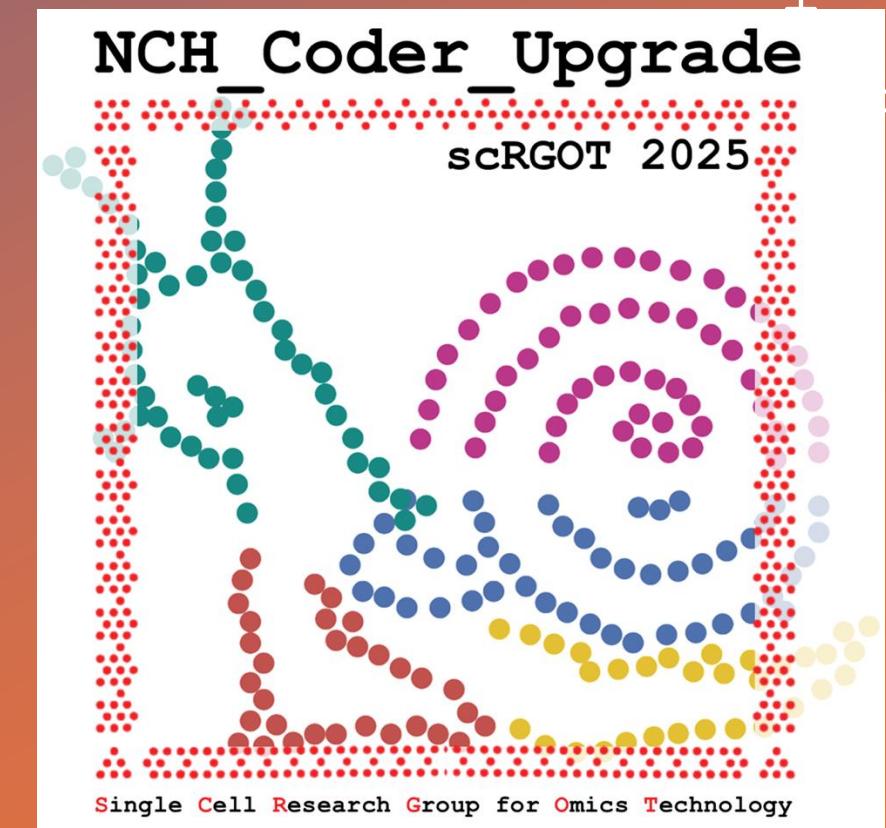
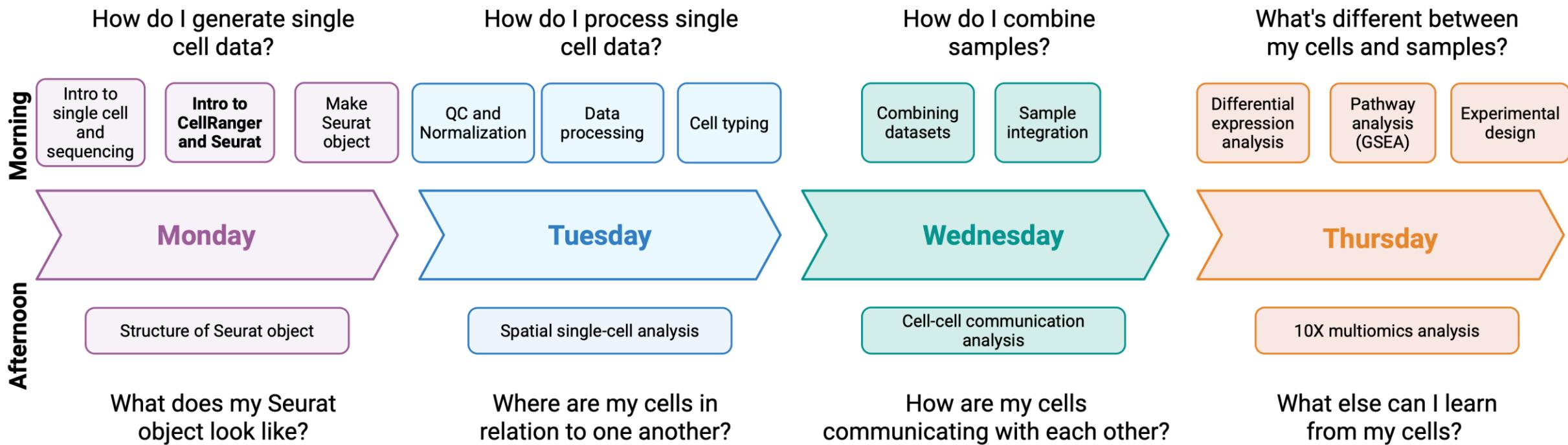


# Introduction to Seurat

Coder Upgrade 2025  
Elizabeth Garfinkle, PhD

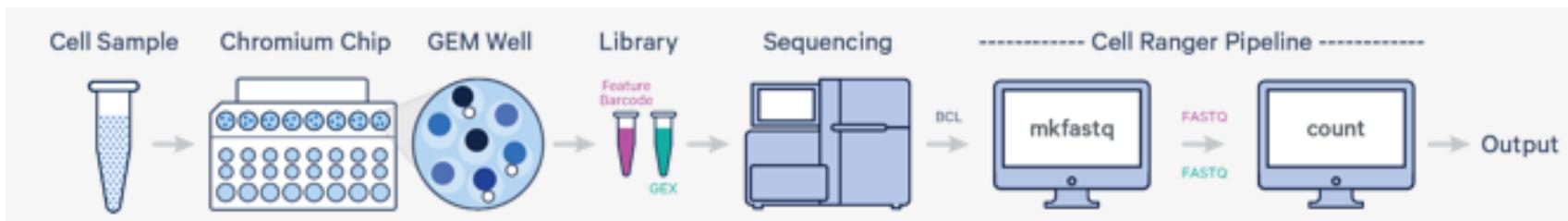


# Coder Upgrade 2025



# Cell Ranger

Set of analyses pipelines that process raw single cell data in preparation for downstream analyses



## **cellranger mkfastq or BCL convert (Illumina)**

**Input:** Raw base call (BCL) files from Illumina sequencer

**Processing:** Demultiplexes files

**Output:** FASTQ files

## **cellranger count**

**Input:** FASTQ files

**Processing:** Alignment (human or mouse), filtering, barcode and UMI (unique molecular identifier) counts

**Output:** Feature-barcode matrices for loading into Seurat for downstream analysis (clustering, gene expression, etc.)

## **cellranger multi**

**Input:** FASTQ files from Fixed RNA profiling samples, samples with GEX and VDJ libraries, or multiplexed samples

**Processing:** Alignment (human or mouse), filtering, barcode and UMI (unique molecular identifier) counts

**Output:** Feature-barcode matrices for loading into Seurat for downstream analysis (clustering, gene expression, etc.)

## **cellranger aggr**

**Input:** Multiple runs of **cellranger count** or **cellranger multi**

**Processing:** Normalizes runs to same sequencing depth

**Output:** Recomputed feature-barcode matrices

# Cell Ranger

## Example **cellranger count**

```
## -j y  
## -cwd  
## -S /bin/bash  
## -N cellranger_count  
## -q all.q  
## -pe smp 16  
## -V  
  
/igm/apps/10X_chromium/cellranger-7.0.0/bin/cellranger count --id=CGITD_Stroma \  
    --transcriptome=/igm/apps/10X_chromium/refdata-gex-mm10-2020-A \  
    --fastqs=/igm/projects/230308_GSL-PY-3079_Dorrance_scRNA-seq_mkfastq/H \  
    --sample=CGITD_Stroma \  
    --include-introns=true \  
    --localcores=16 \  
Include introns  
- Account for 2–40% of all reads  
- Are usable data – arise from polyA tracts in transcripts  
- Increase reads mapped confidently to transcription and media gene per cell, especially for nuclei
```

Reference for alignment – example here is for mouse Transcriptome reference file made with **cellranger mkfastq** -- common mouse and human can be downloaded from 10X

FASTQ file from **cellranger mkfastq**

### Run

```
Martian Runtime - v4.0.8  
  
Running preflight checks (please wait)...  
Checking sample info...  
Checking FASTQ folder...  
Checking reference...  
  
Checking optional arguments...  
...
```

### Successful run

```
Outputs:  
- Run summary HTML: /opt/sample345/outs/web_summary.html  
- Run summary CSV: /opt/sample345/outs/metrics_summary.csv  
- BAM: /opt/sample345/outs/possorted_genome_bam.bam  
- BAM index: /opt/sample345/outs/possorted_genome_bam.bam.bai  
- Filtered feature-barcode matrices MEX: /opt/sample345/outs/filtered_feature_bc_matrix  
- Filtered feature-barcode matrices HDF5: /opt/sample345/outs/filtered_feature_bc_matrix.h5  
- Unfiltered feature-barcode matrices MEX: /opt/sample345/outs/raw_feature_bc_matrix  
- Unfiltered feature-barcode matrices HDF5: /opt/sample345/outs/raw_feature_bc_matrix.h5  
- Secondary analysis output CSV: /opt/sample345/outs/analysis  
- Per-molecule read information: /opt/sample345/outs/molecule_info.h5  
- CRISPR-specific analysis: null  
- Loupe Browser file: /opt/sample345/outs/cloupe.cloupe  
- Feature Reference: null  
- Target Panel File: null  
  
Waiting 6 seconds for UI to do final refresh.  
Pipestance completed successfully!
```

```
yyyy-mm-dd hh:mm:ss Shutting down.  
Saving pipestance info to "tiny/tiny.mri.tgz"
```

# Cell Ranger

## Example `cellranger count`

- Uses STAR (splice-aware) aligner then uses transcript annotation GTF to bucket reads into:
  - Exonic
  - Intronic
  - Intergenic

```
/igm/apps/10X_chromium/refdata-gex-mm10-2020-A/
  └── fasta
      ├── genome.fa
      └── genome.fa.fai
  └── genes
      └── genes.gtf
  └── pickle
      └── genes.pickle
  └── reference.json
  └── star
      ├── chrLength.txt
      ├── chrNameLength.txt
      ├── chrName.txt
      ├── chrStart.txt
      ├── exonGeTrInfo.tab
      ├── exonInfo.tab
      ├── geneInfo.tab
      ├── Genome
      ├── genomeParameters.txt
      ├── SA
      ├── SAindex
      ├── sjdbInfo.txt
      ├── sjdbList.fromGTF.out.tab
      ├── sjdbList.out.tab
      └── transcriptInfo.tab
```

# Cell Ranger – v8

v8.0.0 (Mar 13, 2024)

## New features

- Cell Ranger v8.0 introduces 3' v4 and 5' v3 chemistry to support the analysis of Chromium GEM-X Single Cell Gene Expression v4 and Chromium GEM-X Single Cell Immune Profiling v3 libraries.

Required for samples sequenced on NovaSeqX+

```
#$ -j y
#$ -cwd
#$ -S /bin/bash
#$ -N R_Cellranger
#$ -q all.q
#$ -pe smp 16
#$ -V

/igm/apps/10X_chromium/cellranger-8.0.0/bin/cellranger count --id=24_0268_S0280_150pm \
    --transcriptome=/igm/apps/10X_chromium/refdata-gex-GRCh38-2020-A\
    --fastqs=/igm/projects/Elizabeth/NovaSeqXPlus/L001_150PM/FASTQ/2237:
    --sample=24_0268_S0280\
    --create-bam=true
    --localcores=16 \
    --localmem=64 \
```

Now have to specific “create-bam”

## Transcriptome reference

2024-A (Mar 13, 2024)

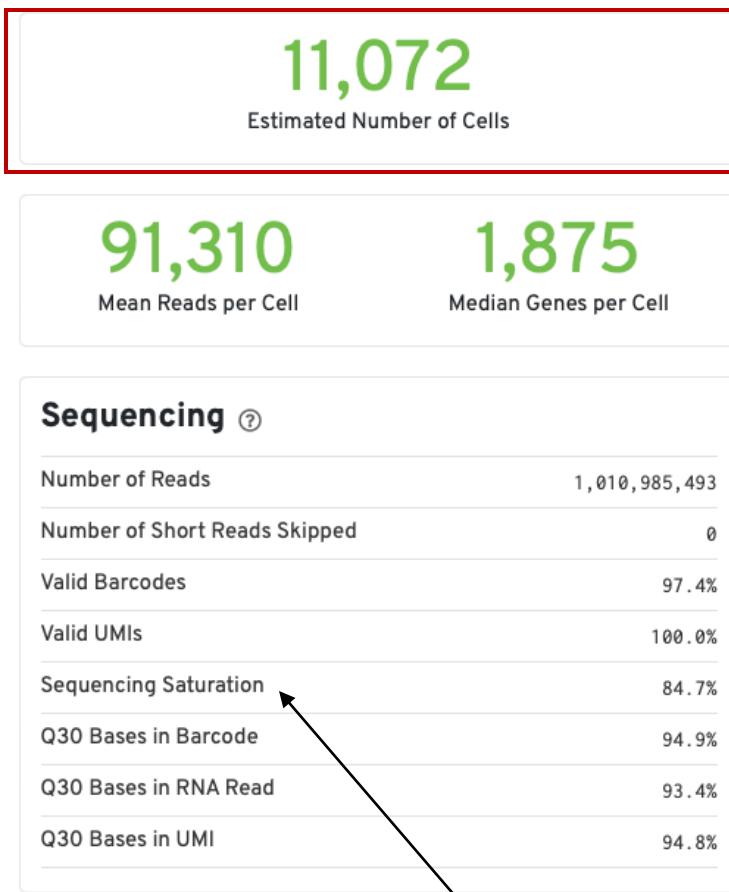
2024-A reference packages are **not** backward compatible with Cell Ranger v5.0.1 and prior.

- Human GRCh38 (GENCODE v44/Ensembl110 annotations)
- Mouse GRCm39 (GENCODE vM33/Ensembl110 annotations)
- Human and mouse GRCh38 and GRCm39

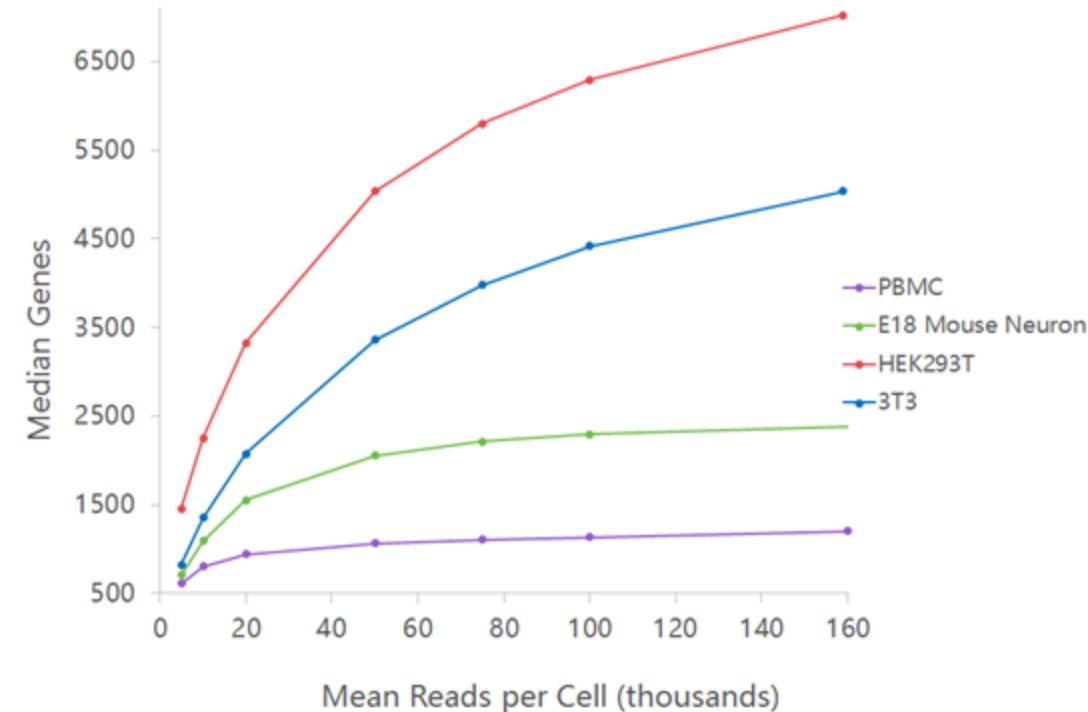
	Human	Mouse
Number of new gene IDs	2339	1746
Number of genes removed	301	335
Number of gene names changed	12913	1905
Number of gene IDs changed (based on gene name)	69	56

# Cell Ranger

Example **cellranger count** – output



Sequencing saturation is dependent on the library complexity and sequencing depth. Different cell types will have different amounts of RNA and thus will differ in the total number of different transcripts in the final library (also known as library complexity)



Level of duplicated reads - measures the fraction of library complexity

Sequencing depth affects sequencing saturation : more reads = more additional unique transcripts you can detect

*Expected to be between 60-80% for most applications*

# Cell Ranger

Example **cellranger count** – output

## Barcode Rank Plot:

All 10x Barcodes detected during sequencing are plotted in decreasing order of the number of UMIs associated with that particular barcode.

The number of UMIs detected in each GEM is used by Cell Ranger to determine which GEMs likely contain a cell.



- Raw feature-barcode matrix contains columns that are empty droplets
- Gene expression counts in these droplets aren't 0 due to technical noise
- Amount of RNA present can be used to identify empty droplets
- Cell filtering algorithm:
  - find first “knee point” in the barcodes vs UMI counts plot

# Cell Ranger

## Example **cellranger count** – output --- poor quality

### Alerts

The analysis detected ⚠ 1 warning.

Alert	Value	Detail
⚠ Low Fraction Reads in Cells	62.4%	Ideal > 70%. Application performance may be affected. Many of the reads were not assigned to cell-associated barcodes. This could be caused by high levels of ambient RNA or by a significant population of cells with a low RNA content, which the algorithm did not call as cells. The latter case can be addressed by inspecting the data to determine the appropriate cell count and using --force-cells.

Summary

Analysis

1,409

Estimated Number of Cells

332,871

Mean Reads per Cell

3,451

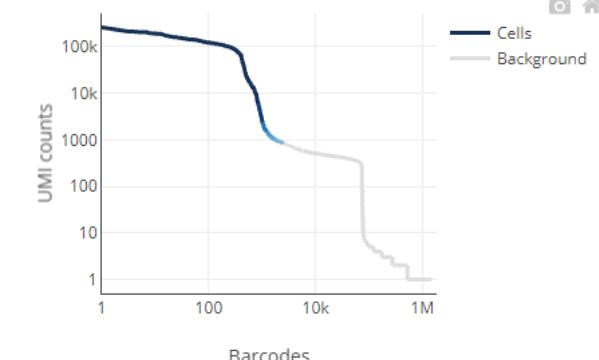
Median Genes per Cell

### Sequencing

Number of Reads	469,014,739
Number of Short Reads Skipped	0
Valid Barcodes	96.2%
Valid UMI	100.0%
Sequencing Saturation	74.0%

### Cells

Barcode Rank Plot



Estimated Number of Cells

1,409

Fraction Reads in Cells

62.4%

Mean Reads per Cell

332,871

# Cell Ranger to Seurat

< > filtered\_feature\_bc\_matrix

Seurat input

Name

- barcodes.tsv.gz
- features.tsv.gz
- matrix.mtx.gz

Each barcode sequence includes a suffix with a dash separator followed by a number:

AAACCCAAGGAGAGTA-1

feature_type	gene	feature_id	AAACCCAAGGAGAGTA-1	AAACGCTTCAGCCCAG-1	...
Gene Expression	MIR1302-2HG	ENSG00000243485	0	0	
Gene Expression	FAM138A	ENSG00000237613	0	0	
Gene Expression	OR4F5	ENSG00000186092	0	0	
Gene Expression	AL627309.1	ENSG00000238009	0	0	
Gene Expression	AL627309.3	ENSG00000239945	0	0	
...					

# What is Seurat?



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

## SATIJA LAB

News   People   Research   Publications   Seurat   Join/Contact   Single Cell Genomics Day

### Welcome to the Satija Lab

Our goal is to understand how cellular heterogeneity encodes the molecular structure, function, and regulation of complex biological systems. Primarily using single cell genomics, we analyze systems by profiling their most fundamental units individually - a 'bottom-up' approach that allows us to study how diverse groups of cells work together to drive biological processes and behaviors.

We're located at the [New York Genome Center](#) in Lower Manhattan, a short walk from our joint appointment at the [NYU Center for Genomics and Systems Biology](#).



# What is Seurat?



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

## SATIJA LAB

News People Research Publications Seurat Join/Contact Single Cell Genomics Day

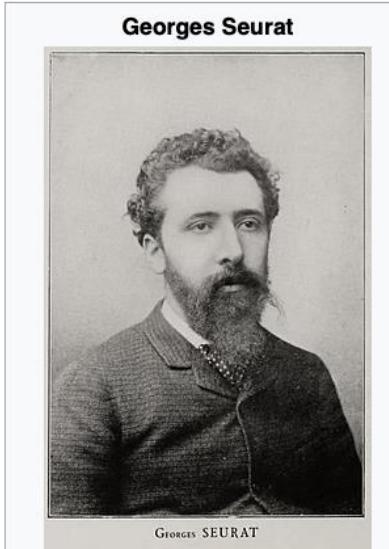
### Welcome to the Satija Lab

Our goal is to understand how cellular heterogeneity encodes the molecular structure, function, and regulation of complex biological systems. Primarily using single cell genomics, we analyze systems by profiling their most fundamental units individually - a 'bottom-up' approach that allows us to study how diverse groups of cells work together to drive biological processes and behaviors.

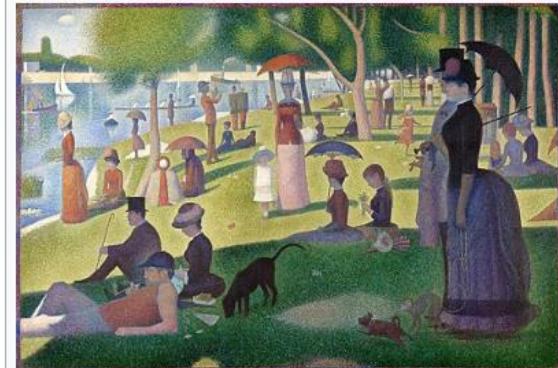
We're located at the [New York Genome Center](#) in Lower Manhattan, a short walk from our joint appointment at the [NYU Center for Genomics and Systems Biology](#).



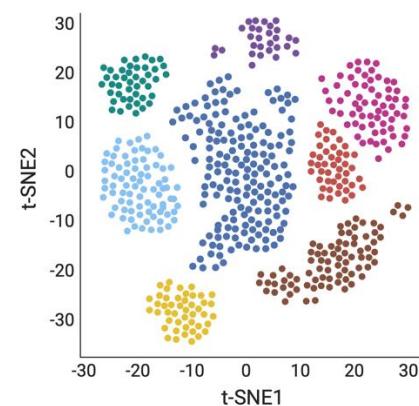
# Who is Seurat?



*A Sunday Afternoon on the Island of La Grande Jatte*



**Artist** Georges Seurat  
**Year** 1884–1886  
**Medium** Oil on canvas  
**Subject** People relaxing at [la Grande Jatte](#) in Paris  
**Dimensions** 207.6 cm × 308 cm (81.7 in × 121.25 in)  
**Location** Art Institute of Chicago, Chicago



*Pointillism*

Wikipedia  
Biorender.com

# What is Seurat?



## SATIJA LAB

News People Research Publications Seurat Join/Contact Single Cell Genomics Day

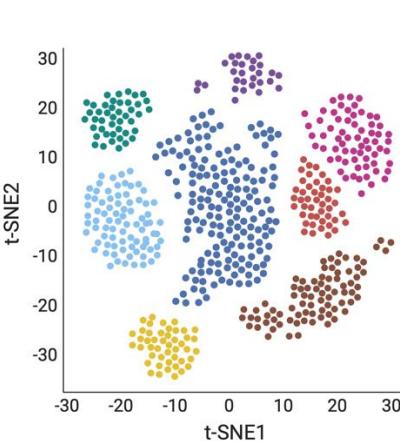
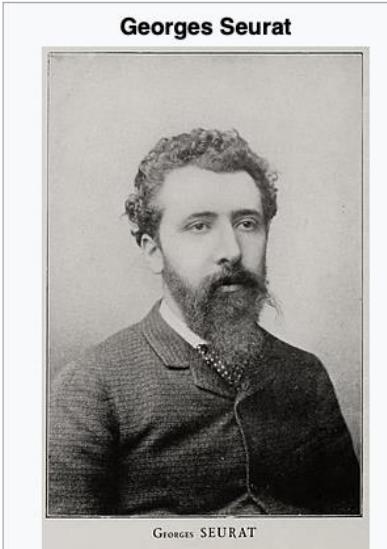
### Welcome to the Satija Lab

Our goal is to understand how cellular heterogeneity encodes the molecular structure, function, and regulation of complex biological systems. Primarily using single cell genomics, we analyze systems by profiling their most fundamental units individually - a 'bottom-up' approach that allows us to study how diverse groups of cells work together to drive biological processes and behaviors.

We're located at the [New York Genome Center](#) in Lower Manhattan, a short walk from our joint appointment at the [NYU Center for Genomics and Systems Biology](#).



# Who is Seurat?



Topiary Park in Columbus, Ohio  
replicates much of the painting.



**Artist** Georges Seurat  
**Year** 1884–1886  
**Medium** Oil on canvas  
**Subject** People relaxing at [la Grande Jatte](#) in Paris  
**Dimensions** 207.6 cm × 308 cm (81.7 in × 121.25 in)  
**Location** Art Institute of Chicago, Chicago

Pointillism

# What is Seurat?



## SATIJA LAB

News People Research Publications Seurat Join/Contact Single Cell Genomics Day

### Welcome to the Satija Lab

Our goal is to understand how cellular heterogeneity encodes the molecular structure, function, and regulation of complex biological systems. Primarily using single cell genomics, we analyze systems by profiling their most fundamental units individually - a 'bottom-up' approach that allows us to study how diverse groups of cells work together to drive biological processes and behaviors.

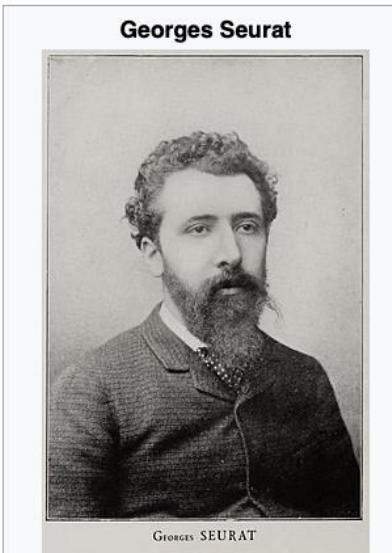
We're located at the [New York Genome Center](#) in Lower Manhattan, a short walk from our joint appointment at the [NYU Center for Genomics and Systems Biology](#).



Other tools: Scanpy (Python)

# Who is Seurat?

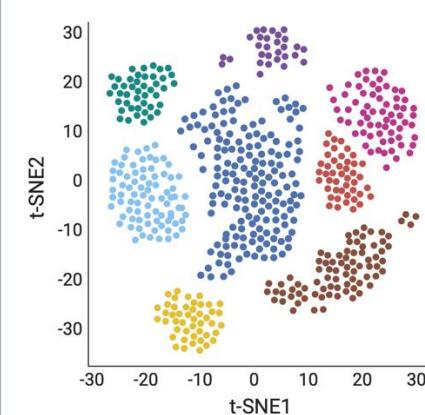
Topiary Park in Columbus, Ohio replicates much of the painting.



**A Sunday Afternoon on the Island of La Grande Jatte**

**Artist** Georges Seurat  
**Year** 1884–1886  
**Medium** Oil on canvas  
**Subject** People relaxing at [la Grande Jatte](#) in Paris  
**Dimensions** 207.6 cm × 308 cm (81.7 in × 121.25 in)  
**Location** Art Institute of Chicago, Chicago

Pointillism



Wikipedia  
Biorender.com

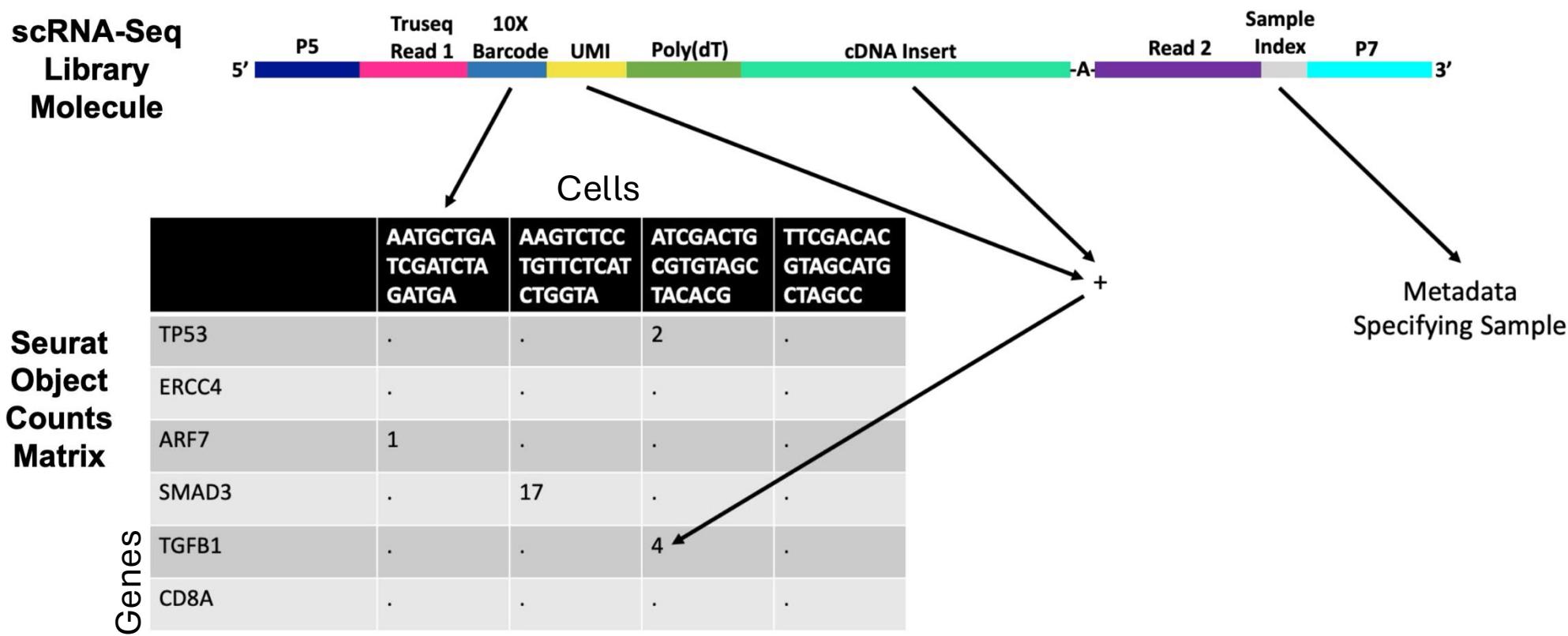
# Cell Ranger to Seurat

< > filtered\_feature\_bc\_matrix

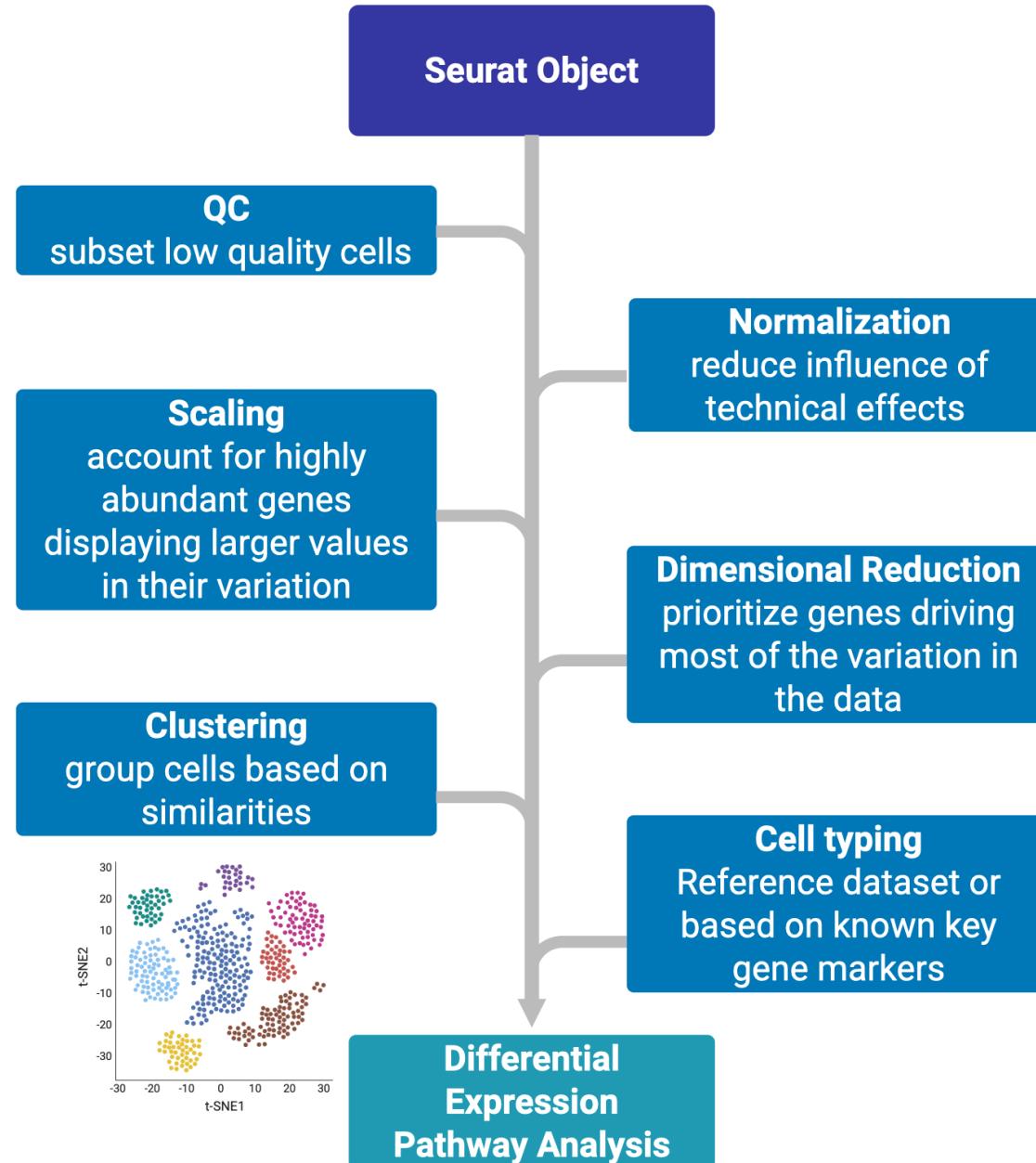
Name

- barcodes.tsv.gz
- features.tsv.gz
- matrix.mtx.gz

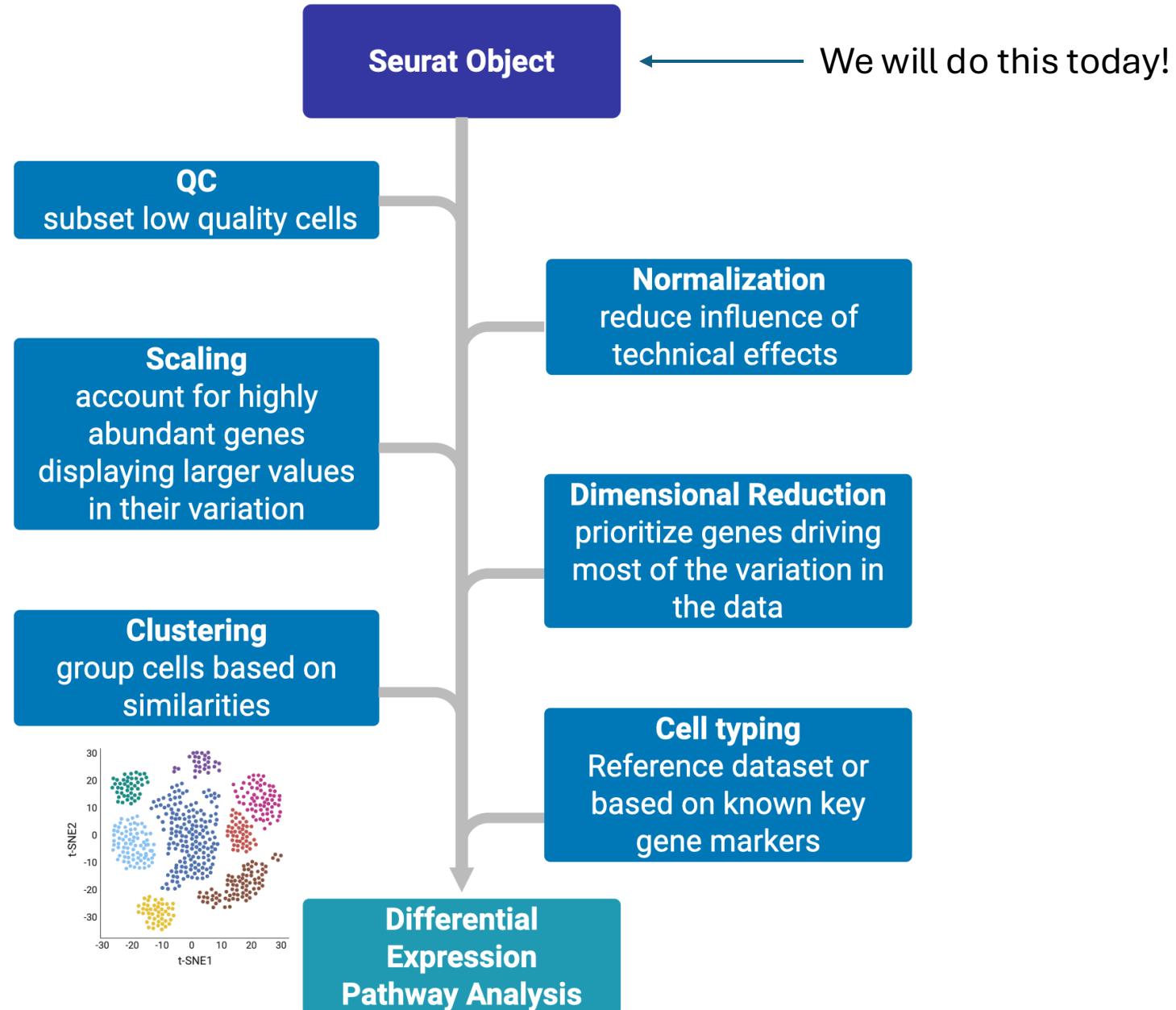
feature_type	gene	feature_id	AAACCCAAGGAGAGTA-1	AAACGCTTCAGCCCAG-1	...
Gene Expression	MIR1302-2HG	ENSG00000243485	0	0	
Gene Expression	FAM138A	ENSG00000237613	0	0	
Gene Expression	OR4F5	ENSG00000186092	0	0	
Gene Expression	AL627309.1	ENSG00000238009	0	0	
Gene Expression	AL627309.3	ENSG00000239945	0	0	
...					



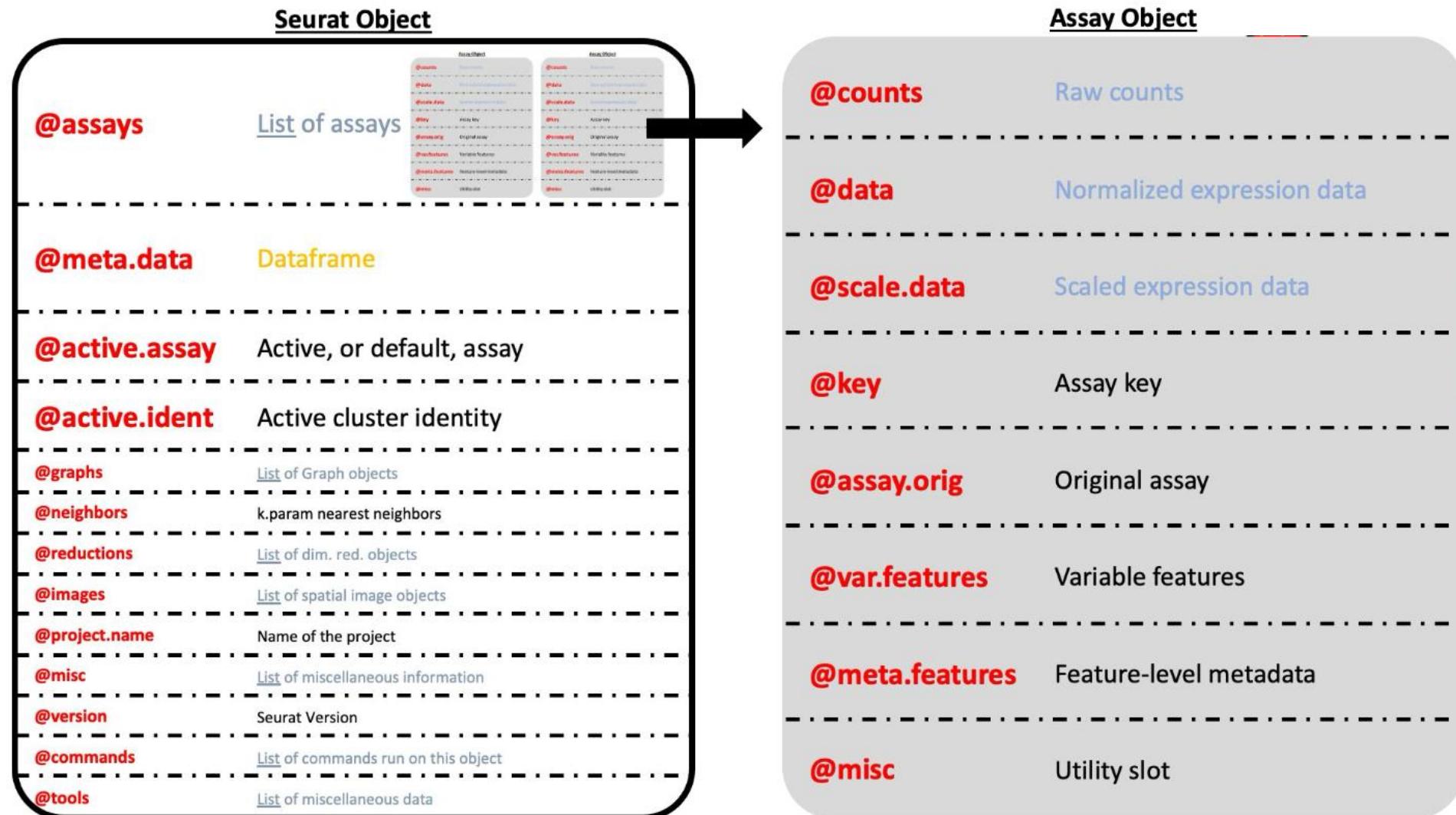
# Seurat analysis workflow overview



# Seurat analysis workflow overview



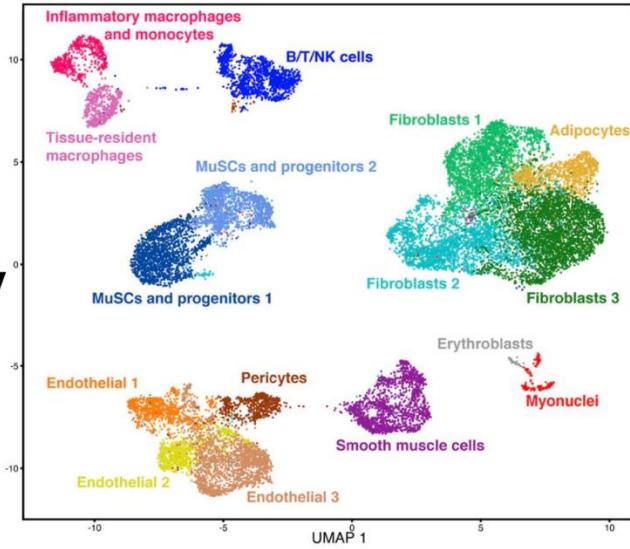
# Seurat object structure



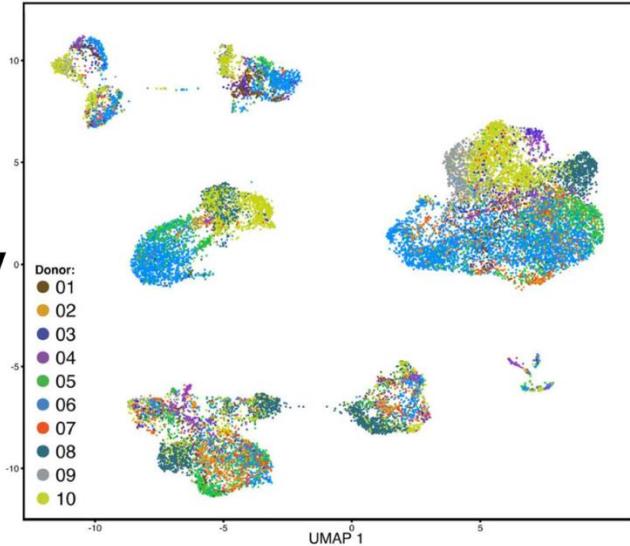
# Metadata

- Additional data associated with specific cells
- Examples:
  - Sample of origin
  - Percent of each cell's total reads aligning to mitochondrial genes
  - Whether or not that cell expressed an interesting gene
  - Total number of features (genes) with at least 1 sequencing read in each cell (nFeatures)
  - Later: cell type annotations
- The same cell barcodes ('AATGTATCTAACTATA') used in the counts matrix are used in the metadata to identify the corresponding cells

**Data grouped by 'Cell Type' Metadata**



**Data grouped by 'Donor' Metadata**



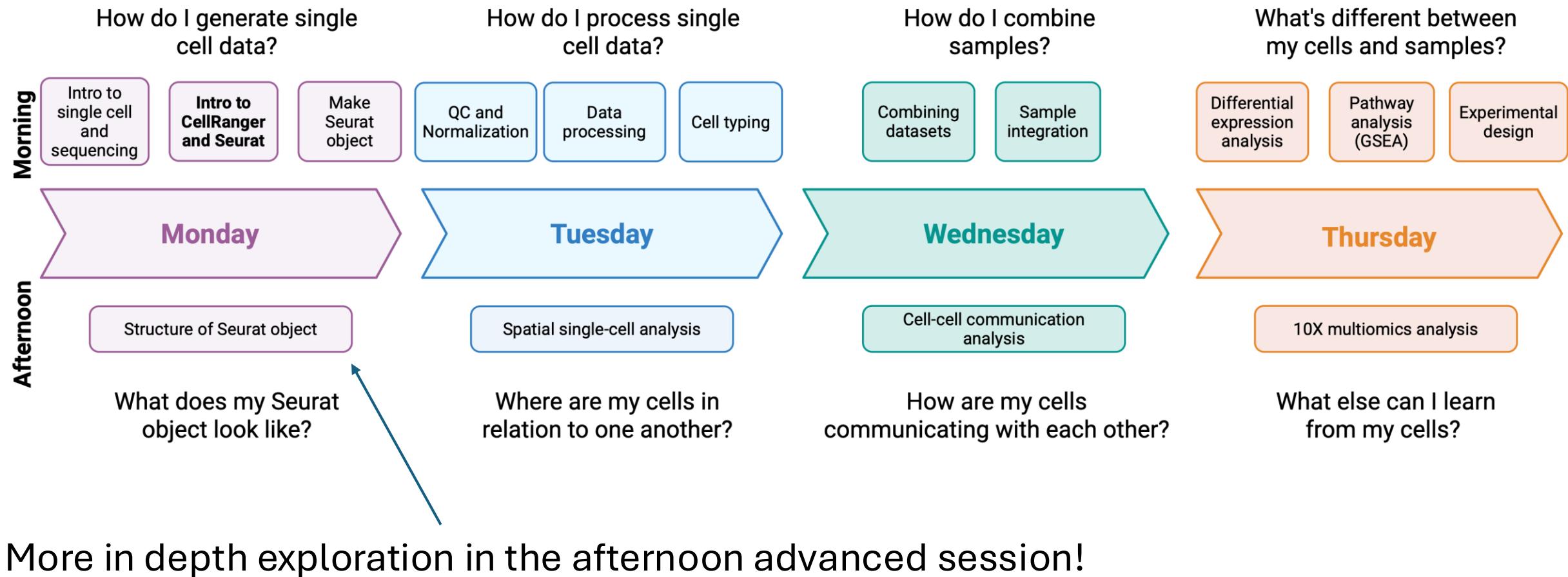
De Micheli, A.J., Spector, J.A., Elemento, O. et al. A reference single-cell transcriptomic atlas of human skeletal muscle tissue reveals bifurcated muscle stem cell populations. *Skeletal Muscle* 10, 19 (2020). <https://doi.org/10.1186/s13395-020-00236-3>

# Metadata

active.idents = the metadata that your cells are currently grouped by

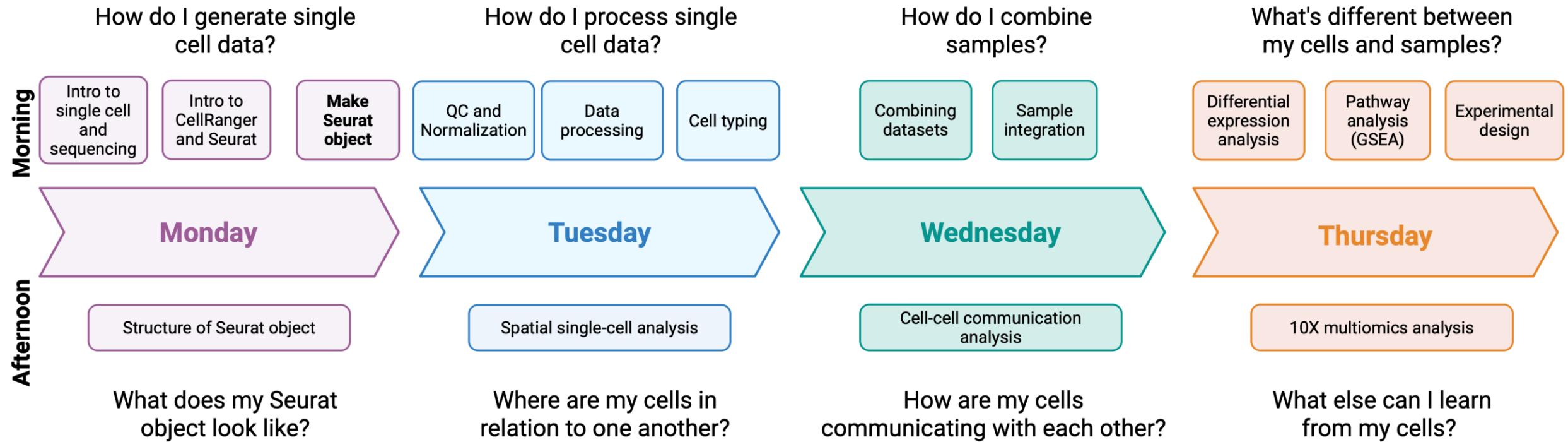
Patients	S4 [16115 x 807] (SeuratObject::S	S4 object of class Seurat
assays	list [2]	List of length 2
meta.data	list [807 x 9] (S3: data.frame)	A data.frame with 807 rows and 9 columns
orig.ident	factor	Factor with 24 levels: "T1", "T10", "T11", "T12", "T13", "T14", ...
nCount_RNA	double [807]	8196 7480 13989 19802 10842 6777 ...
nFeature_RNA	integer [807]	2357 2204 4060 4299 2990 2294 ...
percent.mito	double [807]	4.42 2.61 5.10 3.96 8.49 5.56 ...
percent.mt	double [807]	4.42 2.61 5.10 3.96 8.49 5.56 ...
nCount_SCT	double [807]	8554 8374 10031 9916 9443 8172 ...
nFeature_SCT	integer [807]	2355 2204 3956 3534 2985 2290 ...
SCT_snn_res.0.8	factor	Factor with 11 levels: "0", "1", "2", "3", "4", "5", ...
seurat_clusters	factor	Factor with 11 levels: "0", "1", "2", "3", "4", "5", ...
active.assay	character [1]	'SCT'
active.ident	factor	Factor with 11 levels: "Ductal1", "Macrophage1", "Macrophage2", "Ductal2", "4", ...
graphs	list [2]	List of length 2
neighbors	list [0]	List of length 0
reductions	list [2]	List of length 2
images	list [0]	List of length 0

# Seurat object structure



15 minute break

# Coder Upgrade 2025



# What are our samples?

nature aging



Resource

<https://doi.org/10.1038/s43587-023-00373-6>

## Heterochronic parabiosis reprograms the mouse brain transcriptome by shifting aging signatures in multiple cell types

Received: 19 May 2022

Accepted: 30 January 2023

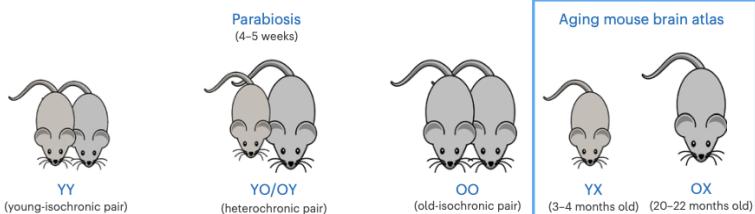
Published online: 9 March 2023

Check for updates

Methodios Ximerakis <sup>1,2,3,8</sup>, Kristina M. Holton <sup>1,2,3,8</sup>, Richard M. Giadone <sup>1,2</sup>, Ceren Ozek <sup>1,2</sup>, Monika Saxena <sup>1,2</sup>, Samara Santiago <sup>1,2</sup>, Xian Adiconis <sup>3,4</sup>, Danielle Dionne <sup>4</sup>, Lan Nguyen <sup>4</sup>, Kavya M. Shah <sup>1,2</sup>, Jill M. Goldstein <sup>1,2</sup>, Caterina Gasperini <sup>1,2</sup>, Ioannis A. Gampierakis <sup>1,2</sup>, Scott L. Lipnick <sup>1,2,3</sup>, Sean K. Simmons <sup>3,4</sup>, Sean M. Buchanan <sup>1,2</sup>, Amy J. Wagers <sup>1,2,5,6</sup>, Aviv Regev <sup>4,7</sup>, Joshua Z. Levin <sup>3,4</sup> & Lee L. Rubin <sup>1,2,3</sup>

Aging is a complex process involving transcriptomic changes associated with deterioration across multiple tissues and organs, including the brain. Recent studies using heterochronic parabiosis have shown that various aspects of aging-associated decline are modifiable or even reversible. To better understand how this occurs, we performed single-cell transcriptomic profiling of young and old mouse brains after parabiosis. For each cell type, we cataloged alterations in gene expression, molecular pathways, transcriptional networks, ligand–receptor interactions and senescence status. Our analyses identified gene signatures, demonstrating that heterochronic parabiosis regulates several hallmarks of aging in a cell-type-specific manner. Brain endothelial cells were found to be especially malleable to this intervention, exhibiting dynamic transcriptional changes that affect vascular structure and function. These findings suggest new strategies for slowing deterioration and driving regeneration in the aging brain through approaches that do not rely on disease-specific mechanisms or actions of individual circulating factors.

a



Sample for training: GSM6925133\_OX1X  
We will use this sample all week

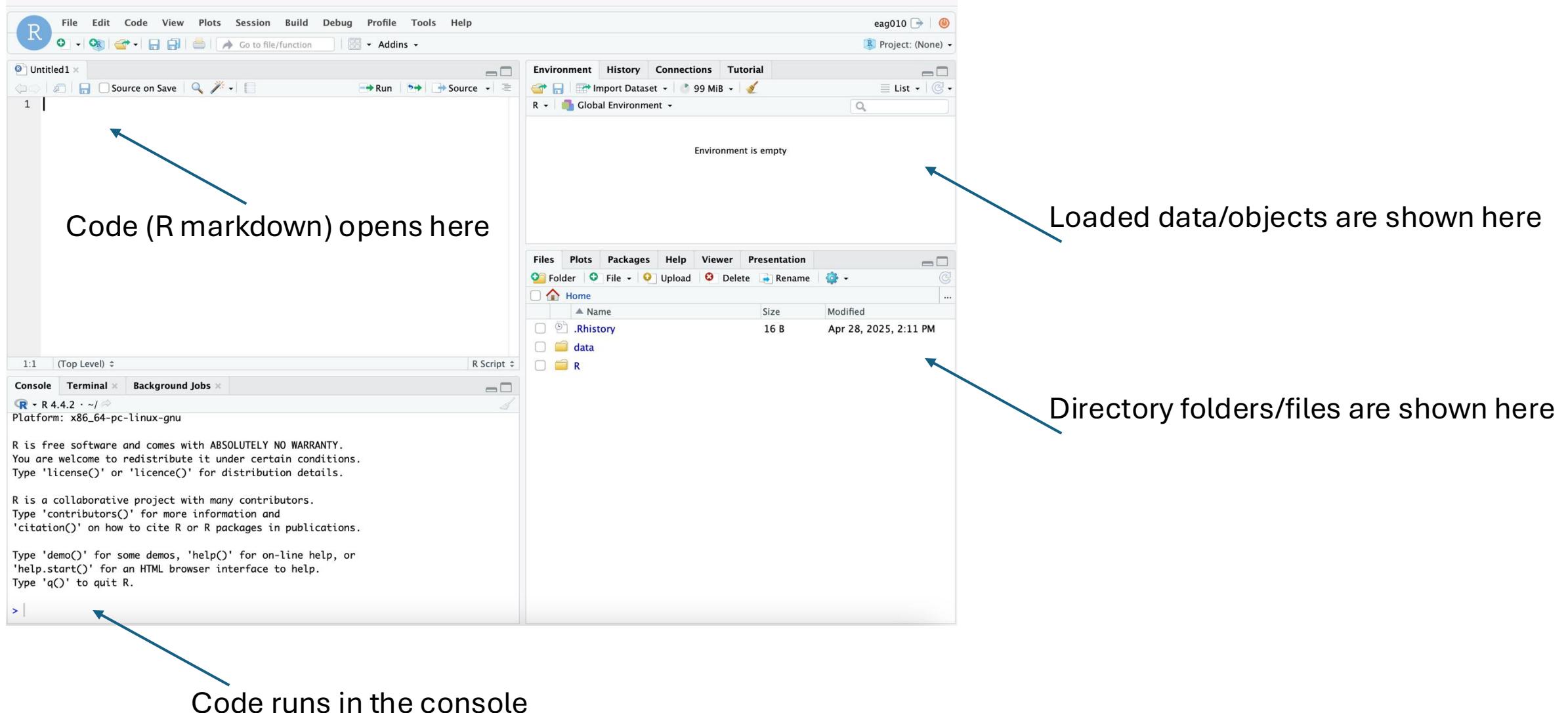
## Results

### Single-cell profiling to study rejuvenation and aging acceleration

We employed high-throughput scRNA-seq to examine the transcriptional profiles of young and old mouse brains after parabiosis (Fig. 1a,b). We generated heterochronic pairs in which 3-to 4-month-old mice were joined with 20- to 22-month-old mice. We also generated age-matched isochronic pairs of young and old mice as controls. All pairs were maintained for 4–5 weeks before tissue collection and analysis. We confirmed successful parabiosis and establishment of blood crosscirculation as previously described<sup>38</sup> (Extended Data Fig. 1). We dissociated the brain tissues using our recently developed protocol<sup>37</sup> and analyzed the transcriptomes of 158,767 single cells (Extended Data Figs. 2 and 3). On stringent filtering and batch effect examination (Methods), we retained 105,329 cells, of which 67,992 cells derived from 34 parabionts (7 isochronic young (YY), 9 heterochronic young (YO), 7 isochronic old (OO), 11 heterochronic old (OY) and 37,337 cells derived from 16 unpaired animals (8 young (YX) and 8 old (OX)) (Fig. 1 and Extended Data Figs. 4 and 5).

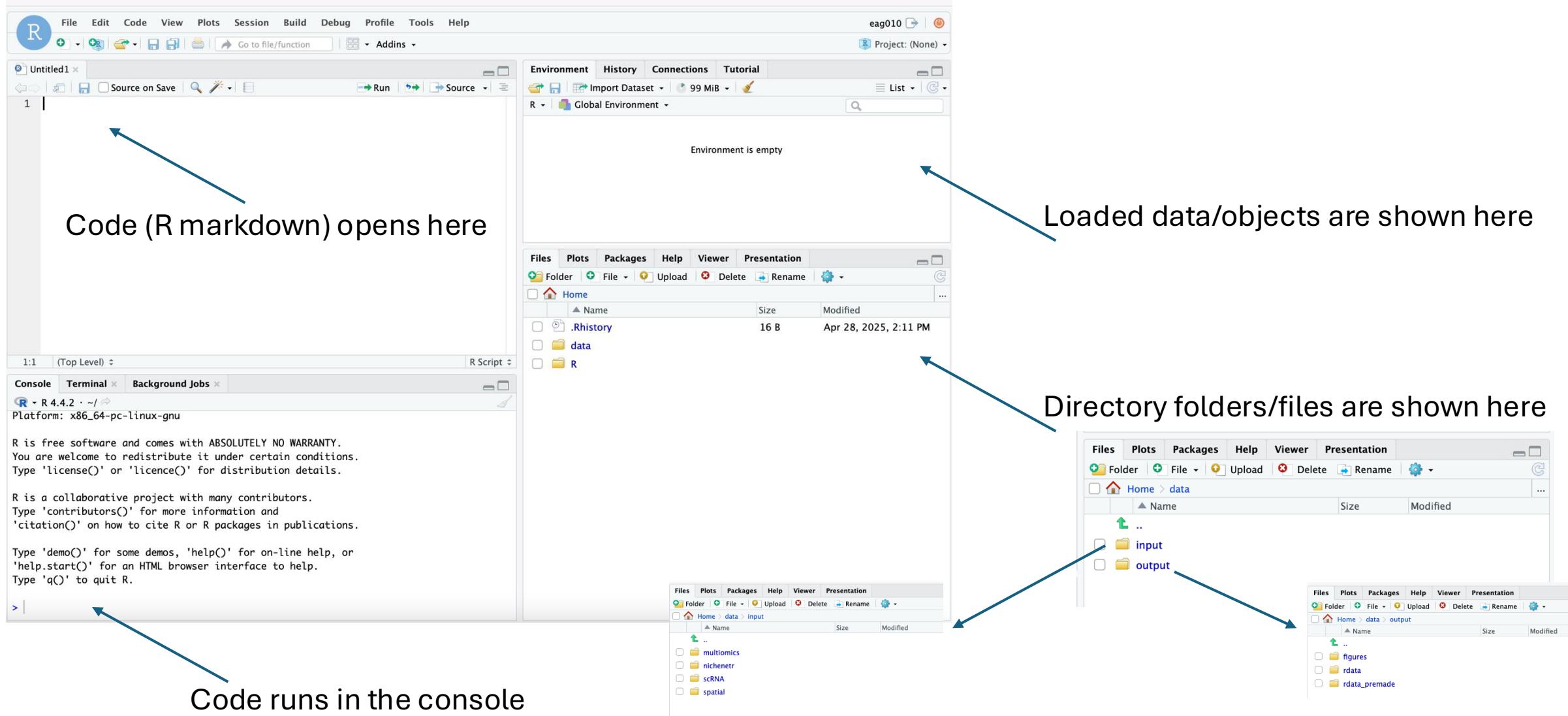
# RStudio

Follow link sent in e-mail



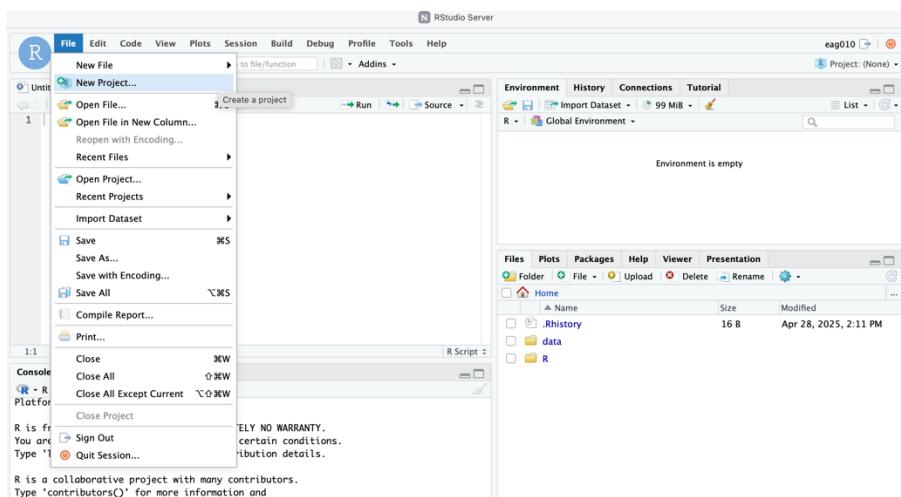
# RStudio

Follow link sent in e-mail

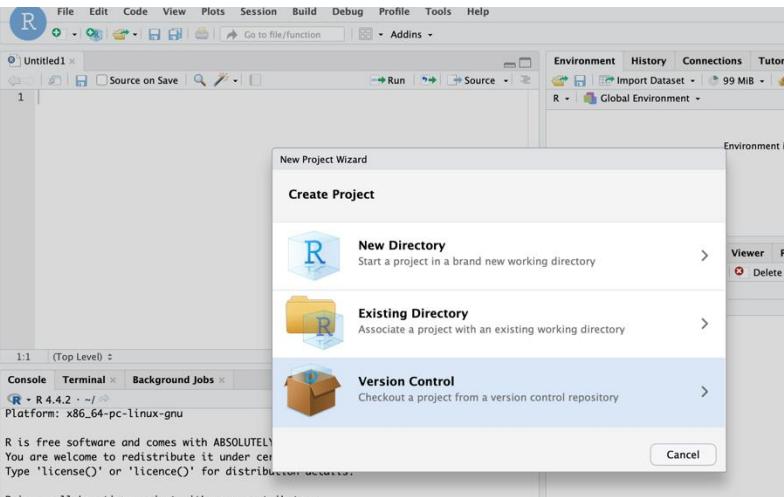


# Setting up a project

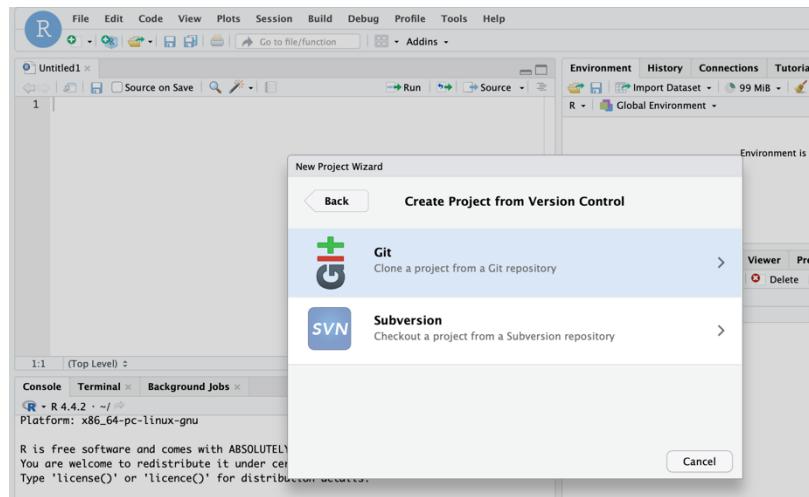
## 1. Go to File then New Project



## 2. Choose Version Control



## 3. Select Git

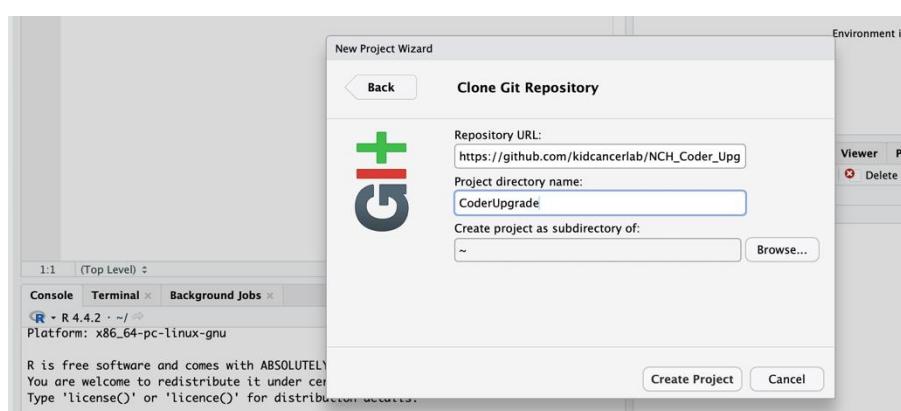


## 4. Paste GitHub Coder Upgrade link:

[https://github.com/kidcancerlab/NCH\\_Coder\\_Upgrade](https://github.com/kidcancerlab/NCH_Coder_Upgrade)

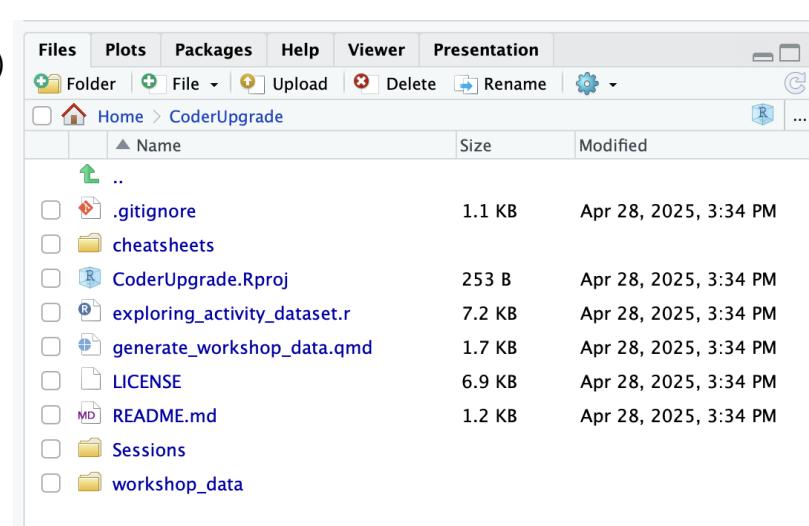
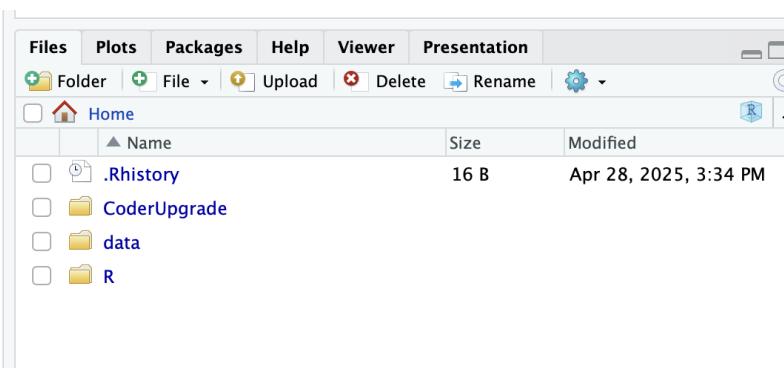
Project directory name: CoderUpgrade

Click Create Project

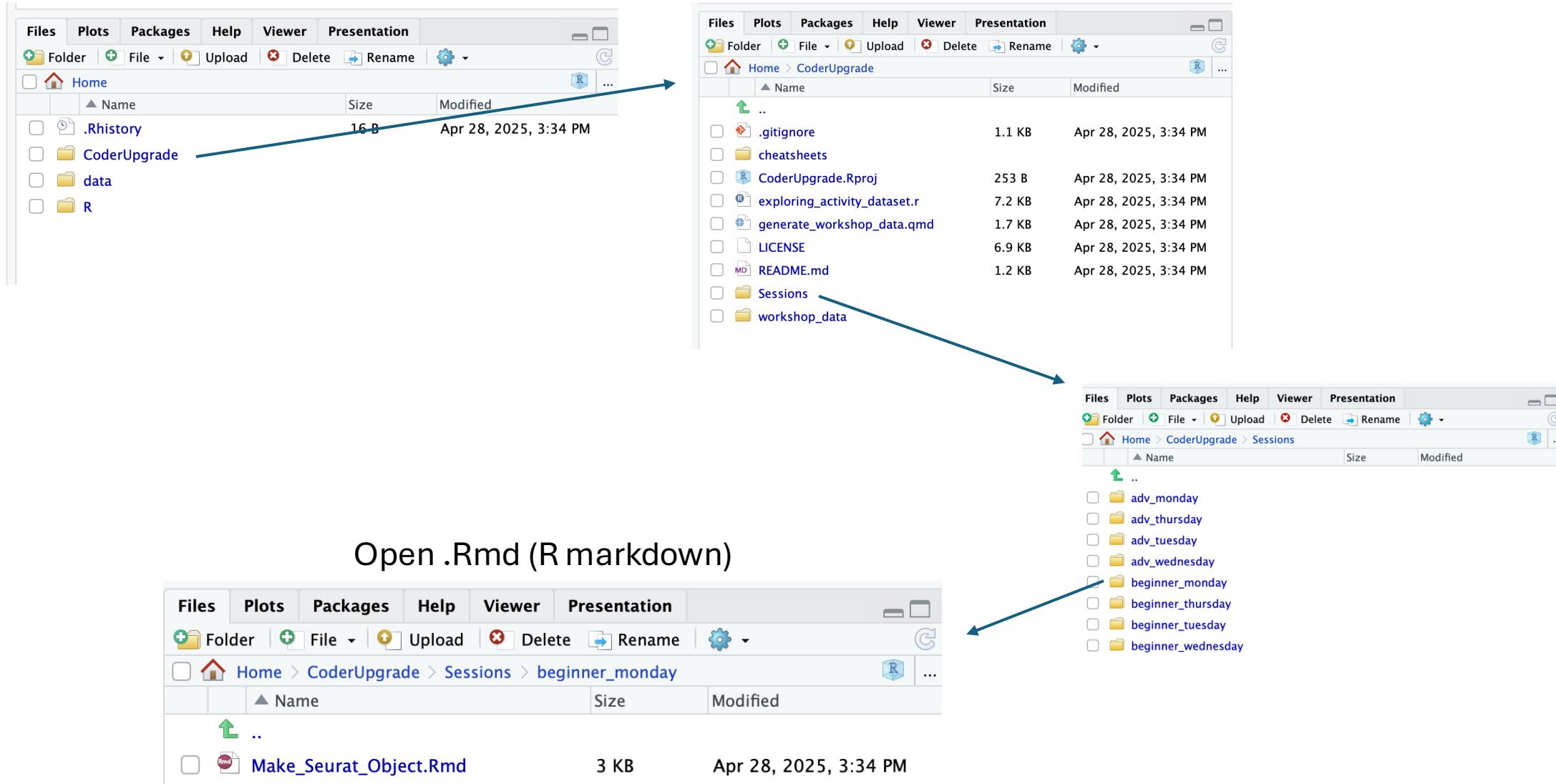


## 5. You should now see a folder called CoderUpgrade

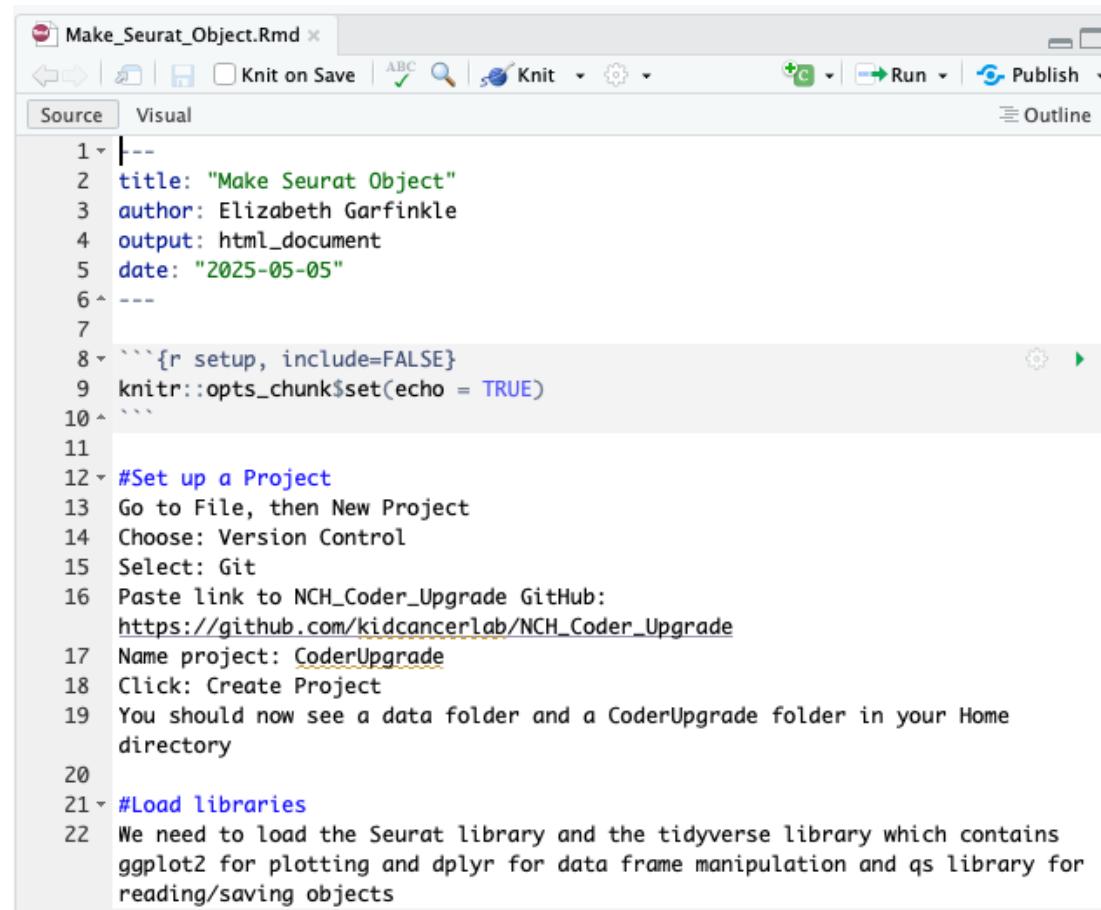
Navigate to Sessions subfolder for each R markdown (lesson)



# Open R markdown – beginner\_monday



# Open R markdown – beginner\_monday



The screenshot shows the RStudio interface with the file "Make\_Seurat\_Object.Rmd" open. The window title is "Make\_Seurat\_Object.Rmd". The toolbar includes icons for back, forward, save, knit, run, and publish. Below the toolbar, tabs for "Source" and "Visual" are visible, with "Source" selected. The main pane displays the R Markdown code:

```
1 ---  
2 title: "Make Seurat Object"  
3 author: Elizabeth Garfinkle  
4 output: html_document  
5 date: "2025-05-05"  
6 ---  
7  
8 ```{r setup, include=FALSE}  
9 knitr::opts_chunk$set(echo = TRUE)  
10 ```  
11  
12 #Set up a Project  
13 Go to File, then New Project  
14 Choose: Version Control  
15 Select: Git  
16 Paste link to NCH_Coder_Upgrade GitHub:  
17 https://github.com/kidcancerlab/NCH\_Coder\_Upgrade  
18 Name project: CoderUpgrade  
19 Click: Create Project  
20 You should now see a data folder and a CoderUpgrade folder in your Home directory  
21 #Load libraries  
22 We need to load the Seurat library and the tidyverse library which contains ggplot2 for plotting and dplyr for data frame manipulation and qs library for reading/saving objects
```

# Loading libraries

Libraries need to be loaded each time you run an R script

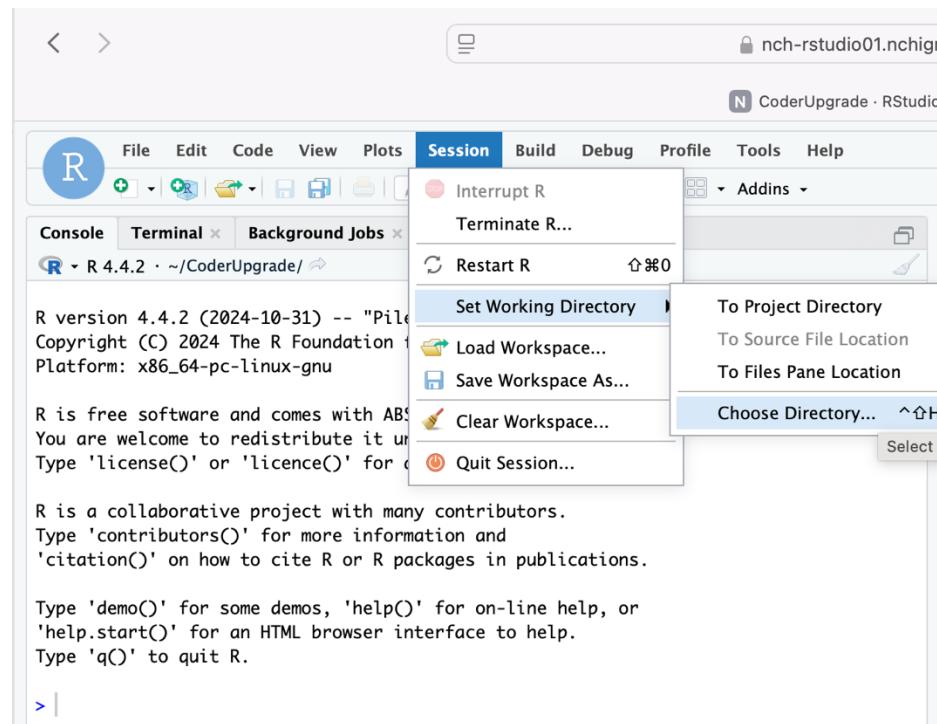
```
#Load libraries
We need to load the Seurat library and the tidyverse library which contains ggplot2 for plotting and
dplyr for data frame manipulation and qs library for reading/saving objects
```{r}
library(Seurat)
library(tidyverse) # This package contains ggplot2 and dplyr
library(qs) # For reading/saving objects
```
```

```

# Setting a working directory

To read/write files, you need to specify your working directory

```
#Check and set working directory
getwd() # This will show you your current working directory
setwd("") # You can set your working directory by putting your desired file path in the quotes or click
Session, Set Working Directory, Choose Directory, then navigate to your file
```{r}
#check what your current working directory is
getwd() #should look like "/rstudio-workshop/home/username/CoderUpgrade"
```
``
```



# Making a Seurat object

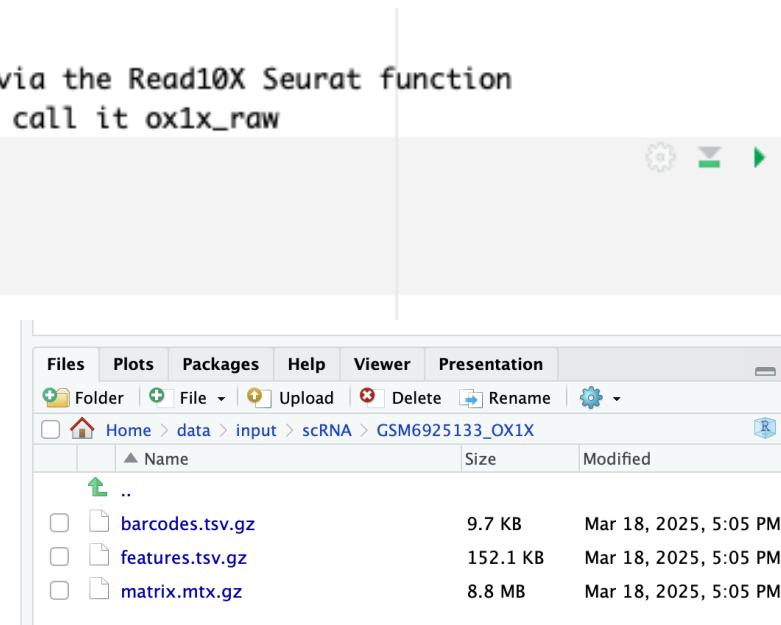
```
#Read 10X data
```

First, read in the filtered feature matrix output files via the Read10X Seurat function

We will start with reading in just one sample, 0X1X, and call it ox1x\_raw

```
```{r}  
ox1x_raw <-  
  Read10X("~/data/input/scRNA/GSM6925133_0X1X")  
...```

```



Name	Size	Modified
..		
barcodes.tsv.gz	9.7 KB	Mar 18, 2025, 5:05 PM
features.tsv.gz	152.1 KB	Mar 18, 2025, 5:05 PM
matrix.mtx.gz	8.8 MB	Mar 18, 2025, 5:05 PM

feature_type	gene	feature_id	AAACCCAAGGAGAGTA-1	AAACGCTTCAGCCAG-1	...
Gene Expression	MIR1302-2HG	ENSG00000243485	0	0	
Gene Expression	FAM138A	ENSG00000237613	0	0	
Gene Expression	OR4F5	ENSG00000186092	0	0	
Gene Expression	AL627309.1	ENSG00000238009	0	0	
Gene Expression	AL627309.3	ENSG00000239945	0	0	
...					

# Making a Seurat object

```
#Make Seurat object
Now we will make a Seurat object via the CreateSeuratObject Seurat function
from the data we just read in that we named ox1x_raw and call it ox1x_sobj
We also set two standard filters:
min.cells = 3
  Here, we are filtering out any genes that are detected in fewer than 3 cells
  to remove genes that are likely not expressed in a meaningful way
min.features = 200
  Here, we are filtering out any cells that have less than 200 features
  (genes) detected, which could indicate poor quality or dying cells
Both of these filters minimize noise/poor quality data for downstream analyses
```{r}
ox1x_sobj <-
  CreateSeuratObject(
    counts = ox1x_raw,
    min.cells = 3,
    min.features = 200
  )
```
```


```

# Metadata

## #Rename sample

By default, the `orig.ident` (original `ident`, aka original sample name) is the name that is assigned when making the Seurat object. Often, this name is short/easy to type but we may want a more meaningful/longer name when making plots for figures or publications. Here, we can reassign the `orig.ident` to have a more descriptive name.

The use of the `$` sign means you are accessing a metadata slot.

```
```{r}
```

```
#Look at the orig.ident name  
View(ox1x_sobj)
```

## #Reassign name

```
ox1x_sobj$orig.ident <- "GSM6925133_OX1X"
```

## #Look at the orig.ident name now

```
View(ox1x_sobj)
```

## #Set orig.ident to be the active ident

```
Idents(ox1x_sobj) <- "orig.ident"
```

## #Check

```
View(ox1x_sobj)
```

```
...
```

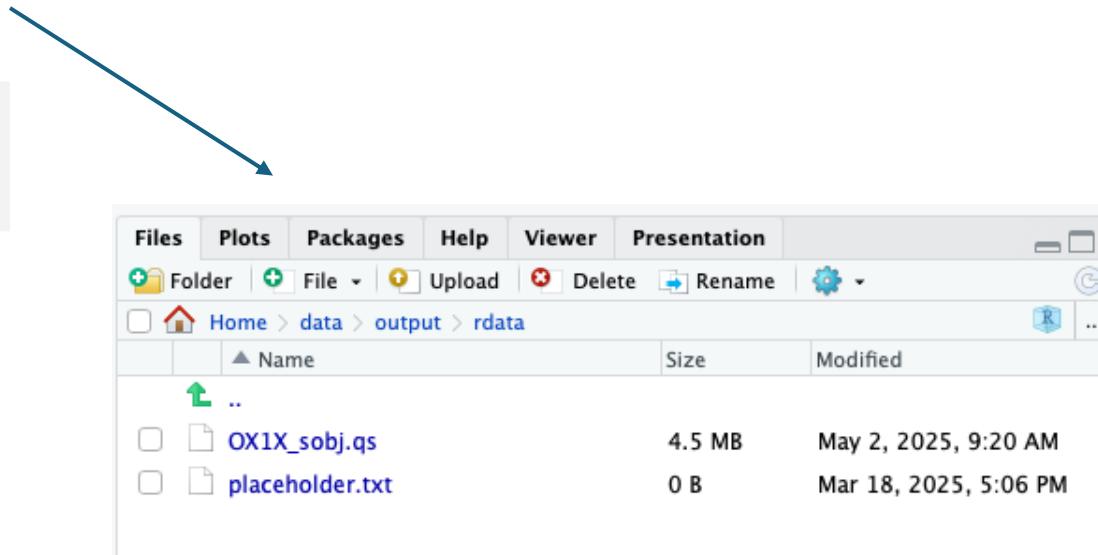
# Saving and reading objects

It's important to save objects as you progress through analysis so that you can go back to different steps for troubleshooting or making changes – this way you don't always have to start from the very beginning

## #Save Seurat object

We will now save our newly made Seurat object using the `qsave` function. The object name you want to save is listed first and then the output file path and name you want to save the object as is listed in quotes. By default, the object will save in your current working directory (important to know what that is at all times!). If you want to save it in a subfolder within your current working directory, you can add the file path before the final name of the saved object. Use `.qs` to save the final object.

```
```{r}
qsave(ox1x_sobj, "~/data/output/rdata/OX1X_sobj.qs") #save to data/output/rdata
directory
````
```



# Saving and reading objects

You can read in saved objects after they are saved and removed from your environment

```
#Remove data and Seurat object
```

Now that we have saved our object, we can remove the data and object in our Environment using rm.

```
```{r}
rm(ox1x_raw, ox1x_sobj)
```

```

```
#We can read in a saved object any time for additional analyses using the
qread function.
```

```
```{r}
ox1x_sobj <-
  qread("~/data/output/rdata_premade/OX1X_sobj.qs") #read in premade object
```

```

# Challenges

```
#We can read in a saved object any time for additional analyses using the
qread function.
```{r}
ox1x_sobj <-
  qread("~/data/output/rdata_premade/0X1X_sobj.qs") #read in premade object

#challenge - read in your own made object from your home directory (hint: give
it a unique name so you don't overwrite the premade object that's already
loaded):
```

#Challenge:
Make a Seurat object from an additional sample in our data set,
GSM6925134_0X2X, and call it ox2x_sobj. After you make the Seurat object,
remove both ox1x_sobj and ox2x_sobj from your Environment.
```

Write and run your code here:

```
```{r}
```

```
```
```

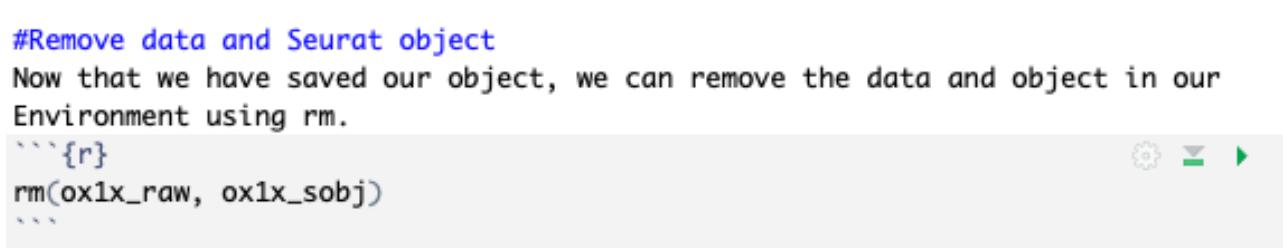
# Clearing environment and logging out

You can use the rm function to remove objects in your environment

```
#Remove data and Seurat object
```

Now that we have saved our object, we can remove the data and object in our Environment using rm.

```
```{r}  
rm(ox1x_raw, ox1x_sobj)  
```
```

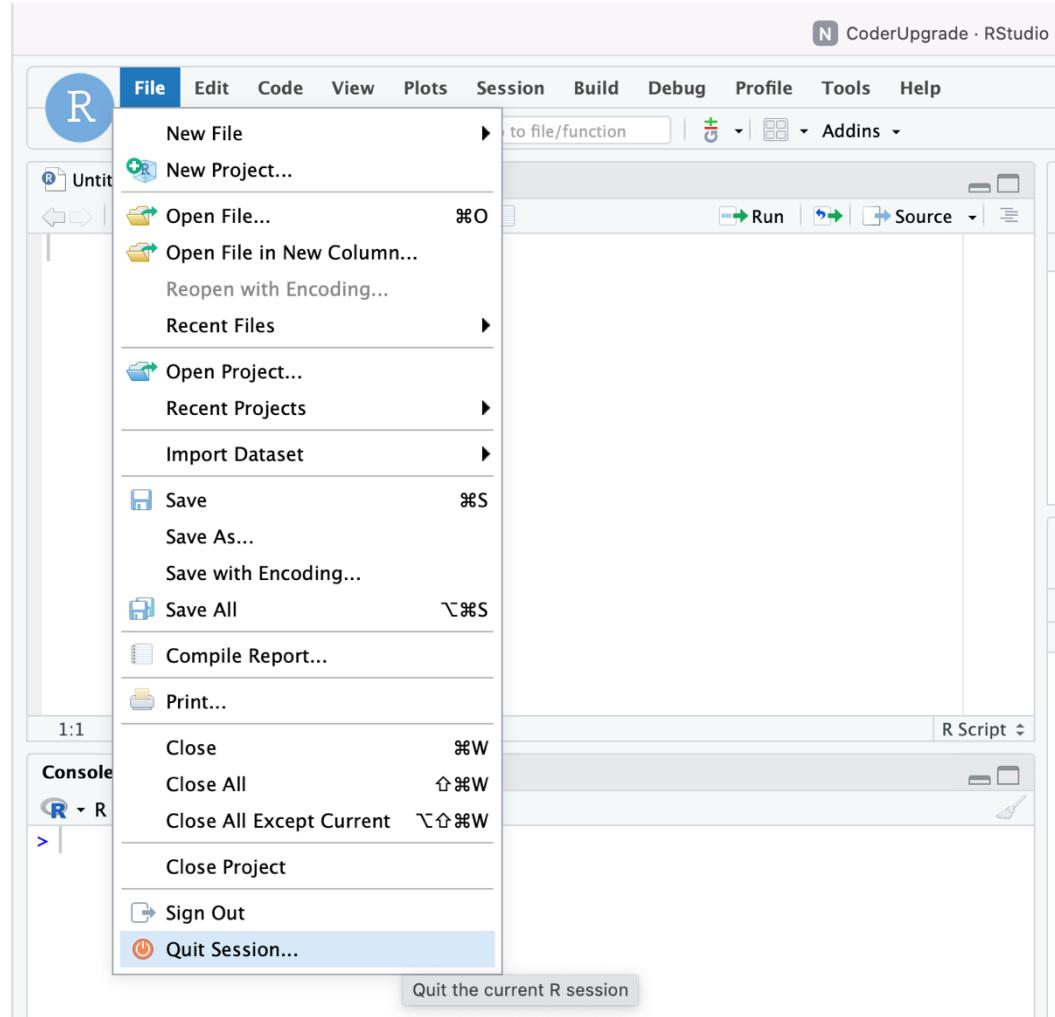


Or you can use the broom icon to remove all objects at once

Two screenshots of the RStudio interface. The top screenshot shows the Environment tab selected. A red box highlights the broom icon in the top right corner of the toolbar. Below it, a button says "Clear objects from the workspace". The bottom screenshot shows the same Environment tab, but the workspace is empty, indicated by the message "Environment is empty". A blue arrow points from the broom icon in the top screenshot down to the empty workspace in the bottom screenshot.

# Clearing environment and logging out

When you are done working in RStudio, log out by quitting the session



# Additional Single Cell Analysis Resources

- CrazyHotTommy GitHub and Blog:

<https://github.com/crazyhottommy/scRNASeq-analysis-notes>

- Nature Reviews Genetics: Best practices for single-cell analysis across modalities: <https://www.nature.com/articles/s41576-023-00586-w>

- Statquest: Clearly explained + visualized statistical concepts relevant to single cell: <https://www.youtube.com/channel/UCtYLUTtgS3k1Fg4y5tAhLbw>

- Further understanding of UMAP with interactive animations:

<https://pair-code.github.io/understanding-umap/>