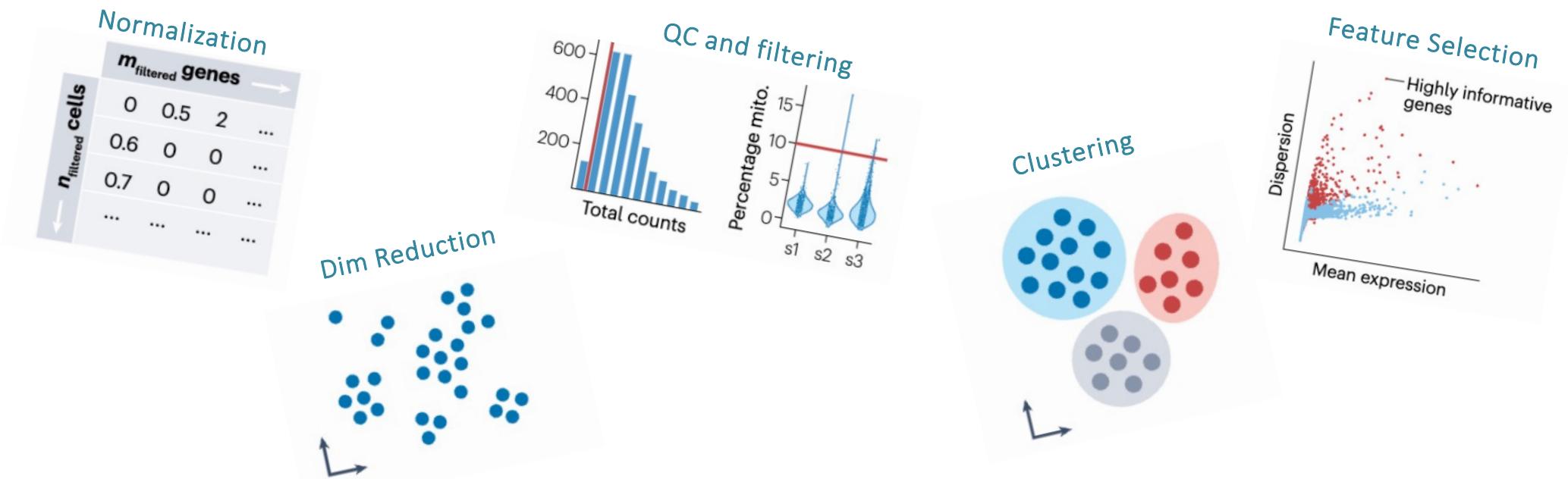
# Outline

- The challenges of studying the human brain: single cell omics approaches
- Single cell transcriptomics
  - From cells to count matrix
  - From count matrix to cell clusters – *more in the practical session*
- External dataset: human fetal cerebral cortex at mid-gestation
- Set up the analytical environment for the hands-on session

# Single cell transcriptomics: overview



## scRNASeq: from count matrix to cell clusters



Modified from <https://www.sc-best-practices.org/>

# scRNASeq: from count matrix to cell clusters

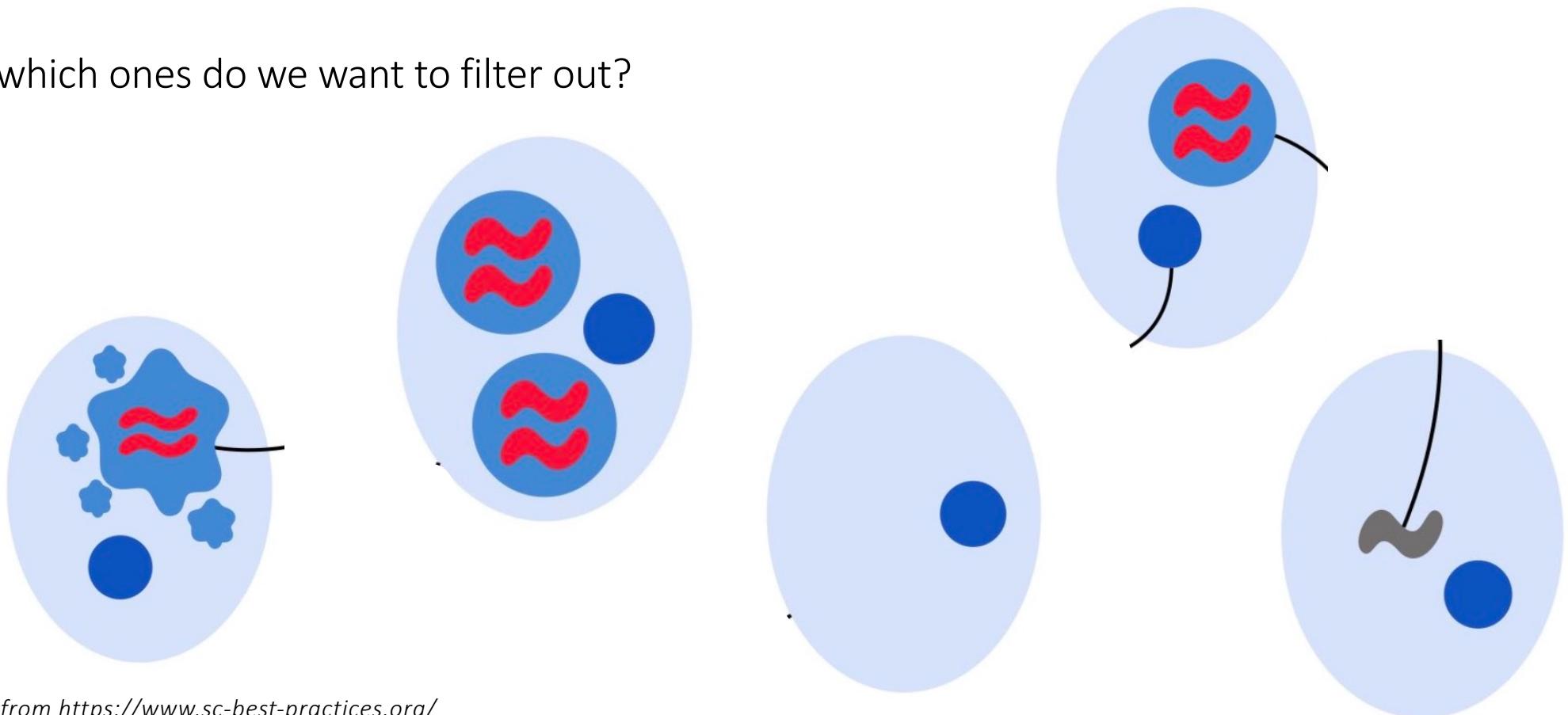


Modified from <https://www.sc-best-practices.org/>

## scRNASeq: quality check and filtering

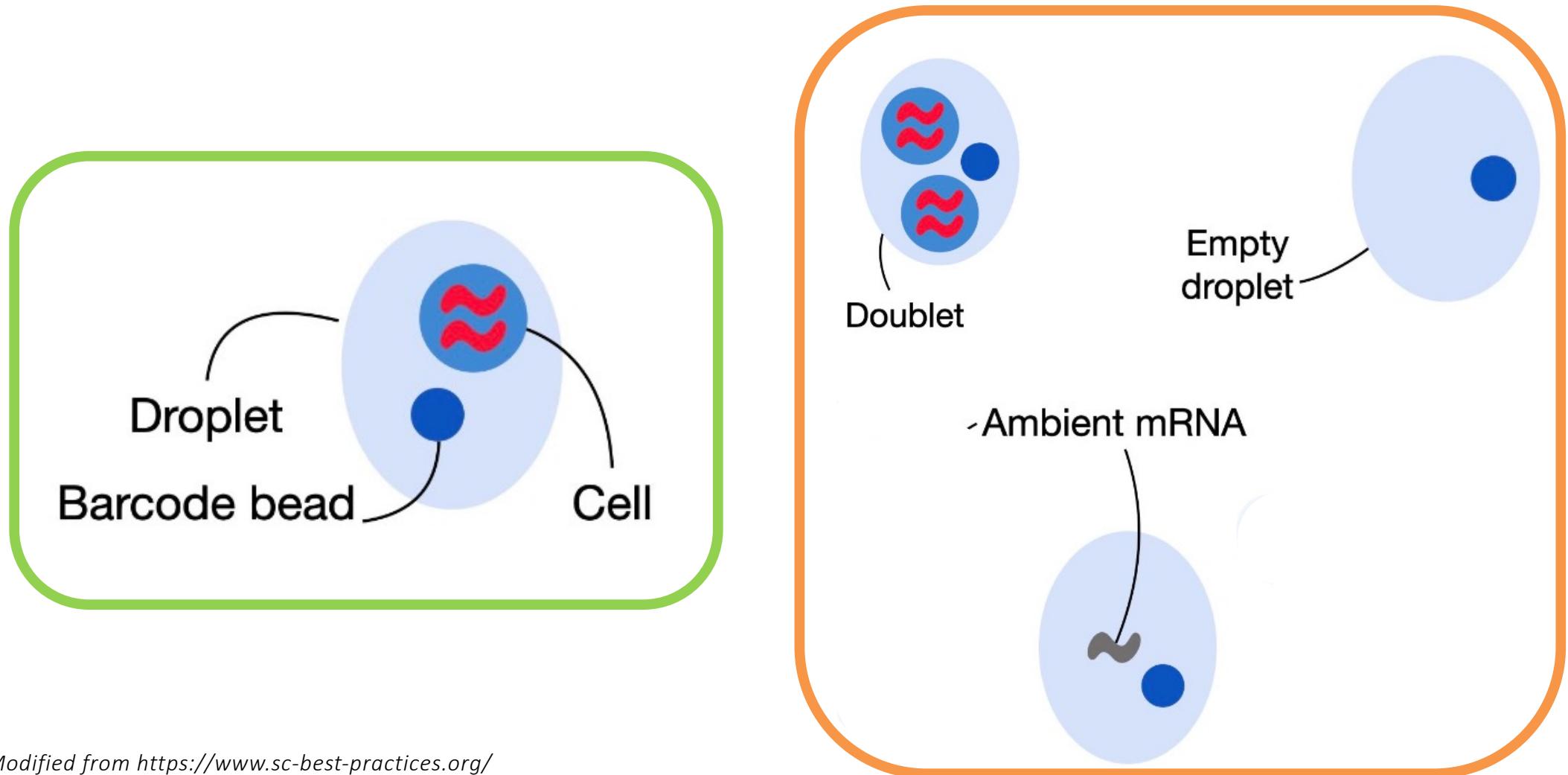
Which droplet(s) do we want to keep?

And which ones do we want to filter out?



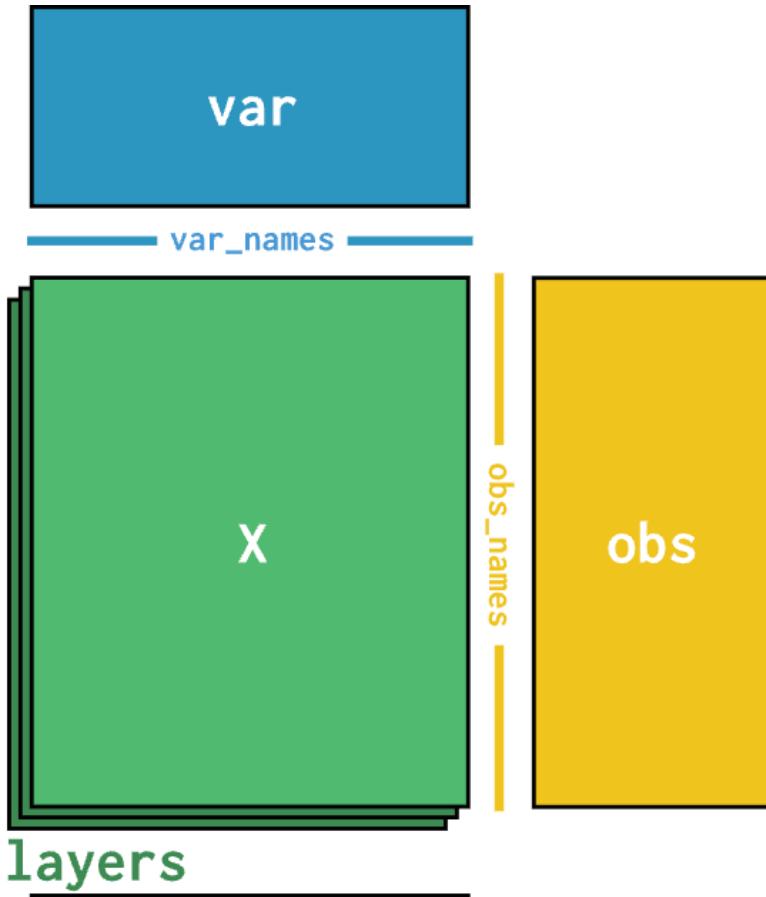
Modified from <https://www.sc-best-practices.org/>

## scRNASeq: from count matrix to cell clusters



Modified from <https://www.sc-best-practices.org/>

## AnnData structure



**X**: count matrix

- sparse format
- Can have different layers (e.g raw counts, normalized counts, etc)

**Obs**: barcode (cell) metadata

**Var**: feature (gene) metadata

## Scanpy framework



# Scanpy – Single-Cell Analysis in Python

Scanpy is a scalable toolkit for analyzing single-cell gene expression data built jointly with [anndata](#). It includes preprocessing, visualization, clustering, trajectory inference and differential expression testing. The Python-based implementation efficiently deals with datasets of more than one million cells.

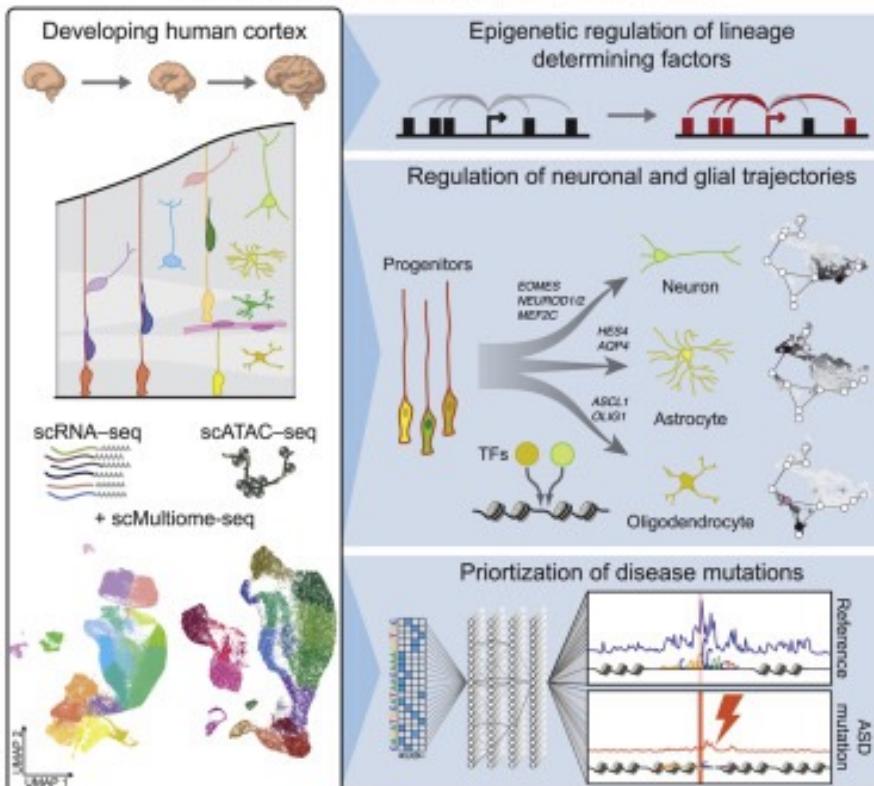
## Outline

- The challenges of studying the human brain: single cell omics approaches
- Single cell transcriptomics
  - From cells to count matrix
  - From count matrix to cell clusters – *more in the practical session*
- External dataset: human fetal cerebral cortex at mid-gestation
- Set up the analytical environment for the hands-on session

# Mid-gestation fetal cortex dataset

'Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution'

## Single-cell transcription and chromatin landscapes of human cortical development



- Samples: **human fetal brain cortex at mid-gestation** (4 subjects from PCW16 to PCW24)
- Sequencing method: single cell RNA Sequencing (Chromium platform - 10x Genomics)
- Obtained number of cells: **~58,000**

## Mid-gestation fetal cortex dataset

# Metadata

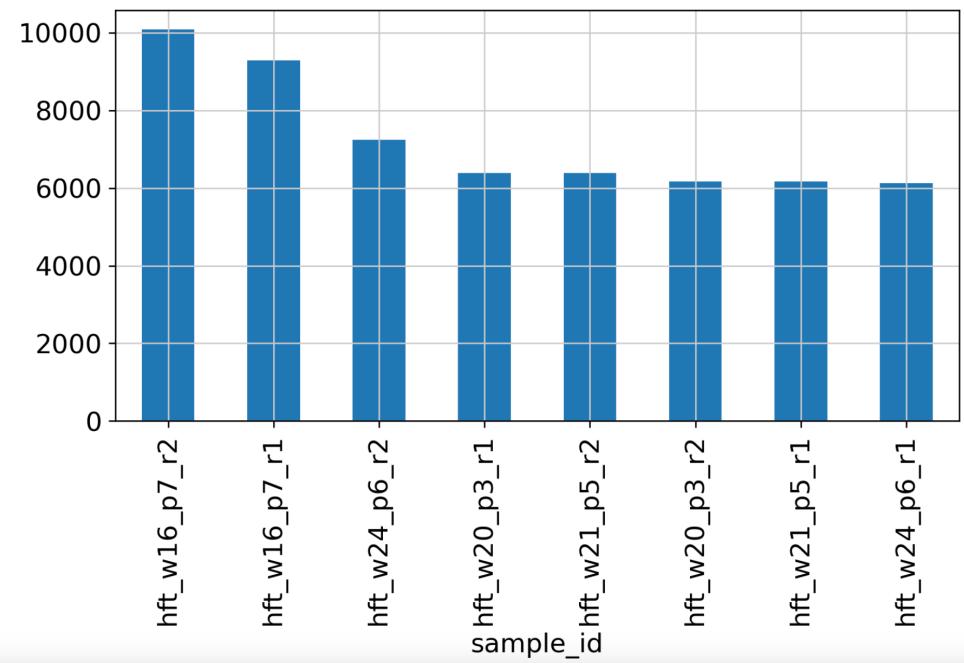
---

Key metadata already made available from the authors:

- Sample ID
- Tissue
- Post-conceptional week
- Cell clusters (seurat\_cluster)
- Assay and Batch
- QC: several metrics

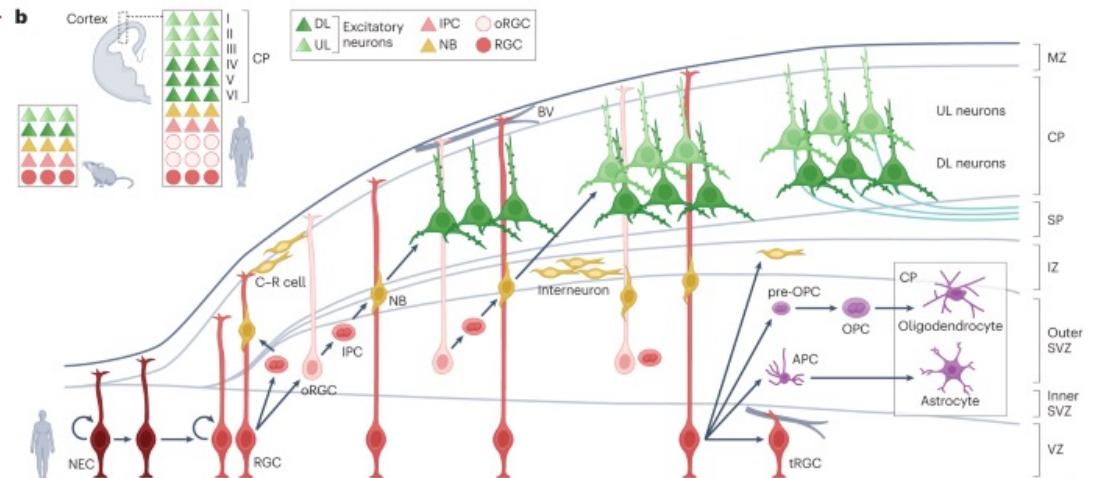
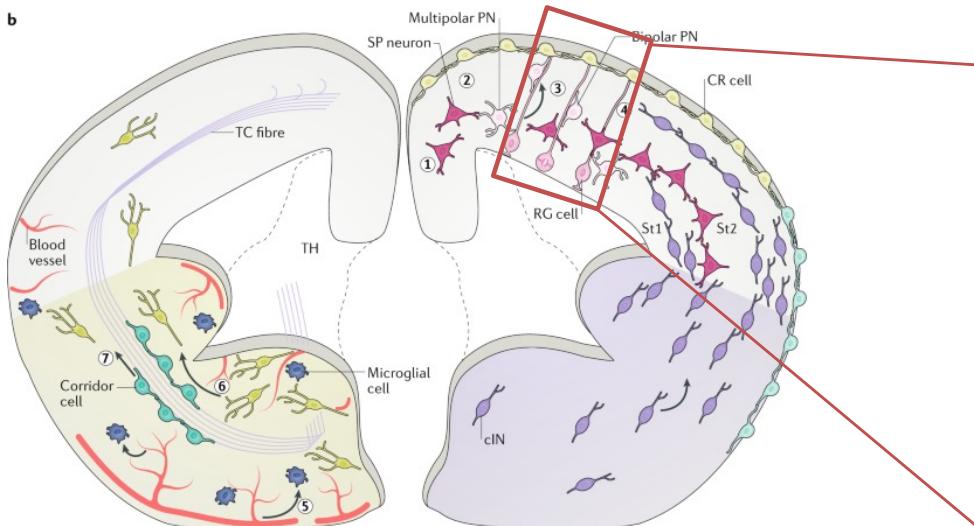
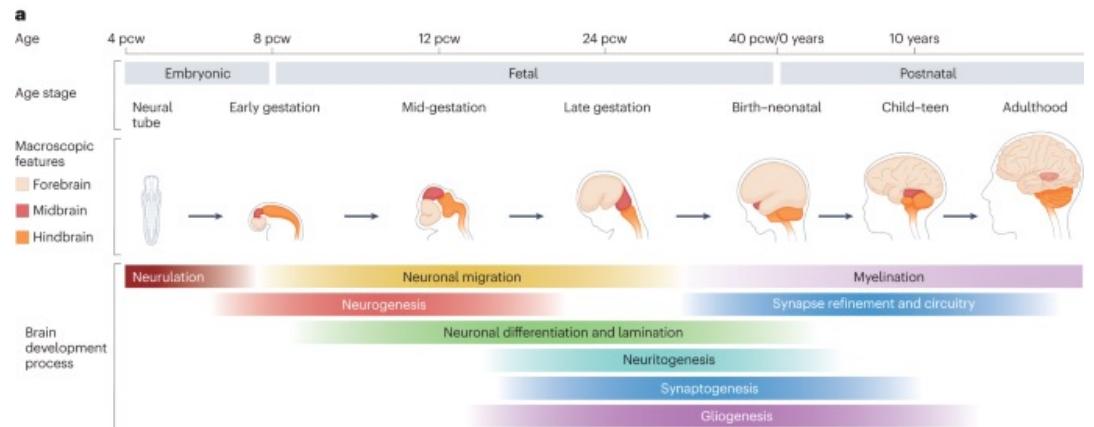
# Mid-gestation fetal cortex dataset

Og_Age	pcw16	pcw20	pcw21	pcw24
sample_id				
hft_w16_p7_r1	9294	0	0	0
hft_w16_p7_r2	10079	0	0	0
hft_w20_p3_r1	0	6393	0	0
hft_w20_p3_r2	0	6174	0	0
hft_w21_p5_r1	0	0	6169	0
hft_w21_p5_r2	0	0	6388	0
hft_w24_p6_r1	0	0	0	6131
hft_w24_p6_r2	0	0	0	7240



# Cell types of the developing cortex

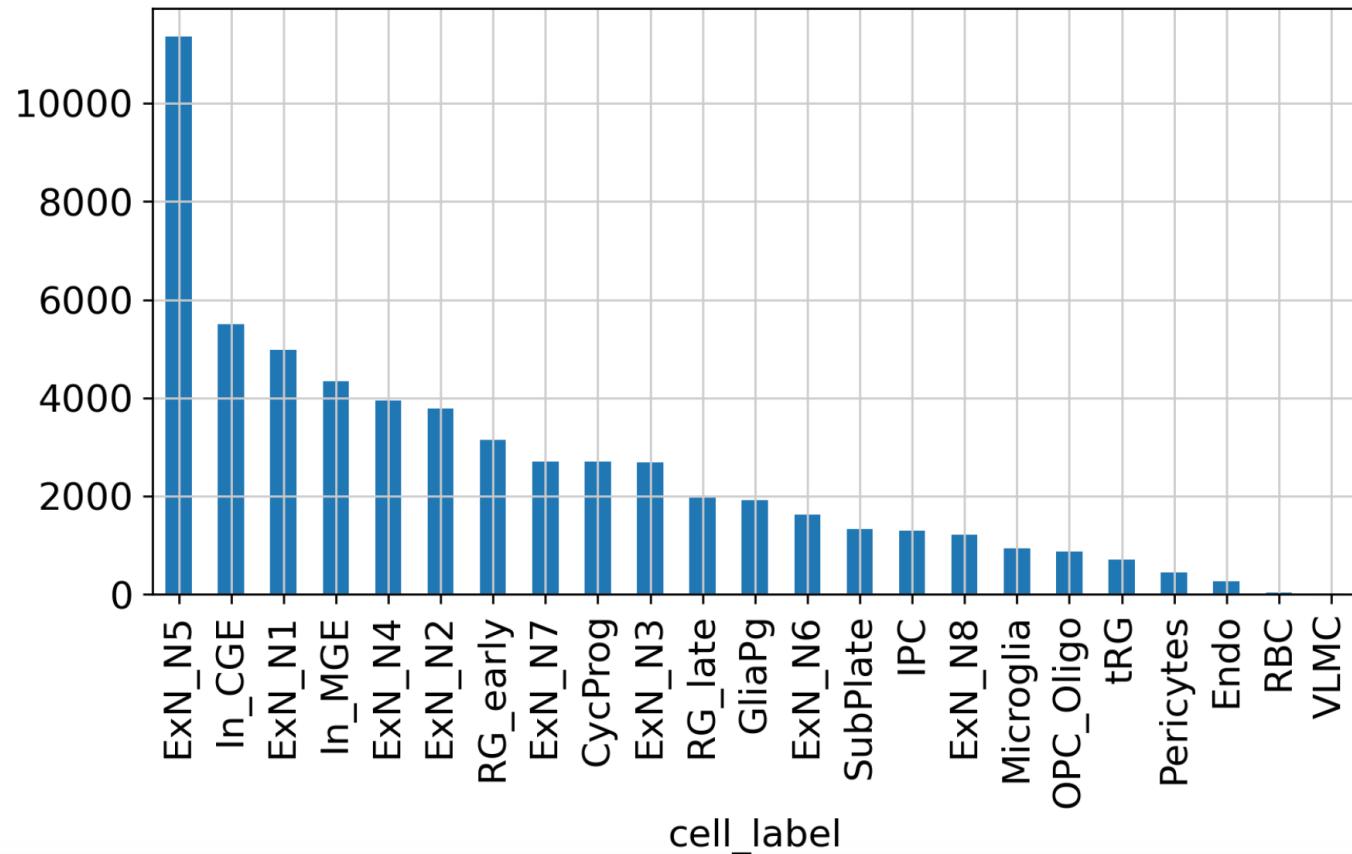
- Excitatory Neurons
- Inhibitory Neurons
- Progenitors
- Glial cells
- Other cell types



From: <https://doi.org/10.1038/s41583-019-0148-y>

From: <https://doi.org/10.1038/s41576-023-00626-5>

## Mid-gestation fetal cortex dataset



- **Excitatory Neurons:** ExN\_N1 – ExN8
- **Inhibitory Neurons:** In\_MGE, In\_CGE
- **Progenitors:** CycProg, RG\_early, RG\_late, IPC, tRG
- **Glial cells:** GliaPg, Microglia, OPC\_Oligo
- **Other cell types:** Pericytes, Endo, RBC, VLMC

*RG: radial glia*

*IPC: intermediate progenitors*

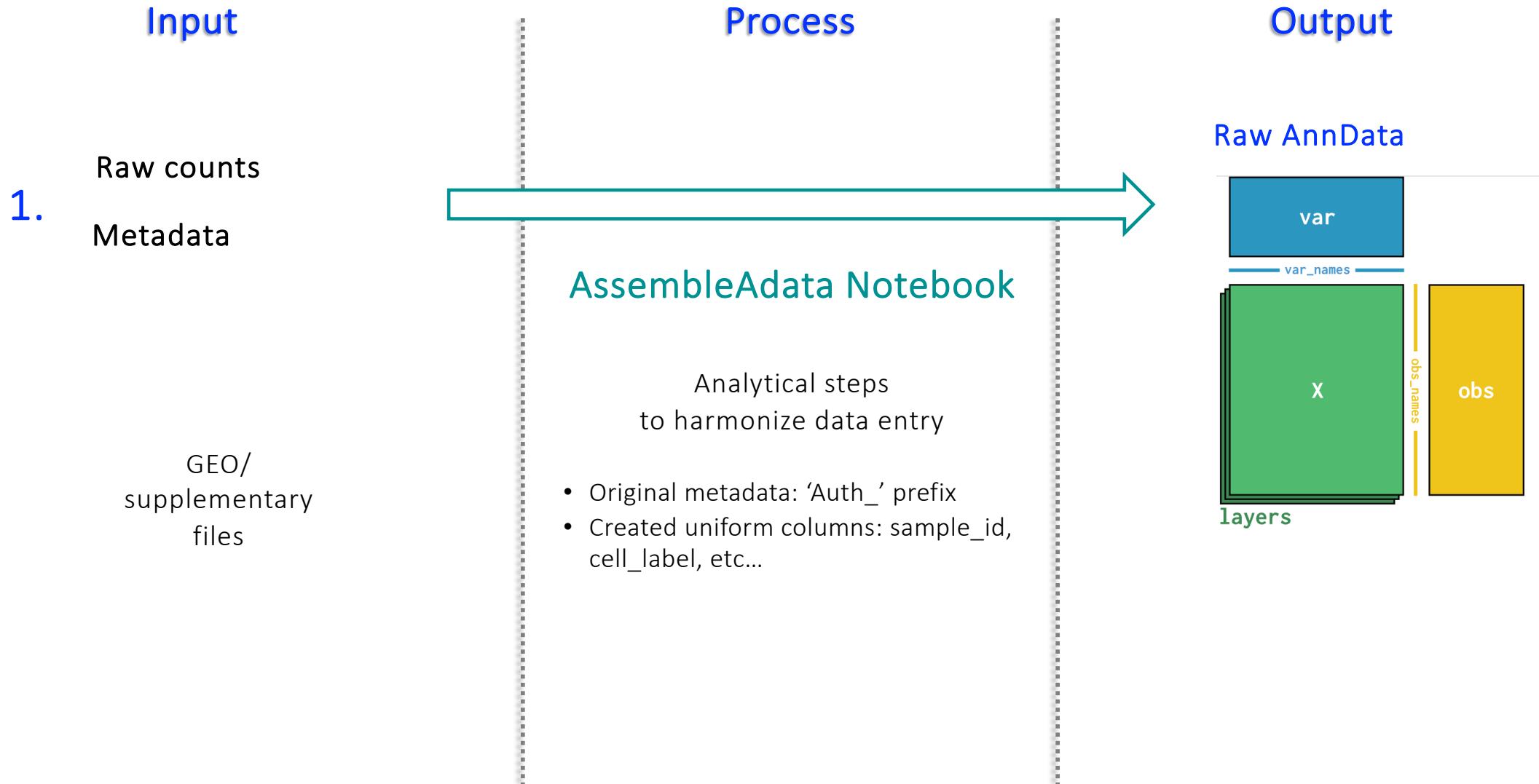
*MGE – CGE: medial/caudal ganglionic eminence*

*OPC: oligodendrocyte progenitor cells*

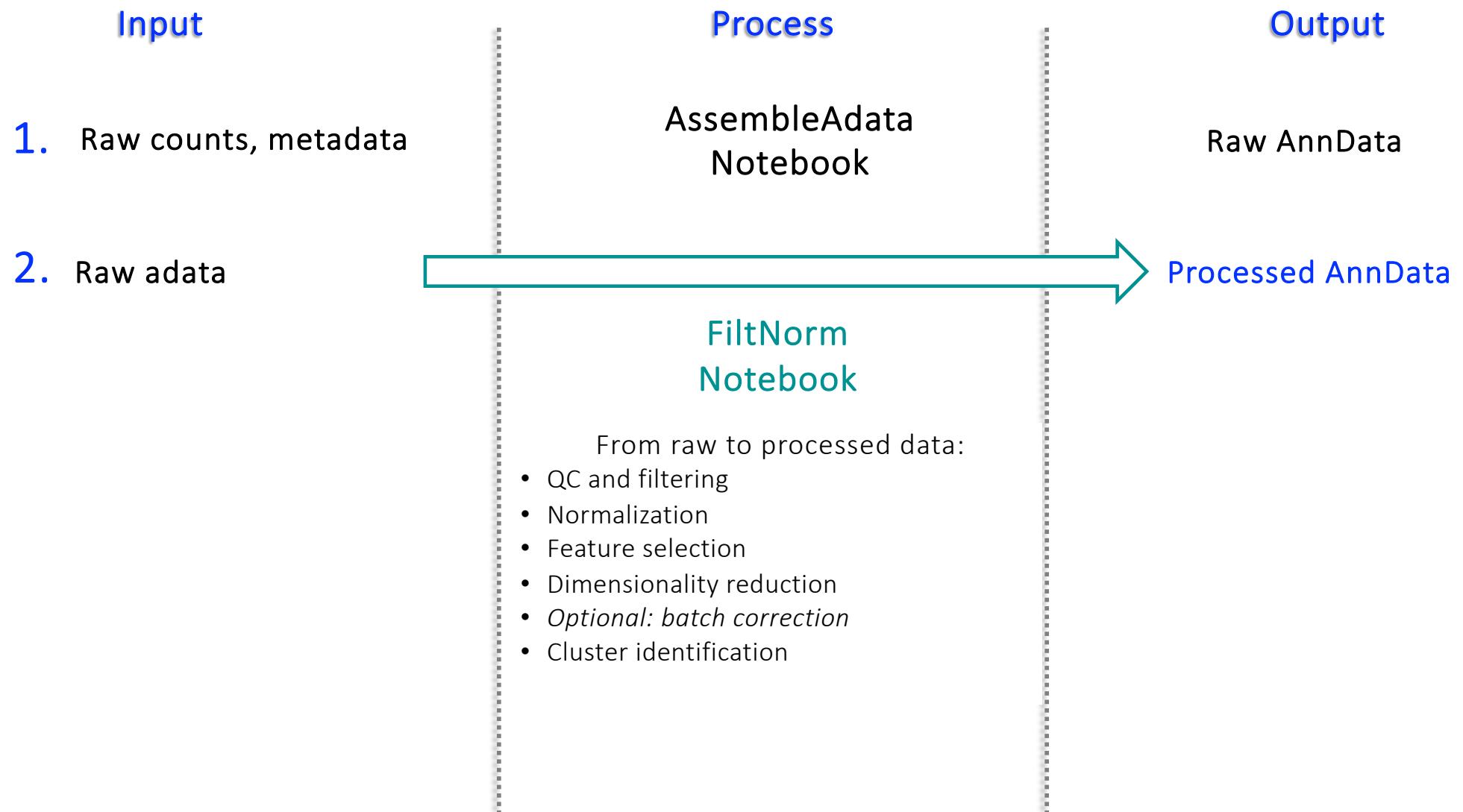
*Endo: endocytes*

*RBC: red blood cells*

# Data flux



## Data flux



## Mid-gestation fetal cortex dataset

[FiltNormBatch notebook](#): from raw to processed data

- QC and filtering
- Normalization
- Feature selection
- Dimensionality reduction
- *Optional: batch correction*
- Cluster identification

[Cluster Notebook](#): cluster characterization

# Outline

- The challenges of studying the human brain: single cell omics approaches
- Single cell transcriptomics
  - From cells to count matrix
  - From count matrix to cell clusters – *more in the practical session*
- External dataset: human fetal cerebral cortex at mid-gestation
- Set up the **analytical environment** for the hands-on session

## BrainOmics2024 GitHub repo

- Access the VDI system
- Open the browser and go to the course repo:

<https://github.com/BrainOmicsCourse/BrainOmics2024>

**Github repository for BrainOmics course 2024**



This repository contains all the notebooks of the hands-on sessions that will be held during the [BrainOmics 2.0](#) course at the [Human Technopole](#) in Milan, Italy from November 18 to 22, 2024.

### Hands-on Sessions

**Day 1: Essential Workflow for single-cell RNASeq computational analysis**

- [WalkThrough](#)
- [Notebook 1: from count matrix to clusters](#)
- [Notebook 2: cluster characterization](#)
- [Resources](#)

## Following Day1 ‘Walkthrough’ section

1. Open a terminal window and clone the course repo in your home

```
cd $HOME  
git clone https://github.com/Brain0micsCourse/Brain0mics2024.git
```

2. Launch the analysis container

### Load Singularity Module

```
module load singularity/3.8.5
```

### Launch the container

```
singularity exec -B /group/brainomics -H $HOME/Brain0mics2024/1_Day1/ \  
/group/brainomics/Container/Day1_container.sif \  
jupyter lab -y --ip=0.0.0.0 --port=8888 --notebook-dir=$HOME/Brain0mics2024/1_Day1/
```

3. Open Jupyter lab in your browser

Or copy and paste one of these URLs:

<http://brainomics-20:8888/lab?token=7b2655f63a1423f0a59923fe7696b24b9089b8e5786f6d09>  
<http://127.0.0.1:8888/lab?token=7b2655f63a1423f0a59923fe7696b24b9089b8e5786f6d09>

## Course Project: organization

---

### Working groups

- Divide in **4 groups** of 4-5 people
  - At each project step, compile a single project notebook for each group
  - Save it also in html to be able to easily share and discuss your results
- *Try to balance for computational-biological knowledge*

### Working sessions

<b>19/11</b>	17.00 - 18.30	<i>group work</i>
<b>20/11</b>	11.45 - 13.00	<i>group work</i>
<b>21/11</b>	17.00 - 18.30	<i>group work</i>
<b>22/11</b>	14.30 - 17.00	<i>discussion</i>

# Course Project: notebooks

## Notebooks

- Exercise organized in two notebooks: Step1 and Step2
  - *For each step, compile a single project notebook for each group*
  - *Save it also in html to be able to easily share and discuss your results*

1_Day1	Update CourseWalkThrough.md
2_Day2	Updated Notebooks and paths
3_Day3	minor fixes NB 3
4_Day4	Create Day4
Project	Updated path in step1 Project



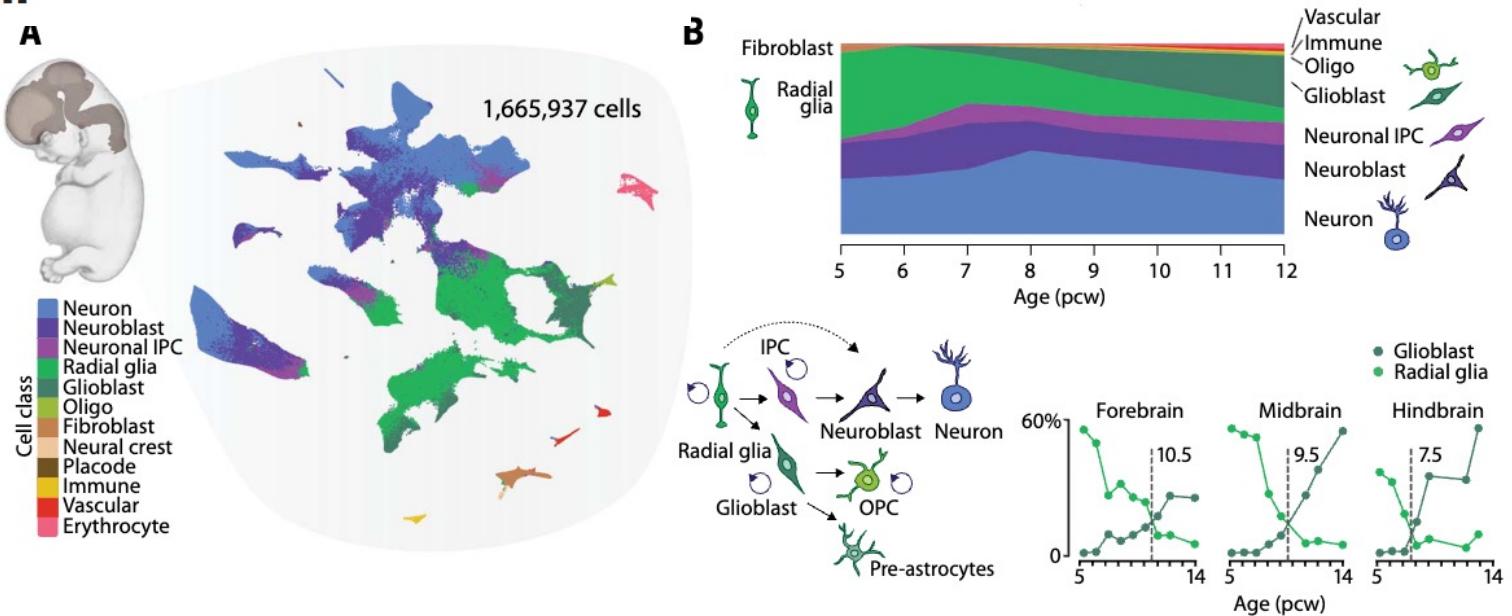
Name	Last commit message
..	
utils	Updated Notebooks Project2
BraunDataset_Step1.ipynb	Updated path in step1 Project
BraunDataset_Step2.ipynb	Updated Notebooks Project2
Helper.py	Added helper file for project

## Container

- Use Day 2 container
- When launching the container and specifying the –H option, check to be able to access Project folder

## Course Project: dataset

### Comprehensive cell atlas of the first-trimester developing human brain



### Features

- 26 brain specimens (5-14 PCW) dissected in 111 distinct biological samples
- snRNASeq profiled 10X Genomics - Chromium Single Cell V2 and V3
- more than 1.6 million cells from 10 brain regions; rich but very complex metadata information

### For our exploration

- selected 3 regions: cerebral cortex, cerebellum and thalamus; subsampled to 75,000 input nuclei