Welcome to the December 2025 Advanced Single-cell RNA-seq Training Workshop!

December 8-12, 2025
Childhood Cancer Data Lab
https://alexslemonade.github.io/2025-december-training/

Childhood Cancer Data Lab x Alex's Lemonade Stand

Tell us about you!

- What’s your name?
- Where do you work?
- What are you studying?
- What is a recent piece of media (book, tv, movie, live performance…) you enjoyed?

Meet your instructors

JOSH

Joshua Shapiro

Data Scientist @ the Data Lab

PhD Ecology & Evolution, *UChicago*
Postdoc Integrative Genomics, *Princeton*

Research interests:

- **Evolutionary genomics**
- **Single cell workflows**

jashapiro

Meet your instructors

STEPHANIE

Stephanie Spielman

Data Scientist @ the Data Lab

PhD Integrative Biology *UT Austin*
Postdoc Computational Molecular Evolution *Temple*

Research interests:

- **Protein evolution and comparative sequence analysis**
- **Data science and bioinformatics education**

sjspielman

Meet your instructors

ALLY

Ally Hawkins

Data Scientist @ the Data Lab

PhD Cancer Biology *University of Michigan*
Postdoc Computational Biology *Cornell*

Research interests:

- **Single cell data analysis**
- **Origins of pediatric solid tumors**

allyhawkins

Other staff you may see (or have seen already!)

JEN

Jen O'Malley

she/her/hers

Scientific Community Manager

- Helps administer Data Lab programs such as workshops
- Engages with the research community

DEEPA

Deepa Prasad

she/her/hers

User Experience Designer

- Talks to researchers about their needs and frustrations
- Designs usable software

We're conducting a study to understand how you use exploratory visualizations to make decisions while working with single-cell transcriptomic data.

We'll be reaching out to researchers who:

- **Have used the Single-cell Pediatric Cancer Atlas (ScPCA) Portal**
- **Are new to using R to visualize data**

Deepa may contact you with more information!

Tell us about you!

- What’s your name?
- Where do you work?
- What are you studying?
- What is a recent piece of media (book, tv, movie, live performance…) you enjoyed?

Code of Conduct

Be kind, have fun

We value the involvement of everyone in the community. We are committed to creating a friendly and respectful place for learning, teaching, and contributing.

- Use welcoming and inclusive language
- Be respectful of different viewpoints and experiences
- Gracefully accept constructive criticism
- Focus on what is best for the community
- Show courtesy and respect towards other community members

Read the full Code of Conduct here:
https://alexslemonade.github.io/2025-december-training/code-of-conduct.html

If you at any time feel harassed or treated inappropriately, please contact ccdl@alexslemonade.org.

What you will learn (and what you won’t)

What you will learn

- We will review working with single-cell RNA-seq data in R with Bioconductor.
- We will introduce working with CITE-seq data.
- We will address integration of multiple single-cell RNA-seq libraries.
- We will describe approaches to differential expression analyses with single-cell data.
- We will cover pathway analysis using GSEA and AUCell.

What you won’t learn

- We won’t address experimental design (e.g., how many replicates you need).
- We won’t compare tools (e.g., Seurat vs. Scater for single-cell RNA-seq data pre-processing).
- We won't cover pre-processing (alignment and quantification) of single-cell data.
- We won't address integrating with other omics types (e.g. joint analysis of single-cell RNA-seq and ATAC-seq).
- We won’t cover every feature or assumption of the tools we do present, but we will try to highlight the features and gotchas that we think are relevant to most users.
- You may not be able to perform every analysis you need for your own work, particularly for complex experimental designs.

How do we pick what we teach?

We want methods to be or to have:

- Useful for a wide range of experimental designs, sample sizes
- Easy to use, well-documented, and consistently updated
- Solid tutorials, a sizeable user base, and responsive authors/maintainers

We have a preference for methods that integrate easily into a single workflow that can be run on a laptop (and our own personal biases as scientists).

Schedule

Monday

Workshop Intro

Module
-Reading, filtering, and normalizing scRNA-seq data

Consultations
-Exercise notebooks

Tuesday

Module
-Integrating scRNA-seq datasets

Consultations
-Exercise notebooks

Wednesday

Module
Differential expression analysis for scRNA-seq

Consultations
-Exercise notebooks
-Your own data

Thursday

Module
Pathway and gene-set analysis

Consultations
-Exercise notebooks
-Your own data

Friday

Consultations
-Exercise notebooks
-Your own data

Presentations

Full schedule: https://alexslemonade.github.io/2025-december-training/workshop/SCHEDULE.html

Daily Schedule Outline

Instruction

Full group

Lectures

- Introduce concepts and background
- Demonstrate usage
- Answer general questions

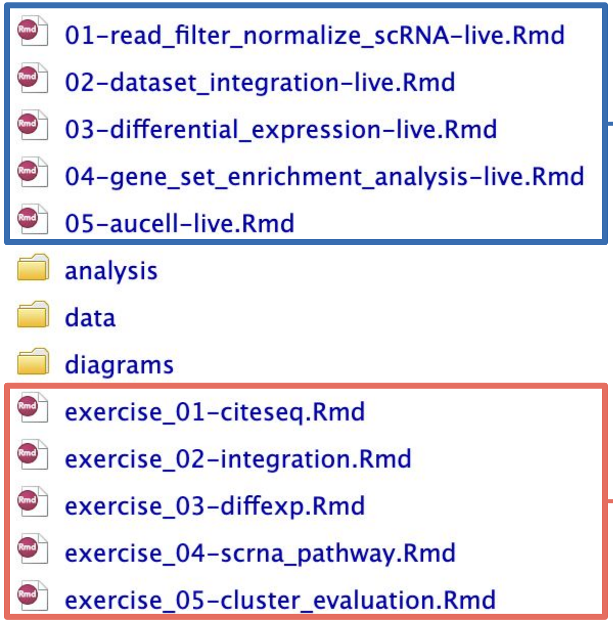
Consultation Period

Exercise notebooks

Your own data

- Ask questions of instructors and other participants
- Practice what you have learned
- Work on exercises individually or in groups
- Work with your own data

Module Layout

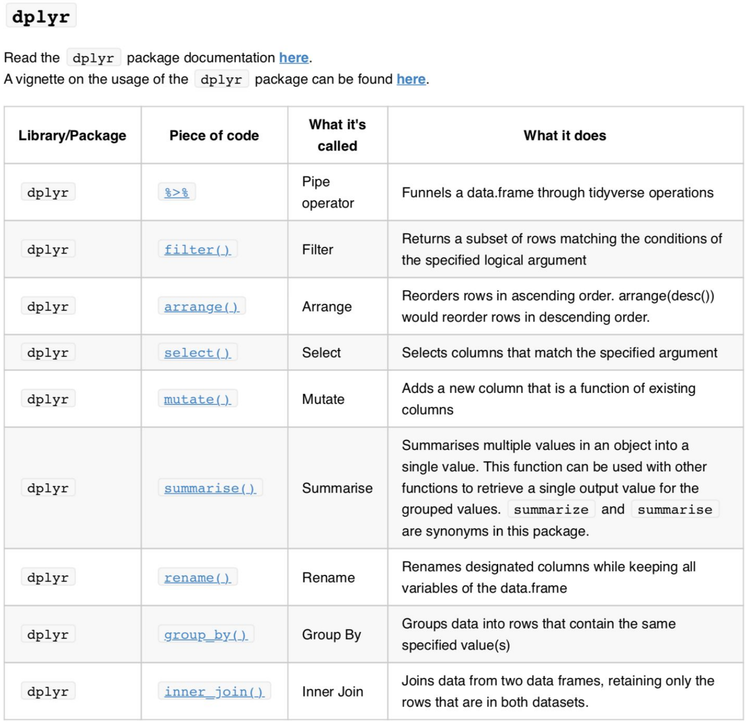


01-read_filter_normalize_scRNA-live.Rmd
02-dataset_integration-live.Rmd
03-differential_expression-live.Rmd
04-gene_set_enrichment_analysis-live.Rmd
05-aucell-live.Rmd
analysis
data
diagrams
exercise_01-citeseq.Rmd
exercise_02-integration.Rmd
exercise_03-diffexp.Rmd
exercise_04-scrna_pathway.Rmd
exercise_05-cluster_evaluation.Rmd

These are **Instruction** notebooks.
We'll walk through these together, adding code during the workshop.

These are **Exercise** notebooks.
Use these to practice what you have learned. We're here to help!

Module cheatsheets cover key functions



`dplyr`

Read the `dplyr` package documentation **here**.
A vignette on the usage of the `dplyr` package can be found **here**.

Library/Package	Piece of code	What it's called	What it does
`dplyr`	`%>%`	Pipe operator	Funnels a data.frame through tidyverse operations
`dplyr`	`filter()`	Filter	Returns a subset of rows matching the conditions of the specified logical argument
`dplyr`	`arrange()`	Arrange	Reorders rows in ascending order. arrange(desc()) would reorder rows in descending order.
`dplyr`	`select()`	Select	Selects columns that match the specified argument
`dplyr`	`mutate()`	Mutate	Adds a new column that is a function of existing columns
`dplyr`	`summarise()`	Summarise	Summarises multiple values in an object into a single value. This function can be used with other functions to retrieve a single output value for the grouped values. `summarize` and `summarise` are synonyms in this package.
`dplyr`	`rename()`	Rename	Renames designated columns while keeping all variables of the data.frame
`dplyr`	`group_by()`	Group By	Groups data into rows that contain the same specified value(s)
`dplyr`	`inner_join()`	Inner Join	Joins data from two data frames, retaining only the rows that are in both datasets.

Friday Afternoon

Your own projects
Exercise notebooks

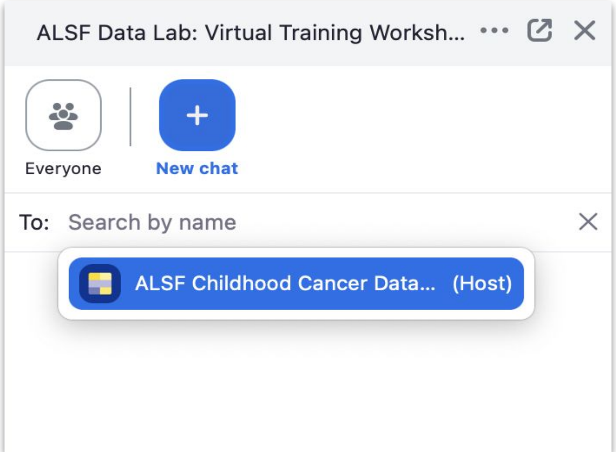
Spend Friday afternoon working with your own data, getting assistance as needed from Data Lab staff and each other.

Presentations

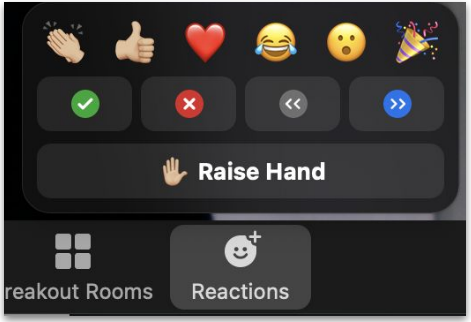
Present what you worked on during the consultation times to the group!

Virtual Training Procedures

General Zoom Etiquette

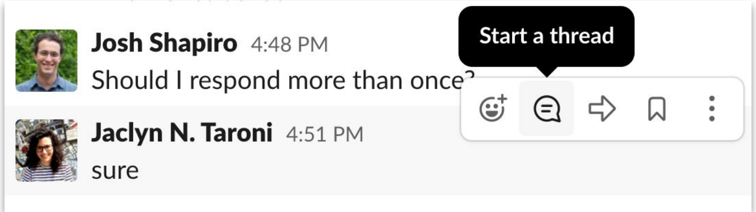


- Keep your microphone muted
- Type questions in the Chat window (directed to the host!)
    - Click on the “Chat” button at the bottom of your window to open the chat.
    - The ability to message everyone in the “Main Chat” will be unavailable.
    - Instead, click “New Chat” and send a direct message to “ALSF Childhood Cancer Data Lab.”
- We encourage you to use zoom reactions!



Using Slack

- Use the **#2025-december-training** Slack channel
- Post public questions, get help with errors and debugging, make comments, and help your fellow participants!
  - Use threads to keep related content together



Josh Shapiro 4:48 PM
Should I respond more than once?

Jaclyn N. Taroni 4:51 PM
sure

Communication during instruction

zoom

- **I have an urgent question** that needs an answer before moving on:
  - ***Raise Hand*** or ***Chat*** with the meeting host

- **I’m stuck with an error** and can’t proceed with the hands-on exercise
  - ***Chat*** with meeting host: Request 1:1 and you will be placed in a breakout room with a Data Lab staff member

slack

- **I have an general question** that does not need an answer right away.
  - ***Post*** in *#2025-december-training*

- **I’m having trouble logging in** to RStudio Server
  - ***Direct Message*** a Data Lab staff member (not the current host or instructor)

Trouble logging into Zoom and Slack? ***Email*** training@ccdatalab.org

Communication at other times (consultation periods)

slack

- I have questions about **previous instruction** or **exercise notebooks**
  - ***Post*** in *#2025-december-training*
  - If you need to share your screen, we will set up a 1:1 or group Zoom call
- **I would like to be paired up with other participants**
  - ***Post*** in *#2025-december-training*; we can set you up in a Zoom breakout room
- I have a question that is **highly specific to my data**
  - ***Direct Message*** a Data Lab staff member
- **I'm having trouble logging in** to RStudio Server
  - ***Direct Message*** a Data Lab staff member

Trouble logging into Zoom and Slack? ***Email*** training@ccdatalab.org