

Introduction to ChIP-seq (Part I)

Harvard Chan Bioinformatics Core

in collaboration with

HMS Research Computing

<https://tinyurl.com/hbc-chipseq>



Shannan Ho Sui
Director



Victor Barrera



Amelie Jule



Zhu Zhus



Radhika Khetani
Training Director



Meeta Mistry



Jihe Liu



Will Gammeltinger



Maria Simoneau



James Billingsley



Sergey Naumenko



Peter Kraft
Faculty Advisor

Consulting

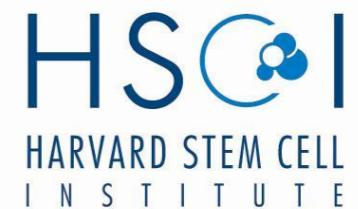
- RNA-seq analysis: bulk, single cell, small RNA
- ChIP-seq and ATAC-seq analysis
- Genome-wide methylation
- WGS, resequencing, exome-seq and CNV studies
- QC & analysis of gene expression arrays
- Functional enrichment analysis
- Grant support

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



**HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH**

NIEHS



Training

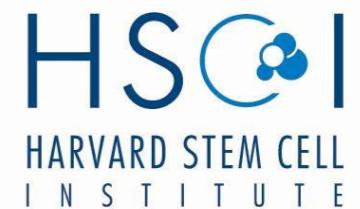
We have divided our short workshops into 2 categories:

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/>





Shannan Ho Sui
Director



Victor Barrera



Amelie Jule



Zhu Zuo



Radhika Khetani
Training Director



Meeta Mistry



Jihe Liu



Will Gammeltinger



Maria Simoneau



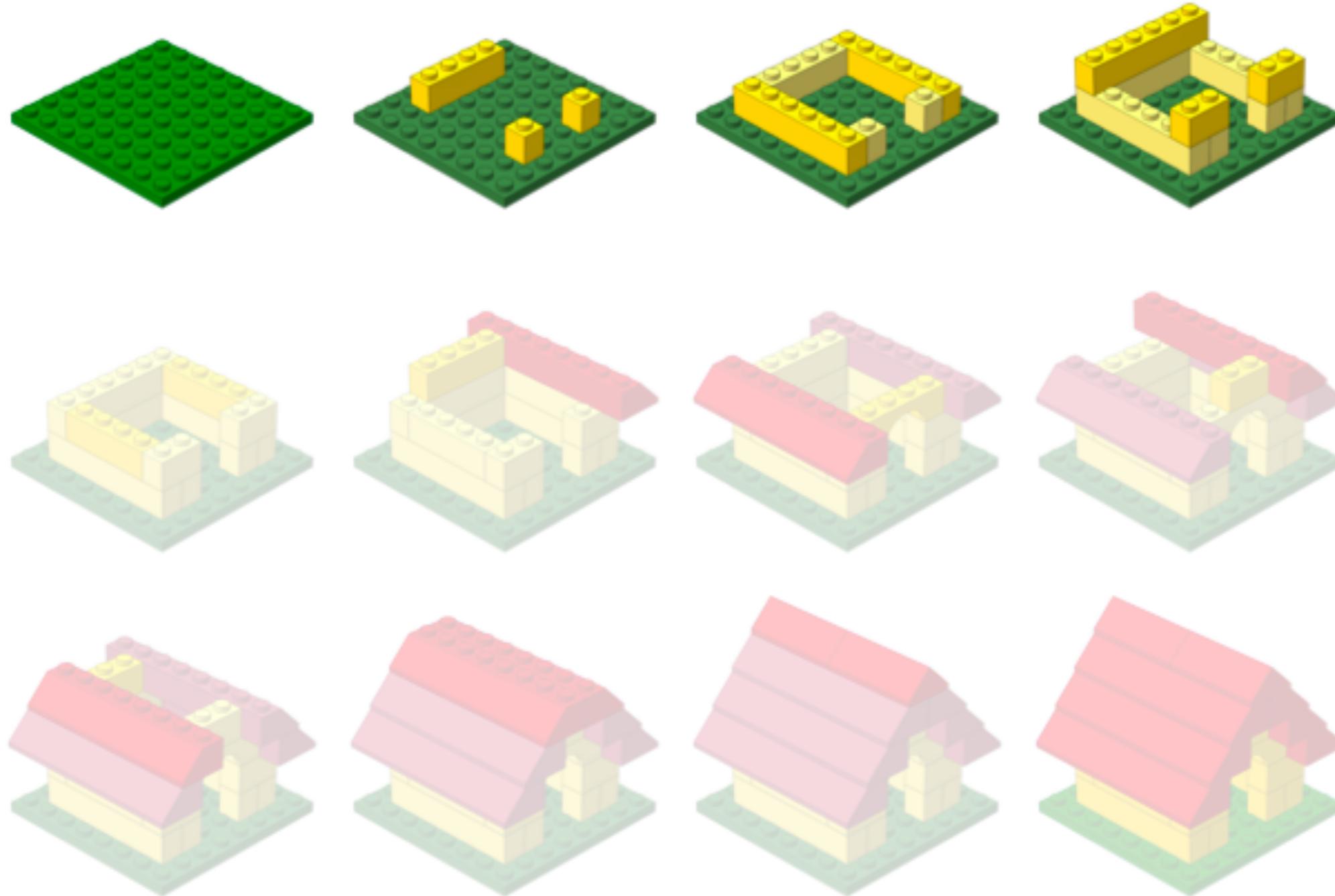
James Billingsley



Sergey Naumenko



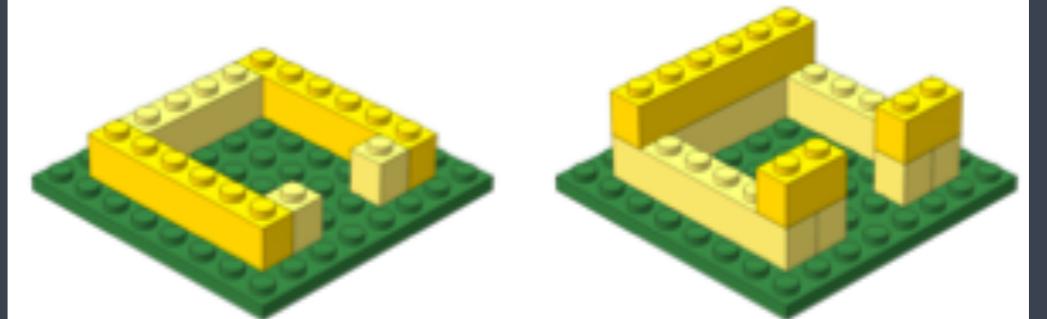
Peter Kraft
Faculty Advisor



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

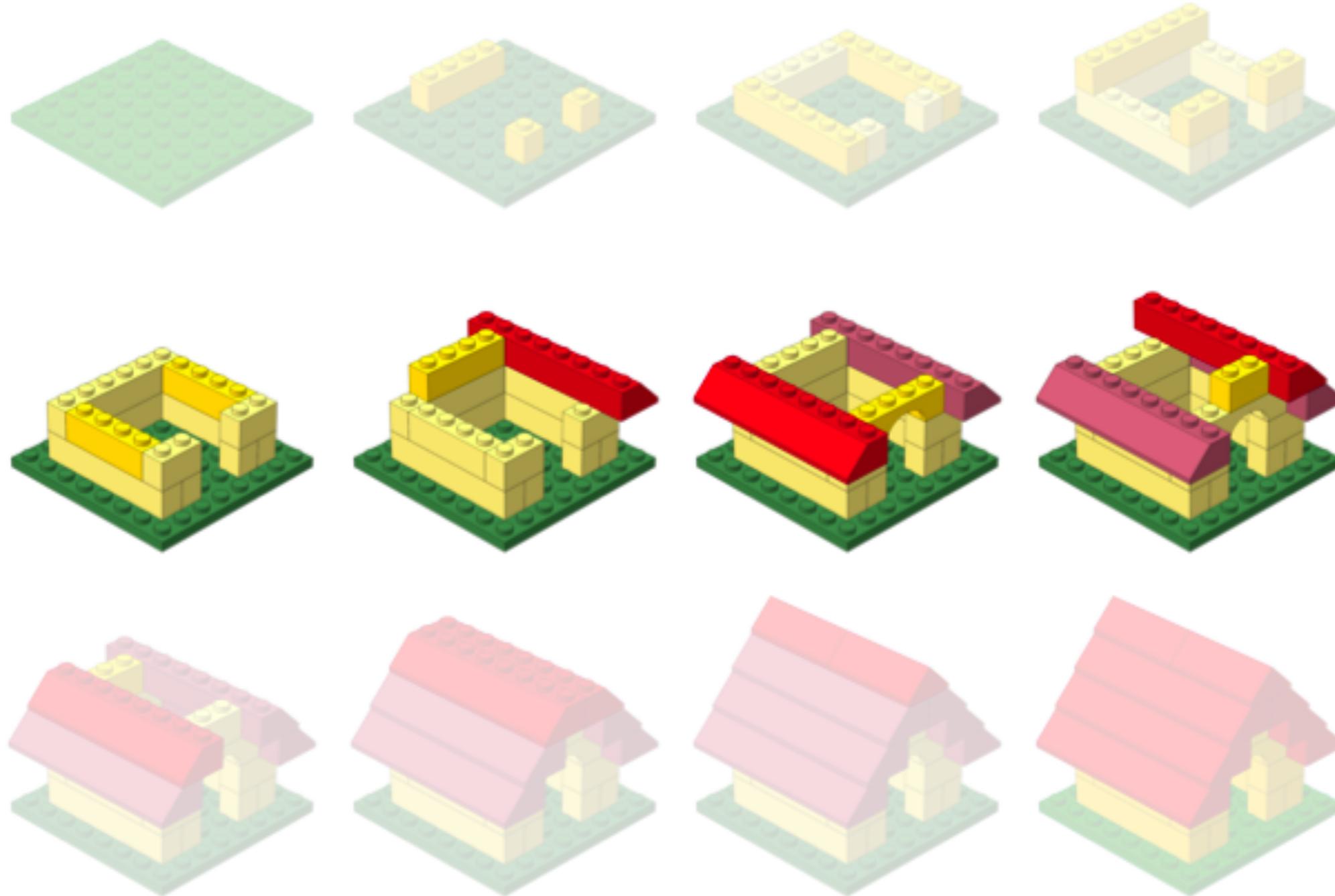
Setting up to perform Bioinformatics analysis

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 BED files
 - Statistical analysis, e.g. differential binding analysis
 - Generating figures from complex data

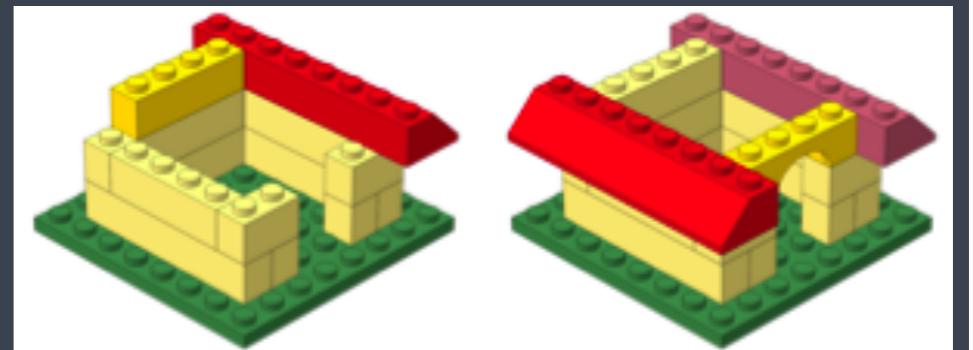
Workshop scope



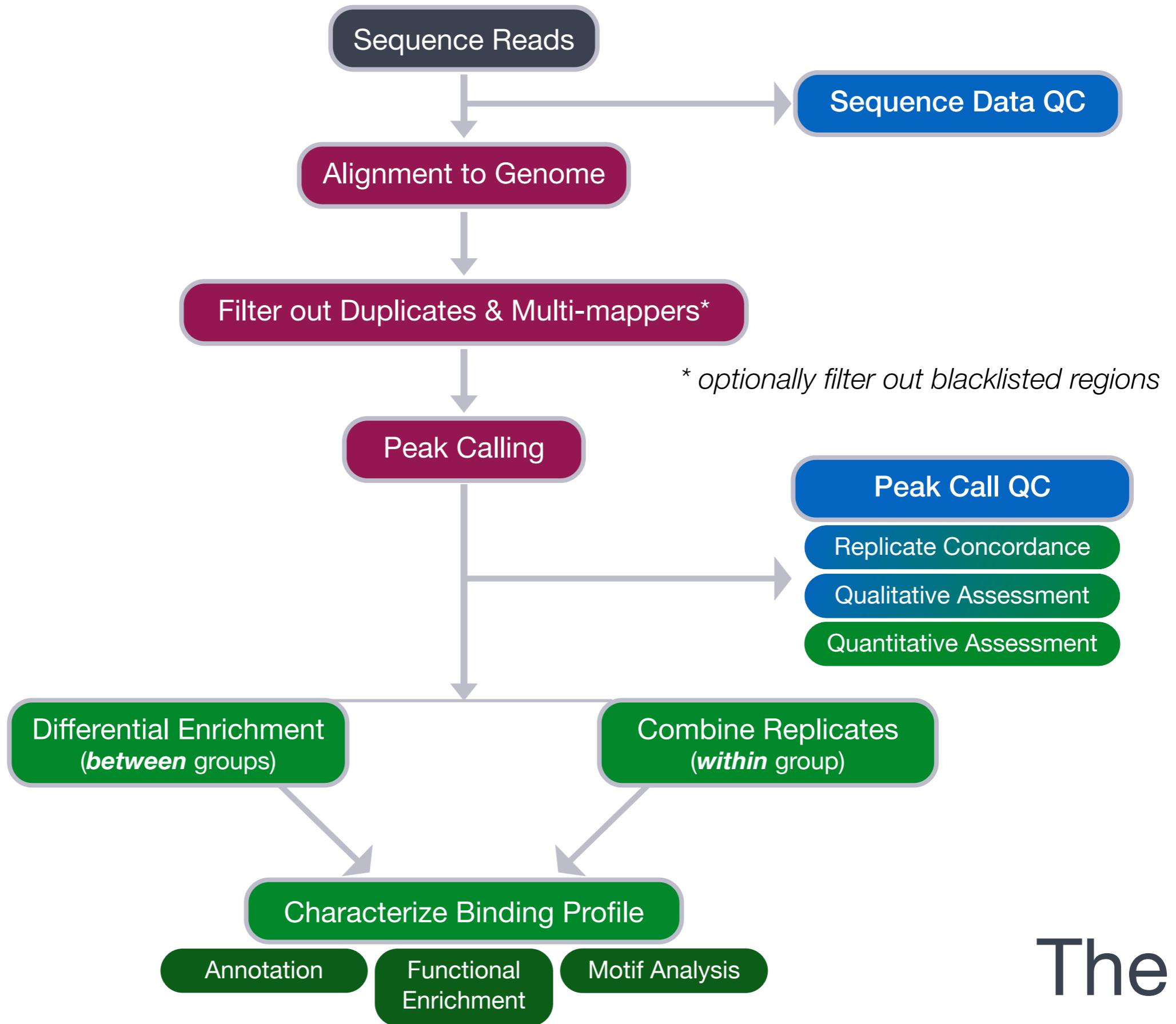
<http://anoved.net/tag/lego/page/3/>

Bioinformatics data analysis

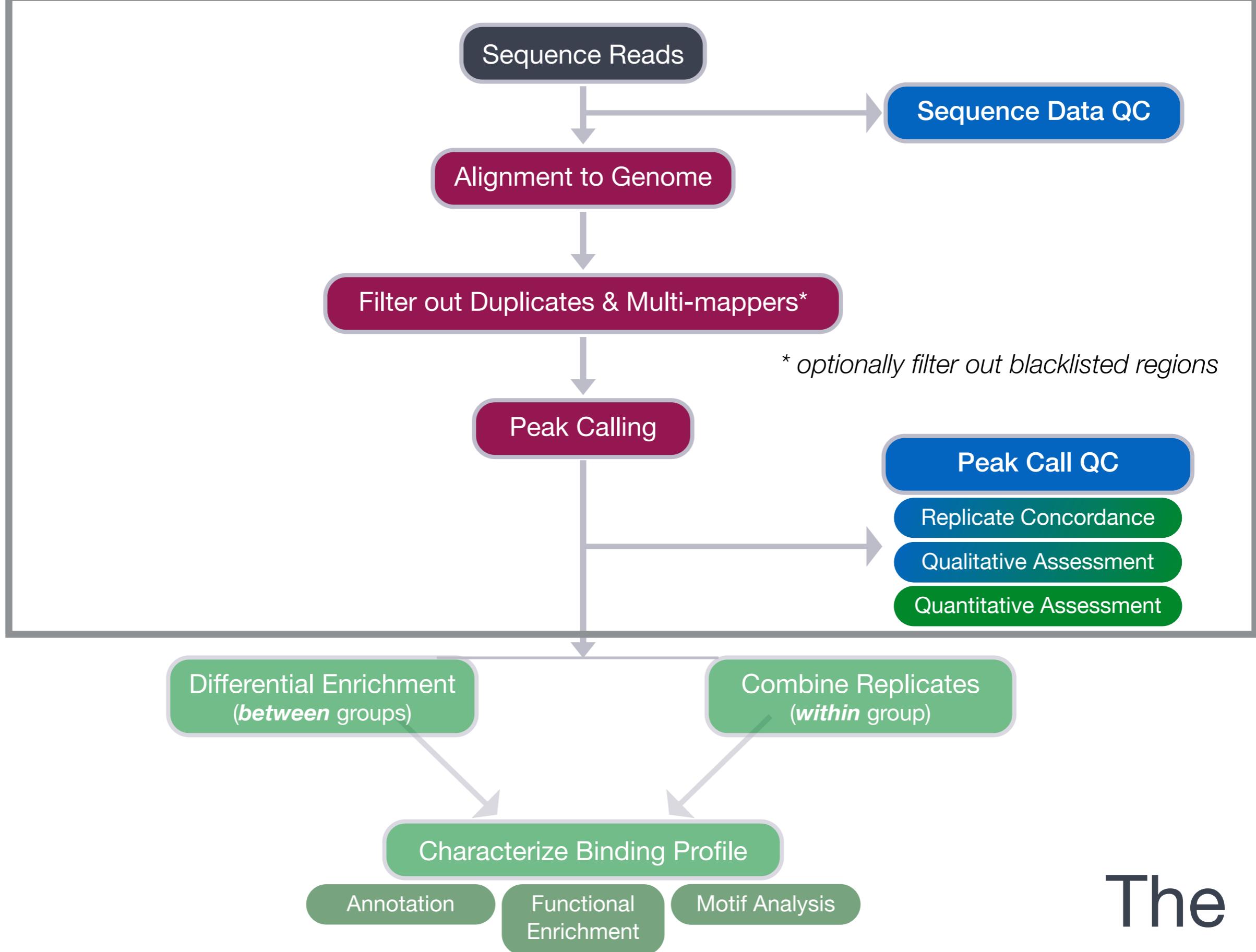
Learning Objectives



- ✓ Describe best practices for designing a good ChIP-seq experiment
- ✓ Describe steps in an ChIP-seq analysis workflow (from sequence data to peak calls)
- ✓ Learn how to handle various file formats encountered with ChIP-seq analysis
- ✓ Implement shell scripts on a high-performance compute cluster to perform the above steps



The Workflow



The Workflow

Boxes in green represent parts of the workflow that will be covered in ChIP-seq Part II

Logistics

Course webpage

<https://tinyurl.com/hbc-chipseq>

Course schedule online

Workshop Schedule

NOTE: The *Basic Data Skills* [Introduction to the command-line interface](#) workshop is a prerequisite.

Pre-reading:

- Please **study the contents** and **work through all the exercises** within the following lessons:
 - [Shell basics review](#)
 - [Best Practices in Research Data Management \(RDM\)](#)

Day 1

Time	Topic	Instructor
09:30 - 09:45	Workshop Introduction	Meeta
09:45 - 11:00	Introduction to ChIP-seq	Dr. Shannan Ho Sui
11:00- 11:05	Break	
11:05 - 11:50	Working in an HPC environment	Radhika

Course materials online



FastQC for quality assessment

[View on GitHub](#)

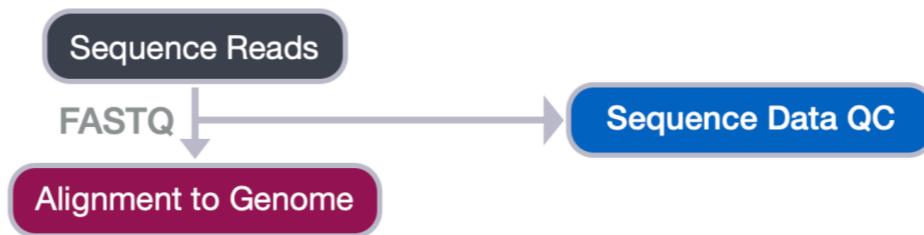
Contributors: Mary Piper, Radhika Khetani, Jihe Liu, Will Gammerdinger

Approximate time:

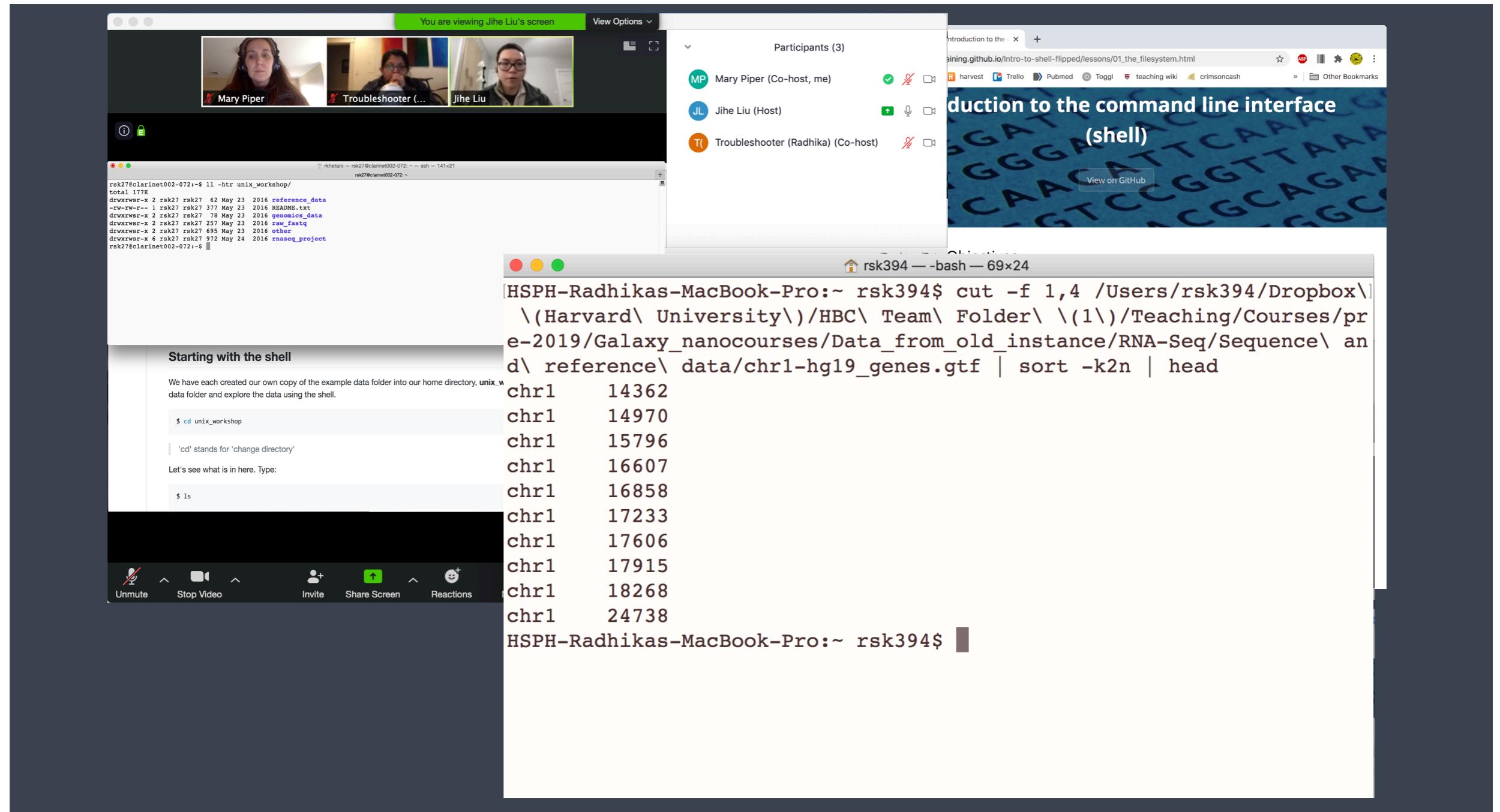
Learning Objectives

- Explain the components of a FASTQ file
- Evaluate the quality of the sequencing data using FastQC

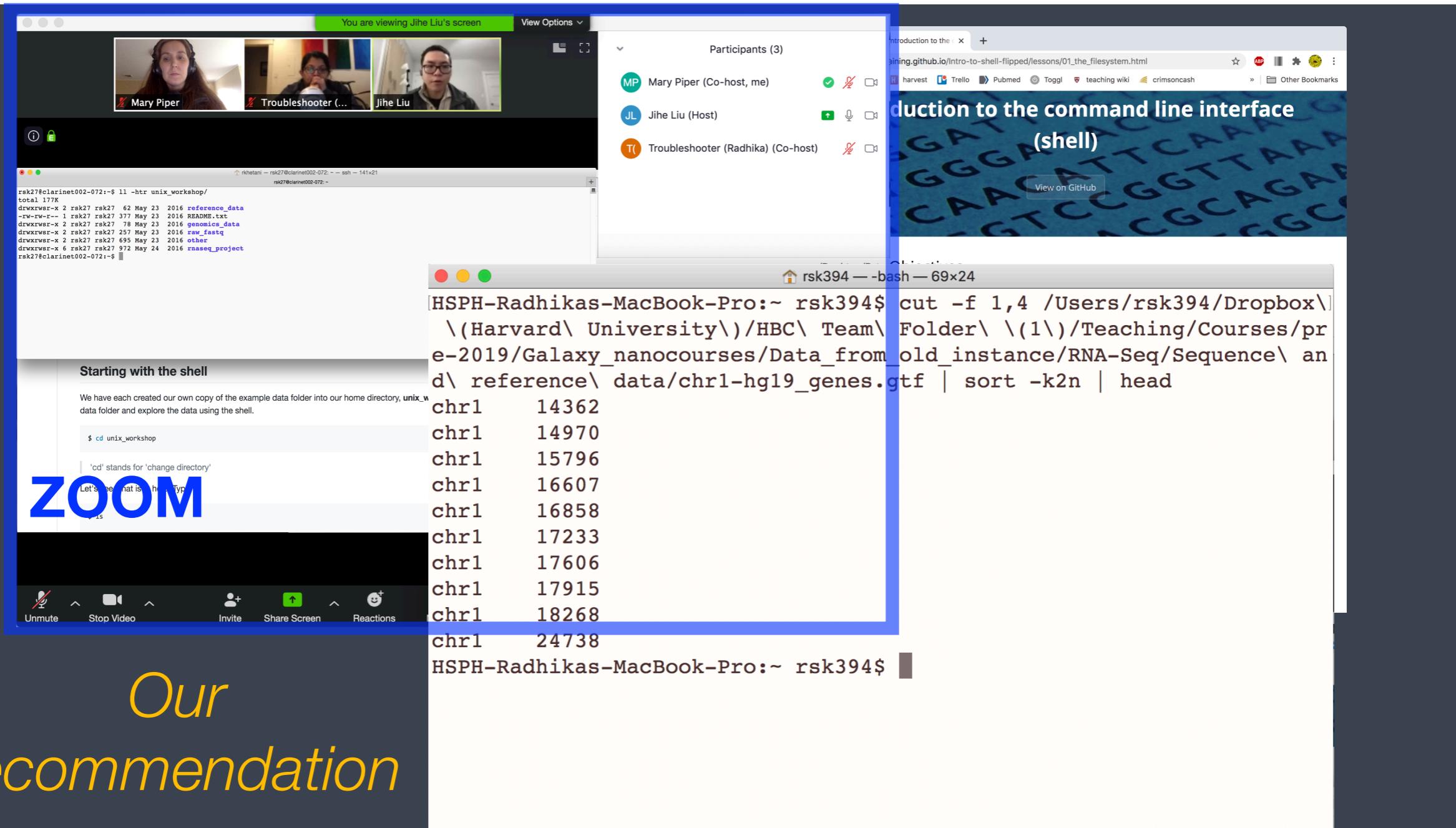
Quality control of sequence reads



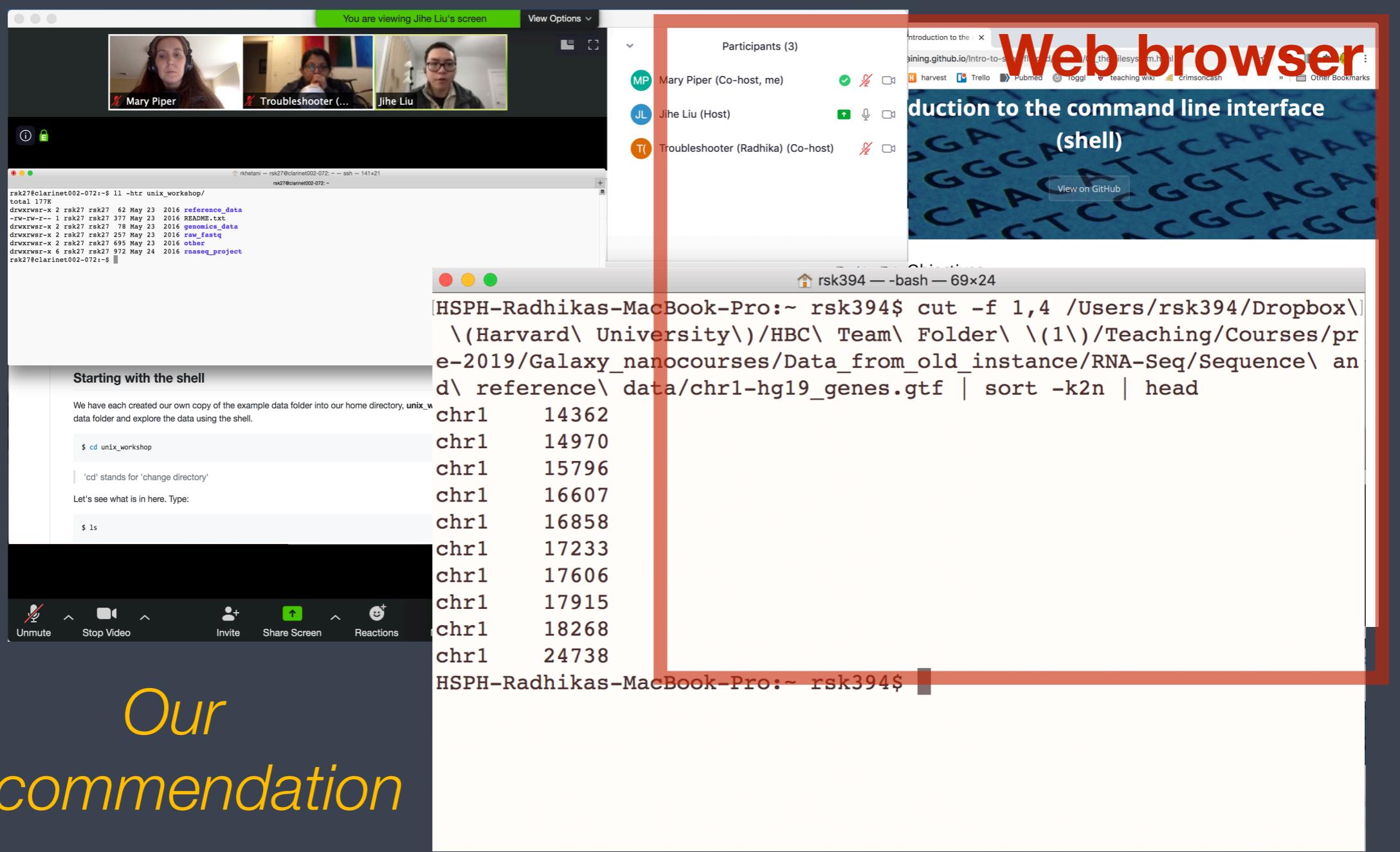
Single screen & 3 windows?



Single screen & 3 windows?



Single screen & 3 windows?



Single screen & 3 windows?

The image shows a video conference interface with three main windows:

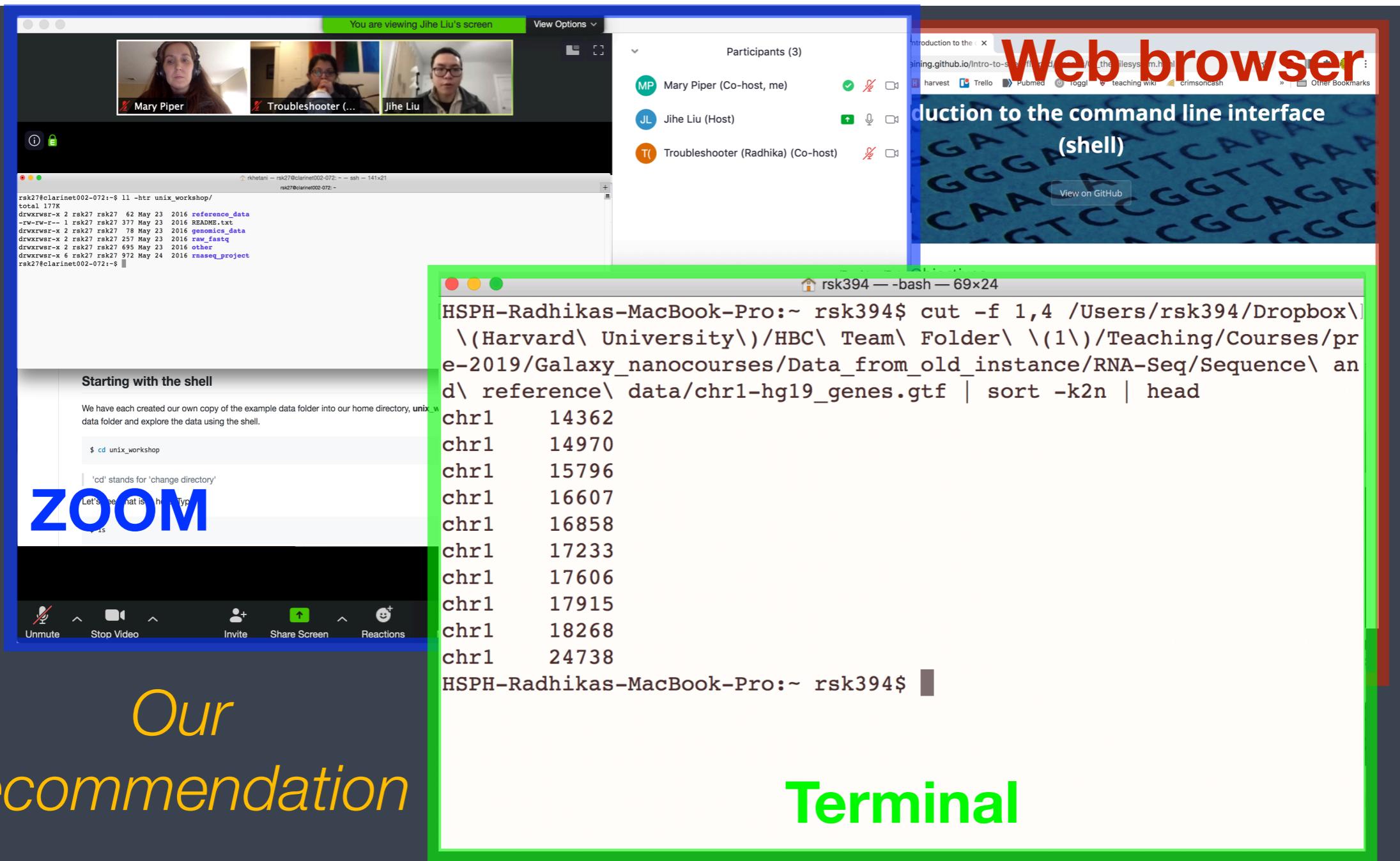
- Video Feed:** Shows three participants: Mary Piper, Troubleshooter (Radhika), and Jihe Liu.
- Participants List:** Shows three participants: Mary Piper (Co-host, me), Jihe Liu (Host), and Troubleshooter (Radhika) (Co-host).
- Terminal Session:** A green-highlighted window showing a command-line interface. The command run is:

```
rsk394 — bash — 69x24
HSPH-Radhikas-MacBook-Pro:~ rsk394$ cut -f 1,4 /Users/rsk394/Dropbox\(\Harvard\ University\)/HBC\ Team\ Folder\ \((1\))/Teaching/Courses/pre-2019/Galaxy_nanocourses/Data_from_old_instance/RNA-Seq/Sequence\ and\ reference\ data/chr1-hg19_genes.gtf | sort -k2n | head
chr1    14362
chr1    14970
chr1    15796
chr1    16607
chr1    16858
chr1    17233
chr1    17606
chr1    17915
chr1    18268
chr1    24738
```

Our recommendation

Terminal

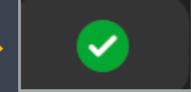
Single screen & 3 windows?

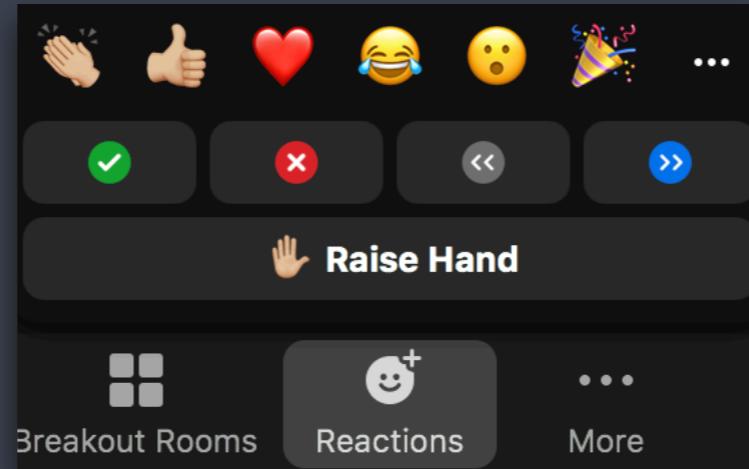


Odds and Ends (1/2)

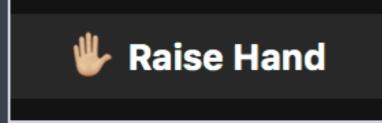
- ❖ Quit/minimize all applications that are not required for class

Odds and Ends (1/2)

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

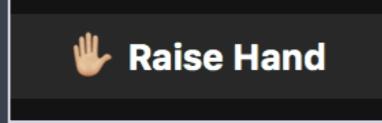


Odds and Ends (2/2)

- ❖ Questions for the presenter?
 - Post the question in the Chat window OR
 -  when the presenter asks for questions
 - Let the Moderator know

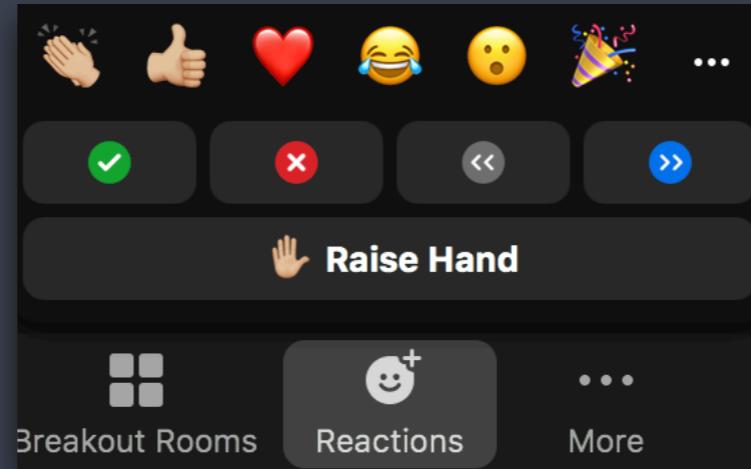
Odds and Ends (2/2)

- ❖ 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 private chat with the Troubleshooter with a description of the problem.



Thanks!

- Kathleen Keating 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 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