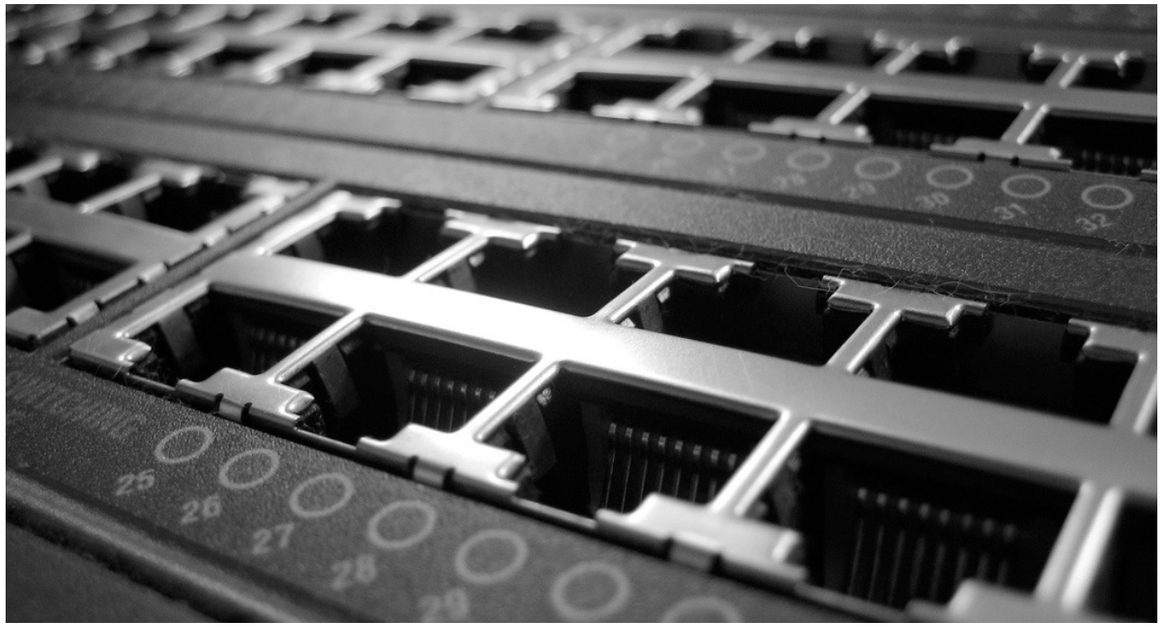


Introduction to bulk RNA-seq (Part I)

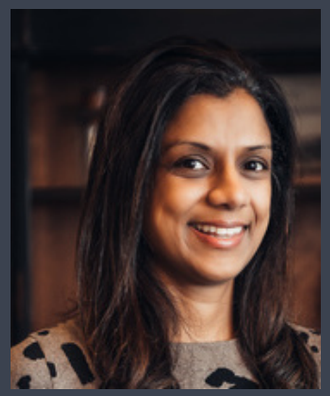
Harvard Chan Bioinformatics Core

in collaboration with

HMS Research Computing

https://tinyurl.com/hbc-rnaseq

Shannan Ho Sui
Director

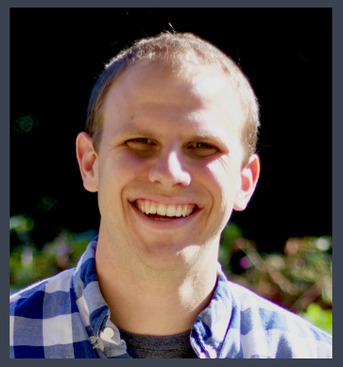


Meeta Mistry
Associate Director

John Quackenbush
Faculty Advisor

Emma Berdan

Heather Wick



Will Gammerdinger

Noor Sohail

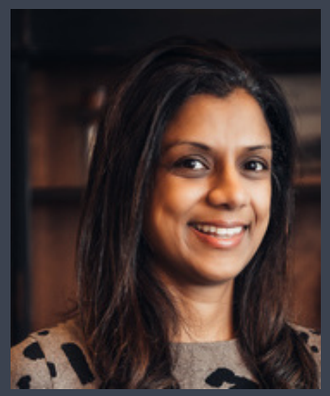
Upendra Bhattarai

James Billingsley

Zhu Zhuo

Maria Simoneau

Shannan Ho Sui
Director

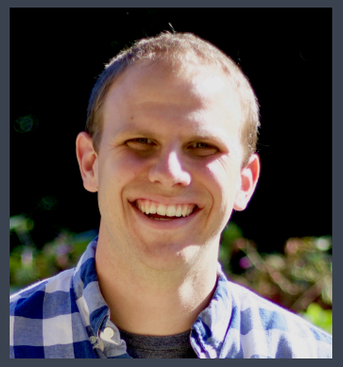


Meeta Mistry
Associate Director

John Quackenbush
Faculty Advisor

Emma Berdan

Heather Wick



Will Gammerdinger

Noor Sohail

Upendra Bhattarai

James Billingsley

Zhu Zhuo

Maria Simoneau

Consulting

- Experimental design help
- 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

HARVARD
CATALYST
THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

HARVARD
MEDICAL SCHOOL

Training

A key component of the H
researchers at Harvard b

HBC's training team is m
training and community b
research projects to ensu

Our hands-on workshops
an emphasis on **experim**
for **wet-lab biologists** ai
data.

We offer three types of w

1. Short, 3-hour monthly
2. Basic Data Skills**
3. Advanced Topics: Ana

**The basic data skills

Our dedicated training team holds workshop to help
or NGS data.

o devote substantial time to material development,
raining team also participate in consultations on
best practices in NGS analysis.

ysis of high-throughput sequencing data, with
and **reproducibility**. Our workshops are designed
cing-based experiments and analysing the resulting

matics)

NGS) data**

r the advanced workshops.

CFAR
HARVARD UNIVERSITY
CENTER FOR AIDS RESEARCH

HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

DF/HCC
DANA-FARBER / HARVARD CANCER CENTER

HARVARD
CATALYST
THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

HARVARD
MEDICAL SCHOOL

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

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

Training

A key component of the H
researchers at Harvard b

HBC's training team is m
training and community b
research projects to ensu

Our hands-on workshops
an emphasis on **experim**
for **wet-lab biologists** ai
data.

We offer three types of w

1. Short, 3-hour monthly
2. Basic Data Skills**
3. Advanced Topics: Ana

**The basic data skills

Our dedicated training team holds workshop to help
or NGS data.

o devote substantial time to material development,
raining team also participate in consultations on
best practices in NGS analysis.

ysis of high-throughput sequencing data, with
and **reproducibility**. Our workshops are designed
cing-based experiments and analysing the resulting

matics)

NGS) data**

r the advanced workshops.

CFAR
HARVARD UNIVERSITY
CENTER FOR AIDS RESEARCH

HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

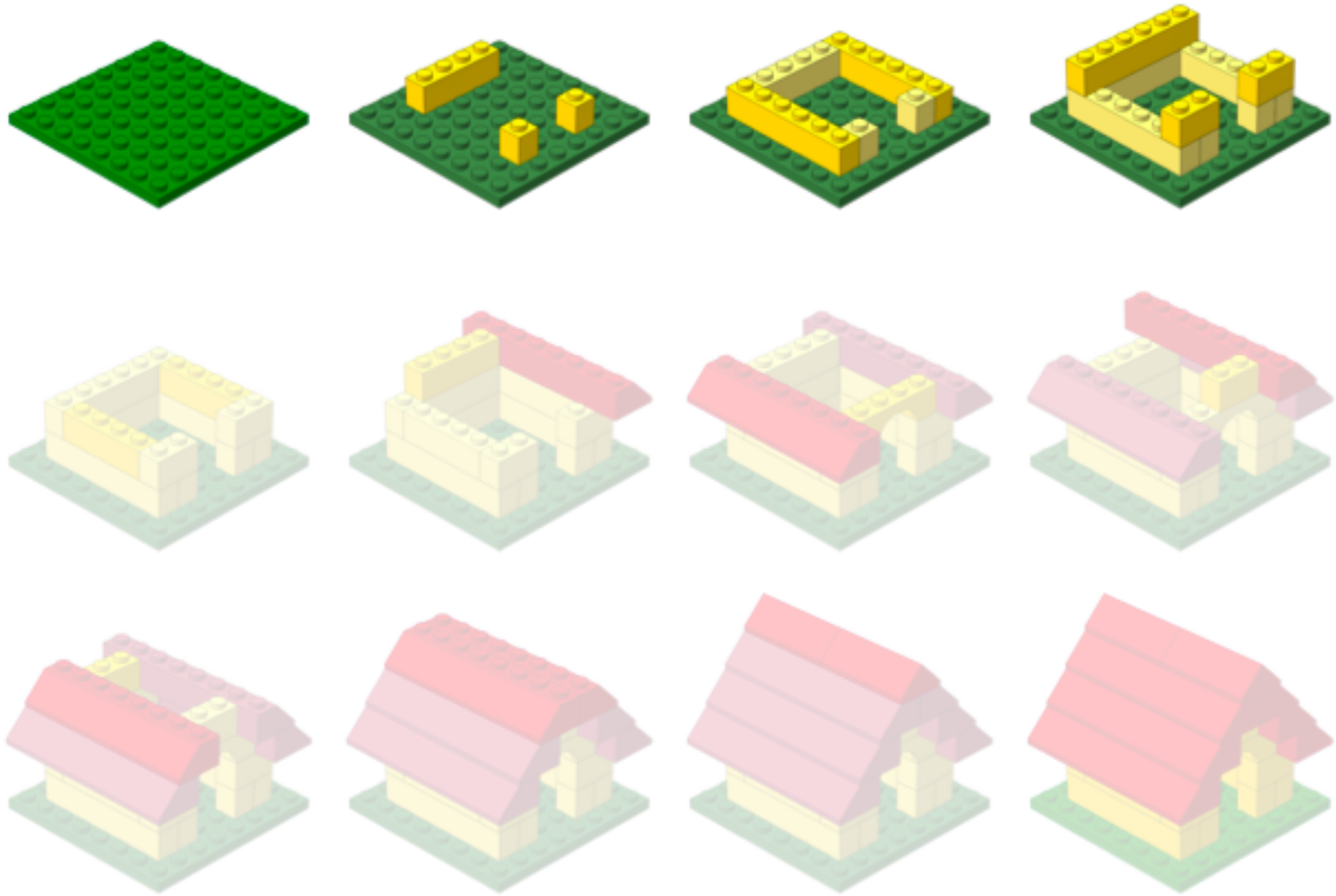
DF/HCC
DANA-FARBER / HARVARD CANCER CENTER

HARVARD
CATALYST
THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

HARVARD
MEDICAL SCHOOL

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

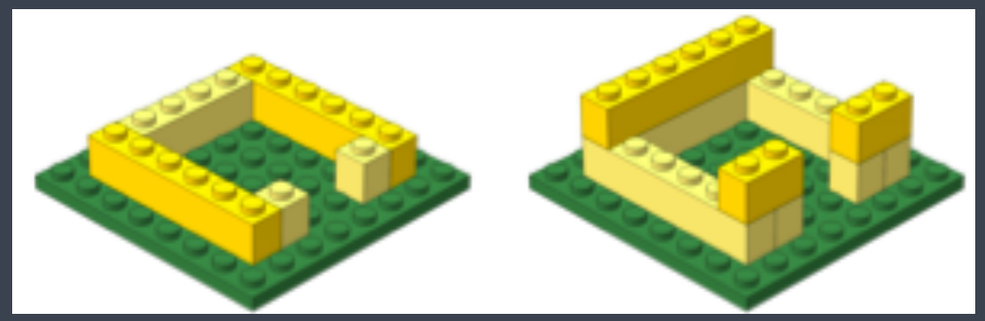
https://hbctraining.github.io/main/



http://anoved.net/tag/lego/page/3/

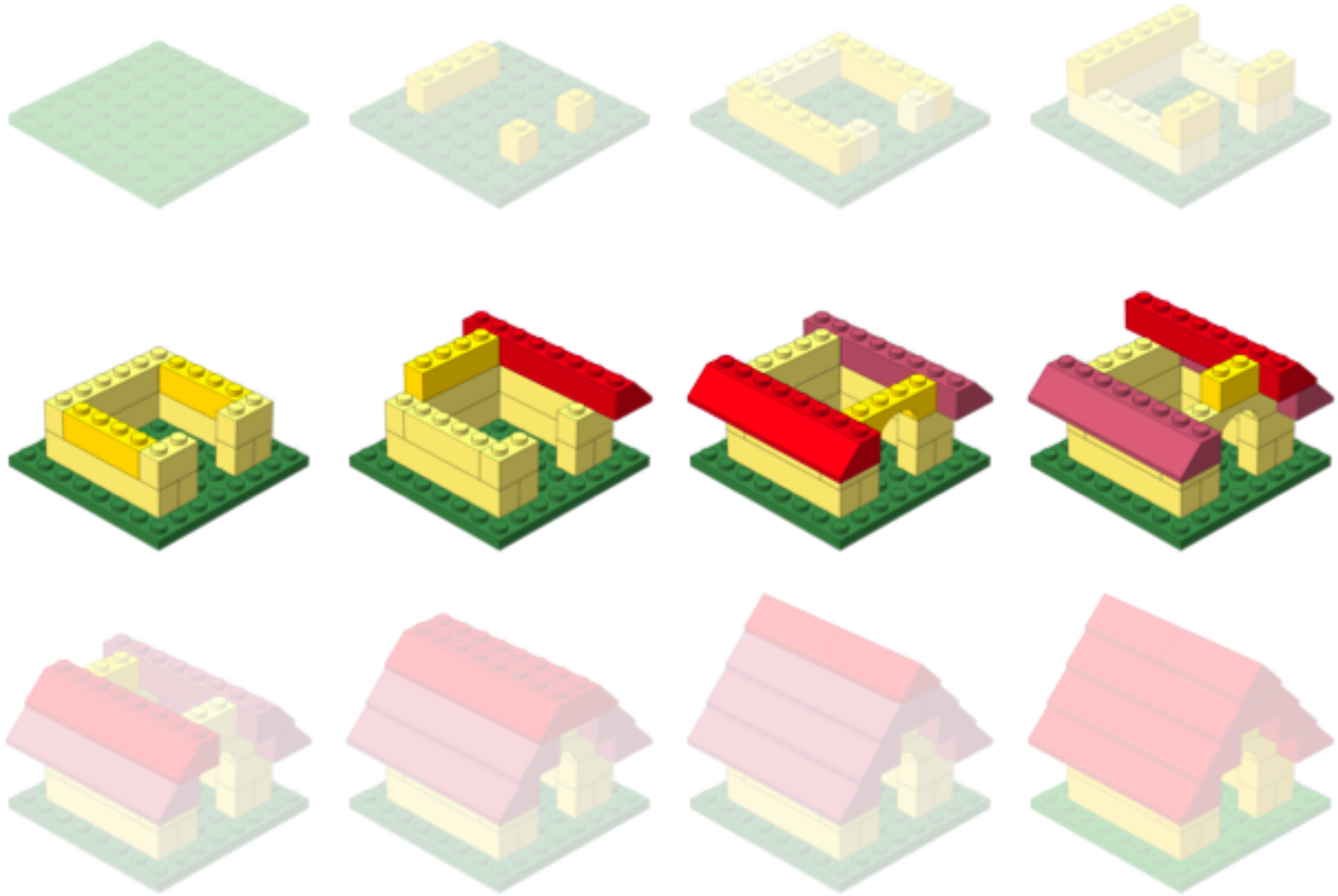
Setting up to perform Bioinformatics analysis

Setting up…



- ✓ Introduction to the command-line interface (shell, Unix, Linux)
  - Dealing with large data files
  - Performing bioinformatics analysis
    - Using tools
    - Accessing and using compute clusters
- ✓ R
  - Parsing and working with smaller results text files
  - Statistical analysis, e.g. differential expression analysis
  - Generating figures from complex data

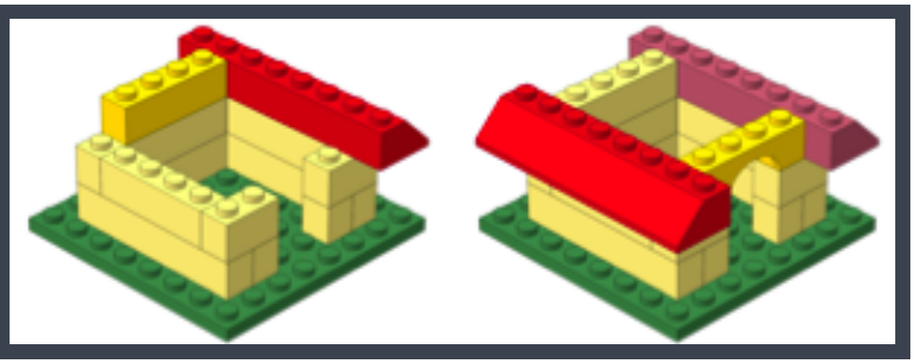
Workshop scope



http://anoved.net/tag/lego/page/3/

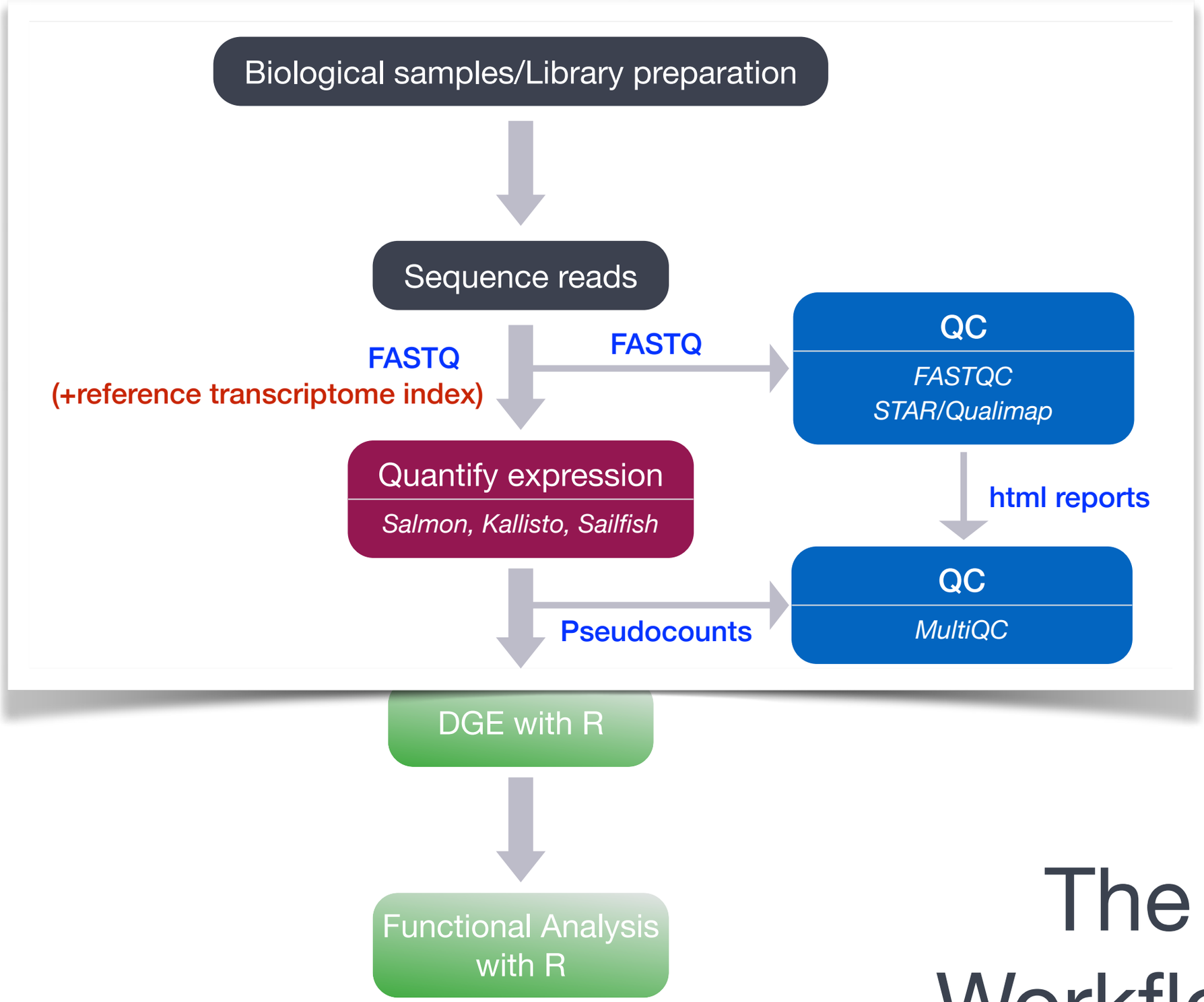
Bioinformatics data analysis

Learning Objectives



- ✓ Describe best practices for designing a bulk RNA-seq experiment
- ✓ Describe steps in an RNA-seq analysis workflow (from sequence data to expression quantification).
- ✓ Implement shell scripts on a high-performance compute cluster to perform the above steps.

We won't be covering how to perform differential gene expression (DGE) analysis on count data in this workshop.



The Workflow

Logistics

Course webpage

https://tinyurl.com/hbc-rnaseq

Course schedule online

Workshop Schedule

> **NOTE:** The *Basic Data Skills* Introduction to the command-line interface workshop is a prerequisite.

Pre-reading

- Shell basics review
- Introduction to RNA-seq

Day 1

Time	Topic	Instructor
09:30 - 09:45	Workshop introduction	Radhika
09:45 - 10:25	Working in an HPC environment	Radhika
10:25 - 11:05	Project Organization and Best Practices in Data Management	Meeta
11:05 - 11:45	Quality Control of Sequence Data: Running FASTQC	Jihe
11:45 - 12:00	Overview of self-learning materials and homework submission	Jihe/Meeta

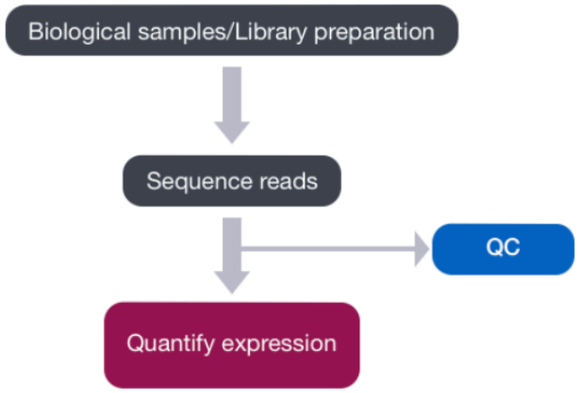
Course materials online

Learning Objectives:

- Understand the quality values in a FASTQ file
- Create a quality report using FASTQC

Quality Control of FASTQ files

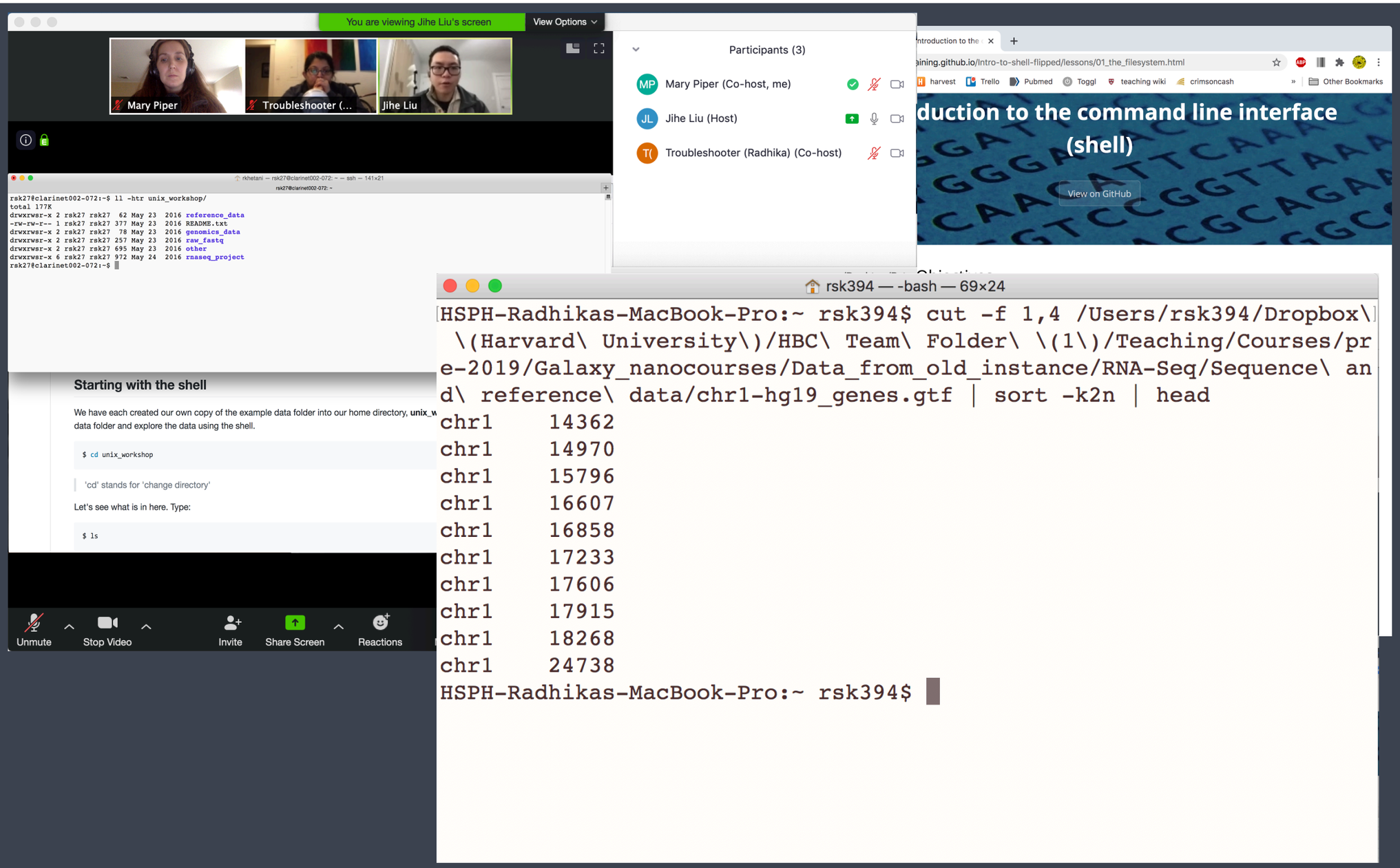
The first step in the RNA-Seq workflow is to take the FASTQ files received from the sequencing facility and assess the quality of the sequence reads.



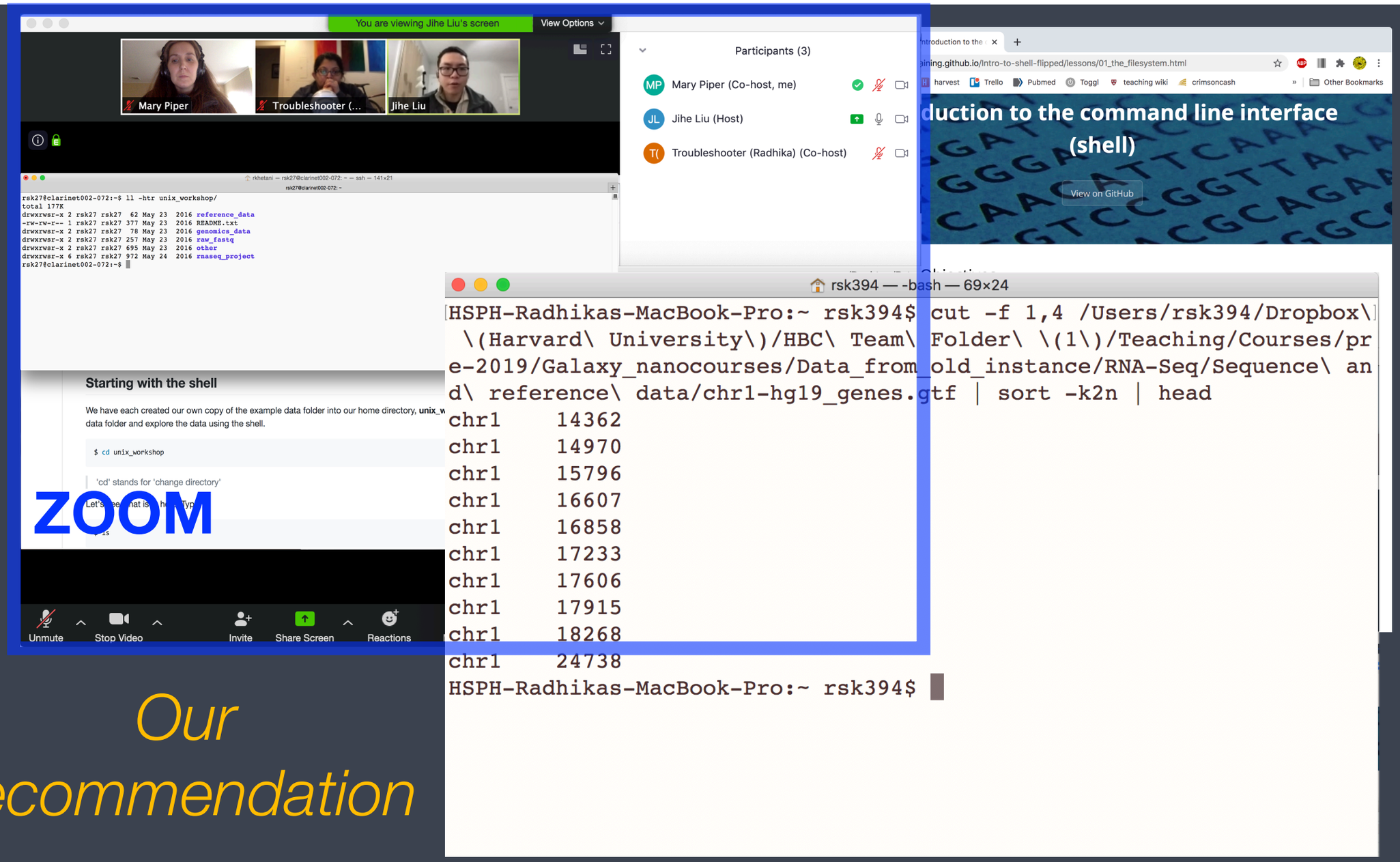
Course participation

- Mandatory review of self-learning lessons and assignments
- Attendance required for all classes
- Your questions and active participation drive learning
- We look forward to all of your questions!

Single screen & 3 windows?

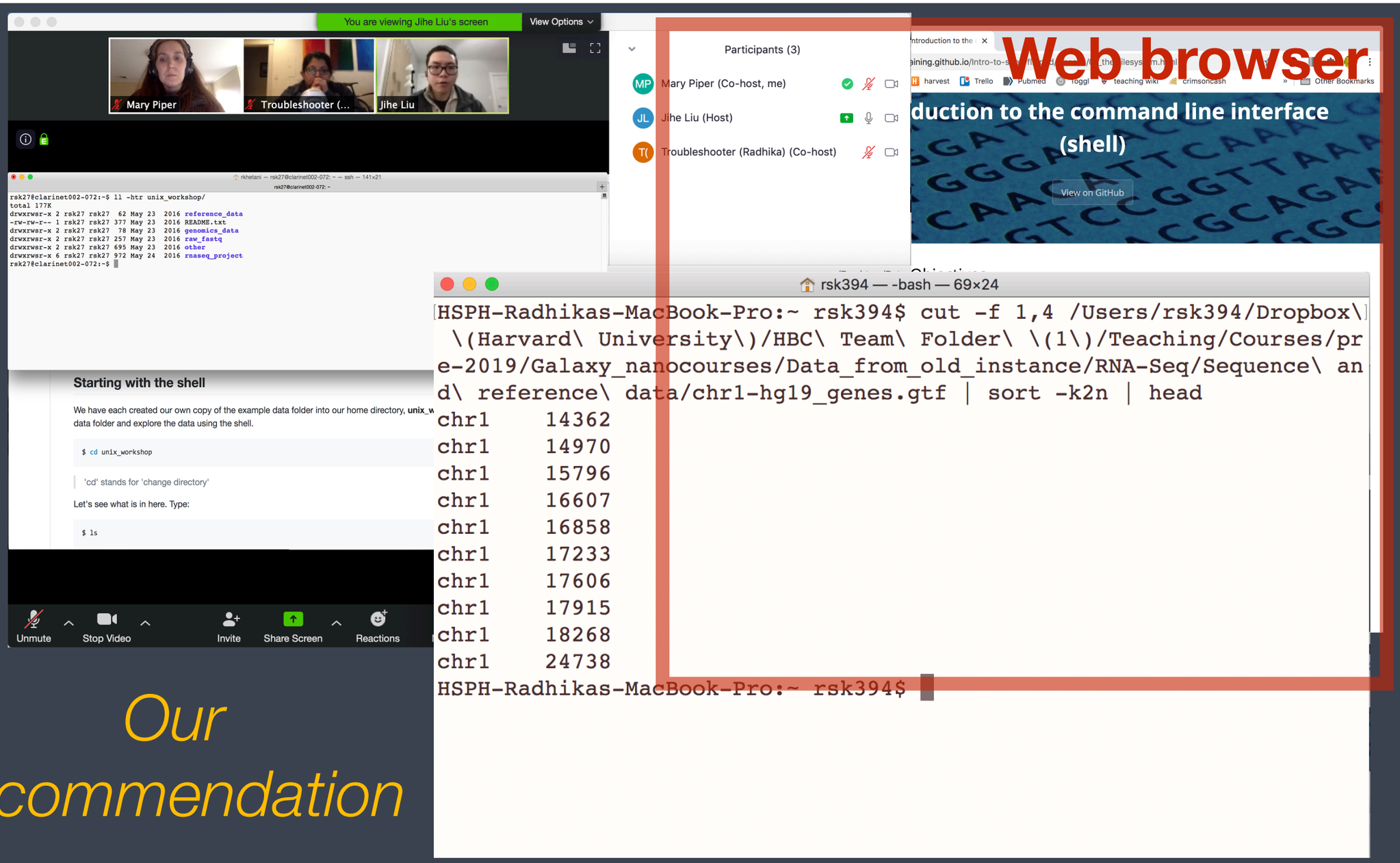


Single screen & 3 windows?



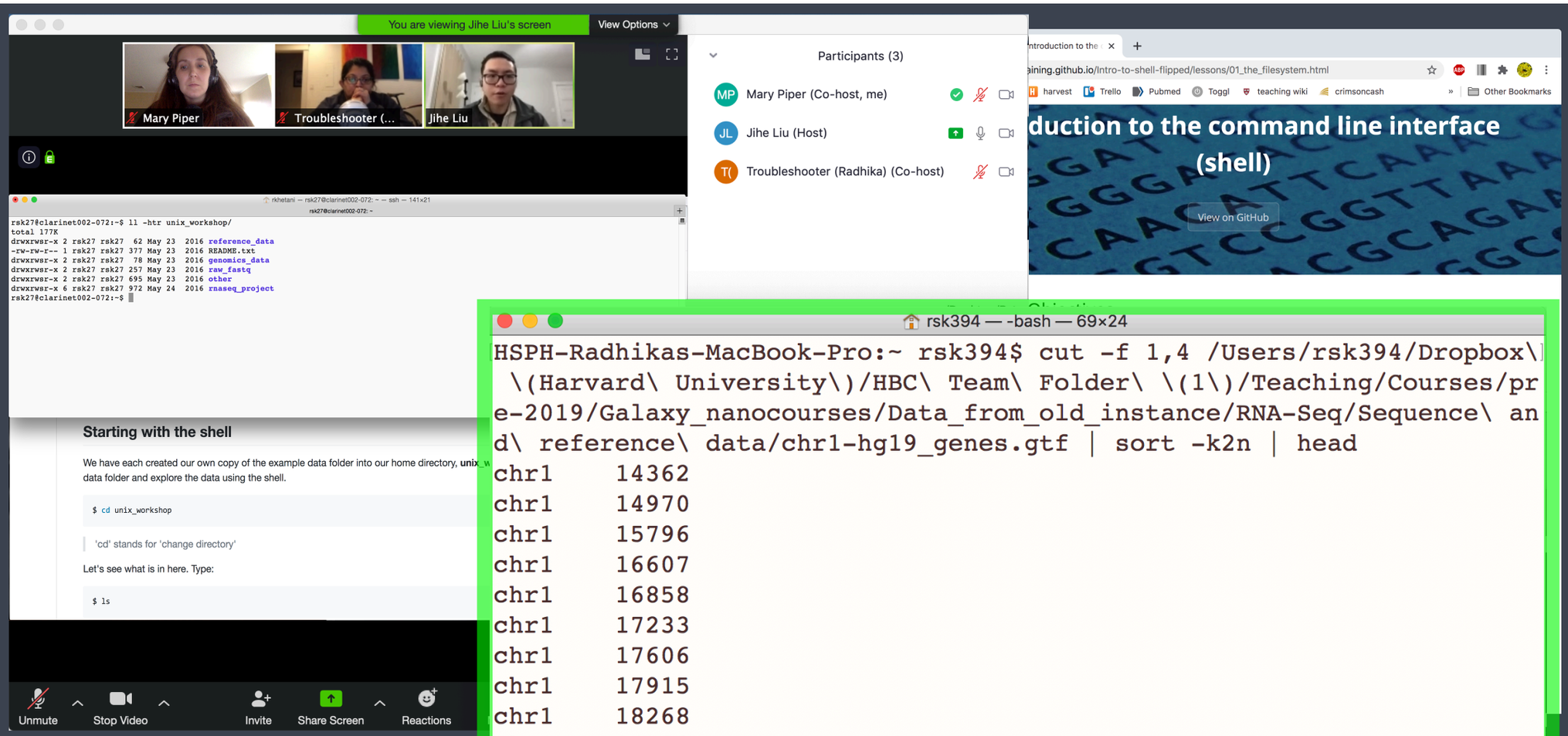
Our recommendation

Single screen & 3 windows?



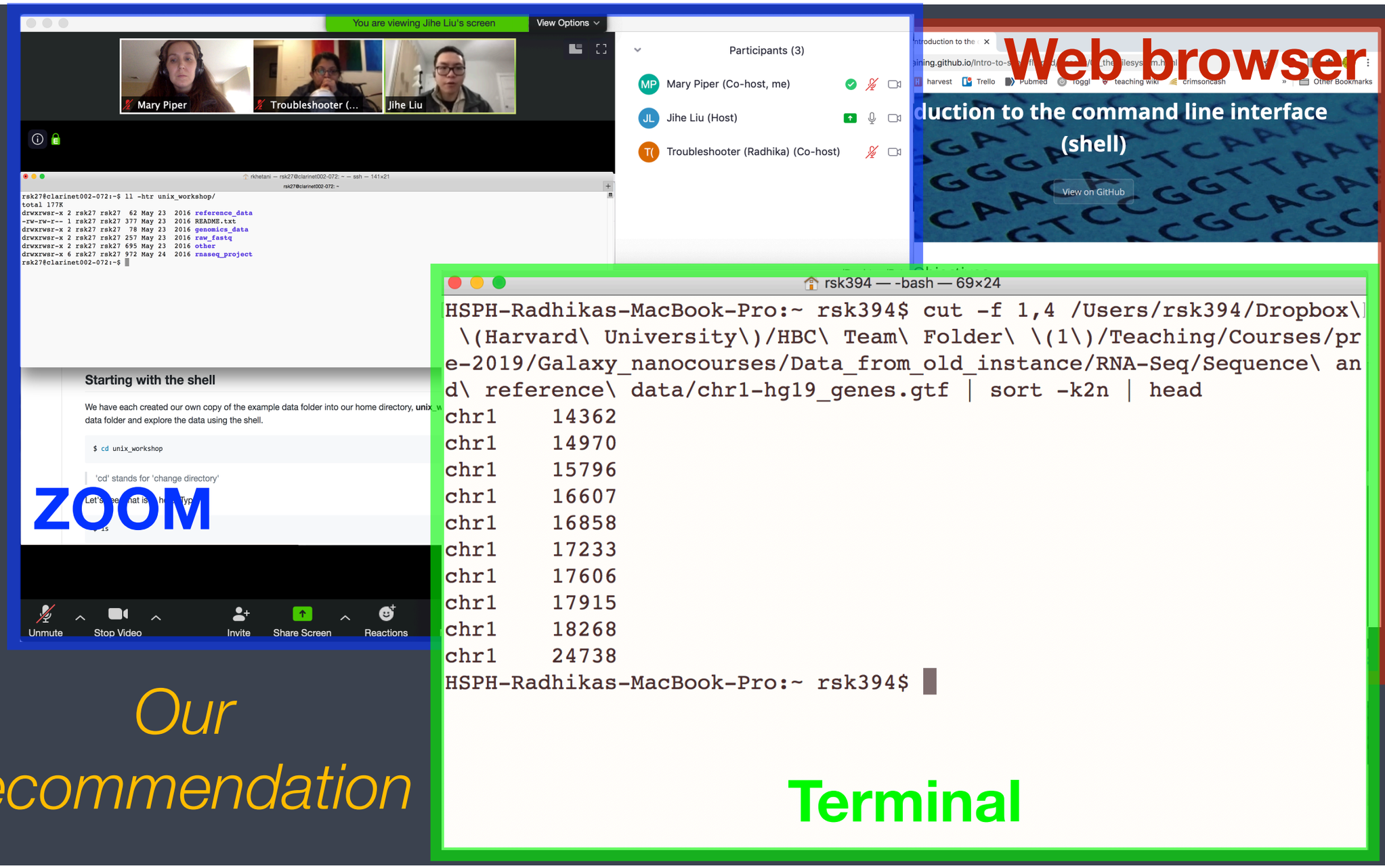
Our recommendation

Single screen & 3 windows?



Our recommendation

Single screen & 3 windows?



Our recommendation

Odds and Ends (1/2)

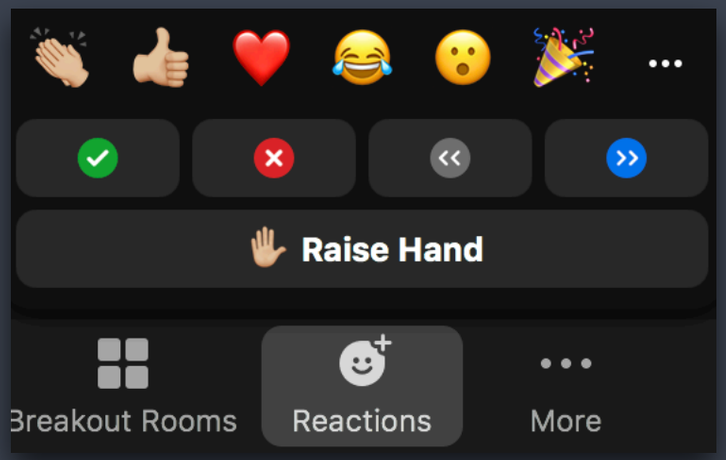
- Quit/minimize all applications that are not required for class

Odds and Ends (1/2)

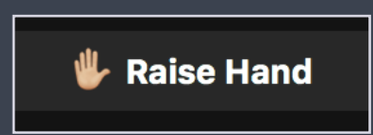
- Quit/minimize all applications that are not required for class
- Captioning is available upon request

Odds and Ends (1/2)

- Quit/minimize all applications that are not required for class
- Captioning is available upon request
- Are you all set?
  -  = "agree", "I'm all set" (equivalent to a green post-it)
  -  = "disagree", "I need help" (equivalent to a red post-it)

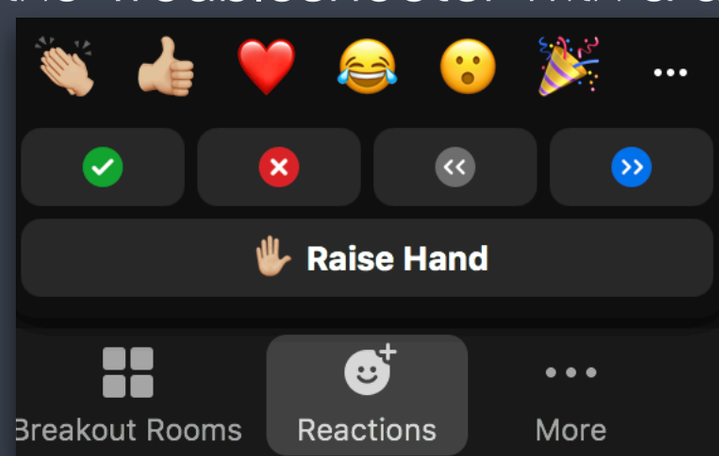


Odds and Ends (2/2)

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

when the presenter asks for questions
  - Let the **Moderator** know

- Technical difficulties with software?
  - Start a <u>private</u> chat with the **Troubleshooter** with a description of the problem.



Thanks!

- Andy Bergman & Kathleen Chappell 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