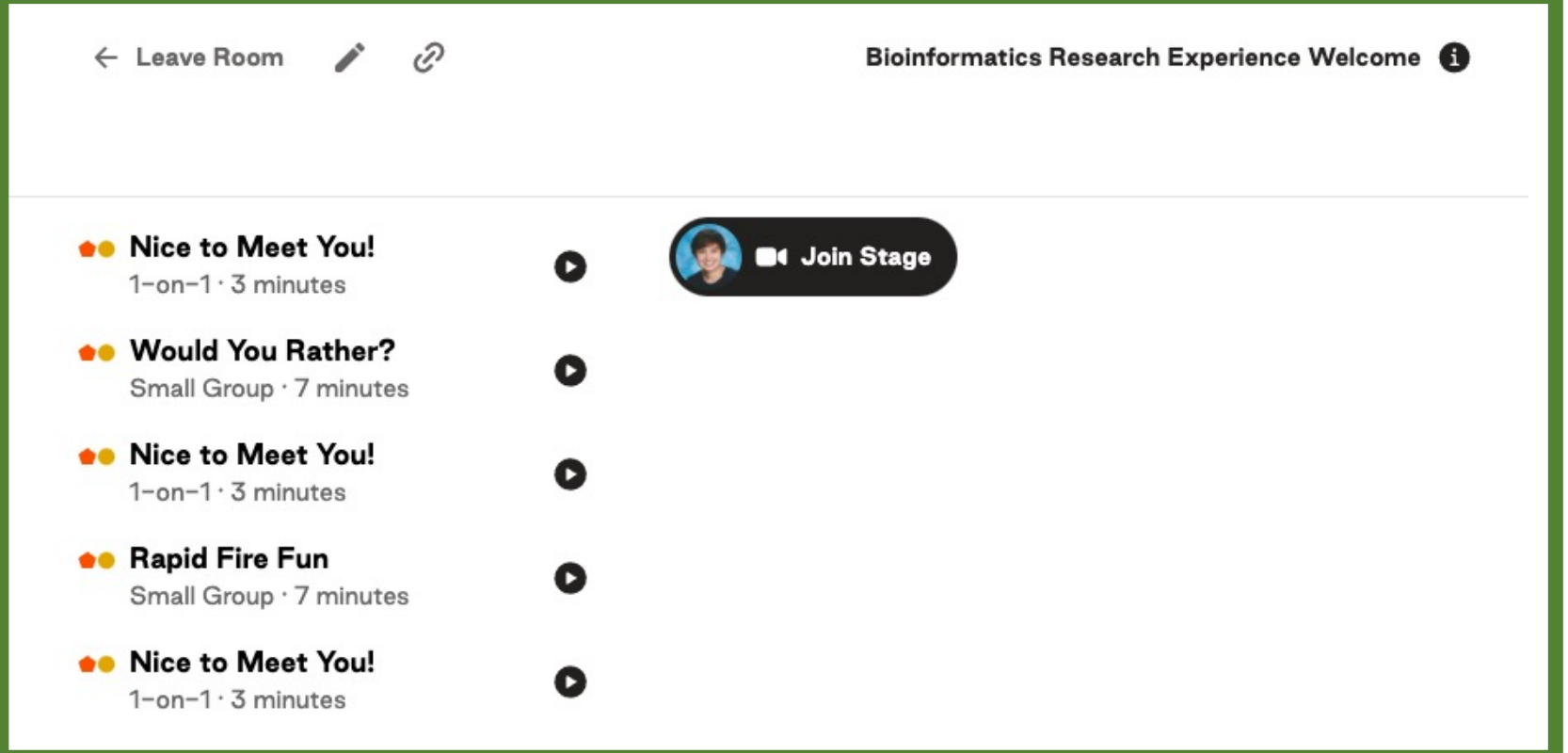
BIOINFORMATICS RESEARCH EXPERIENCE

CORIELL INSTITUTE

Introduction to Coriell

Student Introductions

GATHEROUND



- Get to know each other a bit on Gatheround
- https://gatheround.com/events/Yc92Xyix1NeJbWwIKSax
- Game cycles through a couple of rounds, both 1-on-1 and in small groups where you can answer questions or chat with each other

Introduction to the Bioinformatics Research Experience

BIOINFORMATICS
RESEARCH EXPERIENCE

CORIELL INSTITUTE

Program Expectations

- Expect you to be on time to everything and turn assigned work in on time
- Preferred that you leave your camera on
- If there are any issues, family, medical or otherwise, please let me know ASAP and we'll work out
- Would like to take a picture of everyone on July 23 in their t-shirts

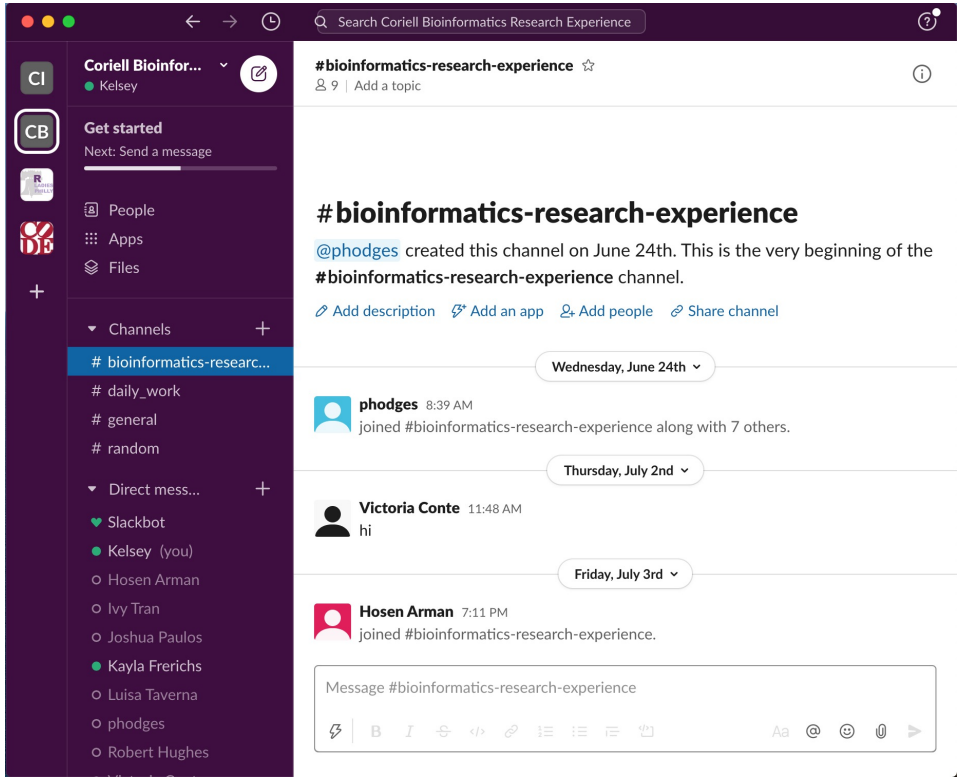
Program Overview

- Have to attend the daily lecture, journal club, the talk and any other presentations. Office hours are optional if you need help with something; can also get help through email or Slack
- There will be work assigned daily that will be due 2 days later
- Starting with general R skills, then move to bioinformatics-specific skills and RNA-seq
- Will have to do 2 journal club presentations, one general data analysis presentation, and one

Daily Schedule	
8AM	
9AM	Daily Lecture
10AM	Journal Club (Weeks 2-4)
11AM	
12PM	Talk
1PM	Office Hours
2PM	
3PM	
4PM	
5PM	
6PM	

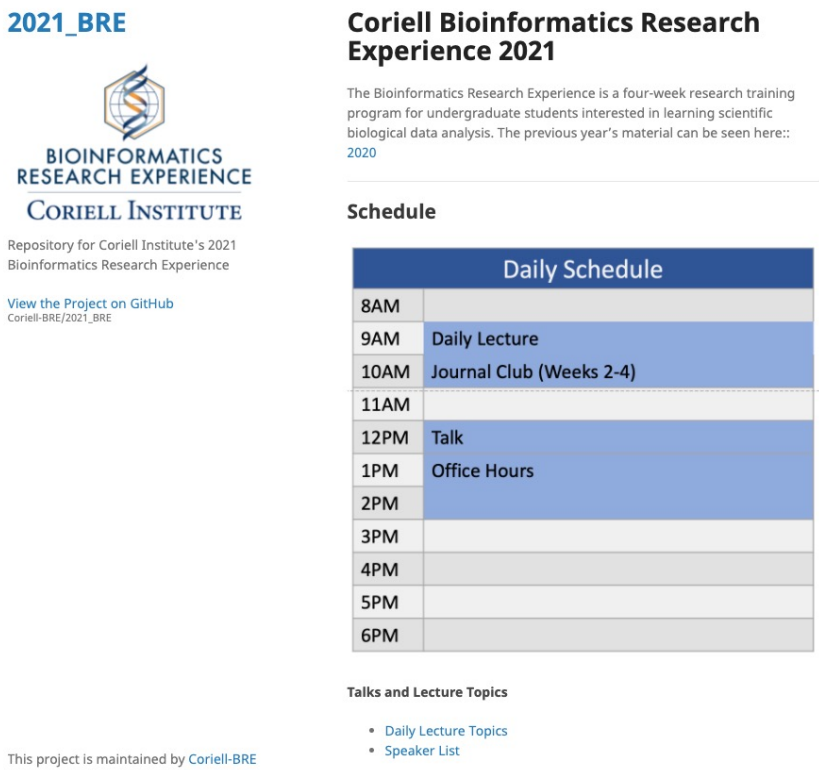
Resources

Slack



You can also email me at kkeith@coriell.org

Bioinformatics Research Experience Website



https://coriell-bre.github.io/2021_BRE/

R and the Tidyverse

R Programming Language

- Open-source scripting language developed for statistical analysis
- Was originally known as S
  - S was developed in 1975
  - S was reimplemented as R in 1993
- Heavily used in bioinformatics, with our own archive Bioconductor, for biology-related packages

The Tidyverse

"The tidyverse is an opinionated collection of R packages designed for data science. All packages share an underlying design philosophy, grammar, and data structures."

- Keep everything simple
  - Use **existing** data structures instead of custom, aka, use tidy data
  - Functions should do one thing well
- Glue the simple things together; simple things put together are more powerful than one complex thing
- Design for humans

R Studio®

Rmarkdown and Rstudio

What is markdown (and Rmarkdown)?

- Markdown is just text, with a few optional symbols that allows a markdown interpreter to make it look good. Goal is to have something that still is human readable even without the interpreter.
- Rmarkdown is markdown for R
  - All the features of markdown, with extras
  - Intention is to make documenting data analysis easy
  - Execute code in Rmarkdown files (cannot do this in markdown)
  - Can also knit Rmarkdown files into other files
    - html, pdf, or Microsoft word reports
    - Can make websites and slides with Rmarkdown

Why use markdown and Rmarkdown?

Bioinformatics is mostly on a Linux machine using the command line terminal. The rest of the universe uses Macs or PCs.

- **This causes a bunch of problems**
  - operating systems don't talk to each other easily
  - files are in proprietary formats (no Microsoft anything on Linux)
  - files are not readable in plain text (and terminal needs them to be!)
- **Markdown solves all the problems**
  - Simple
  - readable by every machine in both GUI and terminal AND humans
  - allows some simple formatting to increase human readability

Rstudio and Rmarkdown DEMO