

Welcome to the June 2023 Data Lab scRNA-Seq Training Workshop!

WIFI: Bala Plaza
Password:
Bala123!

June 13 – 15, 2023
Childhood Cancer Data Lab
<https://alexslemonade.github.io/2023-june-training>

Childhood Cancer
Data  **Lab**

x



Tell us about you!

- What's your name?
- What are you studying?
- What's a recent piece of media (book, movie, TV show, etc.) you enjoyed?



Meet your instructors



JOSH

Joshua Shapiro

Senior 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 reproducible genomics**
- **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

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**

- Helps administer Data Lab offerings such as workshops
- Manages communications



DEEPA
Deepa Prasad
**User Experience
Designer**

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

Tell us about you!

- What's your name?
- What are you studying?
- What's a recent piece of media (book, movie, TV show, etc.) 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.

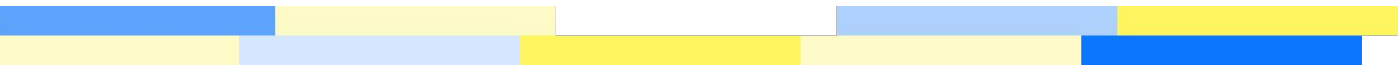
- Use welcoming and inclusive language
- Be respectful of different viewpoints and experiences
- Gracefully accept constructive criticism
- Focus on what is best for the community
- Show courtesy and respect towards other community members

Read the full Code of Conduct here:

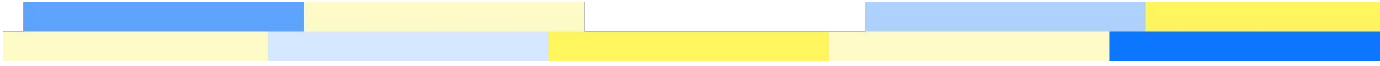
<https://alexslemonade.github.io/2023-june-training/code-of-conduct.html>



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 introduce you to the R programming language, R Notebooks, and some reproducible research practices.

We cover pipelines for the quality control, pre-processing, and initial analysis of single-cell RNA-seq data almost entirely through hands-on exercises.

We generally elect to go *broad* and not *deep*.

Our overarching goals: To prepare you to perform “frontline” analyses of your own data, to get you more comfortable reading documentation/learning new methods on your own, and to give you tools to collaborate more effectively with analysts when needed



What you won't learn


We don't address experimental design (e.g., how many replicates you need).

We won't compare tools (e.g., Bioconductor vs. Seurat for single cell analysis).

We won't cover every feature (or assumption) of the tools we do present.

You may not be able to perform every analysis you need for your own work, particularly for complex experimental designs. For example, in the context of single-cell analysis, we do not cover integrating data from multiple samples.

We present analysis as a series of *linear steps*. In practice, it's **not**. It's important to consult analysis experts when you need to and to keep track of and report what you've done.



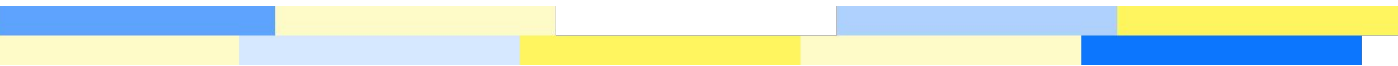
How do we pick what we teach?

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

Intro to R

RStudio Server
Base R
ggplot2 & tidyverse

Intro to Single-cell RNA-seq

Quantification & QC

Consultations

Exercise notebooks
Your own data

Wednesday

Single-cell RNA-seq

Filtering & normalization
Dimensionality reduction
Clustering
Marker identification

Consultations

Exercise notebooks
Your own data

Thursday

Single-cell RNA-seq

Cell-type annotation
Working with CITE-Seq
data

Consultations

Exercise notebooks
Your own data

Presentations

Daily Schedule Outline

Instruction

Full group

Lectures

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

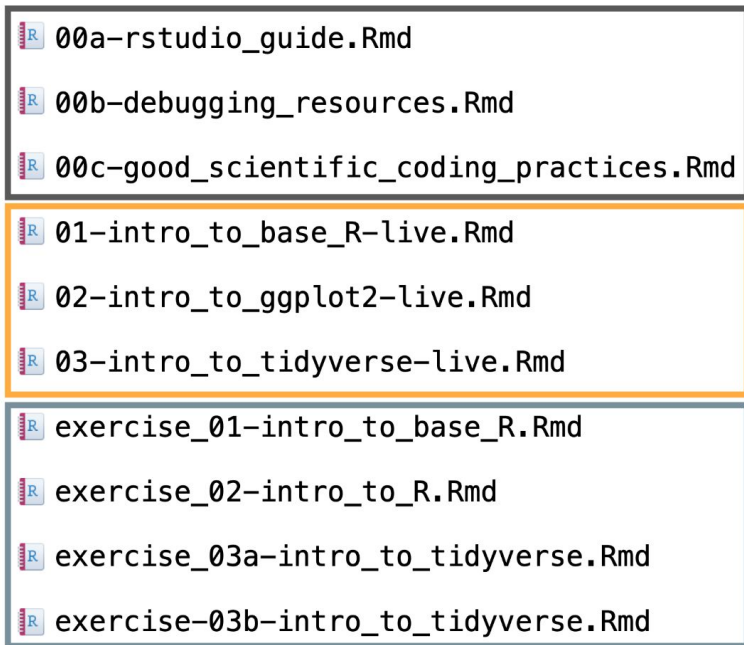
Consultation Period

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



Module cheatsheets cover key functions

<https://github.com/AlexsLemonade/training-modules/tree/2023-june/module-cheatsheets>

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
<code>dplyr</code>	<code>%>%</code>	Pipe operator	Funnels a <code>data.frame</code> through tidyverse operations
<code>dplyr</code>	<code>filter()</code>	Filter	Returns a subset of rows matching the conditions of the specified logical argument
<code>dplyr</code>	<code>arrange()</code>	Arrange	Reorders rows in ascending order. <code>arrange(desc())</code> would reorder rows in descending order.
<code>dplyr</code>	<code>select()</code>	Select	Selects columns that match the specified argument
<code>dplyr</code>	<code>mutate()</code>	Mutate	Adds a new column that is a function of existing columns
<code>dplyr</code>	<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.
<code>dplyr</code>	<code>rename()</code>	Rename	Renames designated columns while keeping all variables of the <code>data.frame</code>
<code>dplyr</code>	<code>group_by()</code>	Group By	Groups data into rows that contain the same specified value(s)
<code>dplyr</code>	<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 time on 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



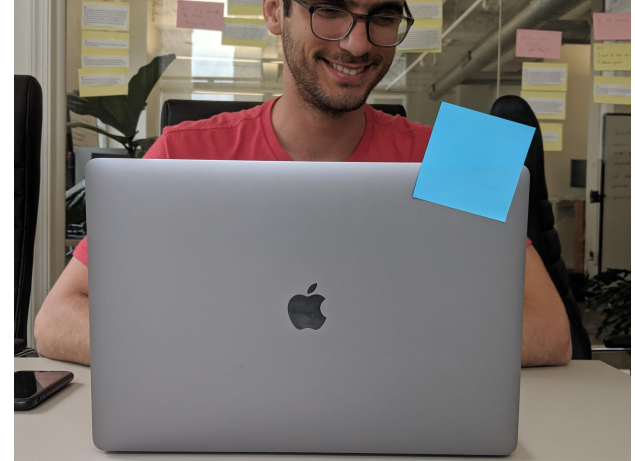
During the session

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
_____.

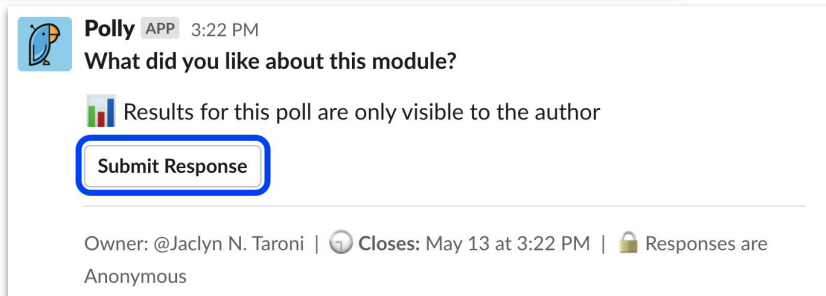
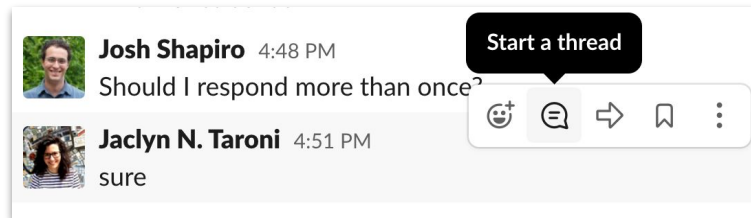
At the end of the session, use one blue note
card and one white note card to write
down:

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-june-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 Single-cell Pediatric Cancer Atlas (ScPCA)

The [Single-cell Pediatric Cancer Atlas \(ScPCA\) Portal](#) is a database of uniformly processed single-cell data from pediatric cancer clinical samples and xenografts.

We are currently seeking contributions from more pediatric cancer researchers!

- We accept 10x Genomics single-cell or single-nuclei profiling of childhood cancer data.
 - The Portal also support bulk RNA-Seq and CITE-Seq data.
- Requirements include processing your data with the [Data Lab's production pipeline](#), submitting the output (project, sample, and cell metadata), and signing a Data Transfer agreement.
- **Researchers that submit data may be eligible to receive a small one-time grant of unrestricted funds to be used for childhood cancer research!**
- Get started by [completing our intake form](#) to tell us more about your data! We will respond within 3 business days to notify you of your eligibility and provide additional information.

For the full guidelines visit: <https://scpca.alexslimonade.org/contribute>

Contact us: scpca@ccdatalab.org



Housekeeping Notes

- Waivers - If you have not yet done so, we have printed copies you can sign!
- Where are the restrooms?
- Where is water/coffee available?
- Snacks!
- Dinner tomorrow (Wednesday) at 6:00PM at Manayunk Brewery located at 4120 Main St., Philadelphia, PA 19127