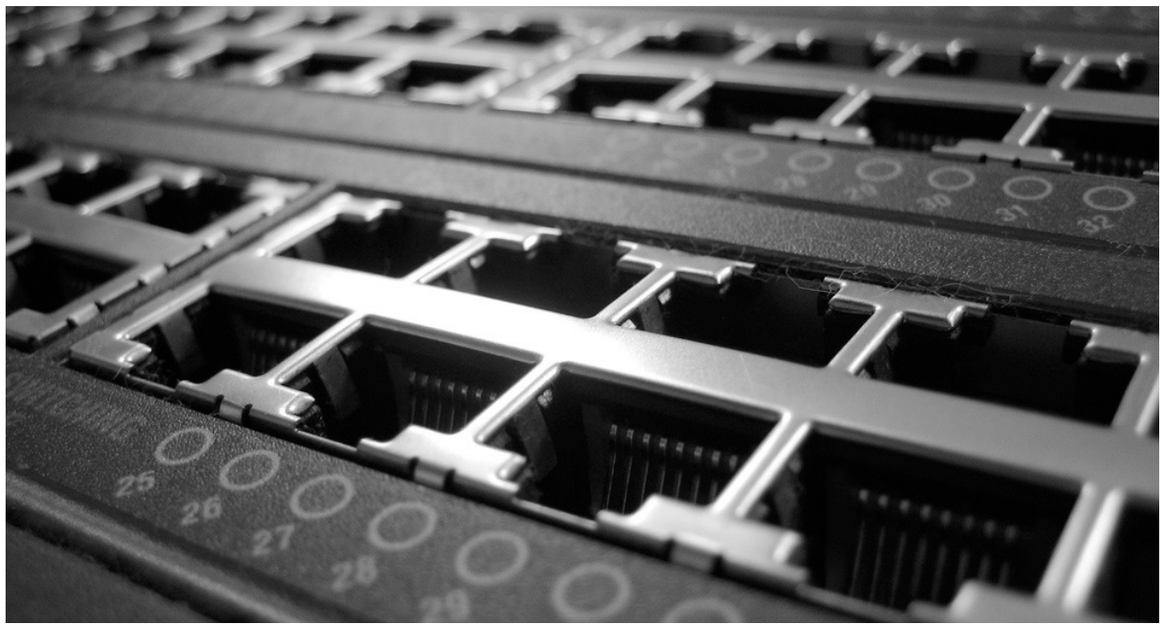


Introduction to the command-line interface

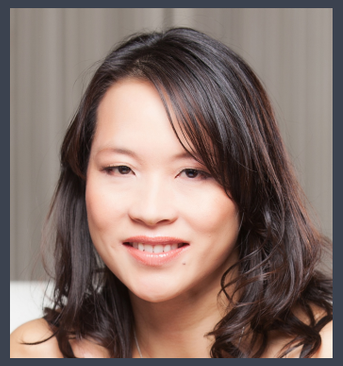
(shell)

Harvard Chan Bioinformatics Core

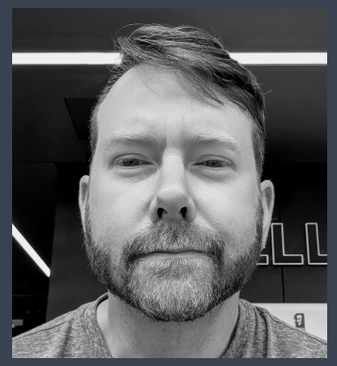
in collaboration with

HMS Research Computing

https://tinyurl.com/hbc-shell-online

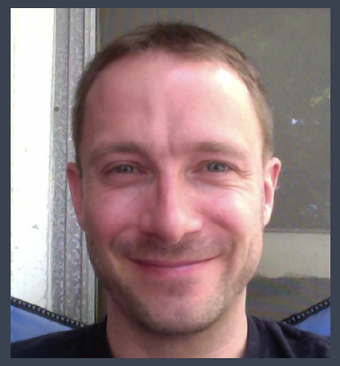


Shannan Ho Sui
Director



John Hutchinson
Associate Director

Victor Barrera



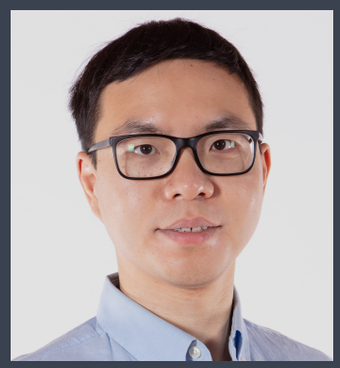
Rory Kirchner

Zhu Zhuo

Preetida Bhetariya

Meeta Mistry

Mary Piper
Assoc. Training Director



Jihe Liu

Radhika Khetani
Training Director

Maria Simoneau
Project coordinator

James Billingsley

Sergey Naumenko

Joon Yoon

Peter Kraft
Faculty Advisor

Consulting

- RNA-seq analysis: bulk, single cell, small RNA
- ChIP-seq and ATAC-seq analysis
- Genome-wide methylation
- WGS, resequencing, exome-seq and CNV studies
- QC & analysis of gene expression arrays
- Functional enrichment analysis
- Grant support

http://bioinformatics.sph.harvard.edu/

CFAR
HARVARD UNIVERSITY
CENTER FOR AIDS RESEARCH

HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

NIEHS

HSCI
HARVARD STEM CELL
INSTITUTE

HARVARD
CATALYST
THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

HARVARD
MEDICAL SCHOOL

Training

We have divided our short workshops into 2 categories:

1. Basic Data Skills - No prior programming knowledge needed (no prerequisites)

2. Advanced Topics: Analysis of high-throughput sequencing (NGS) data - Certain “Basic” workshops required as prerequisites.

Any participants wanting to take an advanced workshop will have to have taken the appropriate basic workshop(s) within the past 6 months.

http://bioinformatics.sph.harvard.edu/training/

https://hbctraining.github.io/main/

DF/HCC
DANA-FARBER / HARVARD CANCER CENTER

HSCI
HARVARD STEM CELL INSTITUTE

HARVARD CATALYST
THE HARVARD CLINICAL AND TRANSLATIONAL SCIENCE CENTER

HARVARD MEDICAL SCHOOL

Training

We have divided our short workshops into 2 categories:

1. Basic Data Skills - No prior programming knowledge needed (no prerequisites)

2. Advanced Topics: Analysis of high-throughput sequencing (NGS) data - Certain “Basic” workshops required as prerequisites.

Any participants wanting to take an advanced workshop will have to have taken the appropriate basic workshop(s) within the past 6 months.

http://bioinformatics.sph.harvard.edu/training/

https://hbctraining.github.io/main/

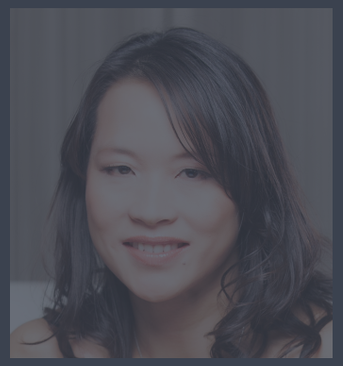
DF/HCC
DANA-FARBER / HARVARD CANCER CENTER

HSCI
HARVARD STEM CELL INSTITUTE

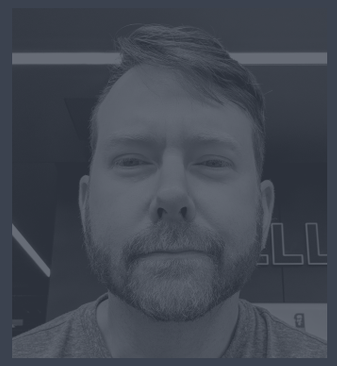
HARVARD CATALYST
THE HARVARD CLINICAL AND TRANSLATIONAL SCIENCE CENTER

HARVARD MEDICAL SCHOOL

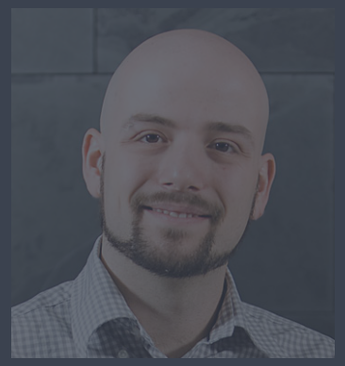
Introductions!



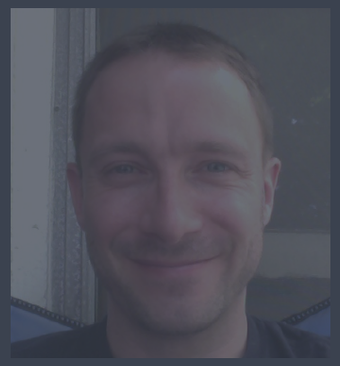
Shannan Ho Sui
Director



John Hutchinson
Associate Director

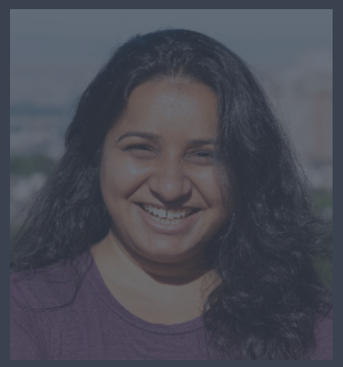


Victor Barrera



Rory Kirchner

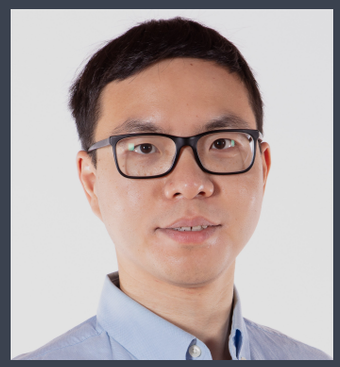
Zhu Zhuo



Preetida Bhetariya

Meeta Mistry

Mary Piper
Assoc. Training Director

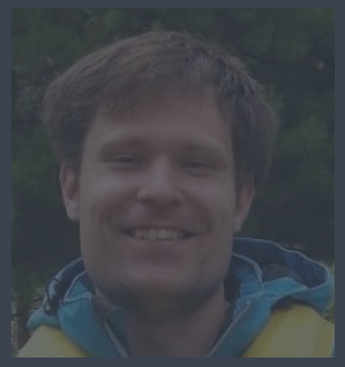


Jihe Liu

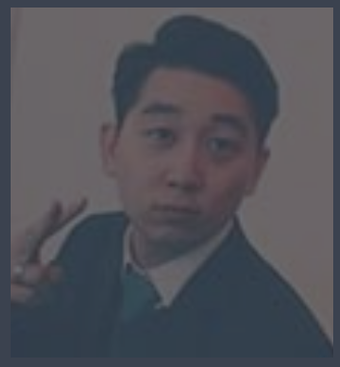
Radhika Khetani
Training Director

Maria Simoneau
Project coordinator

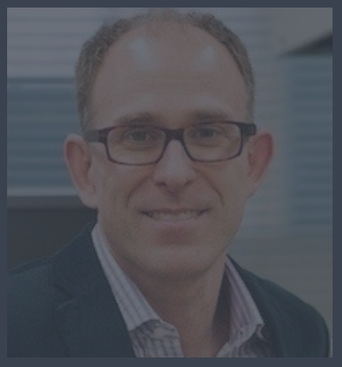
James Billingsley



Sergey Naumenko



Joon Yoon



Peter Kraft
Faculty Advisor

Workshop scope

```
rkhetani — rsk27@clarinet002-072: ~ — ssh — 75×51
rsk27@clarinet002-072: ~
rsk27@clarinet002-072:~$ ll -htr unix_workshop/
total 177K
drwxrwsr-x 2 rsk27 rsk27  62 May 23  2016 reference_data
-rw-rw-r-- 1 rsk27 rsk27 377 May 23  2016 README.txt
drwxrwsr-x 2 rsk27 rsk27  78 May 23  2016 genomics_data
drwxrwsr-x 2 rsk27 rsk27 257 May 23  2016 raw_fastq
drwxrwsr-x 2 rsk27 rsk27 695 May 23  2016 other
drwxrwsr-x 6 rsk27 rsk27 972 May 24  2016 rnaseq_project
rsk27@clarinet002-072:~$
```

“Unix is user-friendly.

It's just very selective about who its friends are.”

http://en.wikipedia.org/wiki/Unix, http://en.wikipedia.org/wiki/Tux

The Unix command-line interface

- Unix is a stable, efficient and powerful operating system
- It can easily coordinate the use and sharing of a computer's (or a system's) resources, i.e. built to allow multi-user functionality
- Can easily handle complex and repetitive tasks easily on large and small datasets

The Unix command-line interface

- Unix is a stable, efficient and powerful operating system
- It can easily coordinate the use and sharing of a computer's (or a system's) resources, i.e. built to allow multi-user functionality
- Can easily handle complex and repetitive tasks easily on large and small datasets
- Usually, written commands are used to work with this OS, instead of the pointing and clicking used with operating systems like Windows and OSX

The Unix command-line interface

- Unix is a stable, efficient and powerful operating system
- It can easily coordinate the use and sharing of a computer's (or a system's) resources, i.e. built to allow multi-user functionality
- Can easily handle complex and repetitive tasks easily on large and small datasets
- Usually, written commands are used to work with this OS, instead of the pointing and clicking used with operating systems like Windows and OSX

The Unix command-line interface

- Unix is a stable, efficient and powerful operating system
- It can easily coordinate the use and sharing of a computer's (or a system's) resources, i.e. built to allow multi-user functionality
- Can easily handle complex and repetitive tasks easily on large and small datasets
- Usually, written commands are used to work with this OS, instead of the pointing and clicking used with operating systems like Windows and OSX

Bioinformatics:

The Unix command-line interface

- Unix is a stable, efficient and powerful operating system
- It can easily coordinate the use and sharing of a computer's (or a system's) resources, i.e. built to allow multi-user functionality
- Can easily handle complex and repetitive tasks easily on large and small datasets
- Usually, written commands are used to work with this OS, instead of the pointing and clicking used with operating systems like Windows and OSX

Bioinformatics:

- A lot of NGS-analysis tools are created for the Unix OS

The Unix command-line interface

- Unix is a stable, efficient and powerful operating system
- It can easily coordinate the use and sharing of a computer's (or a system's) resources, i.e. built to allow multi-user functionality
- Can easily handle complex and repetitive tasks easily on large and small datasets
- Usually, written commands are used to work with this OS, instead of the pointing and clicking used with operating systems like Windows and OSX

Bioinformatics:

- A lot of NGS-analysis tools are created for the Unix OS
- High-performance compute clusters which are necessary to analyze large datasets require a working knowledge of Unix

Linux

- Linux is a free, open-source operating system based on Unix
- It has the same components as the original, but the open source community is involved in active development of various distinct distributions of Linux

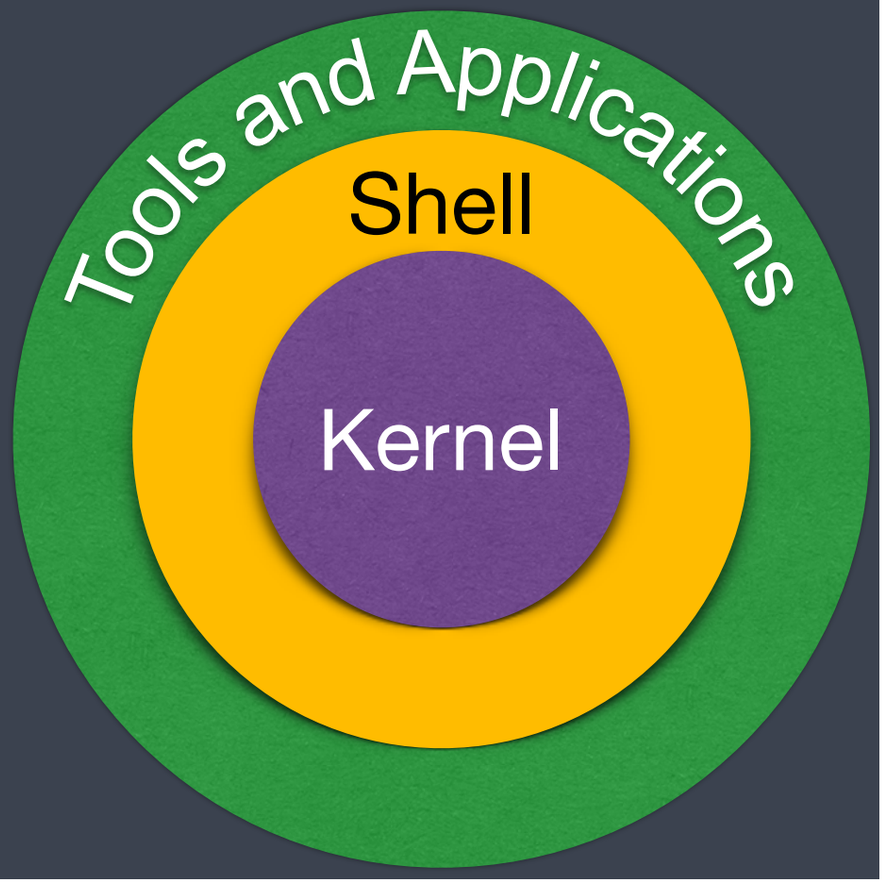
Linux

- Linux is a free, open-source operating system based on Unix
- It has the same components as the original, but the open source community is involved in active development of various distinct distributions of Linux

Components

The Unix/Linux system is functionally organized at 3 levels:

- The kernel, which schedules tasks and manages storage: ***the brain of the system***
- The shell, ***an interpreter*** that helps interprets our input for the kernel
- Utilities, tools and applications, which use the shell to communicate with the kernel



The “shell”

- The shell is **an interpreter**

The “shell”

- The shell is **an interpreter**
- It is independent of the operating system

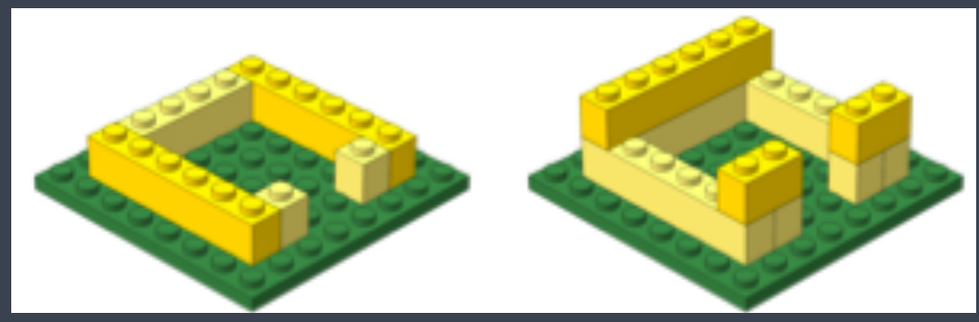
The “shell”

- The shell is **an interpreter**
- It is independent of the operating system
- Dozens of shells have been developed throughout UNIX history, and a lot of them are still in use

The “shell”

- The shell is **an interpreter**
- It is independent of the operating system
- Dozens of shells have been developed throughout UNIX history, and a lot of them are still in use
- The most commonly used shell is **bash**

Learning Objectives



- ✓ Learn what a “shell” is and become comfortable with the command-line interface
  - Find your way around a filesystem using written commands
  - Work with small and large data files
  - Become more efficient when performing repetitive tasks
- ✓ Understand what a computational cluster is and why we need it

Logistics

Course webpage

https://tinyurl.com/hbc-shell-online

Course schedule online

Workshop Schedule

Day 1

Time	Topic	Instructor
10:00 - 10:30	Workshop introduction	Radhika
10:30 - 11:45	Introduction to Shell	Radhika
11:45 - 12:00	Overview of self-learning materials and homework submission	Jihe

Self Learning #1

- Wildcards and shortcuts in bash
- Searching and redirection
- Examining and creating files
- Shell scripts and variables in bash

Assignment #1

- All exercise questions from the self-learning lessons have been put together in a text file (download for local access).
    - The text file can be opened with any text editor application (i.e. Notepad++, TextWrangler) on your local computer
- Add your solutions to the exercises in the downloaded .txt file and **upload the saved text file** to Dropbox **day before the next class**.

Course materials online

Learning Objectives

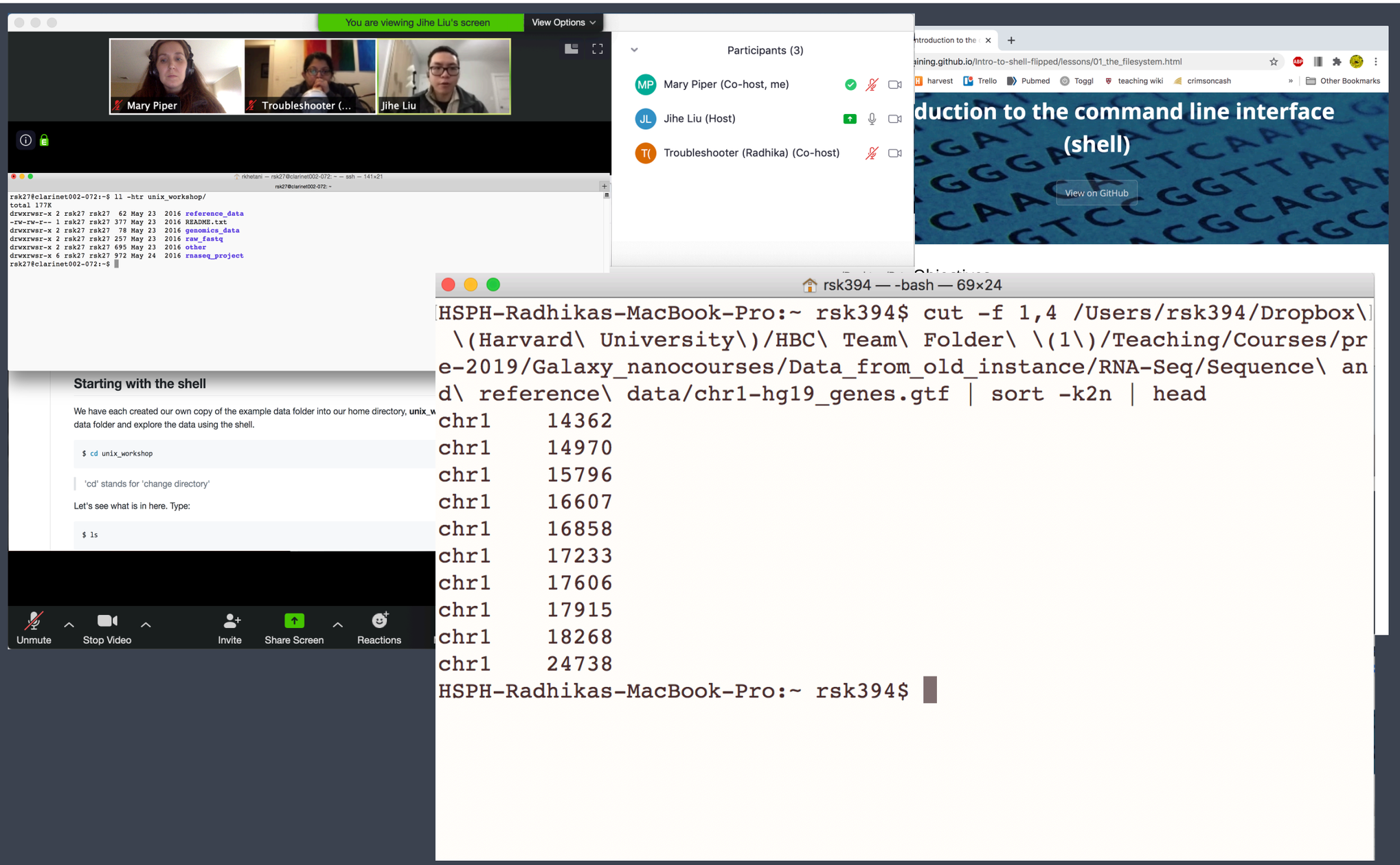
- How do you access the shell?
- How do you use it?
  - Getting around the Unix file system
  - looking at files
  - manipulating files
  - automating tasks
- What is it good for?

Setting up

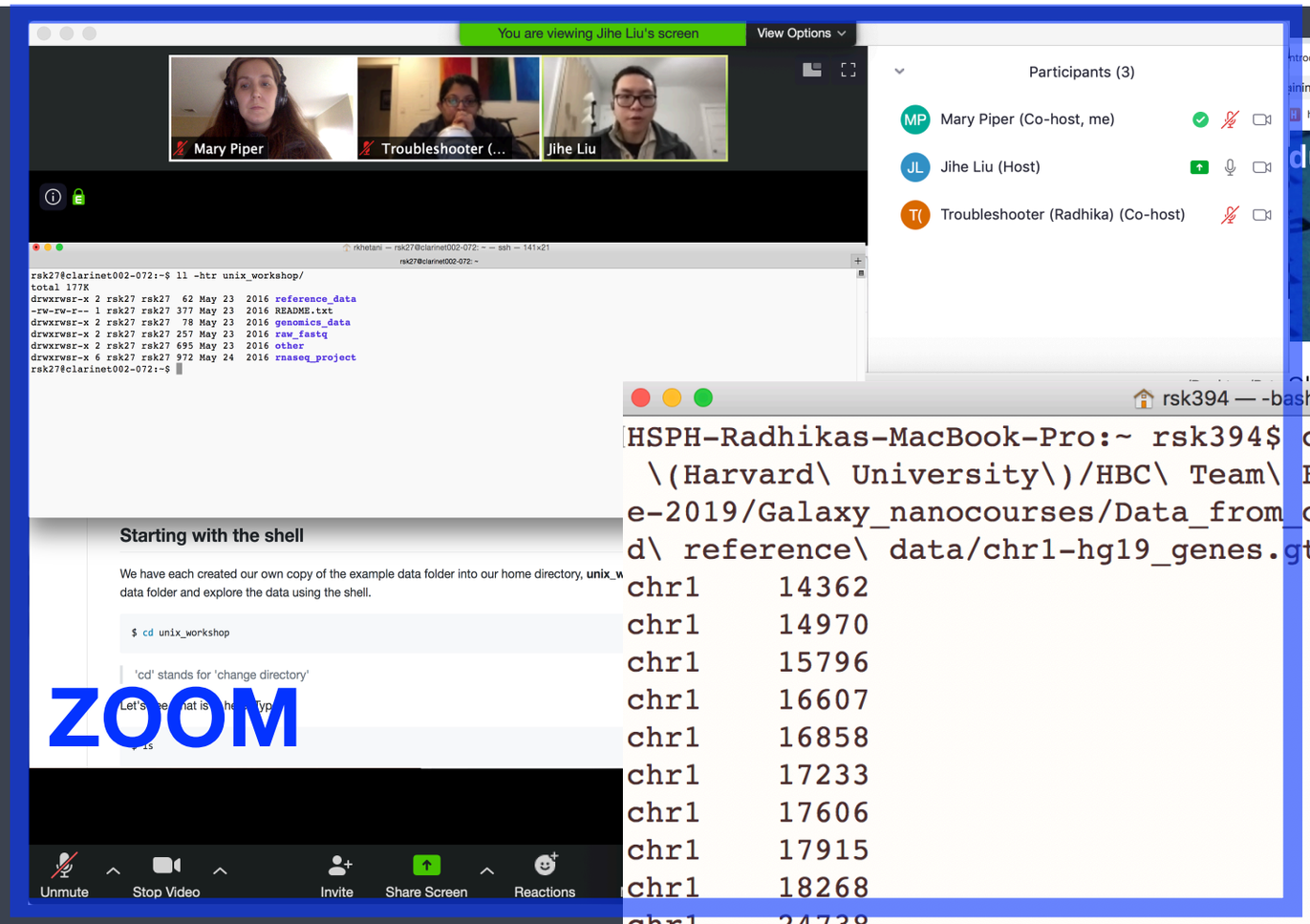
We will spend most of our time learning about the basics of the shell command-line interface (CLI) by exploring experimental data on the **O2** cluster. So, we will need to log in to this remote compute cluster first before we can start with the basics.

Let's take a quick look at the basic architecture of a cluster environment and some cluster-specific jargon prior to logging in.

Single screen & 3 windows?

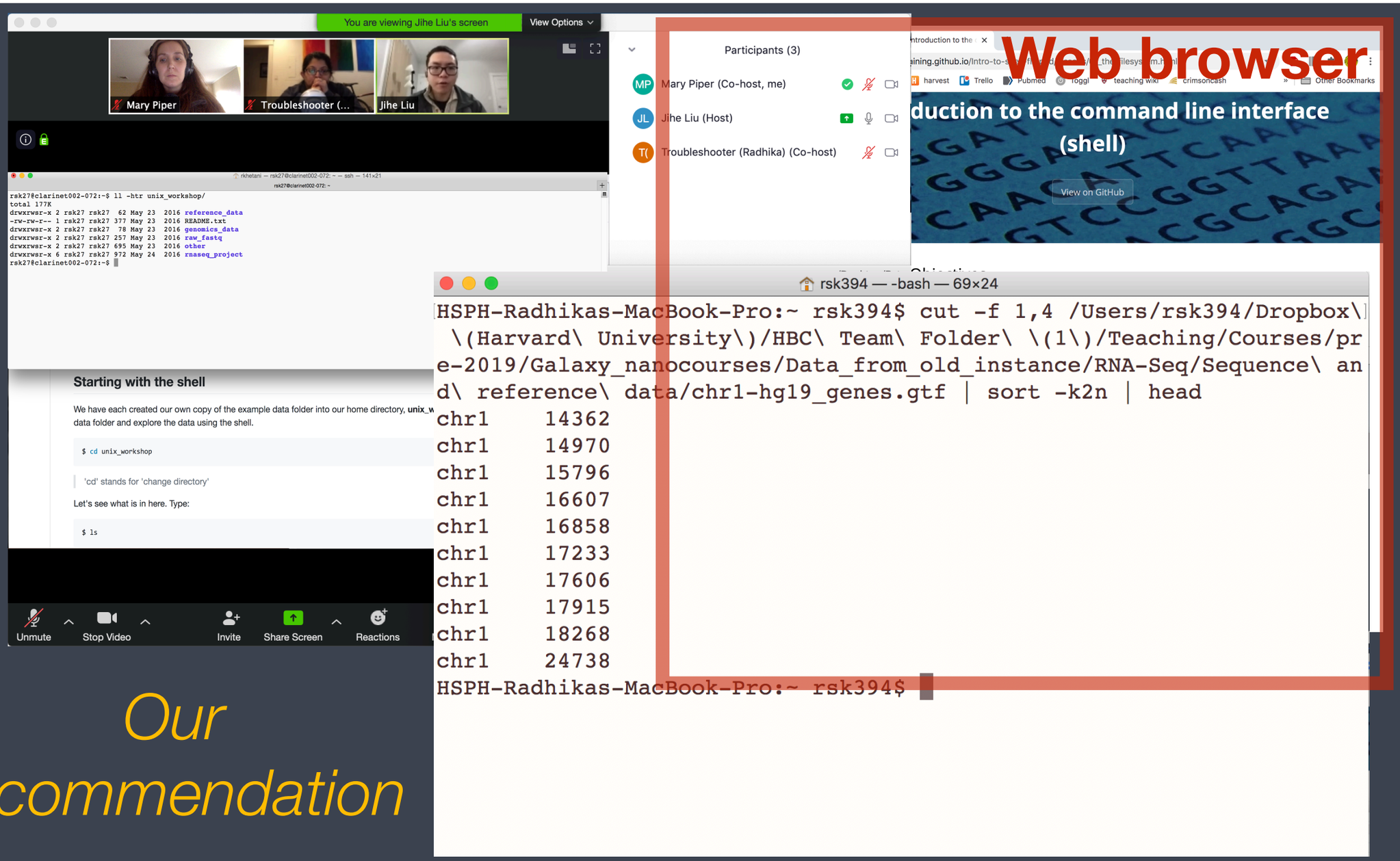


Single screen & 3 windows?



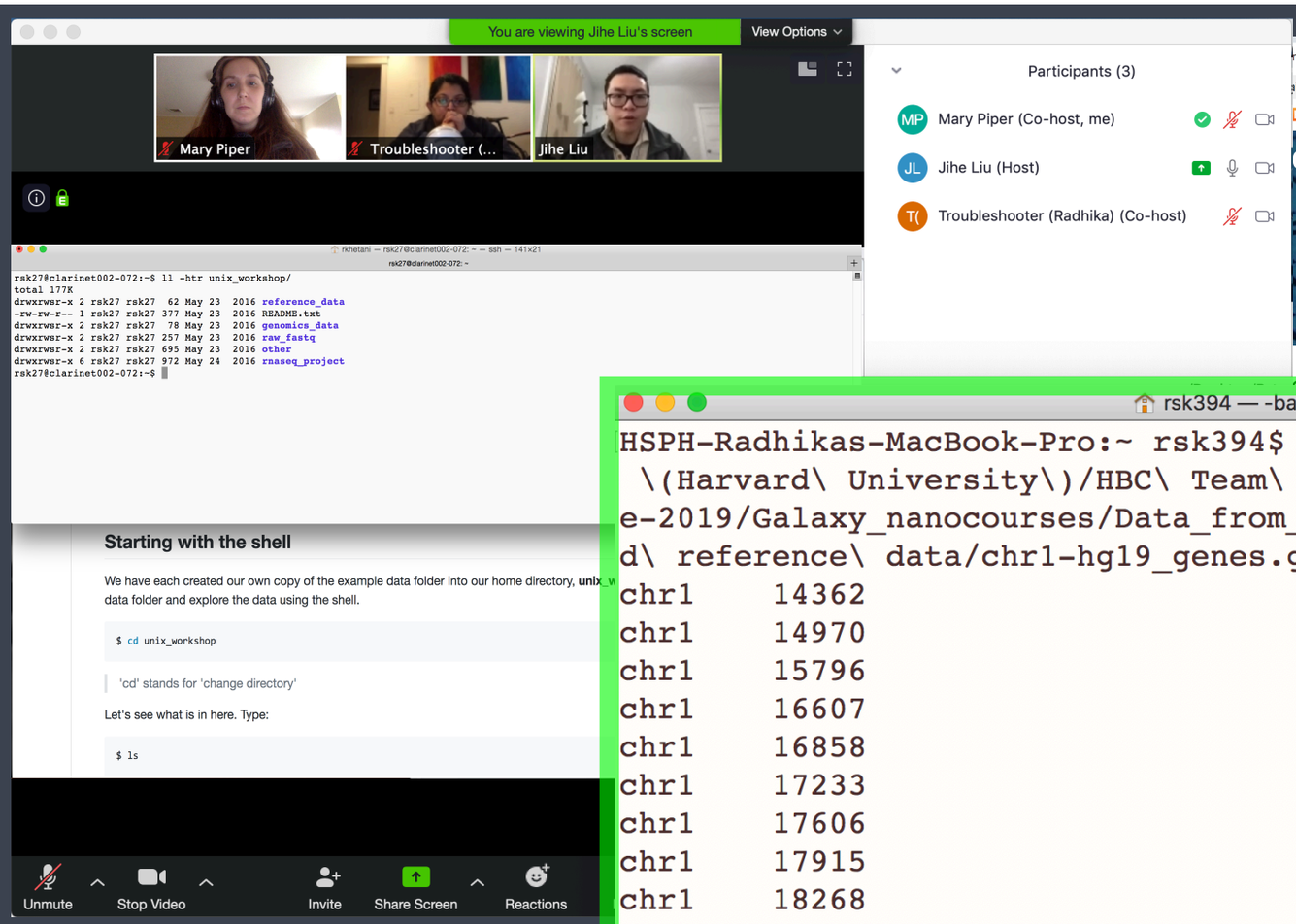
Our recommendation

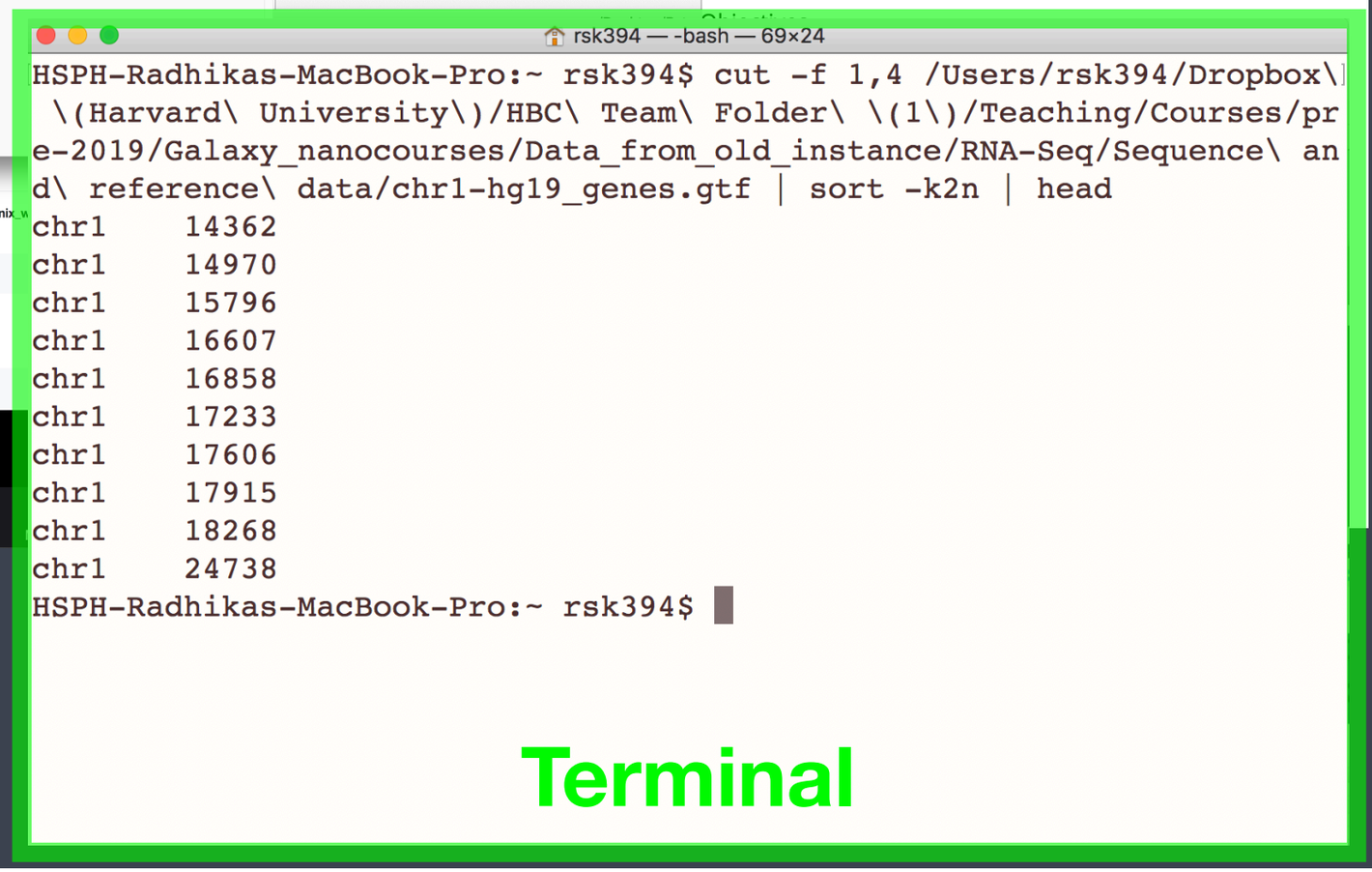
Single screen & 3 windows?



Our recommendation

Single screen & 3 windows?

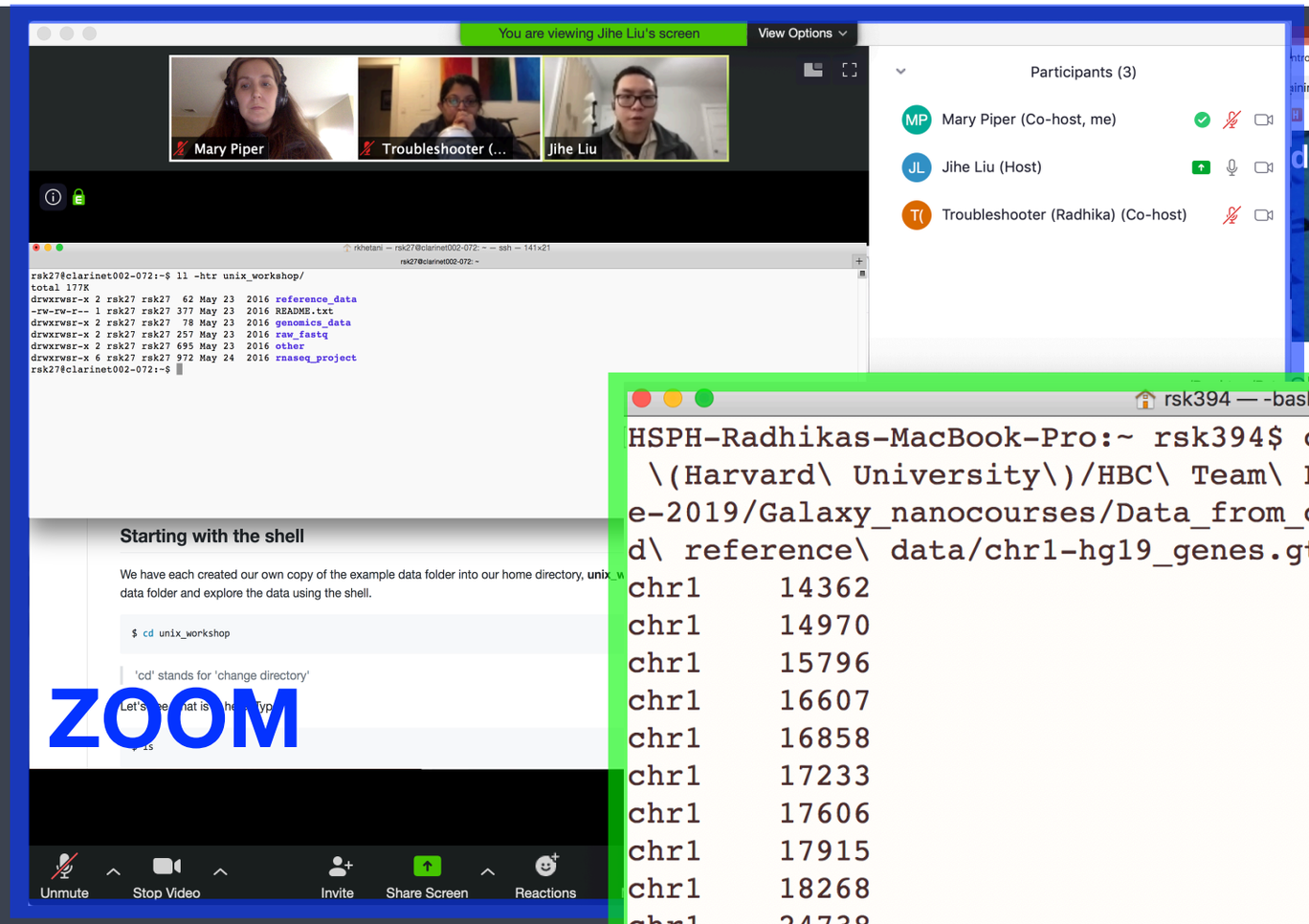


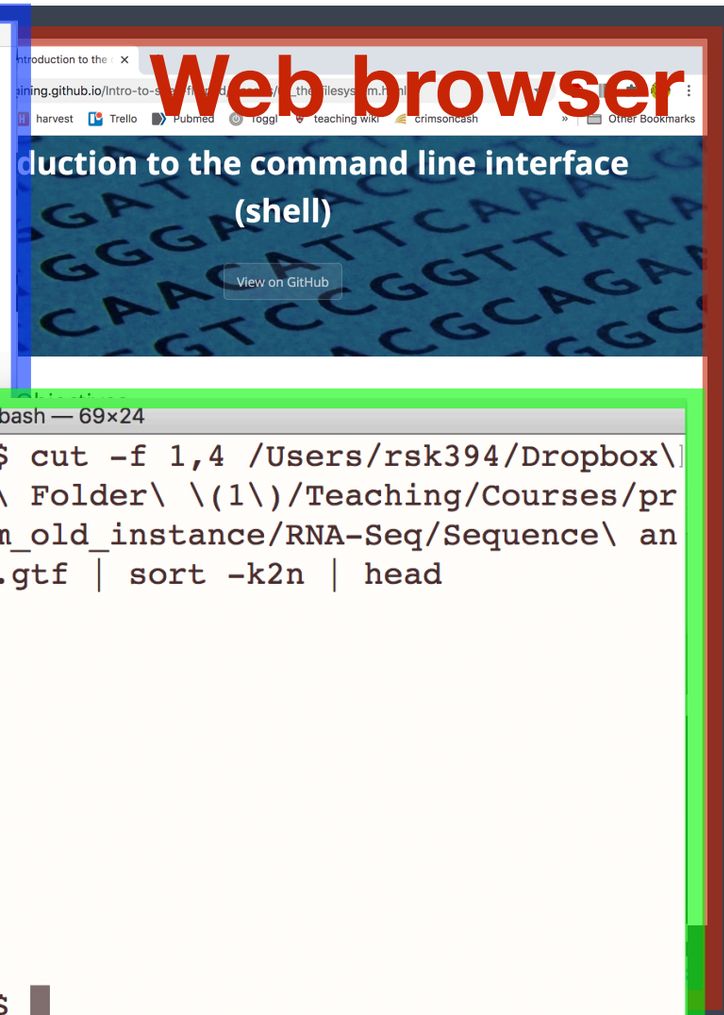


Our recommendation

Terminal

Single screen & 3 windows?



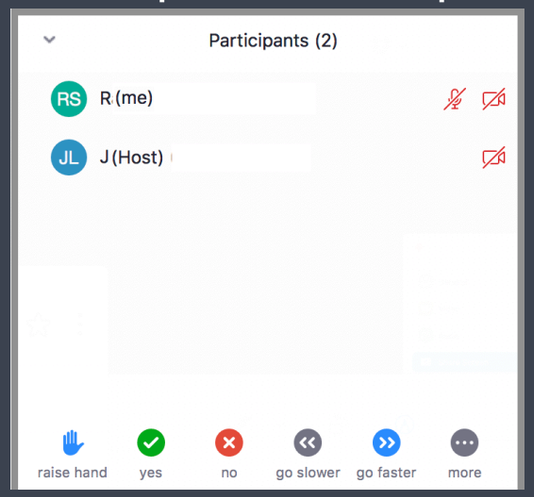


Our recommendation

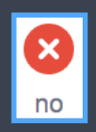
Terminal

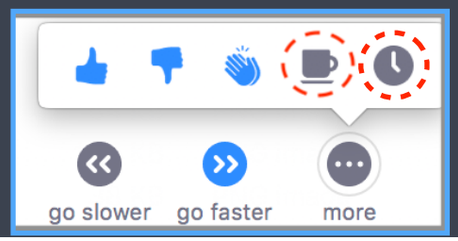
Odds and Ends (1/2)

- Quit/minimize all applications that are not required for class
- Click on “Participants” to open that panel in Zoom



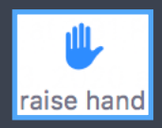
- 

= "agree", "I'm all set" (equivalent to a **green post-it**)
- 

= "disagree", "I need help" (equivalent to a **red post-it**)
- 

If you are away from the computer use the coffee cup or clock icon

Odds and Ends (2/2)

- Questions for the presenter?
  - Post the question in the Chat window OR
  - Raise your hand 

when the presenter asks for questions

- Technical difficulties with logging in or running commands?
  - Start a *private* chat with the *Troubleshooter* with a description of the problem.

Thanks!

- **Kathleen Keating** and **Andy Bergman** from HMS-RC
- Data Carpentry

Contact us!

HBC training team: hbctraining@hsph.harvard.edu

O2 (HMS-RC): rchelp@hms.harvard.edu

HBC consulting: bioinformatics@hsph.harvard.edu

Twitter

HBC: @bioinfocore

HMS-RC: @hms_rc