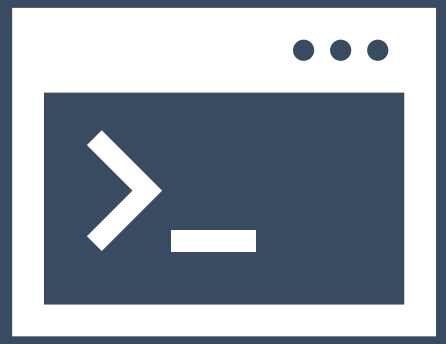
HBC
Harvard Chan Bioinformatics Core

Intro to Bulk RNA-seq (Part I)

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



Harvard Chan Bioinformatics Core

Introductions!

Shannan Ho Sui
Director

Meeta Mistry
Associate Director

Lorena Pantano
Director of Bioinformatics Platform

John Quackenbush
Faculty Advisor

Upen Bhattarai

Heather Wick

Will Gammerdinger

Noor Sohail

Alex Bartlett

Elizabeth Partan

Emma Berdan

James Billingsley

Zhu Zhuo

Maria Simoneau

Shannan Ho Sui
Director

Meeta Mistry
Associate Director

Lorena Pantano
Director of Bioinformatics Platform

John Quackenbush
Faculty Advisor

Upen Bhattarai

Heather Wick

Will Gammerdinger

Noor Sohail

Alex Bartlett

Elizabeth Bartan

Emma Berdan

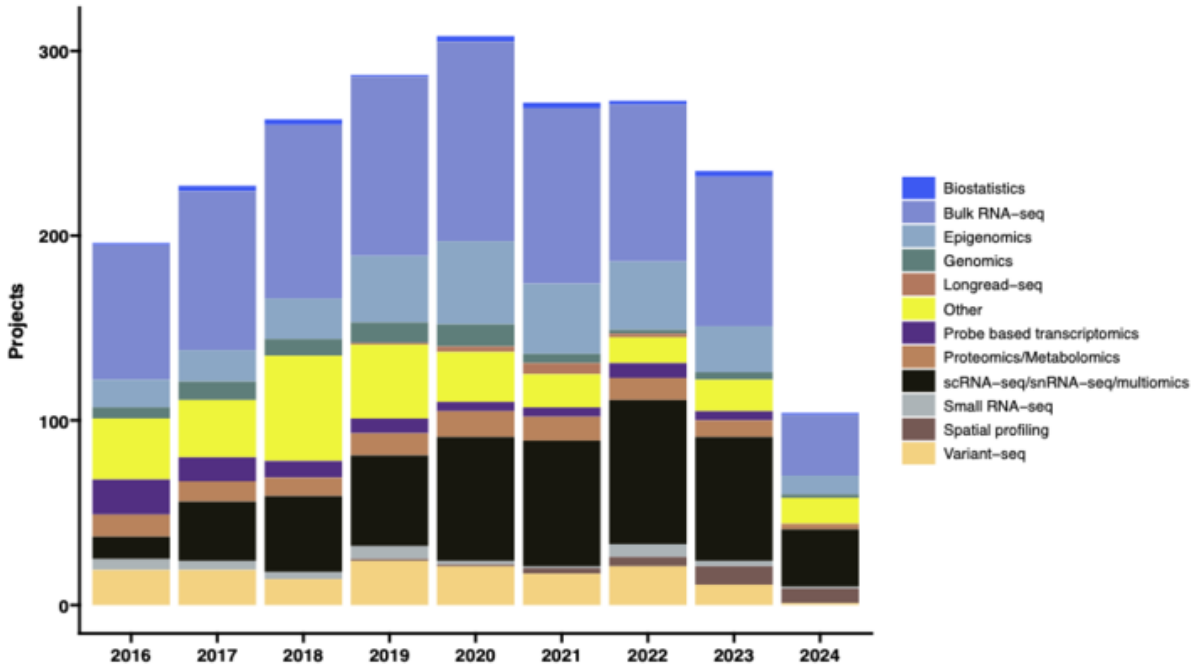
James Billingsley

Zhu Zhuo

Maria Simoneau

Consulting

- Transcriptomics: Bulk, single cell, small RNA
- Epigenomics: ChIP-seq, CUT&RUN, ATAC-seq, DNA methylation
- Variant discovery: WGS, resequencing, exome-seq and CNV
- Multiomics integration
- Spatial biology
- Experimental design and grant support



https://bioinformatics.sph.harvard.edu/services

Consulting

- Transcriptomics: Bulk, single cell, small RNA
- Epigenomics: ChIP-seq, CUT&RUN, ATAC-seq, DNA methylation
- Variant discovery: WGS, resequencing, exome-seq and CNV
- Multiomics integration
- Spatial biology
- Experimental design and grant support

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 HBC's mission is its training initiative. Our dedicated training team holds workshop to help researchers at Harvard better understand analytical methods for NGS data.

HBC's training team is made up of four PhD-level scientists who devote substantial time to material development, training and community building/outreach. All members of the training team also participate in consultations on research projects to ensure they remain up-to-date on current best practices in NGS analysis.

Our hands-on workshops focus on **basic data skills** and **analysis of high-throughput sequencing data**, with an emphasis on **experimental design**, current **best practices** and **reproducibility**. Our workshops are designed for **wet-lab biologists** aiming to independently design sequencing-based experiments and analysing the resulting data.

We offer three types of workshops:

1. Short, 3-hour monthly workshops (*Current topics in bioinformatics*)
2. Basic Data Skills**
3. Advanced Topics: Analysis of high-throughput sequencing (NGS) data**

***The basic data skills workshops serve as the foundation for the advanced workshops.*

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

Training

A key component of the HBC's mission is its training initiative. Our dedicated training team holds work
researchers at Harvard better understand analytical methods for NGS data.

HBC's training team is made up of four PhD-level scientists who devote substantial time to material de
training and community building/outreach. All members of the training team also participate in consult
research projects to ensure they remain up-to-date on current best practices in NGS analysis.

Our hands-on workshops focus on **basic data skills** and **analysis of high-throughput sequencing**
an emphasis on **experimental design**, current **best practices** and **reproducibility**. Our workshops a
for **wet-lab biologists** aiming to independently design sequencing-based experiments and analysing
data.

We offer three types of workshops:

1. Short, 3-hour monthly workshops (*Current topics in bioinformatics*)
2. Basic Data Skills**
3. Advanced Topics: Analysis of high-throughput sequencing (NGS) data**

***The basic data skills workshops serve as the foundation for the advanced workshops.*

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

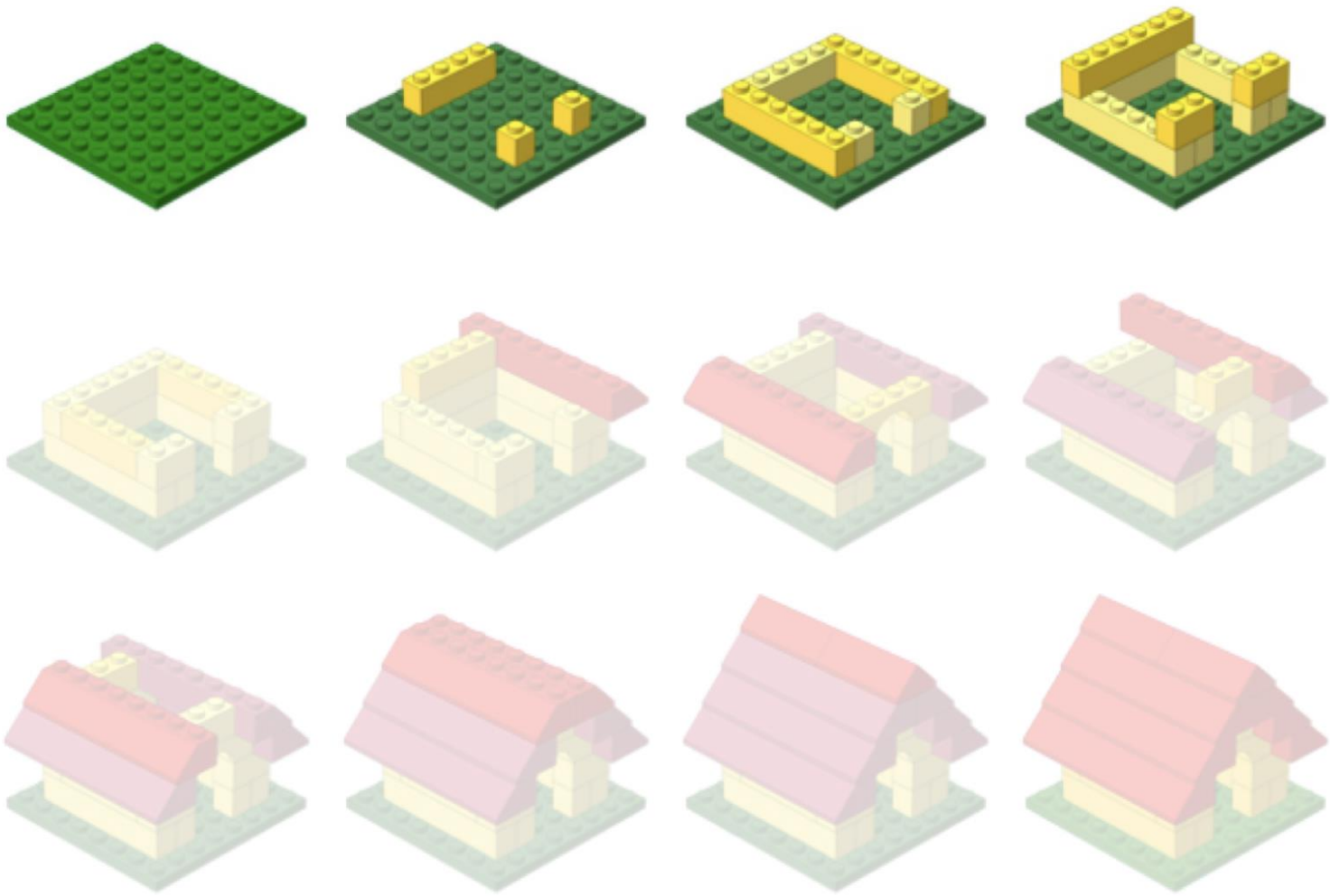
CFAR
HARVARD UNIVERSITY
CENTER FOR AIDS RESEARCH

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

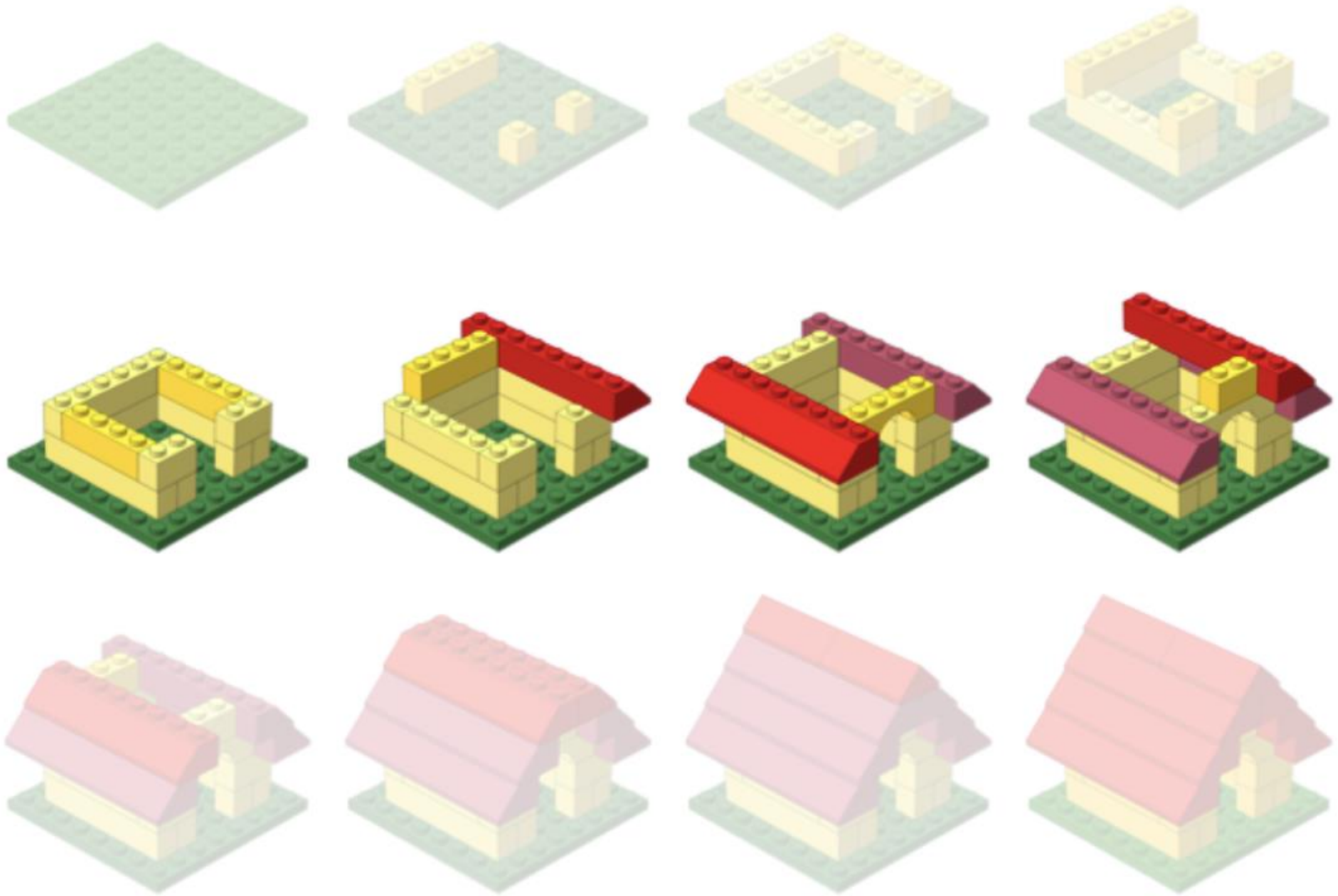


Learning Bioinformatics

Setting up

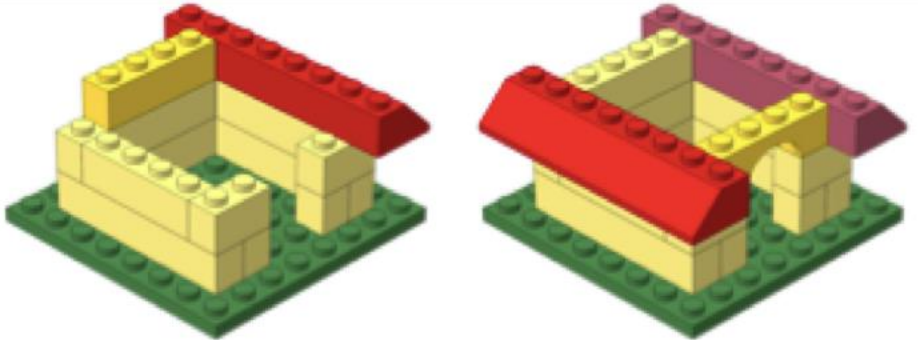
- Shell for Bioinformatics
  - Dealing with large data files
  - Performing Bioinformatic Analyses
    - Using tools
    - Accessing and using computer clusters
- Introduction to R
  - Parsing and working with smaller results text files
  - Statistical analyses, e.g. differential expression analysis
  - Generating figures from complex data

Workshop scope



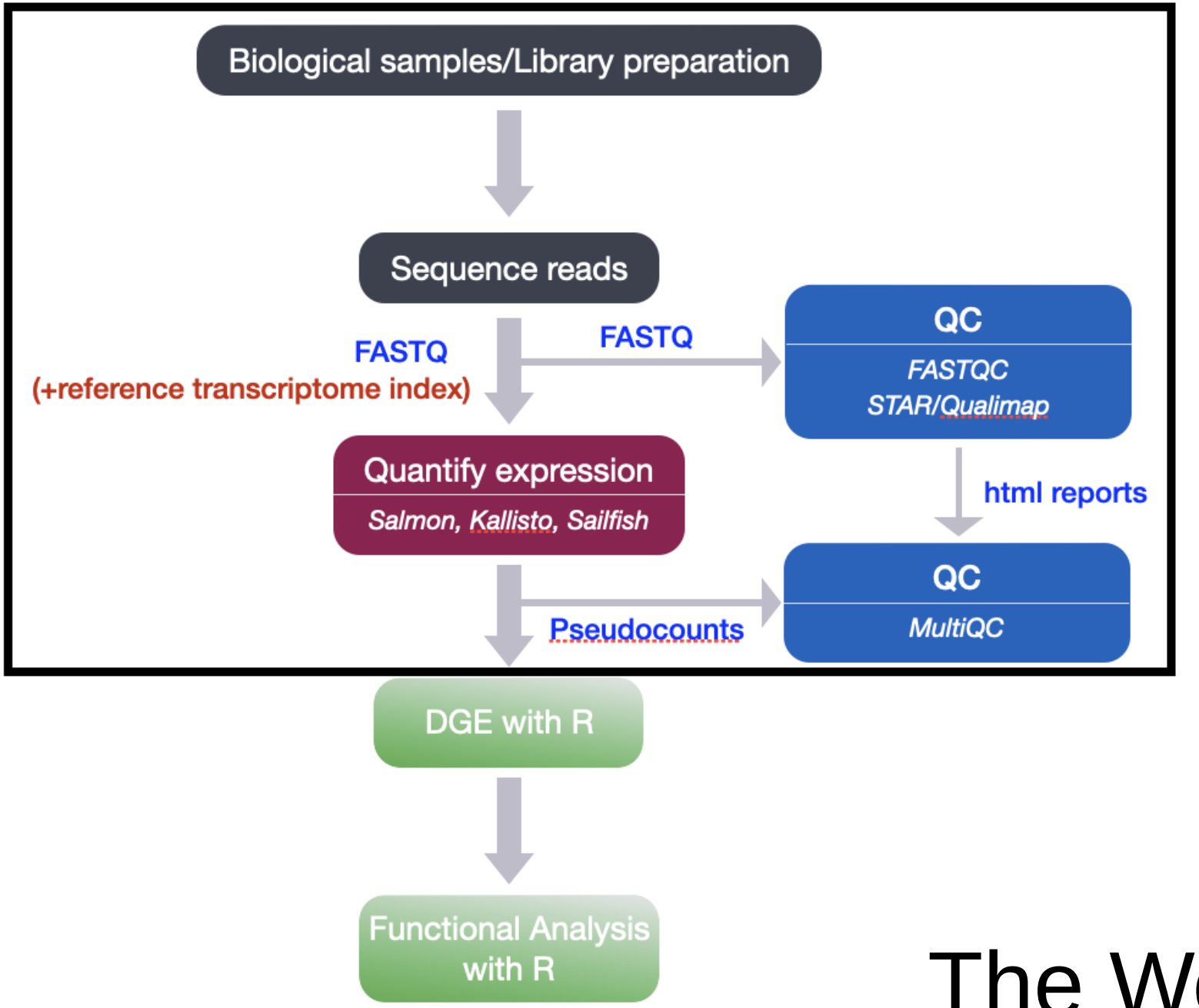
Bioinformatics Data Analysis

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

Day 1

Time	Topic	Instructor
09:30 - 09:45	Workshop Introduction	Will
09:45 - 10:25	Working in an HPC environment - Review	Upen
10:25 - 11:05	Project Organization (using Data Management best practices)	Will
11:05 - 11:45	Quality Control of Sequence Data: Running FASTQC	Upen
11:45 - 12:00	Overview of self-learning materials and homework submission	Will

Before the next class:

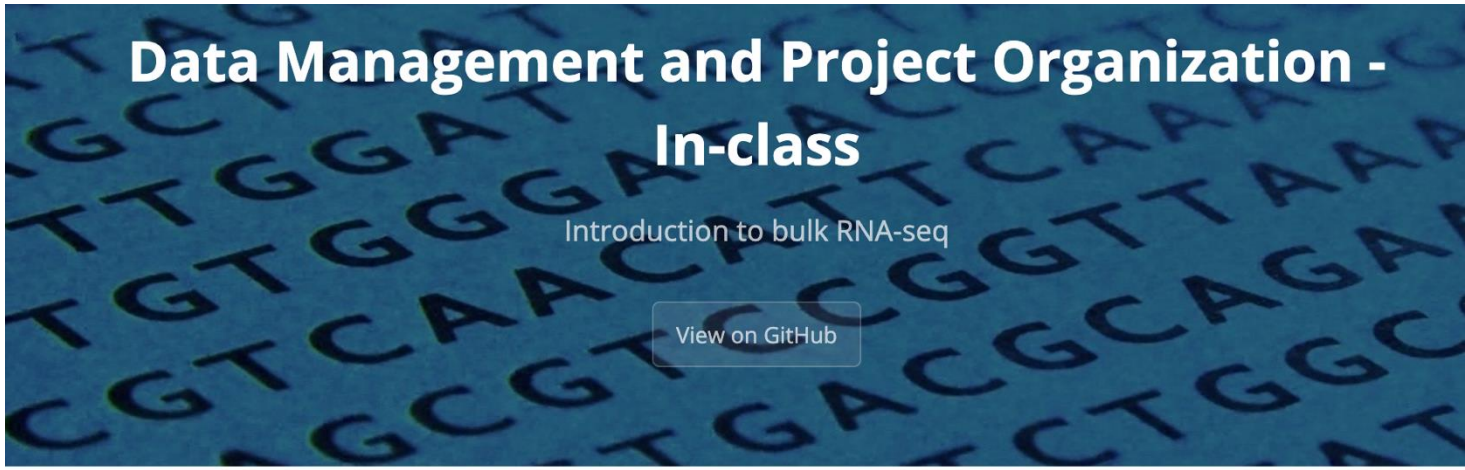
1. Please **study the contents** and **work through all the code** within the following lessons:

- Experimental design considerations
- Quality Control of Sequence Data: Running FASTQC on multiple samples
- Quality Control of Sequence Data: Evaluating FASTQC reports

https://tinyurl.com/hbc-rnaseq

Course materials

- We continuously update our materials to reflect changes in the field/software



Learning Objectives

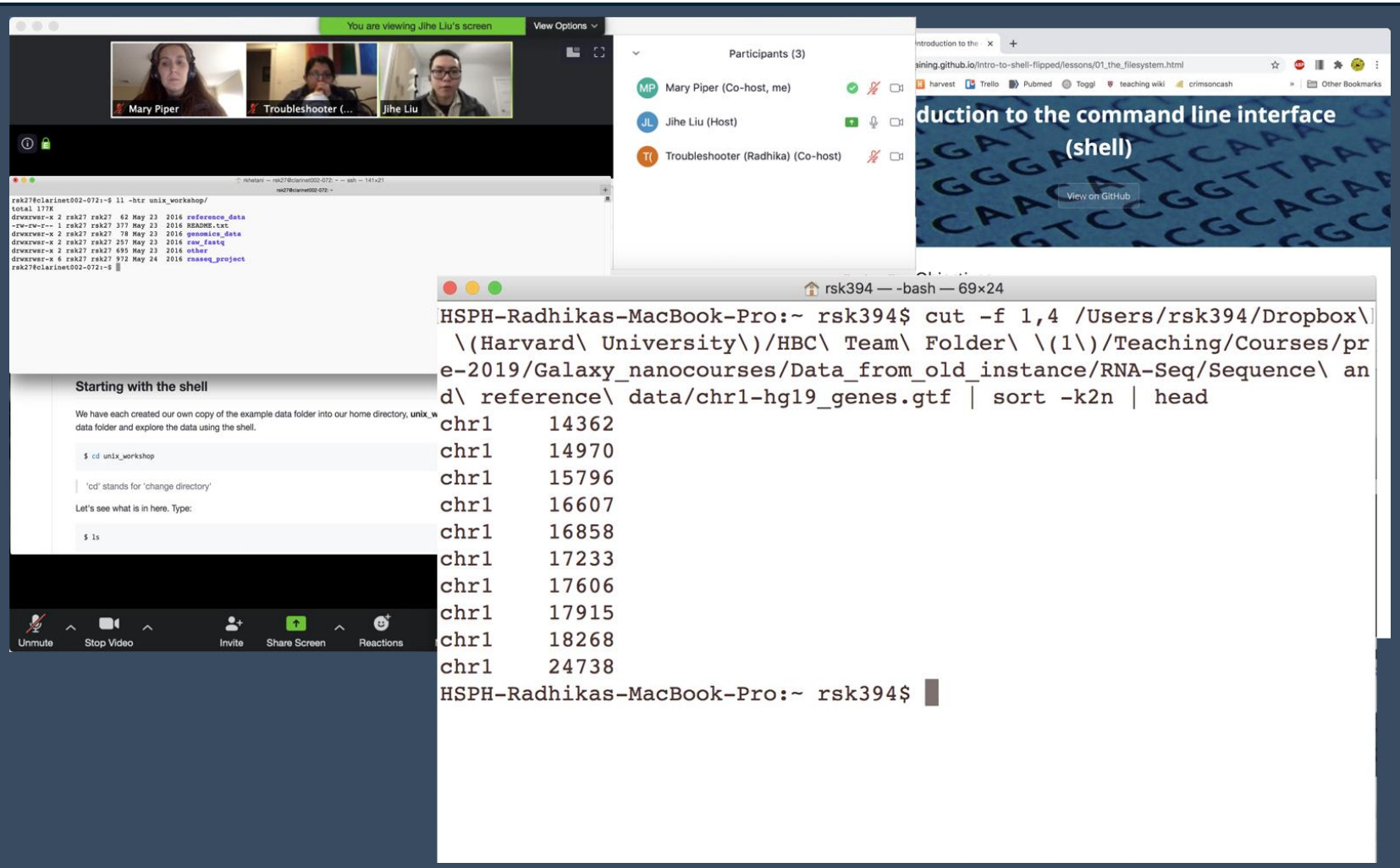
- Describe the example RNA-seq experiment and its objectives.
- Demonstrate strategies for good data management and project organization.

The Dataset

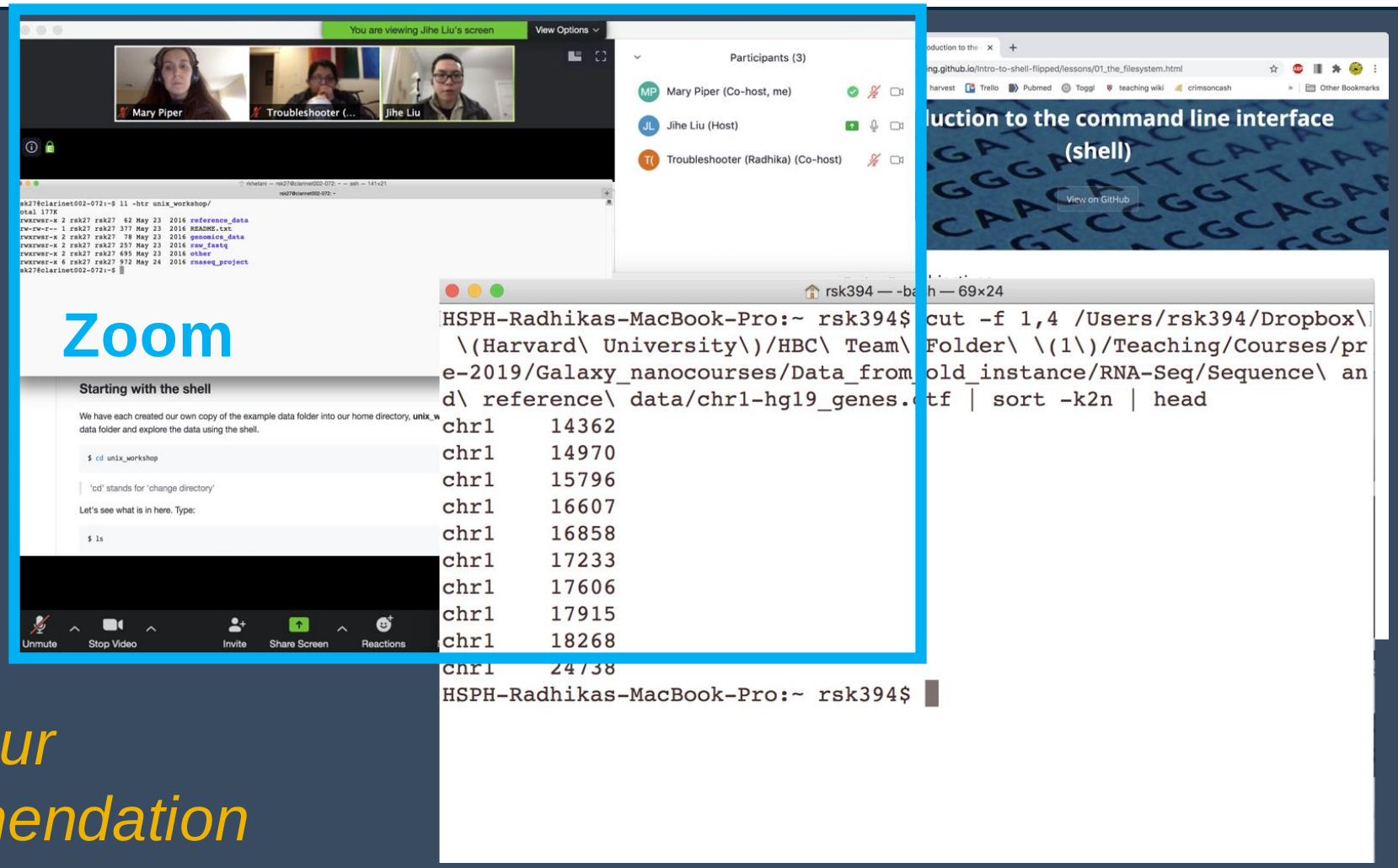
The dataset we are using for this workshop is part of a larger study described in Kenny PJ et al., *Cell Rep* 2014. The authors are investigating interactions between various genes involved in Fragile X syndrome, a disease of aberrant protein production, which results in cognitive impairment and autistic-like features. **The authors sought to show that RNA helicase MOV10 regulates the translation of RNAs involved**

https://tinyurl.com/hbc-rnaseq

Single Screen & 3 Windows

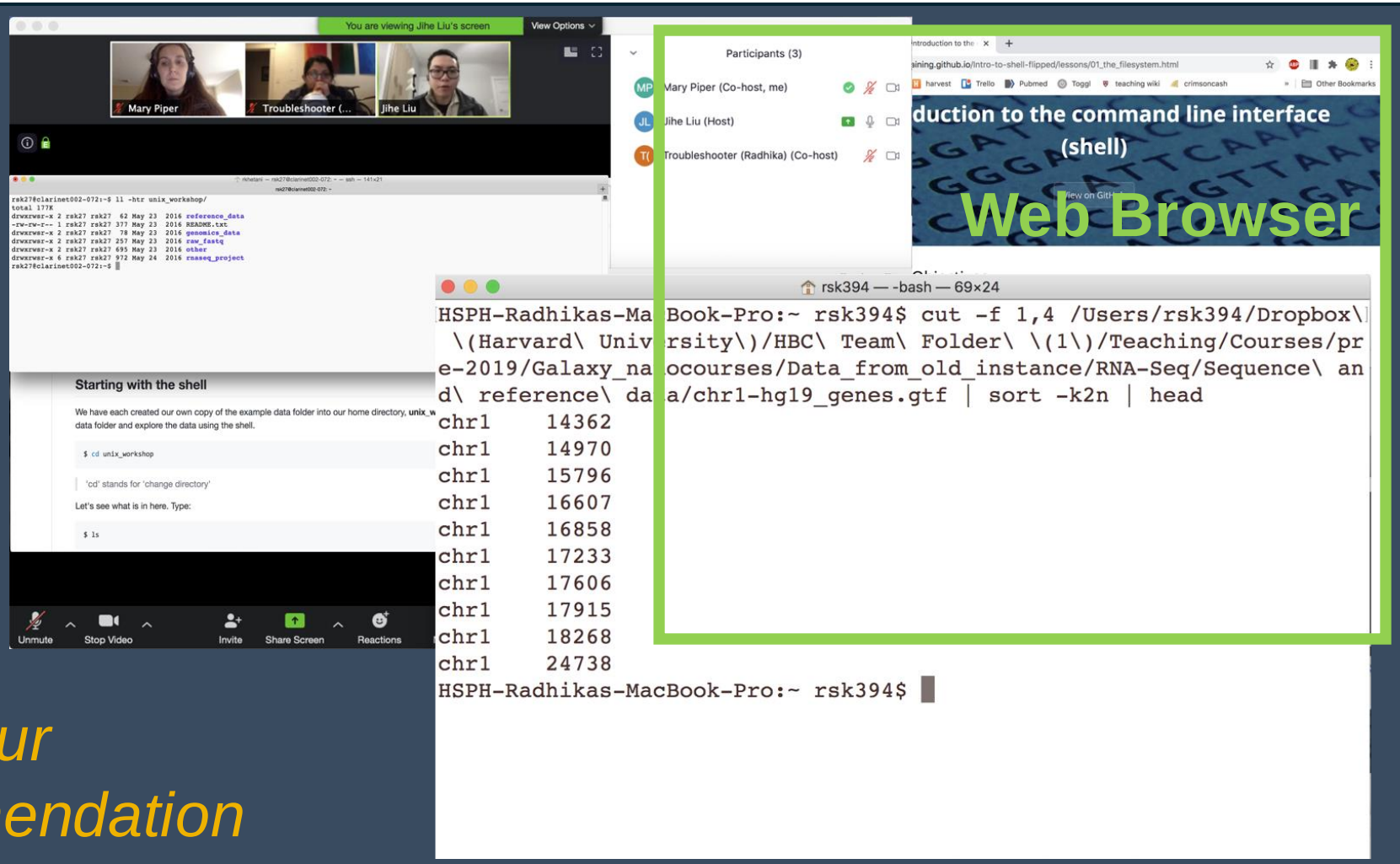


Single Screen & 3 Windows



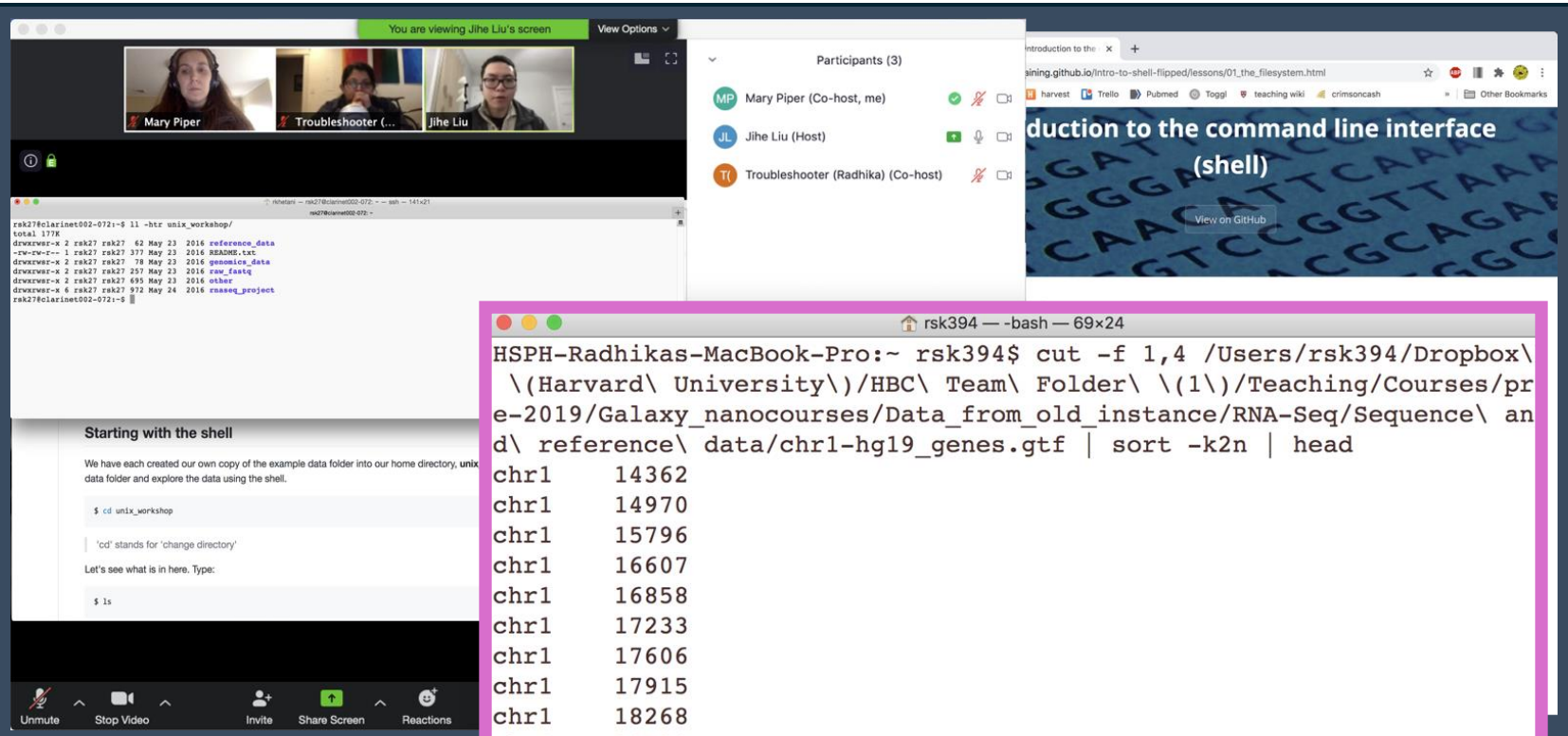
Our Recommendation

Single Screen & 3 Windows



Our Recommendation

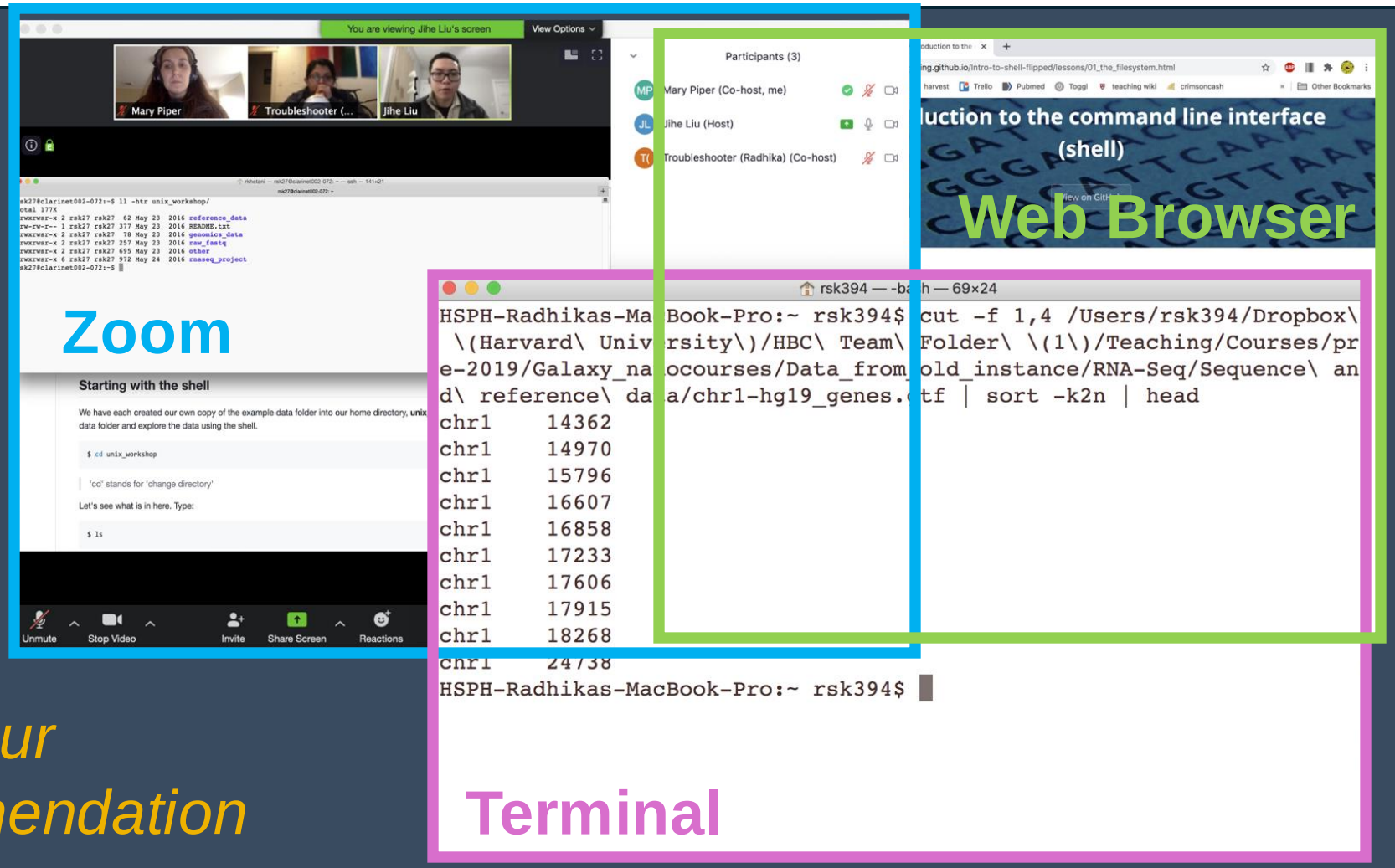
Single Screen & 3 Windows



Our Recommendation

Terminal

Single Screen & 3 Windows



Our Recommendation

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!**

Course participation

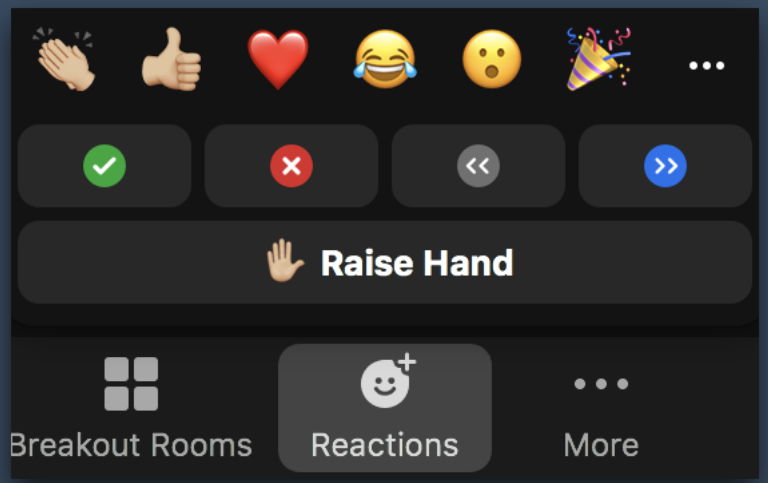
- At-home lessons and exercises after each session
- Cover material not previously discussed
- Provides us feedback to help pace the course appropriately
- 3-5 hours to complete
- Homework load is heavier in the beginning of this workshop series and tapers off

Using AI for Assignments

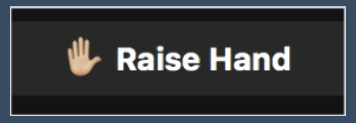
- Do
  - Try to resolve error messages with it
  - Test code written by AI on a dataset where you have expected results
  - Take the time to review the generated code line-by-line
- Don’t
  - Implement it in replacement to learning
  - Write code that you don’t understand
  - Assume the output from an AI process is correct

Odds & Ends

- Quit/minimize all applications that are not required for class
- Are you all set?
  - = "agree", "I'm all set"
  - = "disagree", "I need help"

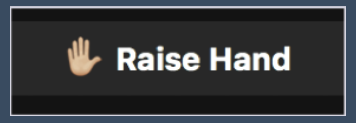


Odds & Ends

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

 when the presenter asks for questions
  - Let the Troubleshooter know

Odds & Ends

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

 when the presenter asks for questions
  - Let the Troubleshooter know
- Technical difficulties with software?
  - Start a private chat with the Troubleshooter with a description of the problem

Thanks!

- Kathleen Chappell and Andy Bergman from HMS-RC
- Data Carpentry

These materials have been developed by members of the teaching team at the Harvard Chan Bioinformatics Core (HBC). These are open access materials distributed under the terms of the Creative Commons Attribution license (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Contact Us

HBC
Harvard Chan Bioinformatics Core

- *HBC training team:* hbctraining@hsph.harvard.edu
- *HBC consulting:* bioinformatics@hsph.harvard.edu
- *O2 (HMS-RC):* rchelp@hms.harvard.edu