Welcome to the May 2020 Virtual CCDL RNA-Seq Training Workshop!

May 4-8, 2020
Childhood Cancer Data Lab

Meet your instructors

JOSH
Joshua Shapiro

Data Scientist @ the CCDL

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

Research interests:

- **Evolutionary Genomics**

jashapiro

Meet your instructors

JACLYN
Jaclyn Taroni

Principal Data Scientist @ the CCDL

PhD Genetics *Dartmouth*
Postdoc Computational Biology *UPenn*

Research interests:

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

jaclyn-taroni

Meet your helpers

CANDACE

Candace Savonen

Biological Data Analyst @ the CCDL

Masters Neuroscience at *Michigan State University*

Research interests:

- **Neurogenomics**
- **Single-cell transcriptomics**

cansavvy

Meet your helpers

CHANTE

Chante Bethell

Biological Data Analyst @ the CCDL

Bachelor's in Bioinformatics from *Rowan University*

Research interests:

- **Functional motifs in the proteome**

cbethell

Other staff you may see

KURT
Kurt Wheeler

Data Engineer
@ CCDL

- Builds scalable systems
- Manages servers

DEEPA
Deepa Prasad

User Experience Designer
@ CCDL

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

TRISH
Trish Adkins

Writer @ ALSF

- Edits the ALSF blog
- Interviewer of Scientists, Translator of Science for ALSF Supporters

SHANNON
Shannon O'Connor

Writer @ ALSF

- Crafts emails, stories, you name it
- Interviews scientists, occasionally

Tell us about you!

- What’s your name?
- What are you studying?
- What’s one thing you are proud of?

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:
http://bit.ly/CCDL-training-code

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

Monday

Intro to RStudio Server

Intro to R and the Tidyverse

Wednesday

Bulk RNA-seq
Pre-processing
Differential expression

Friday

Consultations
Your own data
Exercise notebooks

Presentations

Tuesday

Consultations
Exercise notebooks
Your own data

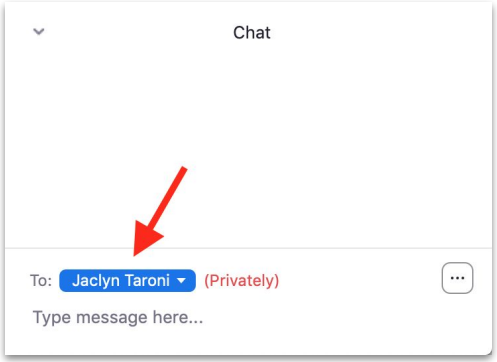
Thursday

Consultations
Exercise notebooks
Your own data

Full schedule: https://github.com/AlexsLemonade/2020-may-training/wiki/Schedule

Virtual Training Procedures

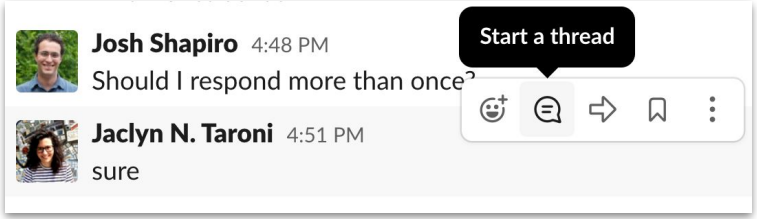
General Zoom Etiquette



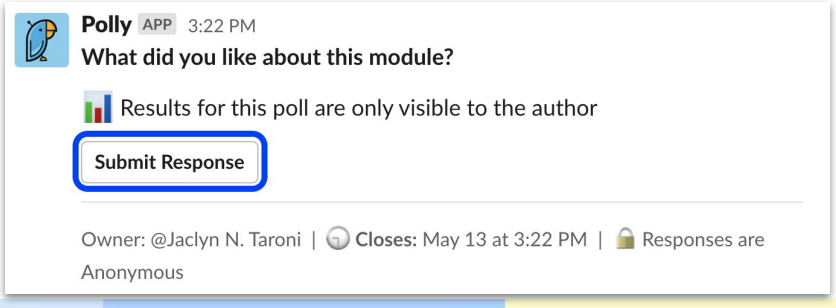
- Keep your microphone muted
- Type questions in the Chat window (directed to the host)
  - Click on the “Chat” button at the bottom of your window to open the chat.
- Use the Zoom status buttons to tell us how you are doing!
  - Click on the “Participants” button at the bottom of your zoom window to see these buttons

Using Slack

- Use the **#2020-may-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



- We will post questions at the end of each session with Polly



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, processing, and downstream analysis of bulk and 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., edgeR vs. DESeq2 for differential gene expression).

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 to perform for your own work, particularly for complex experimental designs.

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

Monday

Intro to RStudio Server

Intro to R and the Tidyverse

Wednesday

Bulk RNA-seq
Pre-processing
Differential expression

Friday

Consultations
Your own data
Exercise notebooks

Presentations

Tuesday

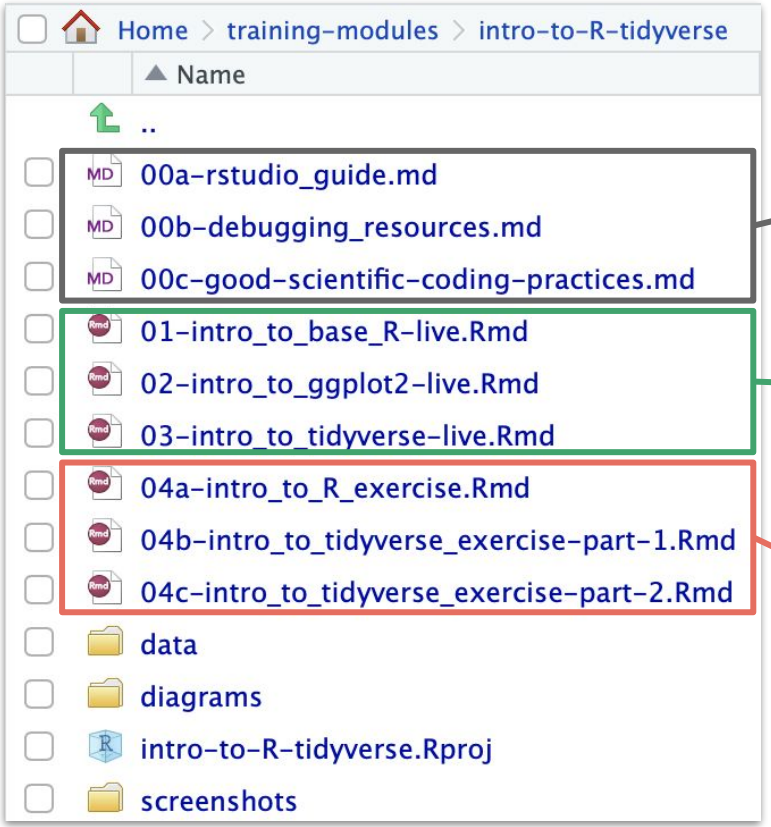
Consultations
Exercise notebooks
Your own data

Thursday

Consultations
Exercise notebooks
Your own data

Full schedule: https://github.com/AlexsLemonade/2020-may-training/wiki/Schedule

Module Layout



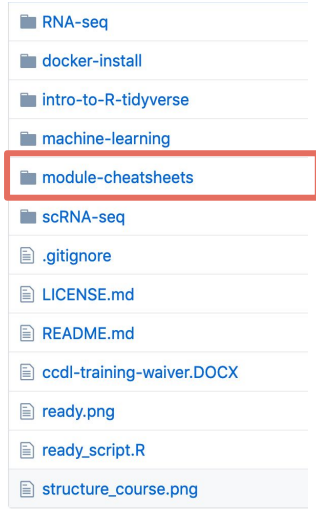
This is a reference document.
We will not go through this.

We'll walk through these notebooks together, step-by-step

You will practice what you have learned. We're here to help!

Module cheatsheets cover key functions

https://github.com/AlexsLemonade/training-modules



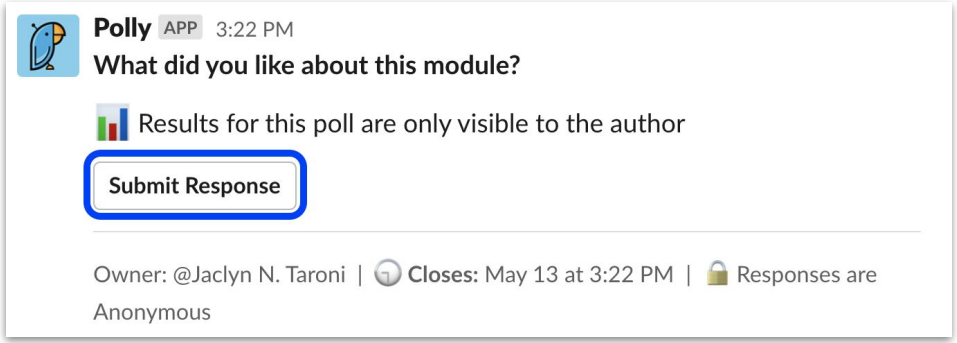
`dplyr`

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.

We want your feedback!

At the end of each module,
we will post a few questions
in the Slack channel.



- The most difficult or confusing point of the module ("muddiest point")
  Responses to this question will be anonymously posted in the channel

- What did you like about the module?
- How we can improve the module?
  These responses will be collected anonymously (and not posted).

Consultation Days

- We will post additional material at the start of each consultation day, responding to your **muddiest point** questions.

- Spend your time as you like, reviewing previous day presentations, working through exercise notebooks, or analyzing your own data.
  - We will ask the day before (using Polly in Slack) what you want to work on so we can be ready to assist you with your plans.
  - Work in groups, if you like! We can assist setting up groups with similar types of data or interests.

- CCDL staff are available in the Slack channel with Zoom meetings as needed.

Friday

Own data/Exercise notebooks	Spend Friday working with your own data, getting assistance as needed from CCDL staff and each other.
Presentations	Present what you worked on during the consultation times to the group!

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 CCDL staff member

slack

- **I have an general question** that does not need an answer right away.
  - ***Post*** in *#2020-may-training*
- **I'm having trouble logging in** to RStudio Server
  - ***Direct Message*** a CCDL staff member (not the current host or instructor)

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

Communication at other times (consultation days)

- I have questions about **yesterday's instruction** or **exercise notebooks**
  - ***Post*** in *#2020-may-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 *#2020-may-training*; we will set you up in a Zoom breakout room
- I have a question that is **highly specific to my data**
  - ***Direct Message*** a CCDL staff member
- **I'm having trouble logging in** to RStudio Server
  - ***Direct Message*** a CCDL staff member

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