Introduction to Exercises

Susanne Reinsbach
susanne.reinsbach@scilifelab.se

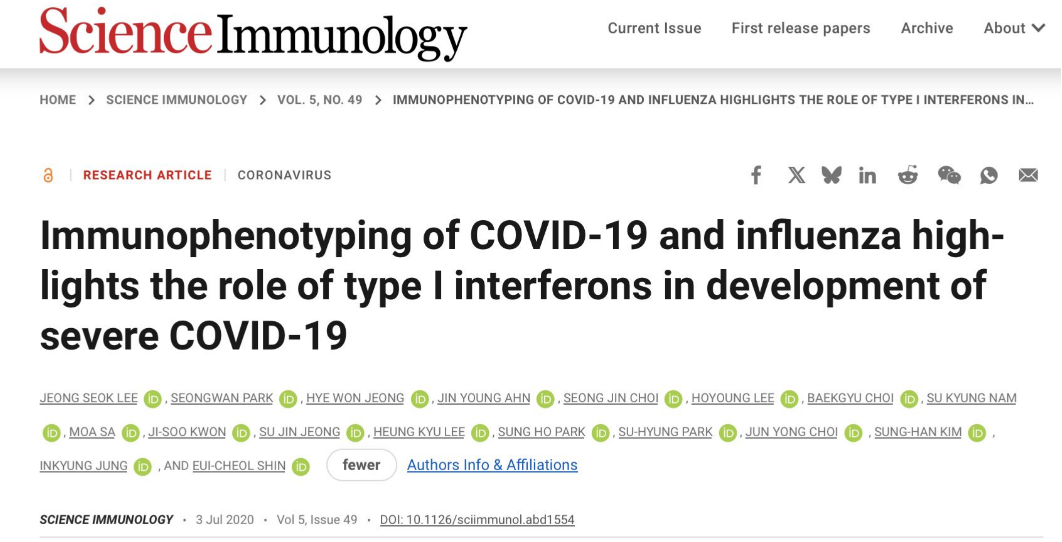
31.03.2025

NATIONAL BIOINFORMATICS
INFRASTRUCTURE SWEDEN

Outline

- Data set
- Analysis steps
- Toolkits and data structures
- Labs setup
- Practical information
- Demo

Data set



HOME > SCIENCE IMMUNOLOGY > VOL. 5, NO. 49 > IMMUNOPHENOTYPING OF COVID-19 AND INFLUENZA HIGHLIGHTS THE ROLE OF TYPE I INTERFERONS IN…

RESEARCH ARTICLE | CORONAVIRUS

Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19

JEONG SEOK LEE, SEONGWAN PARK, HYE WON JEONG, JIN YOUNG AHN, SEONG JIN CHOI, HOYOUNG LEE, BAEKGYU CHOI, SU KYUNG NAM, MOA SA, JI-SOO KWON, SU JIN JEONG, HEUNG KYU LEE, SUNG HO PARK, SU-HYUNG PARK, JUN YONG CHOI, SUNG-HAN KIM, INKYUNG JUNG, AND EUI-CHEOL SHIN

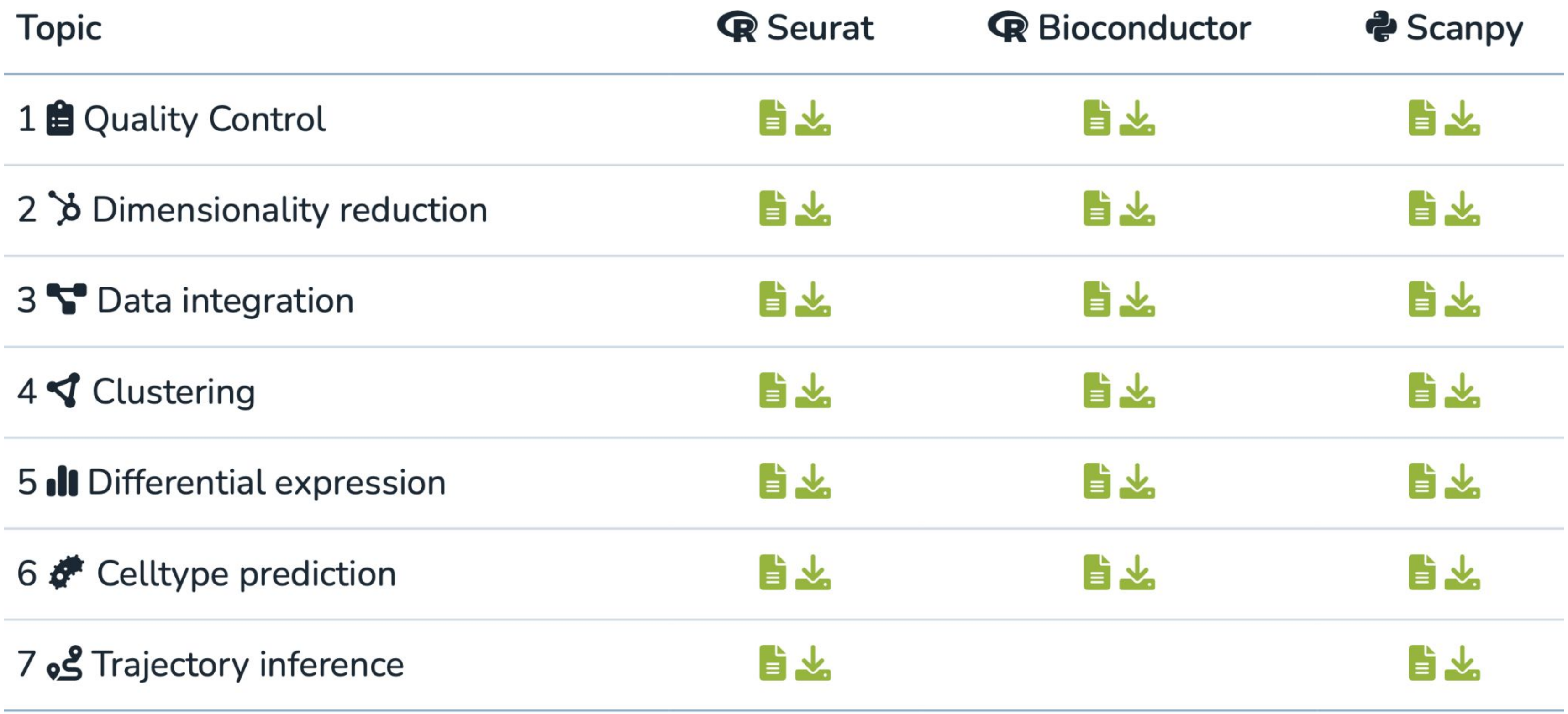
SCIENCE IMMUNOLOGY • 3 Jul 2020 • Vol 5, Issue 49 • DOI: 10.1126/sciimmunol.abd1554

- Selected 4 controls and 4 severe COVID-19 cases
- Cells are subsampled to 1500 cells per subject
- Aim: Which cell types and genes are altered when comparing blood immune cells from healthy vs disease?
- Note: Trajectory and spatial transcriptomics lab have different data sets

https://nbisweden.github.io/workshop-scRNAseq/other/data.html

https://www.science.org/doi/10.1126/sciimmunol.abd1554

Overview analysis steps



Topic	Seurat	Bioconductor	Scanpy
1 Quality Control	📄 ⬇	📄 ⬇	📄 ⬇
2 Dimensionality reduction	📄 ⬇	📄 ⬇	📄 ⬇
3 Data integration	📄 ⬇	📄 ⬇	📄 ⬇
4 Clustering	📄 ⬇	📄 ⬇	📄 ⬇
5 Differential expression	📄 ⬇	📄 ⬇	📄 ⬇
6 Celltype prediction	📄 ⬇	📄 ⬇	📄 ⬇
7 Trajectory inference	📄 ⬇		📄 ⬇

https://nbisweden.github.io/workshop-scRNAseq/home_contents.html

Three main toolkits

	Toolkits		
	Seurat v5	Bioconductor	scanpy
Language	R	R	Python
Data structure	`SeuratObject`	`SingleCellExperiment Object`	`AnnData`
Key features	+ Actively maintained (by Satija Lab) + Open source + Single R package + Specifically designed for single cell analysis (spatial and multimodal analysis)	+ Community-developed + Open source + Interoperability across single-cell related packages - Scalability	+ Community-developed (part of `scverse` project) + Open source + Interoperability with Python ML libraries + Scalable

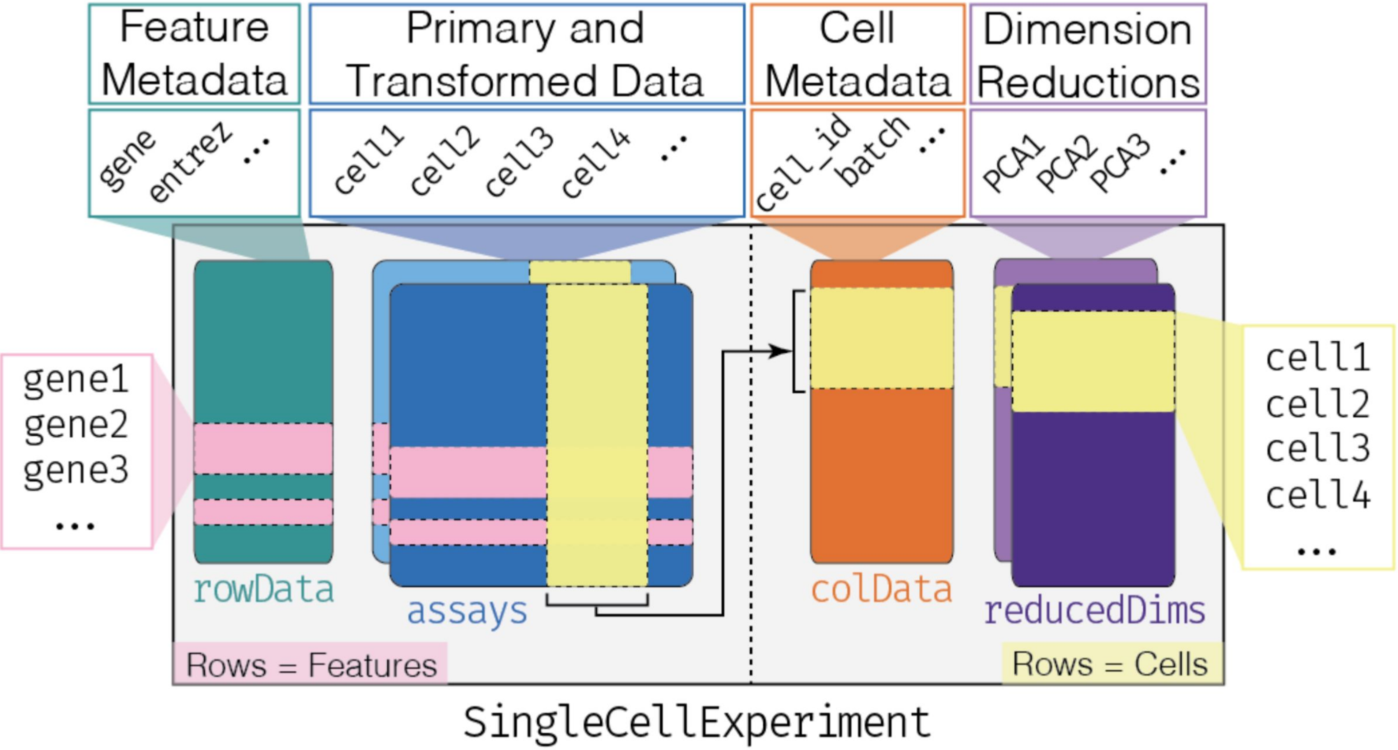
SeuratObject

```
▾ alldata
  ..@ assays: List of assays for count data
  .. ..$ RNA: Formal class 'Assay5'
  .. .. ..@ layers: List of data matrices (features x barcodes)
  .. .. .. .. ..$ counts: raw counts
  .. .. .. .. ..$ data: normalized counts
  .. .. .. .. ..$ scale: scaled counts
  ..@ meta.data: data.frame containing metadata for barcodes/cells
  ..@ active.assay: "RNA"
  ..@ reductions: List of dimensional reduction objects
  .. ..$ pca: (components x barcodes)
```

```
> slotNames(alldata)
[1]   "assays”     "meta.data”    "active.assay”    "active.ident"    "graphs"     "neighbors"
[7]   "reductions”    "images"     "project.name"    "misc”     "version"     "commands"
[13]  "tools"
```

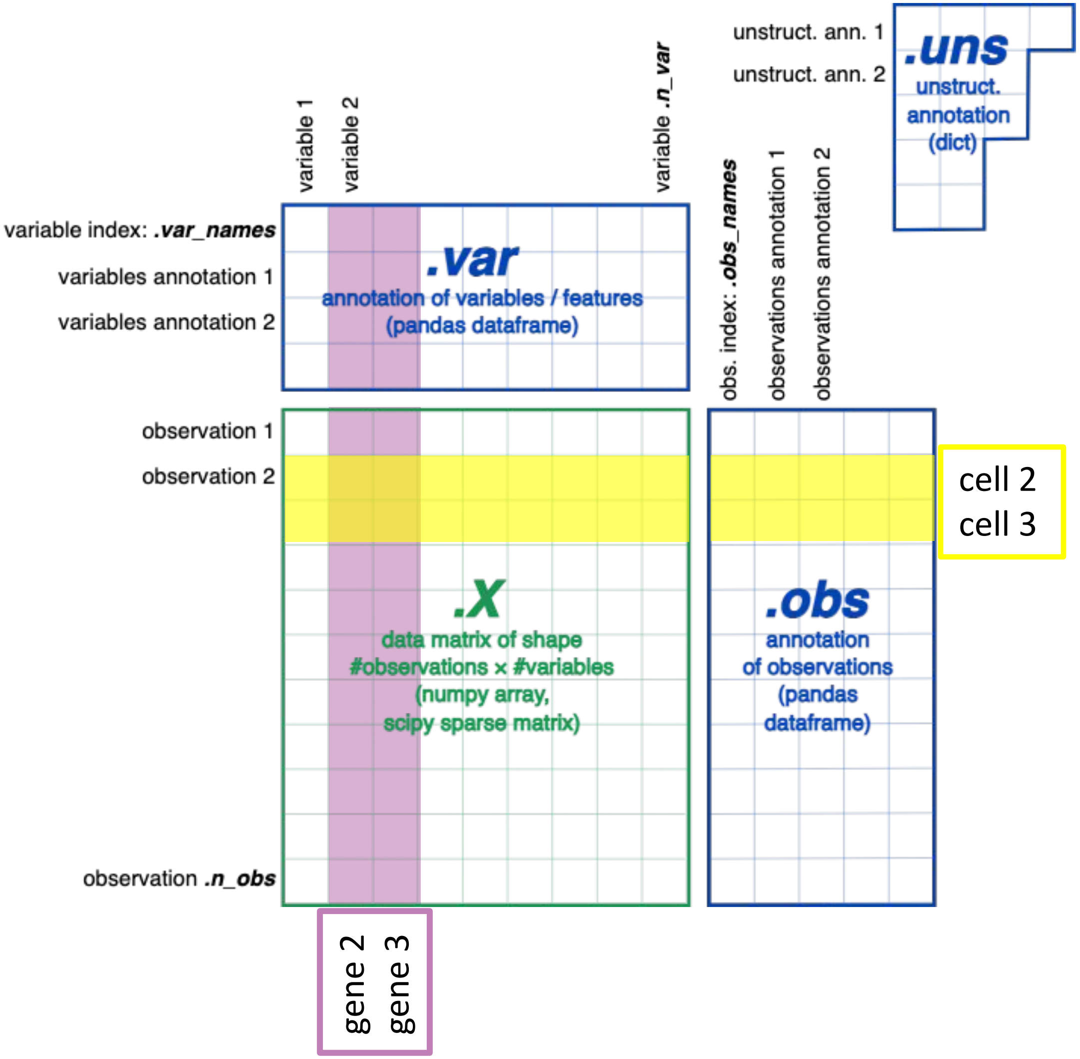
https://satijalab.org/seurat/articles/essential_commands.html

SingleCellExperiment Object



https://bioconductor.org/books/3.13/OSCA.intro/the-singlecellexperiment-class.html

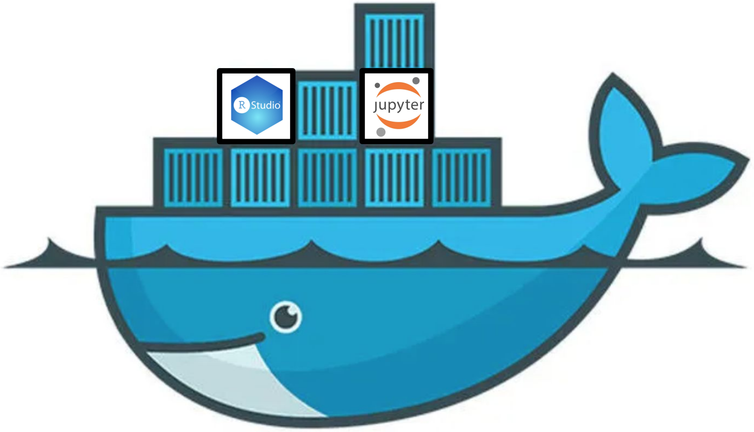
AnnData Object



https://scanpy.readthedocs.io/en/stable/usage-principles.html

NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

Containerized environments



- For each toolkit a Docker image is created
- A Docker **container** is a running instance of a Docker **image**
- Docker containers are isolated environments with all the dependencies required for the analysis
- Choose a toolkit and run the Docker containers on **SciLifeLab Serve** or **locally** and access them through your browser

https://nbisweden.github.io/workshop-scRNAseq/other/containers.html

NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

NBIS course

NBIS / Elixir workshop

Tools for Reproducible Research

How to work reproducibly with control and structuring of project code, environment and workflow management

Updated: 03-03-2025 at 12:38:14 .

https://nbisweden.github.io/workshop-reproducible-research

Running the labs

- The code for the exercises needs to be downloaded from the terminal in Rstudio or JupyterLab
- The code is available as Quarto notebooks (`.qmd`) for Seurat and Bioconductor and Jupyter notebooks (`.ipynb`) for Scanpy
- Please report errors in the code
  - Slack channel #**exercises**
  - Open an issue in the GitHub repository
  - The corrected code needs to be downloaded again using the download script

Getting help

- Ask for help in the Slack channel #**exercises** and a TA will help answering your question
- **Before** asking, please try to do some troubleshooting
  - Read the error messages carefully, most of the times the answer can be found there
  - Ask Google (or ChatGPT)
  - Check forums like SEQanswers, Stack Exchange, Bioconductor Forum or GitHub issues for the relevant package

NATIONAL BIOINFORMATICS
INFRASTRUCTURE SWEDEN

Practical information

- For the exercises and the discussion points you are encouraged to talk with other course participants in your breakout room
- A summary of each lab will be given in the last 10 minutes of each exercise
- If you have time, experiment with different options in the algorithms

Demo