

Welcome to the January 2023 Advanced Single-cell RNA-Seq Training Workshop!

January 31 - February 2, 2023

Childhood Cancer Data Lab

<https://alexslemonade.github.io/2023-january-training/>

Childhood Cancer
Data  **Lab**

x



Tell us about you!

- What's your name?
- What are you studying?
- Do you have any pets? If not, what kind of pet would you choose to have?



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 phylogenetic modeling**
- **Data science and bioinformatics education**



sjspielman

Meet your instructors



CHANTE

Chante Bethell

Biological Data Analyst @ the Data Lab

Bachelor's in Bioinformatics from *Rowan University*

Research interests:

- **Functional motifs in the proteome**



cbethell

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

Meet your instructors



JACLYN

Jaclyn Taroni

Director @ the Data Lab

PhD Genetics *Dartmouth*

Postdoc Computational Biology *UPenn*

Research interests:

- **Transcriptomics in rare, complex diseases**
- **Unsupervised pattern extraction**



jaclyn-taroni

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



JEN
Jen O'Malley
Scientific Community
Manager

- Help administer Data Lab offerings such as training workshops
- Manages communications



DEEPA
Deepa Prasad
User Experience
Designer

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



DAVID
David Mejia
Developer

- Makes websites
- Debugs

Tell us about you!

- What's your name?
- What are you studying?
- Do you have any pets? If not, what kind of pet would you choose to have?





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.

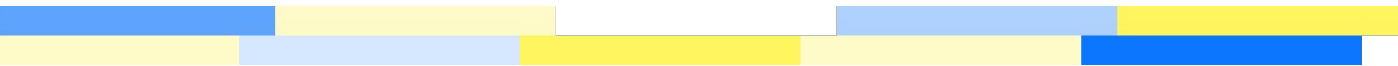
- Use welcoming and inclusive language
- Be respectful of different viewpoints and experiences
- Gracefully accept constructive criticism
- Focus on what is best for the community
- Show courtesy and respect towards other community members

Read the full Code of Conduct here:

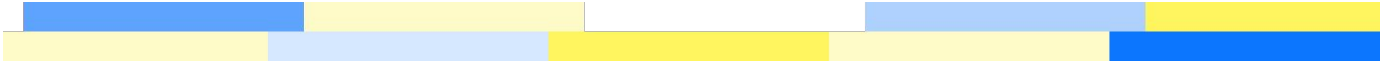
<https://alexslimonade.github.io/2023-january-training/code-of-conduct.html>



If you at any time feel harassed or treated inappropriately, please contact
ccd1@alexslemonade.org.



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



What you will learn

- We will cover cell-type identification using single-cell RNA-seq.
- 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 introduce working with CITE-seq data.



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.



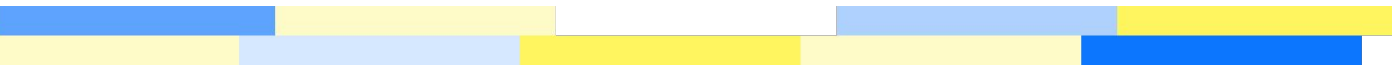
How do we pick what we teach?

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



Tuesday

Workshop Intro

Modules

- Reading, filtering, and normalizing scRNA-seq data
- Assigning cell types

Consultations

Exercise notebooks

Wednesday

Module

Integrating scRNA-Seq datasets

Consultations

- Exercise notebooks
- Your own data

Thursday

Module

Differential expression analysis for single-cell RNA-seq

Consultations

- Exercise notebooks
- Your own data

Presentations

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

00a-rstudio_guide.md
00b-debugging_resources.md
00c-good-scientific-coding-practices.md

These are reference documents.
Go through these on your own.

01-intro_to_base_R-live.Rmd
01-intro_to_base_R.Rmd
01-intro_to_base_R.nb.html
02-intro_to_ggplot2-live.Rmd
02-intro_to_ggplot2.Rmd
02-intro_to_ggplot2.nb.html
03-intro_to_tidyverse-live.Rmd
03-intro_to_tidyverse.Rmd
03-intro_to_tidyverse.nb.html

These are **Instruction** notebooks.
We'll walk through these together,
step-by-step, during the workshop.

exercise_01-intro_to_base_R.Rmd
exercise_02-intro_to_R.Rmd
exercise_03a-intro_to_tidyverse.Rmd
exercise_03b-intro_to_tidyverse.Rmd

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	<code>%>%</code>	Pipe operator	Funnels a data.frame through tidyverse operations
dplyr	<code>filter()</code>	Filter	Returns a subset of rows matching the conditions of the specified logical argument
dplyr	<code>arrange()</code>	Arrange	Reorders rows in ascending order. <code>arrange(desc())</code> would reorder rows in descending order.
dplyr	<code>select()</code>	Select	Selects columns that match the specified argument
dplyr	<code>mutate()</code>	Mutate	Adds a new column that is a function of existing columns
dplyr	<code>summarise()</code>	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. <code>summarize</code> and <code>summarise</code> are synonyms in this package.
dplyr	<code>rename()</code>	Rename	Renames designated columns while keeping all variables of the data.frame
dplyr	<code>group_by()</code>	Group By	Groups data into rows that contain the same specified value(s)
dplyr	<code>inner_join()</code>	Inner Join	Joins data from two data frames, retaining only the rows that are in both datasets.

Thursday

Your own projects Exercise notebooks

Spend Thursday 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!



Training Procedures



We're going to use sticky notes and note cards...

- As an alternative to raising your hands for help
- To give feedback about the session



During the session

Use the yellow sticky note to let us know that you need help or having some trouble following along.

One of our helpers will come over and help you.



Use the blue sticky note to let us know when a long-running step has completed.



At the end of each session

At the end of each module,
write down your muddiest point on
a white note card:

I do not understand
_____.

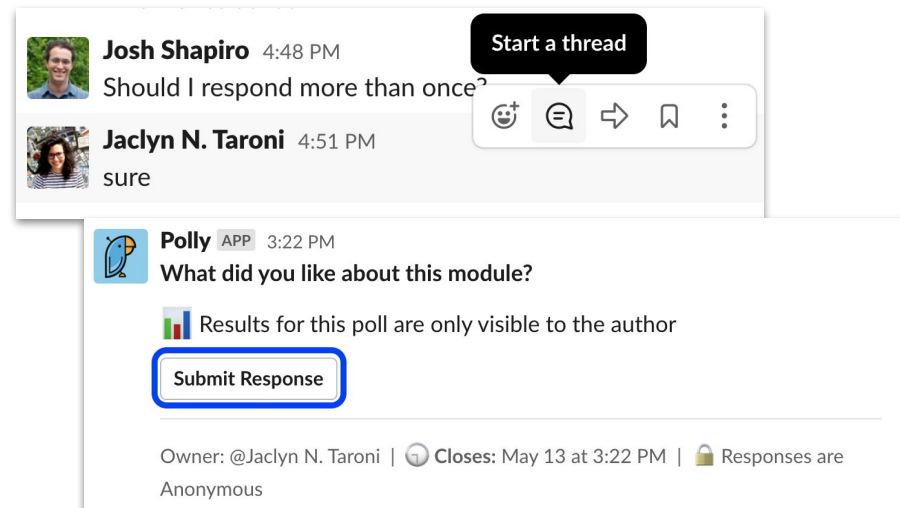
At the end of the session, use one
blue note card and one white note
card to write down:

What did you like about this
session?

How might we improve the
session?

We encourage you to use Slack

- You have been added to the **#2023-january-training** Slack channel
- Post public questions, get help with errors and debugging, make comments, and help others!
 - Use threads to keep related content together
- Stay in touch after the workshop!



The screenshot shows a Slack interface. At the top, a thread is visible with two messages: one from Josh Shapiro asking "Should I respond more than once?" and another from Jaclyn N. Taroni replying "sure". A "Start a thread" button is overlaid on the thread. Below the thread, a poll created by Polly (an app) is shown. The poll asks "What did you like about this module?" and states that results are only visible to the author. A "Submit Response" button is highlighted with a blue border. At the bottom, the poll's metadata is displayed: "Owner: @Jaclyn N. Taroni | Closes: May 13 at 3:22 PM | Responses are Anonymous".

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

Jaclyn N. Taroni 4:51 PM
sure

Start a thread

Polly APP 3:22 PM
What did you like about this module?

Results for this poll are only visible to the author

Submit Response

Owner: @Jaclyn N. Taroni | Closes: May 13 at 3:22 PM | Responses are Anonymous

Housekeeping Notes

- Waivers - If you have not yet done so, we have QR codes you can scan to sign our waivers
- Where are the restrooms?
- Where is water/coffee available?
- Snacks!
- Dinner tomorrow at 6:30PM at Manayunk Brewery located at 4120 Main St., Philadelphia, PA 19127