Basic Single Cell Lab

Jenny Drnevich, PhD

Assistant Director, HPCBio

Roy J. Carver Biotechnology Center

June 25, 2026

Lab assistants: Roberto and Eman

Learning objectives

1. Examine 10x Genomics's web_summary.html report from the output of Cell Ranger
2. Walk through the steps of a standard Seurat analysis
    1. Read explanations of why each step done
    2. Run codes to do each step
    3. Explore how to interact with a Seurat object
3. Try both manual and computational cell type calling
4. Get links for examples of more complex single cell and spatial analyses

Web_summary.html file

This public one is from cellranger multi not cellranger count but all the same information is in there:

Cells tab on left:

1. Any warnings at top?
2. Cell metrics:
   1. **# called** - If >> targeted, high background. If << targeted, poor-quality cells
   2. **mean reads** - If >> targeted, few good cells. If << targeted,
   3. **median genes** - variable between tissue/cell types but ideally > 1000
   4. **total genes** - summed over all genes; 20K+ for mammalian cells.
   5. **median UMIs** - compare with mean reads (PCR dups) and median genes (shows ~UMIs per gene)
   6. **% mapped reads in cells** - if >> 90% can indicate ambient background RNA
3. t-SNE
   1. On left shaded by UMI counts and show strong correlation due to no normalization
   2. right shows first clusters

10k Mouse E18 Combined Cortex, Hippocampus and Subventricular Zone Cells, Chromium GEM-X Single Cell 3' Gene Expression Dataset by Cell Ranger 8.0.0, 10x Genomics, Inc, (2024, Apr 15). Used with permission of 10x Genomics, Inc.

Web_summary.html file

This public one is from cellranger multi not cellranger count but all the same information is in there:

Library tab on left:

1. Cell Statistics (same as Cells tab)
2. Sequencing and Mapping metrics: click on ? for explanations
3. Metrics Per Physical Library
    1. **Sequencing saturation -** "The fraction of reads originating from an already-observed UMI." Indicates whether more sequencing would result in substantially more reads.
4. Plots
    1. **GEX Barcode Rank Plot -** should see sharp "cliff"
    2. **Sequencing Saturation & Median Gene per Cell -** show downsampled values to observe curve and extrapolate what extra sequencing would gain.

10k Mouse E18 Combined Cortex, Hippocampus and Subventricular Zone Cells, Chromium GEM-X Single Cell 3' Gene Expression Dataset by Cell Ranger 8.0.0, 10x Genomics, Inc, (2024, Apr 15). Used with permission of 10x Genomics, Inc.

Log on to Biocluster's Jupyter Hub

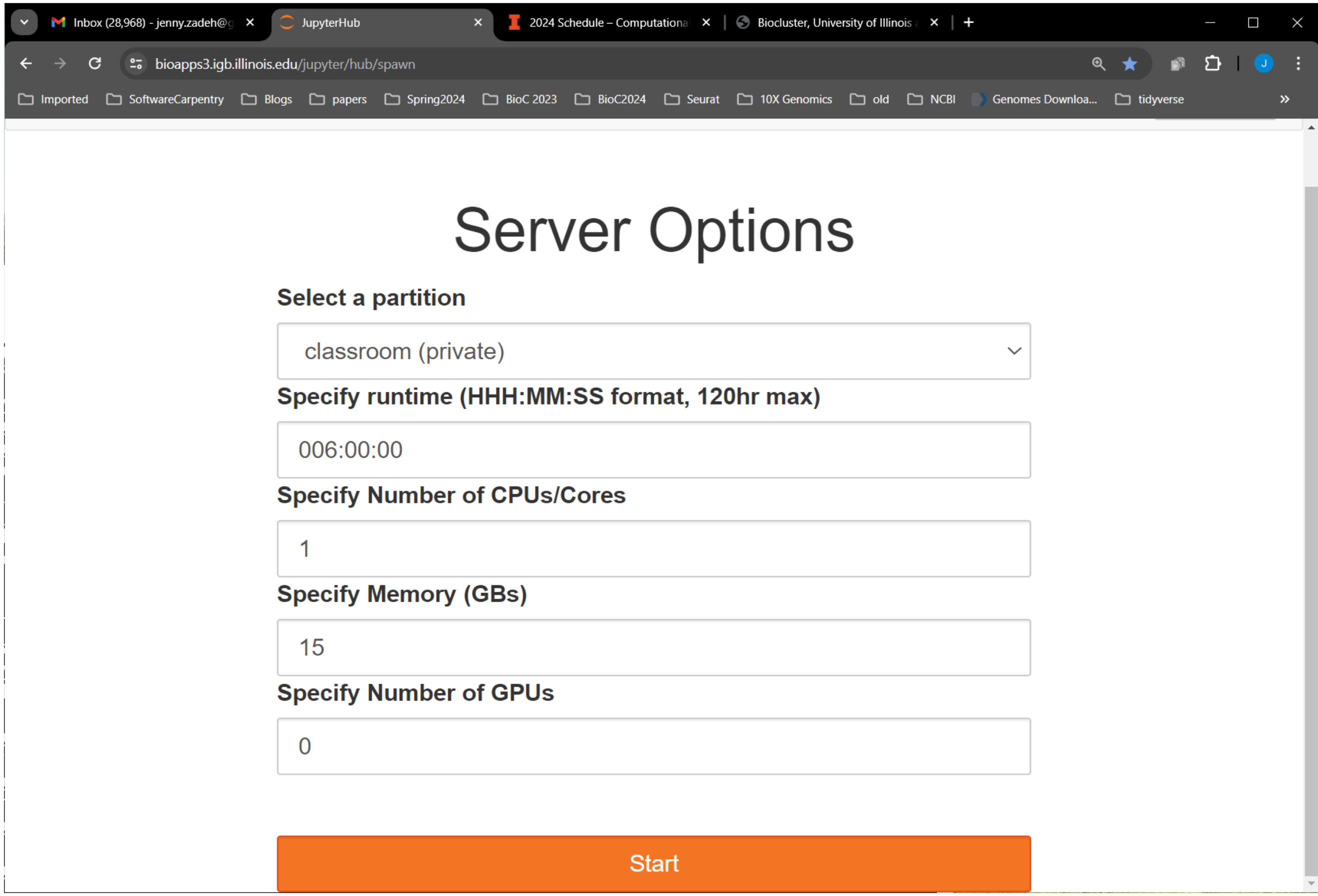
Server options:

Classroom partition

Runtime: **6:00**

Other defaults fine

Click "Start"



Server Options

Select a partition

classroom (private)

Specify runtime (HHH:MM:SS format, 120hr max)

006:00:00

Specify Number of CPUs/Cores

1

Specify Memory (GBs)

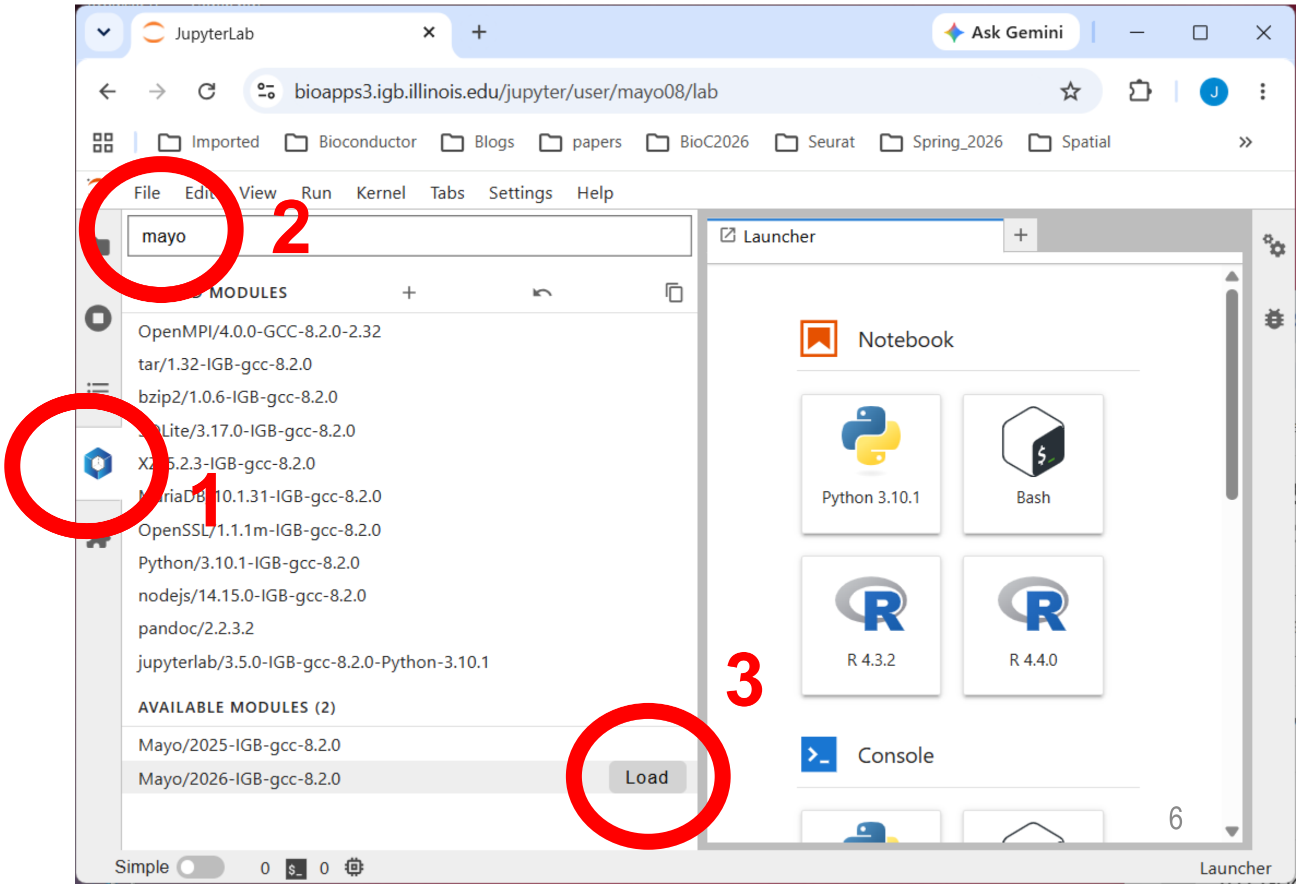
15

Specify Number of GPUs

0

Load Mayo module

1. Click on blue hexagon
2. Search for Mayo at top
3. Click on "Load" by Mayo/2026 module at bottom.



mayo

MODULES

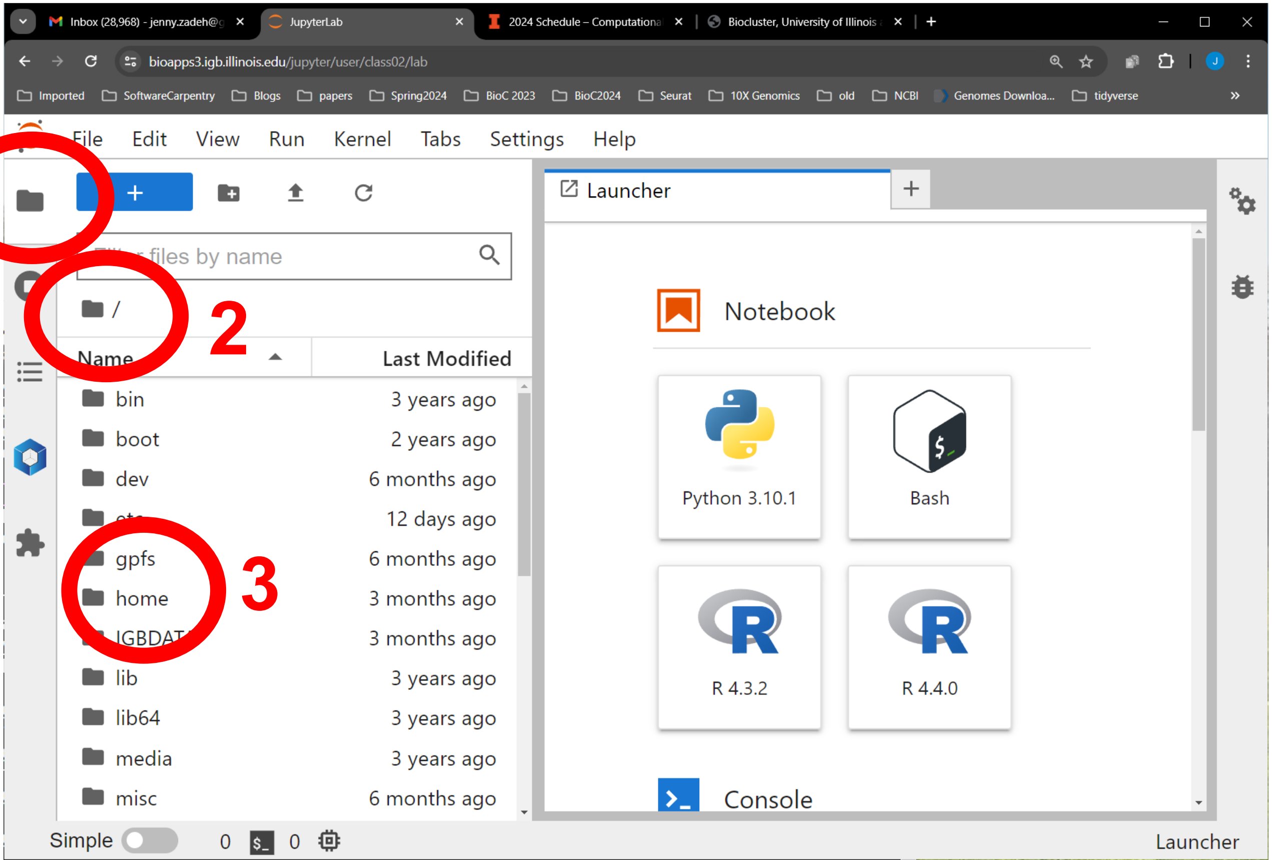
OpenMPI/4.0.0-GCC-8.2.0-2.32
tar/1.32-IGB-gcc-8.2.0
bzip2/1.0.6-IGB-gcc-8.2.0
Lite/3.17.0-IGB-gcc-8.2.0
X1 5.2.3-IGB-gcc-8.2.0
riaDB 10.1.31-IGB-gcc-8.2.0
OpenSSL/1.1.1m-IGB-gcc-8.2.0
Python/3.10.1-IGB-gcc-8.2.0
nodejs/14.15.0-IGB-gcc-8.2.0
pandoc/2.2.3.2
jupyterlab/3.5.0-IGB-gcc-8.2.0-Python-3.10.1

AVAILABLE MODULES (2)

Mayo/2025-IGB-gcc-8.2.0
Mayo/2026-IGB-gcc-8.2.0 Load

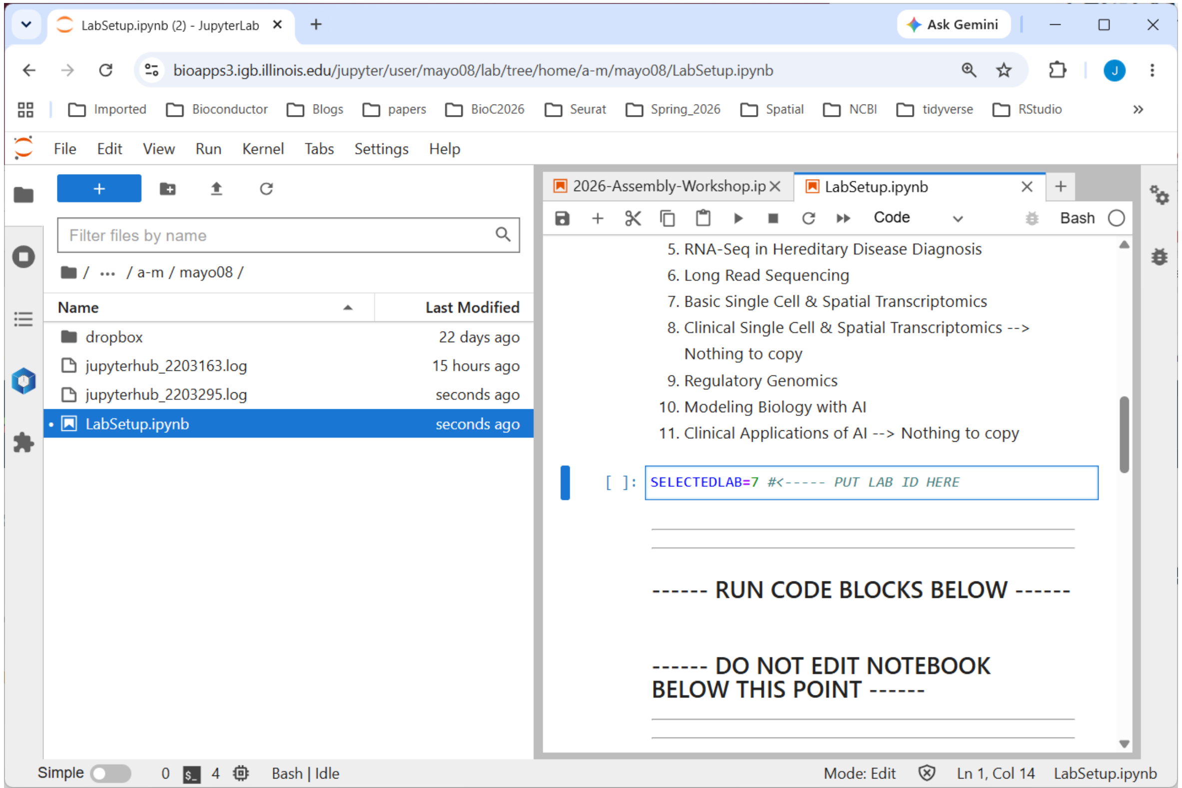
Navigate to proper directory

1. Click on gray folder on very left
2. Click on second gray folder under search box to get to root directory
3. Double click on "home" directory



Continue navigating by double clicking:

- a-m -> mayoXX -> LabSetup.ipynb
- Scroll down and put in **7**
- Run all the code chunks



bioapps3.igb.illinois.edu/jupyter/user/mayo08/lab/tree/home/a-m/mayo08/LabSetup.ipynb

Filter files by name

/ … / a-m / mayo08 /

Name	Last Modified
dropbox	22 days ago
jupyterhub_2203163.log	15 hours ago
jupyterhub_2203295.log	seconds ago
LabSetup.ipynb	seconds ago

5. RNA-Seq in Hereditary Disease Diagnosis
6. Long Read Sequencing
7. Basic Single Cell & Spatial Transcriptomics
8. Clinical Single Cell & Spatial Transcriptomics --> Nothing to copy
9. Regulatory Genomics
10. Modeling Biology with AI
11. Clinical Applications of AI --> Nothing to copy

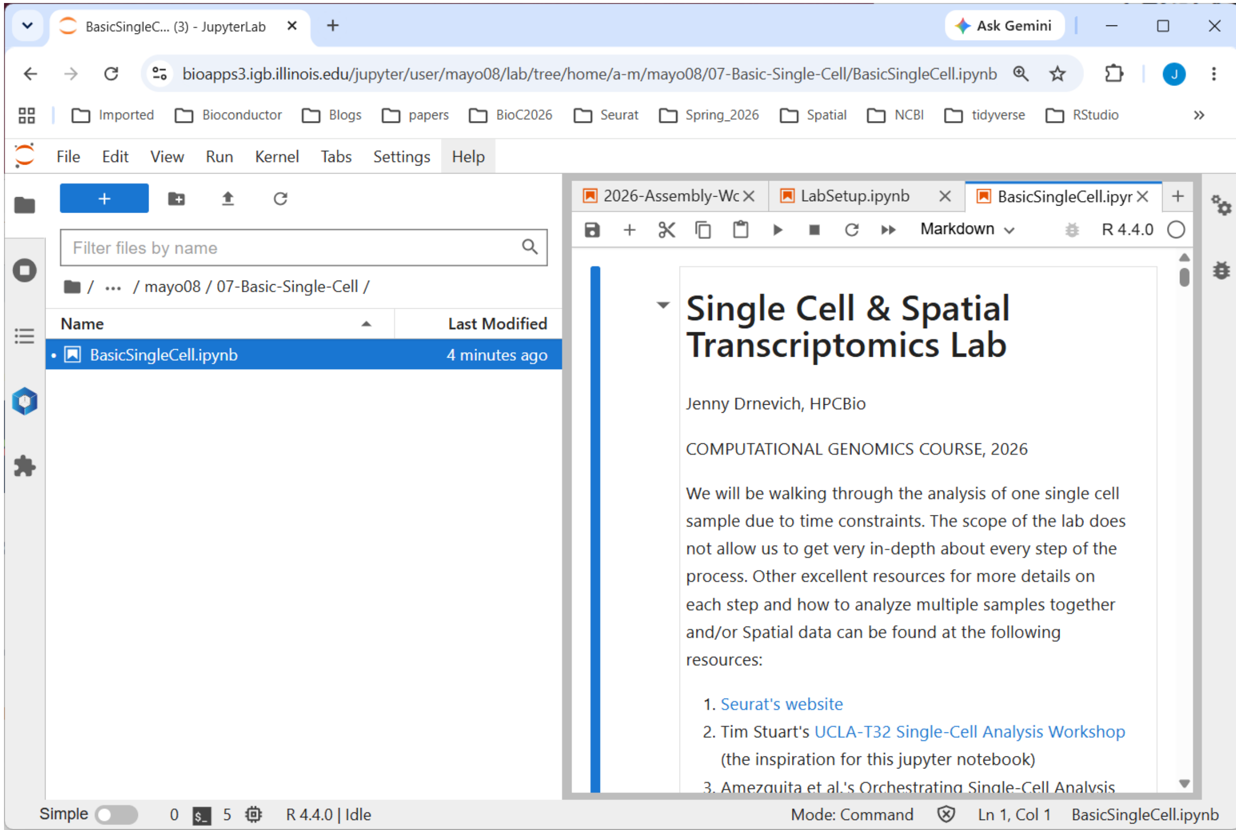
```
[ ]: SELECTEDLAB=7 #<----- PUT LAB ID HERE
```

------ RUN CODE BLOCKS BELOW ------

------ DO NOT EDIT NOTEBOOK BELOW THIS POINT ------

Open 07-Basic-Single-Cell/BasicSingleCell.ipynb

- Double click on the file to open in right pane
- This is a self-paced lab with explanations and runnable code boxes
- If the second code box gives you an error, see next slide



Single Cell & Spatial Transcriptomics Lab

Jenny Drnevich, HPCBio

COMPUTATIONAL GENOMICS COURSE, 2026

We will be walking through the analysis of one single cell sample due to time constraints. The scope of the lab does not allow us to get very in-depth about every step of the process. Other excellent resources for more details on each step and how to analyze multiple samples together and/or Spatial data can be found at the following resources:

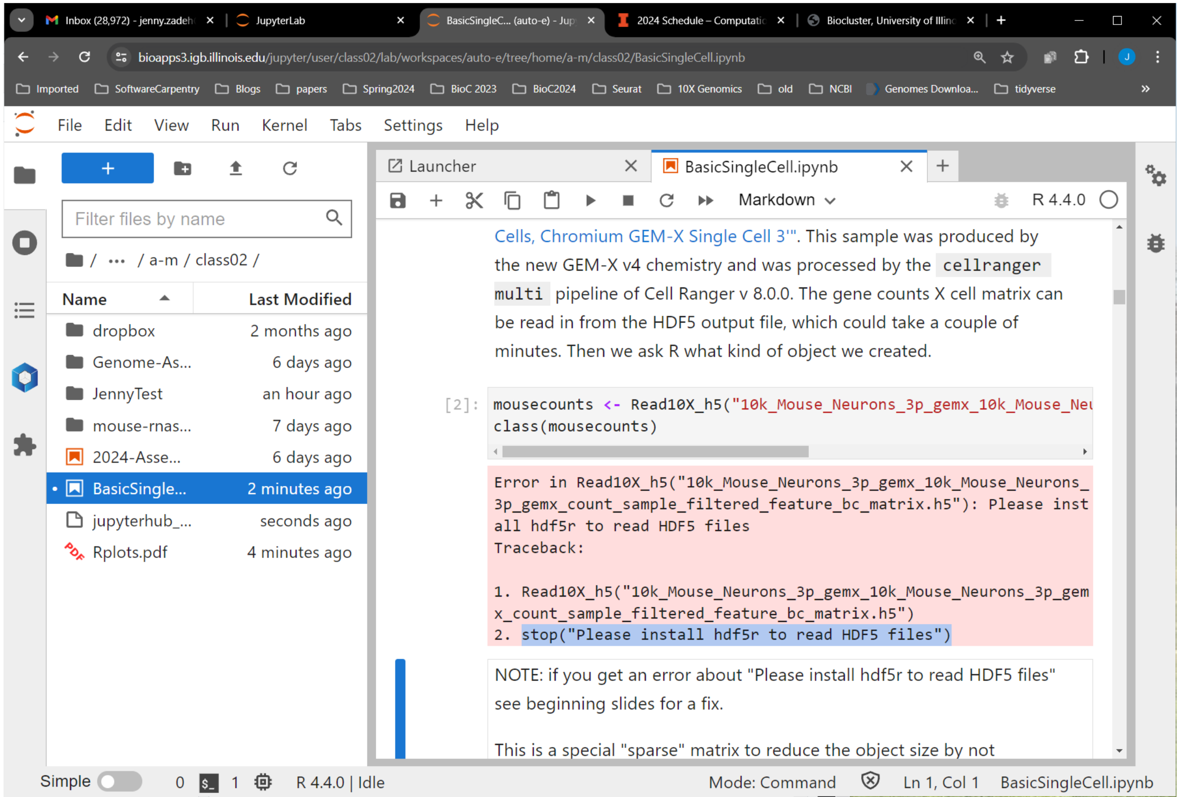
1. Seurat's website
2. Tim Stuart's UCLA-T32 Single-Cell Analysis Workshop (the inspiration for this jupyter notebook)
3. Amezquita et al.'s Orchestrating Single-Cell Analysis

6/24/2026

stop("Please install hdf5r to read HDF5 files")

If you get this error in the second code box then you did not load the Mayo module before opening the .ipynb.

To fix, follow next slides



Name	Last Modified
dropbox	2 months ago
Genome-As...	6 days ago
JennyTest	an hour ago
mouse-rnas...	7 days ago
2024-Asse...	6 days ago
BasicSingle...	2 minutes ago
jupyterhub_...	seconds ago
Rplots.pdf	4 minutes ago

Cells, Chromium GEM-X Single Cell 3'". This sample was produced by the new GEM-X v4 chemistry and was processed by the `cellranger multi` pipeline of Cell Ranger v 8.0.0. The gene counts X cell matrix can be read in from the HDF5 output file, which could take a couple of minutes. Then we ask R what kind of object we created.

```
[2]: mousecounts <- Read10X_h5("10k_Mouse_Neurons_3p_gemx_10k_Mouse_Neu
     class(mousecounts)
```

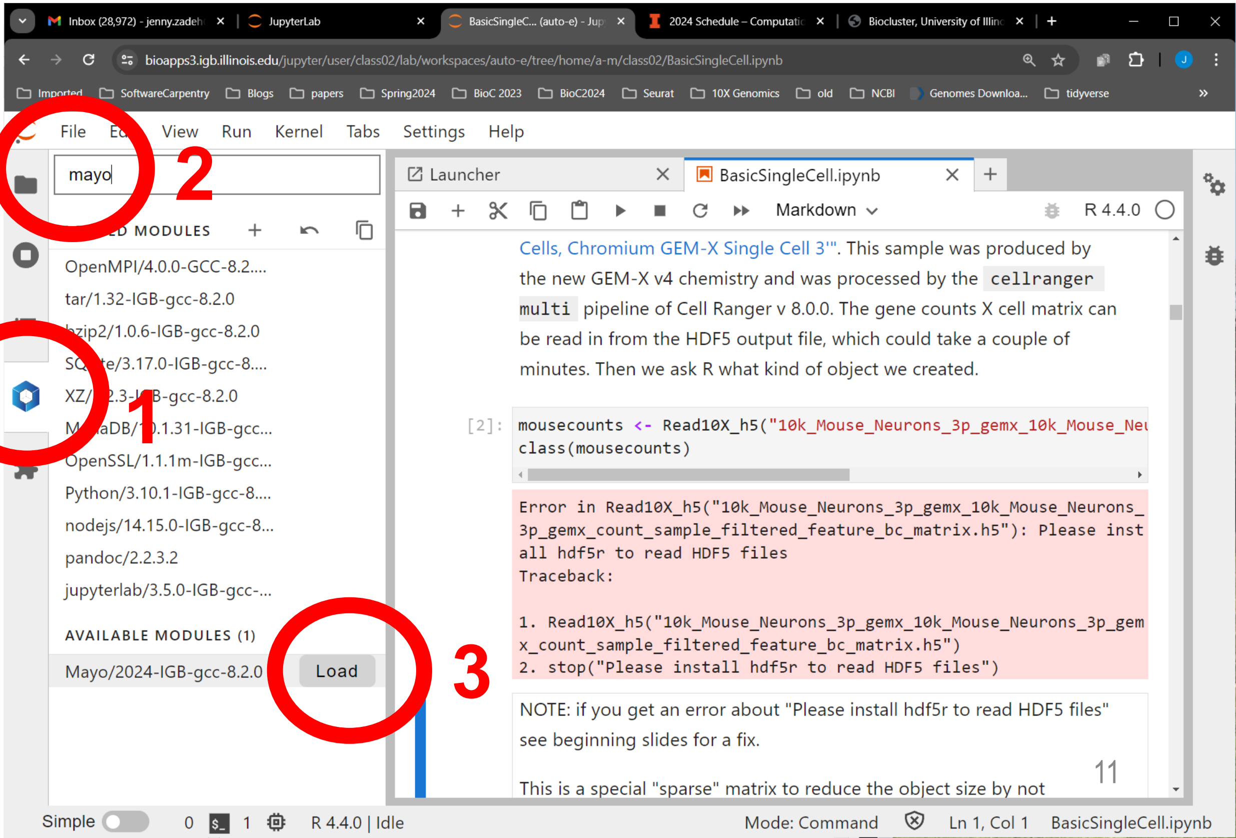
```
Error in Read10X_h5("10k_Mouse_Neurons_3p_gemx_10k_Mouse_Neurons_
3p_gemx_count_sample_filtered_feature_bc_matrix.h5"): Please inst
all hdf5r to read HDF5 files
Traceback:

1. Read10X_h5("10k_Mouse_Neurons_3p_gemx_10k_Mouse_Neurons_3p_gem
x_count_sample_filtered_feature_bc_matrix.h5")
2. stop("Please install hdf5r to read HDF5 files")
```

NOTE: if you get an error about "Please install hdf5r to read HDF5 files" see beginning slides for a fix.

This is a special "sparse" matrix to reduce the object size by not

Load Mayo module

1. Click on blue hexagon
2. Search for Mayo at top
3. Click on "Load" by Mayo module at bottom.



Cells, Chromium GEM-X Single Cell 3'". This sample was produced by the new GEM-X v4 chemistry and was processed by the `cellranger multi` pipeline of Cell Ranger v 8.0.0. The gene counts X cell matrix can be read in from the HDF5 output file, which could take a couple of minutes. Then we ask R what kind of object we created.

```
[2]: mousecounts <- Read10X_h5("10k_Mouse_Neurons_3p_gemx_10k_Mouse_Neu
     class(mousecounts)
```

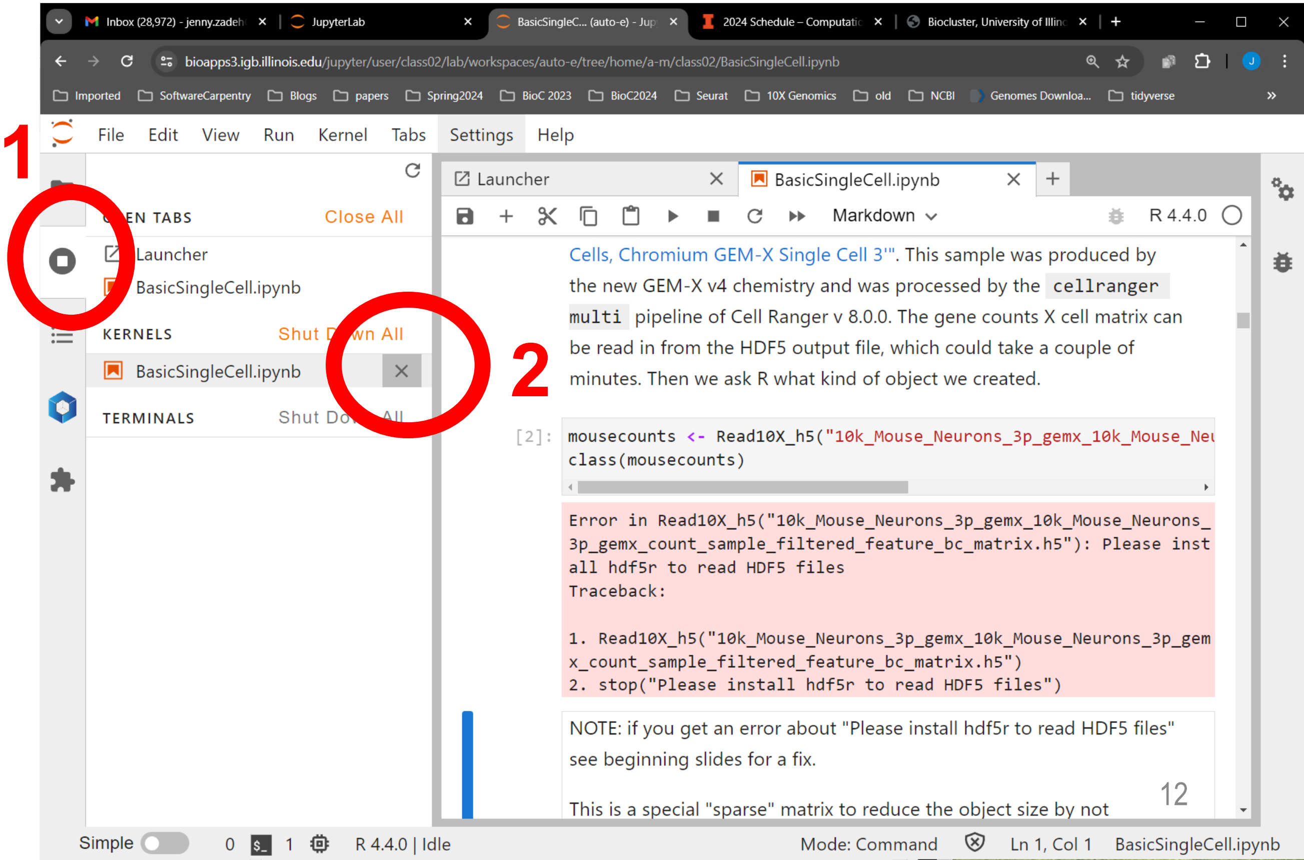
```
Error in Read10X_h5("10k_Mouse_Neurons_3p_gemx_10k_Mouse_Neurons_
3p_gemx_count_sample_filtered_feature_bc_matrix.h5"): Please inst
all hdf5r to read HDF5 files
Traceback:

1. Read10X_h5("10k_Mouse_Neurons_3p_gemx_10k_Mouse_Neurons_3p_gem
x_count_sample_filtered_feature_bc_matrix.h5")
2. stop("Please install hdf5r to read HDF5 files")
```

NOTE: if you get an error about "Please install hdf5r to read HDF5 files" see beginning slides for a fix.

This is a special "sparse" matrix to reduce the object size by not

Stop Kernel

1. Click on gray circle with white square on left
2. Under KERNELS click on the X by BasicSingleCell.ipynb



Cells, Chromium GEM-X Single Cell 3'". This sample was produced by the new GEM-X v4 chemistry and was processed by the `cellranger multi` pipeline of Cell Ranger v 8.0.0. The gene counts X cell matrix can be read in from the HDF5 output file, which could take a couple of minutes. Then we ask R what kind of object we created.

```
[2]: mousecounts <- Read10X_h5("10k_Mouse_Neurons_3p_gemx_10k_Mouse_Neu
     class(mousecounts)
```

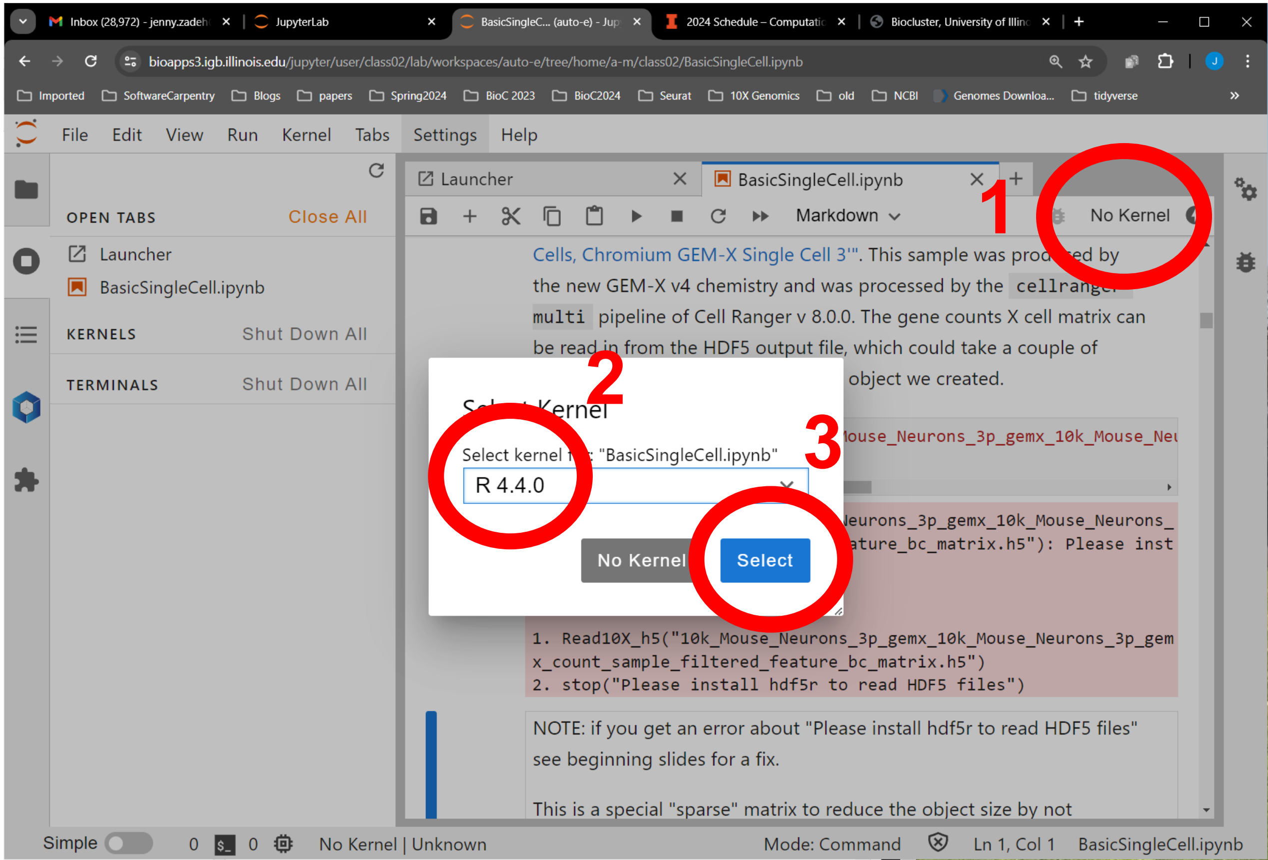
```
Error in Read10X_h5("10k_Mouse_Neurons_3p_gemx_10k_Mouse_Neurons_
3p_gemx_count_sample_filtered_feature_bc_matrix.h5"): Please inst
all hdf5r to read HDF5 files
Traceback:

1. Read10X_h5("10k_Mouse_Neurons_3p_gemx_10k_Mouse_Neurons_3p_gem
x_count_sample_filtered_feature_bc_matrix.h5")
2. stop("Please install hdf5r to read HDF5 files")
```

NOTE: if you get an error about "Please install hdf5r to read HDF5 files" see beginning slides for a fix.

This is a special "sparse" matrix to reduce the object size by not

Re-start Kernel

1. Click on the "No Kernel" in the top right of the .ipynb to get the pop-up
2. Pick R 4.4.0 from the drop down menu
3. Click on "Select"



Cells, Chromium GEM-X Single Cell 3'". This sample was prod ced by the new GEM-X v4 chemistry and was processed by the `cellrange` `multi` pipeline of Cell Ranger v 8.0.0. The gene counts X cell matrix can be read in from the HDF5 output file, which could take a couple of … object we created.

Select Kernel

Select kernel for: "BasicSingleCell.ipynb"

R 4.4.0

No Kernel Select

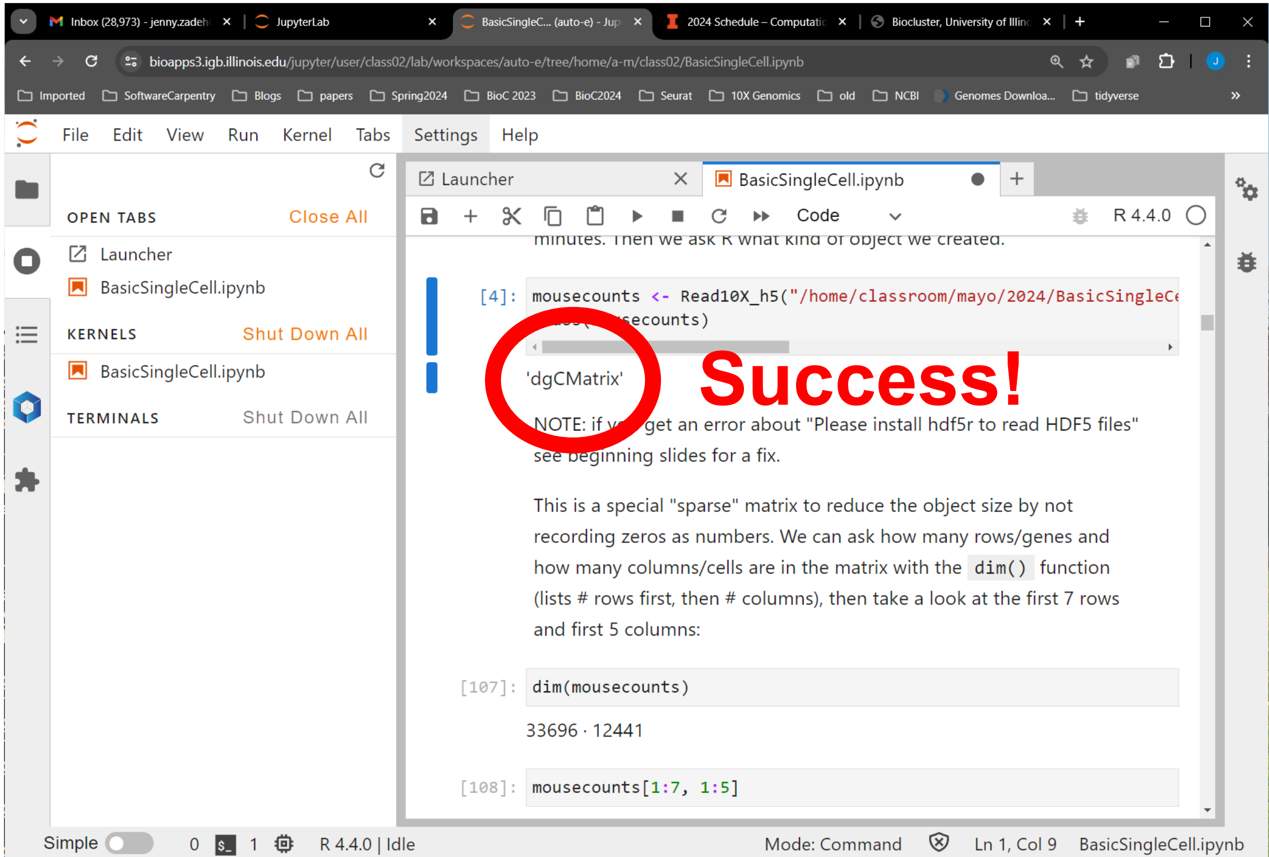
```
louse_Neurons_3p_gemx_10k_Mouse_Neu
leurons_3p_gemx_10k_Mouse_Neurons_
ature_bc_matrix.h5"): Please inst

1. Read10X_h5("10k_Mouse_Neurons_3p_gemx_10k_Mouse_Neurons_3p_gem
x_count_sample_filtered_feature_bc_matrix.h5")
2. stop("Please install hdf5r to read HDF5 files")
```

NOTE: if you get an error about "Please install hdf5r to read HDF5 files" see beginning slides for a fix.

This is a special "sparse" matrix to reduce the object size by not

Re-run the code boxes

You need to re-run the **first code box** to load the packages from the library, then you can successfully run the second code box



minutes. Then we ask R what kind of object we created.

```
[4]: mousecounts <- Read10X_h5("/home/classroom/mayo/2024/BasicSingleCe
        secounts)
```

```
'dgCMatrix'
```

NOTE: if y get an error about "Please install hdf5r to read HDF5 files" see beginning slides for a fix.

This is a special "sparse" matrix to reduce the object size by not recording zeros as numbers. We can ask how many rows/genes and how many columns/cells are in the matrix with the `dim()` function (lists # rows first, then # columns), then take a look at the first 7 rows and first 5 columns:

```
[107]: dim(mousecounts)
```

```
33696 · 12441
```

```
[108]: mousecounts[1:7, 1:5]
```

Review step in downstream analysis

- First check the web_summary.html file
- Read in UMI count data and put into a Seurat object
- QC filtering of cells (high MT, too low or too high UMIs) and genes (low expression)
- Pipeline: normalization, PCA, cluster calling, UMAP
- Cluster marker gene detection
- Cell type annotation: manual vs. computational