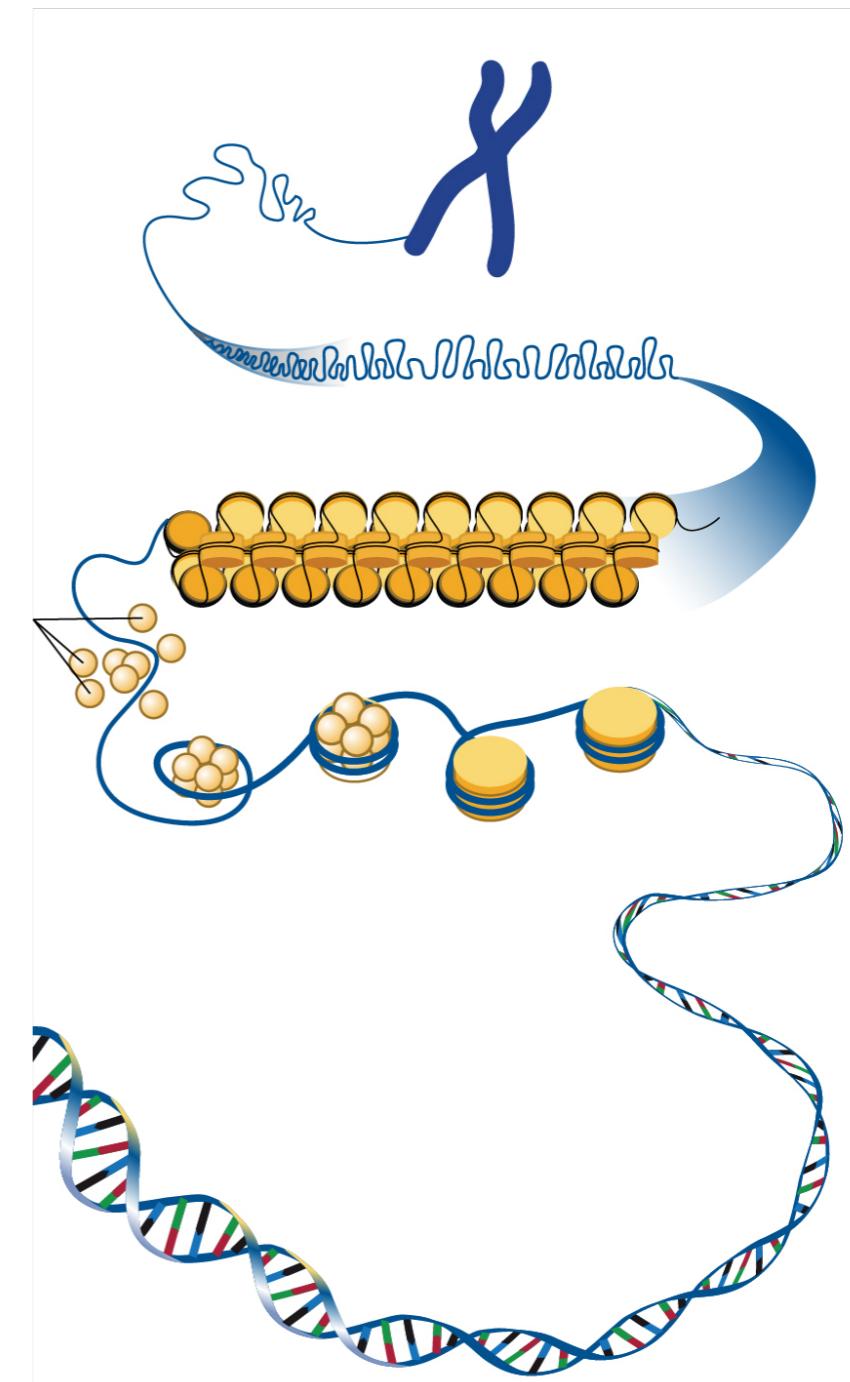


Understanding chromatin biology using high throughput sequencing methods

Harvard Chan Bioinformatics Core
in collaboration with
HMS Research Computing

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





Shannan Ho Sui
Director



Victor Barrera



Amelie Jule



Zhu Zhuo



James Billingsley



Radhika Khetani
Director of Education



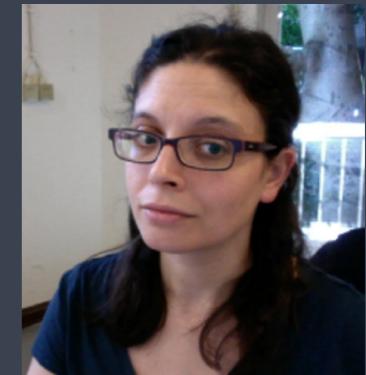
Meeta Mistry



Jihe Liu



Will Gammerdinger



Emma Berdan



Sergey Naumenko



Maria Simoneau



We are hiring!
Email bioinformatics@hspf.harvard.edu
for more information



Peter Kraft
Faculty Advisor

Consulting

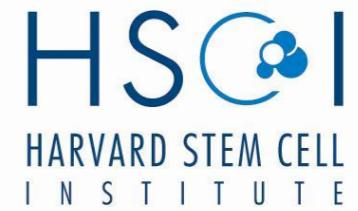
- RNA-seq analysis: bulk, single cell, small RNA
- ChIP-seq and ATAC-seq analysis
- Genome-wide methylation
- WGS, resequencing, exome-seq and CNV studies
- QC & analysis of gene expression arrays
- Functional enrichment analysis
- Grant support

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



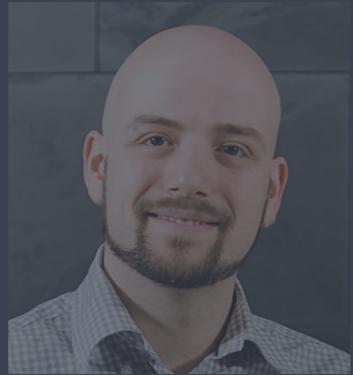
**HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH**

NIEHS





Shannan Ho Sui
Director



Victor Barrera



Amelie Jule



Zhu Zhuo



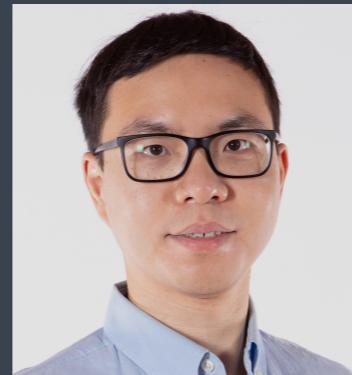
James Billingsley



Radhika Khetani
Director of Education



Meeta Mistry



Jihe Liu



Will Gammerdinger



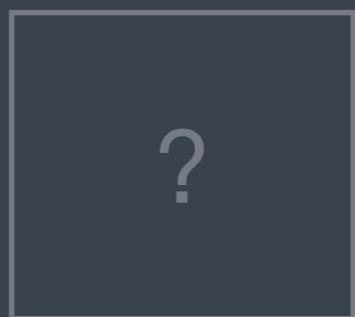
Emma Berdan



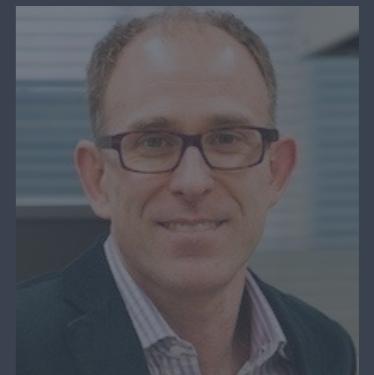
Sergey Naumenko



Maria Simoneau



We are hiring!
Email bioinformatics@hsph.harvard.edu
for more information



Peter Kraft
Faculty Advisor

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.*

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

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

Training

A key component of the HBC's mission is to train researchers at Harvard and beyond.

HBC's training team is made up of scientists, educators, and community leaders who work on research projects to ensure our training is relevant.

Our hands-on workshops focus on **bioinformatics**, with an emphasis on **experimental design** and **data analysis**. We also offer **wet-lab biologists** and **computational biologists** training in how to analyze their data.

We offer three types of workshops:

1. Short, 3-hour monthly workshops
2. Basic Data Skills**
3. Advanced Topics: Analysis of high-throughput sequencing data

***The basic data skills workshops are designed for the non-computer scientist.*



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

DF/HCC
DANA-FARBER / HARVARD CANCER CENTER

HSCI
HARVARD STEM CELL
INSTITUTE

 **HARVARD
CATALYST**
THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER



**HARVARD
MEDICAL SCHOOL**

Our dedicated training team holds workshops to help researchers learn how to analyze **bioinformatics** or **NGS** data.

We also devote substantial time to material development, including **workshop materials**. Our training team also participate in consultations on **best practices in NGS analysis**.

Workshops focus on **bioinformatics**, with an emphasis on **experimental design** and **reproducibility**. Our workshops are designed for **wet-lab biologists** and **computational biologists** to learn how to analyze their experiments and analyzing the resulting data.

bioinformatics)

(**NGS**) data**

and **bioinformatics** for the advanced workshops.

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

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

Training

A key component of the HBC's mission is to train researchers at Harvard and beyond.

HBC's training team is made up of scientists, educators, and community leaders who work on research projects to ensure our training is relevant.

Our hands-on workshops focus on **bioinformatics**, with an emphasis on **experimental design** and **data analysis**. We also offer **wet-lab biologists** and **clinicians** training in how to analyze their data.

We offer three types of workshops:

1. Short, 3-hour monthly workshops
2. Basic Data Skills**
3. Advanced Topics: Analysis of high-throughput sequencing data

***The basic data skills workshops are designed for the non-bioinformatician.*



**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**

Our dedicated training team holds workshops to help researchers learn how to analyze their data, including NGS data.

The training team also devote substantial time to material development, including creating training modules and participating in consultations on best practices in NGS analysis.

Workshops focus on the analysis of high-throughput sequencing data, with an emphasis on **experimental design**, **bioinformatics**, and **reproducibility**. Our workshops are designed for researchers performing sequencing-based experiments and analyzing the resulting data.

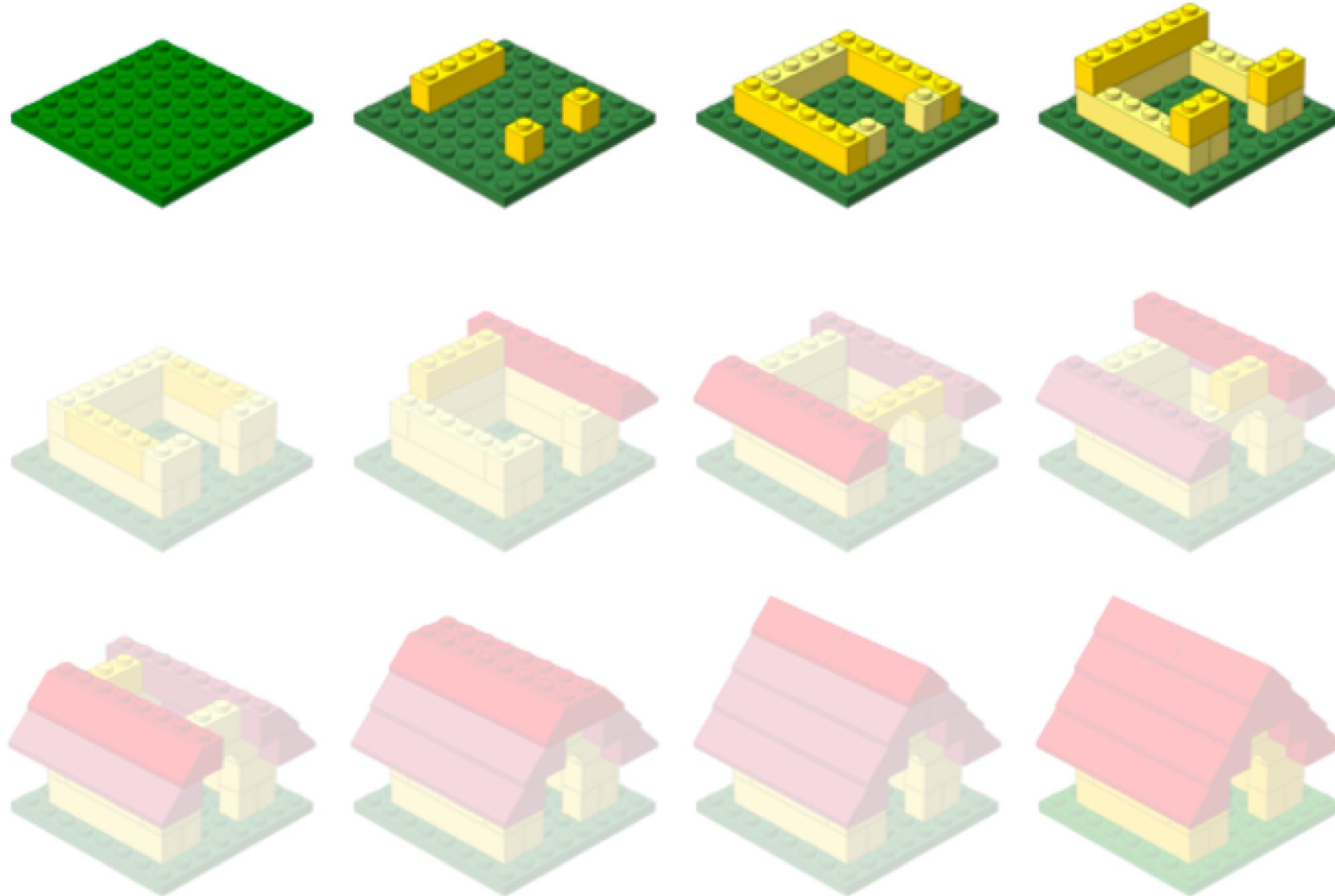
bioinformatics)

(**NGS**) data**

and **bioinformatics** for the advanced workshops.

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

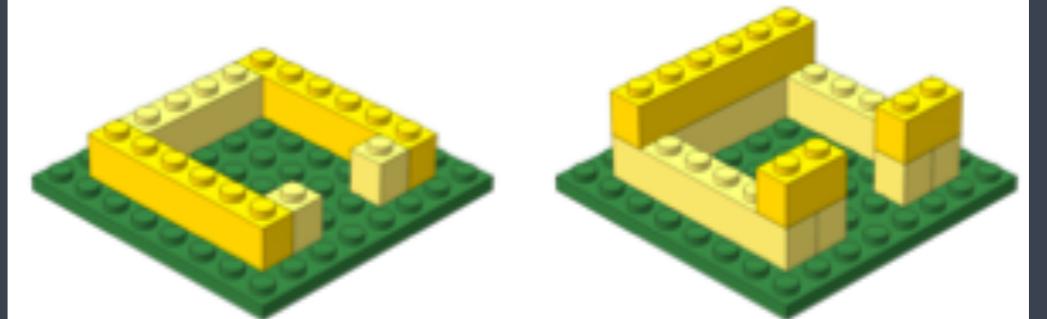
<https://hbctraining.github.io/main/>



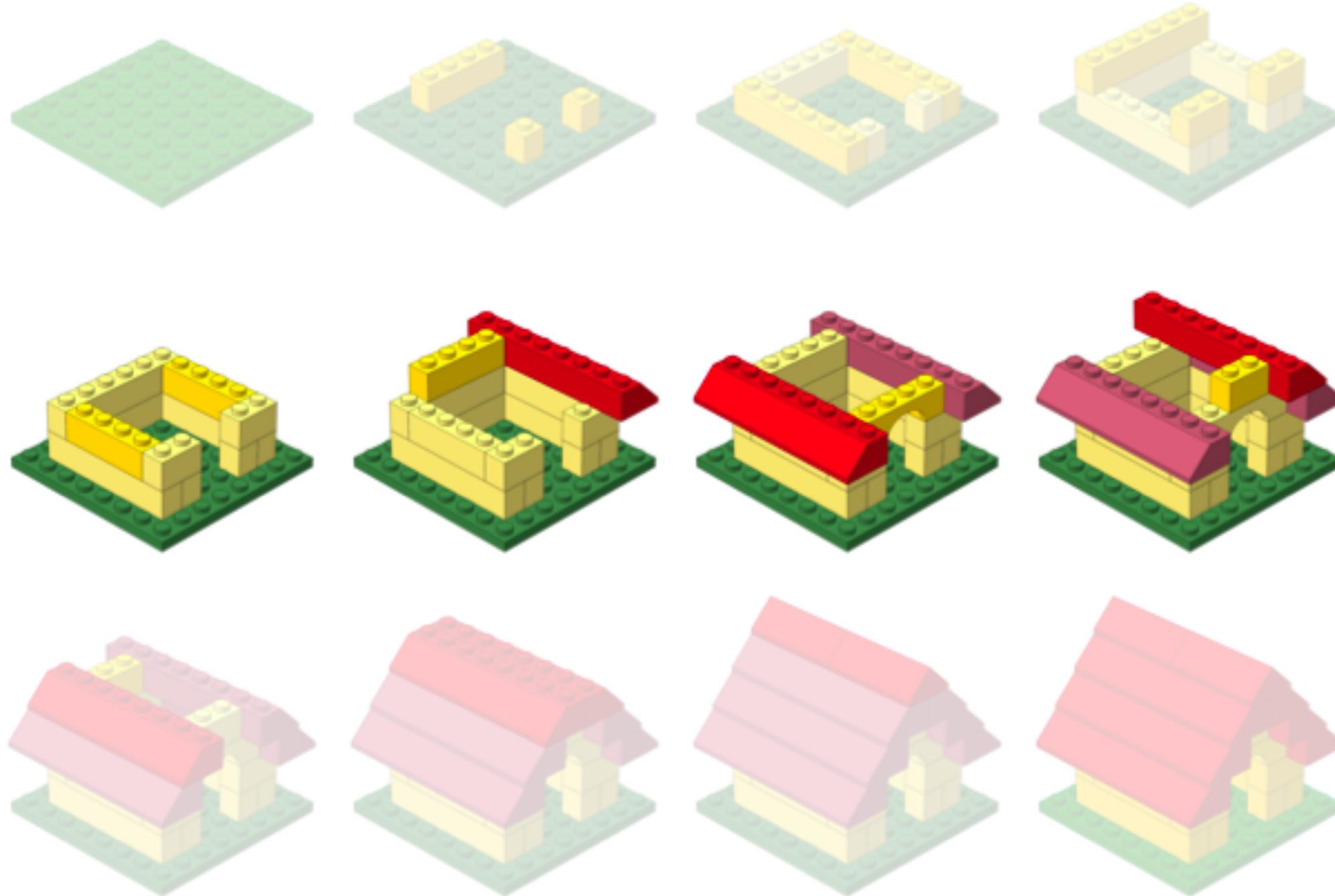
<http://anoved.net/tag/lego/page/3/>

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



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

Bioinformatics data analysis

Workshop scope

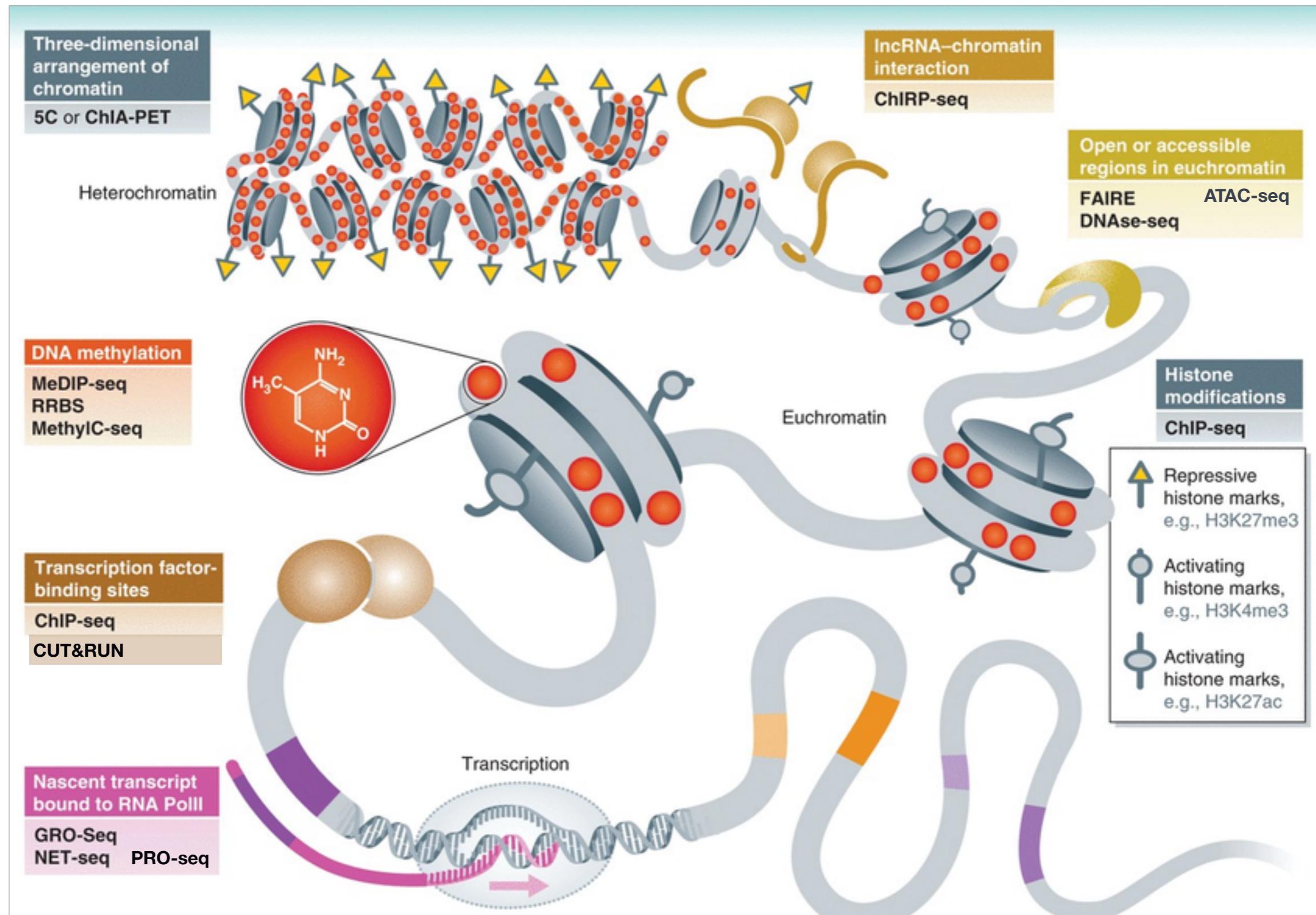
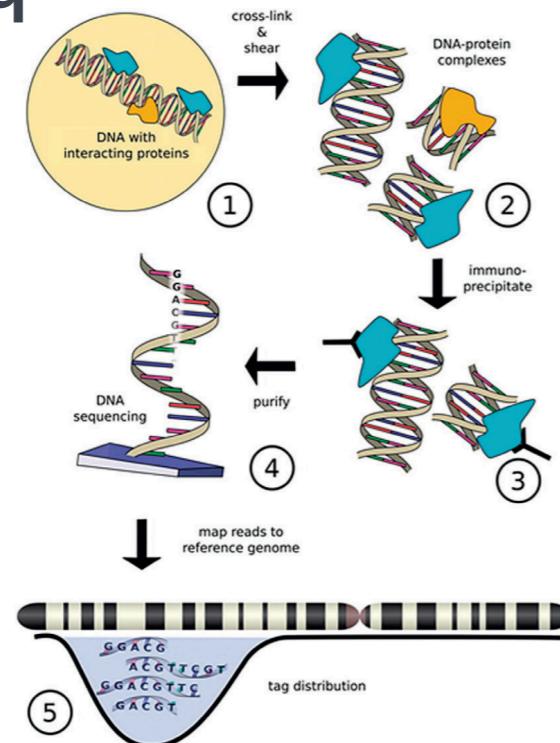


Figure adapted from Soon WW, Hariharan M, Snyder MP, "High throughput sequencing for biology and medicine". Molecular Systems Biology 9:640 2013

Genomic methods for profiling chromatin

ChIP-seq



CUT&RUN

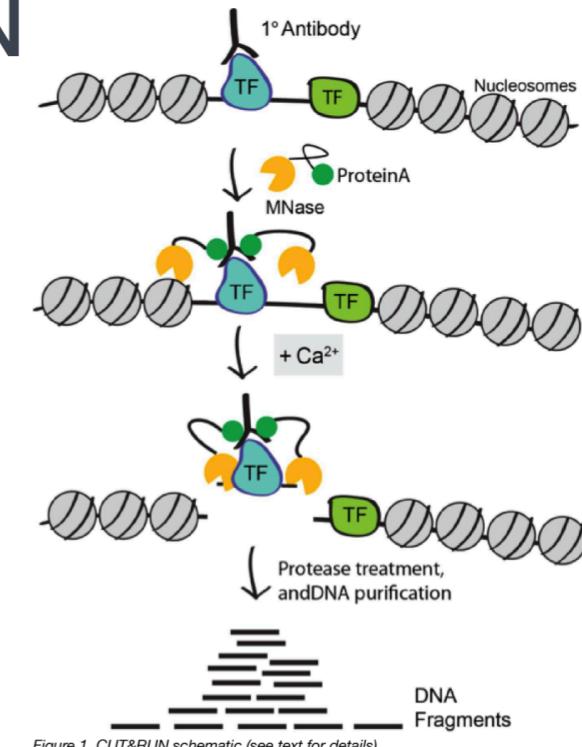
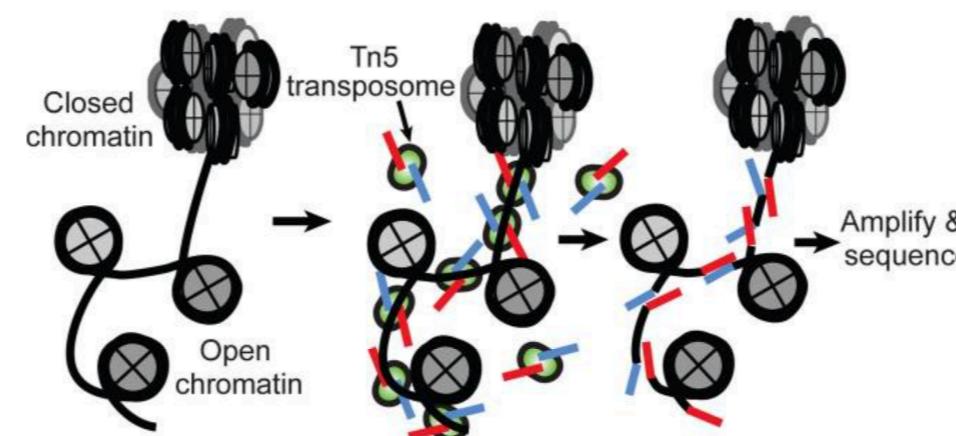
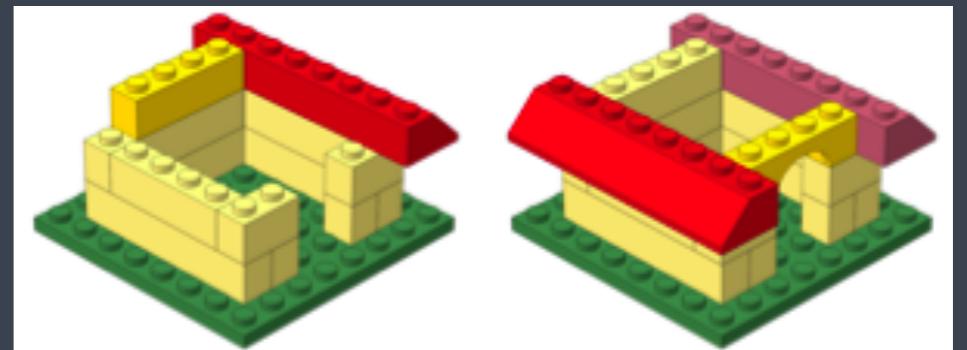


Figure 1. CUT&RUN schematic (see text for details).

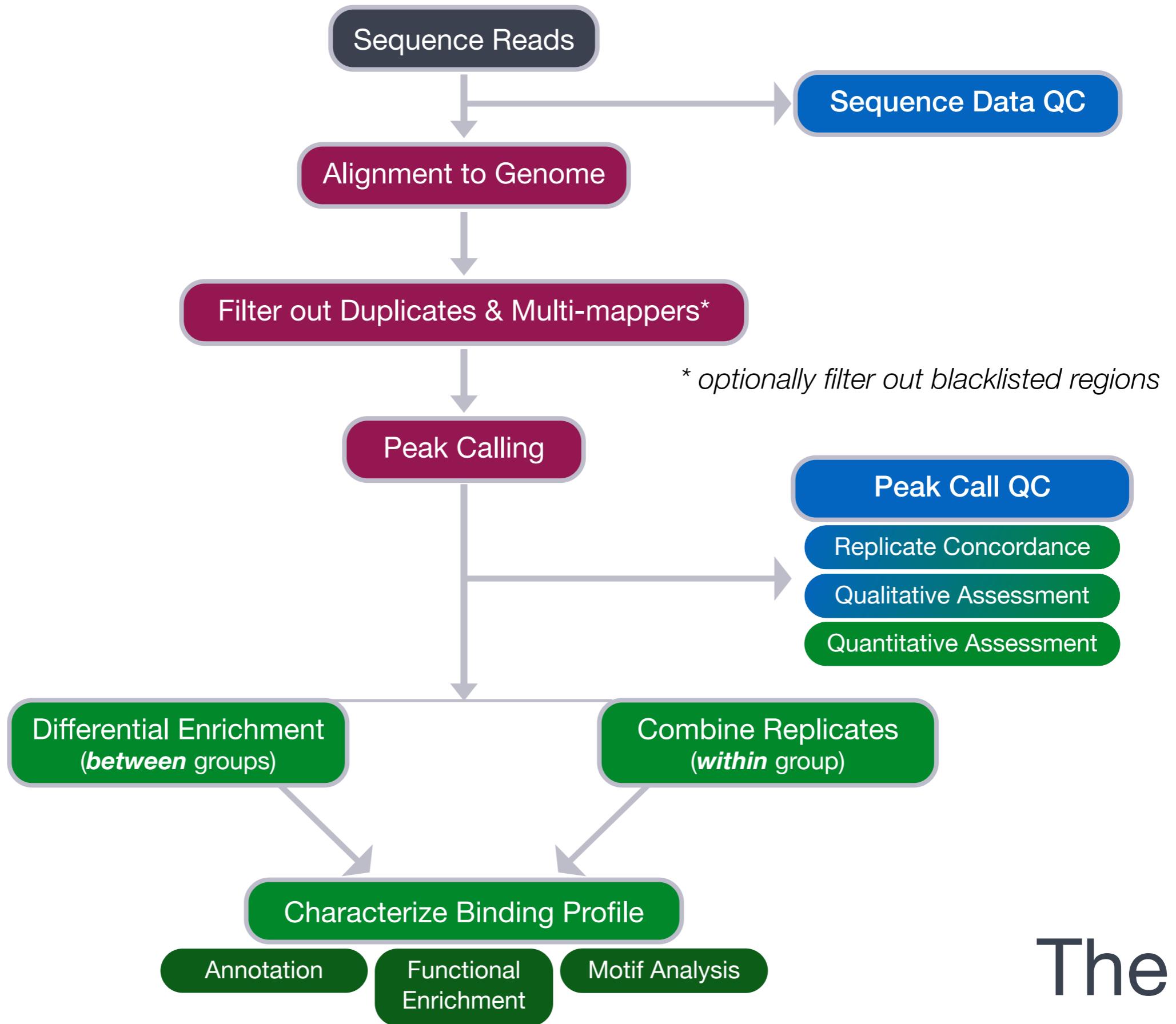
ATAC-seq



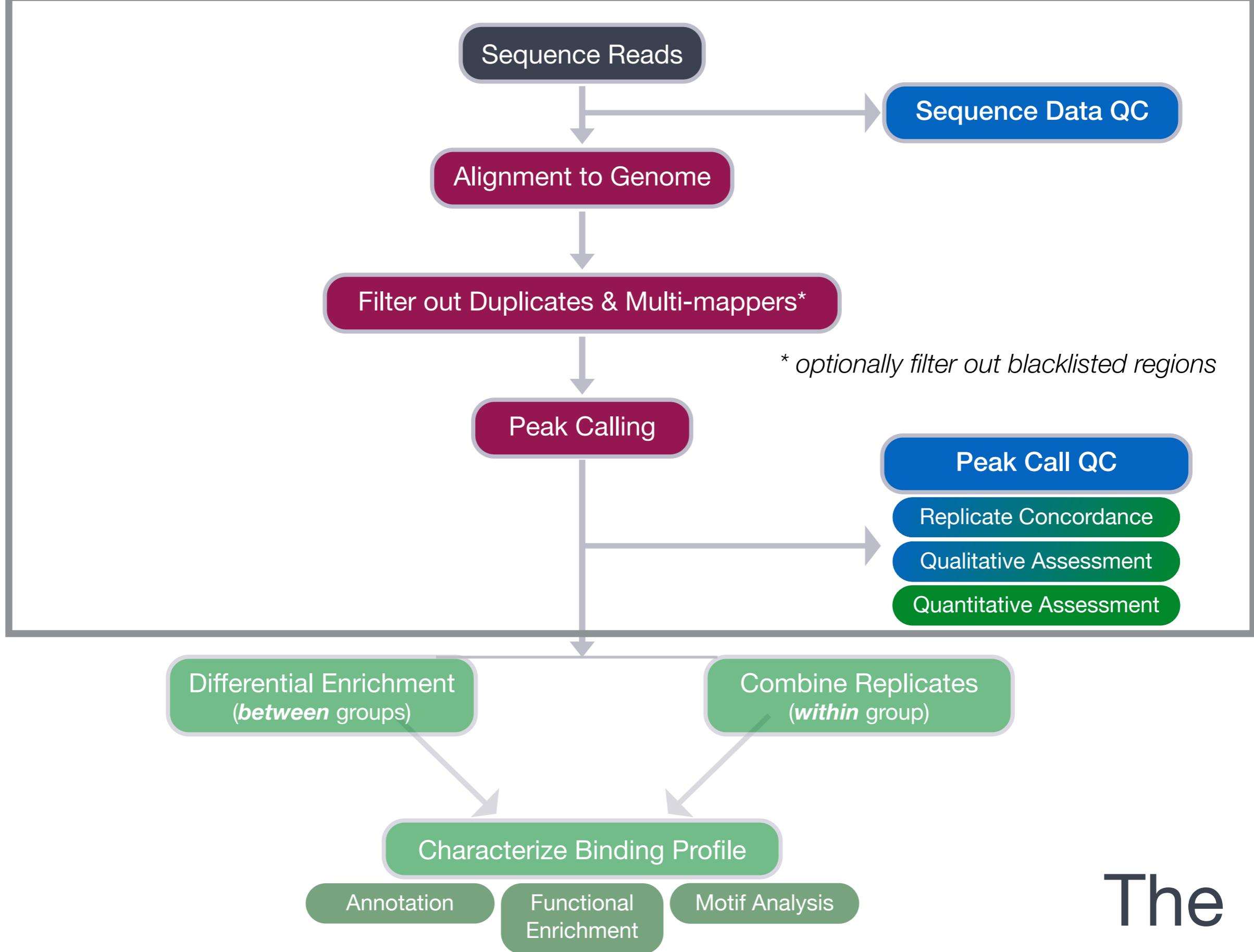
Learning Objectives



- ✓ Describe important considerations for setting up a successful ChIP-seq, CUT&RUN or ATAC-seq experiment
- ✓ Describe the steps in an ChIP-seq analysis workflow (from sequence data to peak calls) and contrast any differences for CUT&RUN and ATAC-seq analyses
- ✓ Learn how to handle various file formats encountered when analyzing ChIP-seq and related data
- ✓ 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 not be covered in this workshop

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 \(RDM\)](#)
 - [Working in an HPC environment](#)
 - [A review of high-throughput sequencing methods for understanding chromatin biology](#)

Day 1

| Time | Topic | Instructor |
|---------------|--|--------------------|
| 09:30 - 09:45 | Workshop Introduction | Meeta |
| 09:45 - 11:00 | Understanding chromatin biology using high-throughput sequencing | Dr. Shannan Ho Sui |
| 11:00-11:05 | Break | |
| 11:05 - 11:20 | HPC review Q&A | Jihe |

Course materials online



FastQC for quality assessment

[View on GitHub](#)

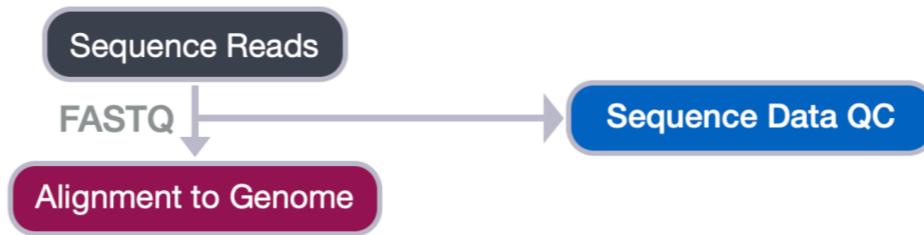
Contributors: Mary Piper, Radhika Khetani, Jihe Liu, Will Gammerdinger

Approximate time:

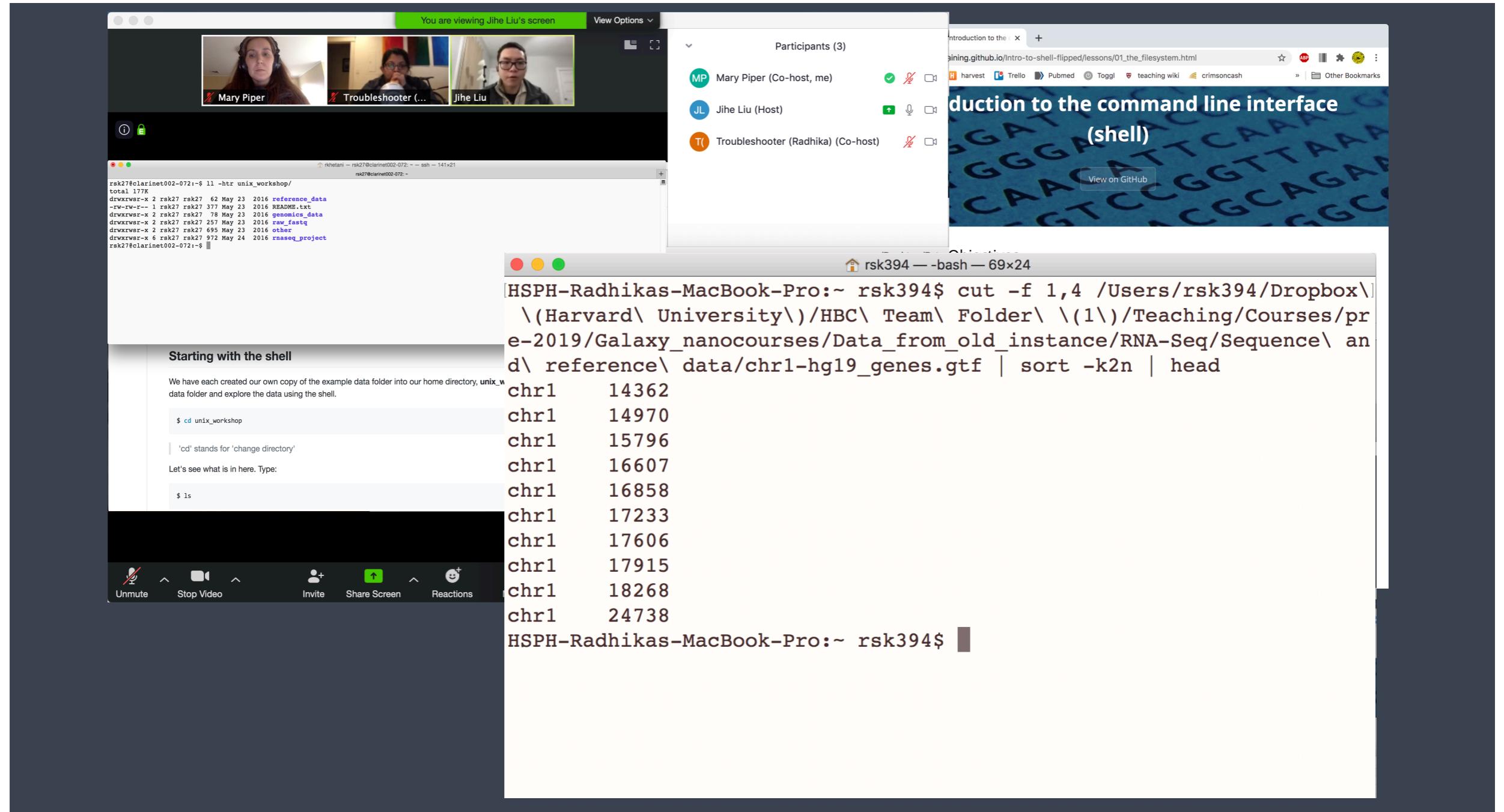
Learning Objectives

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

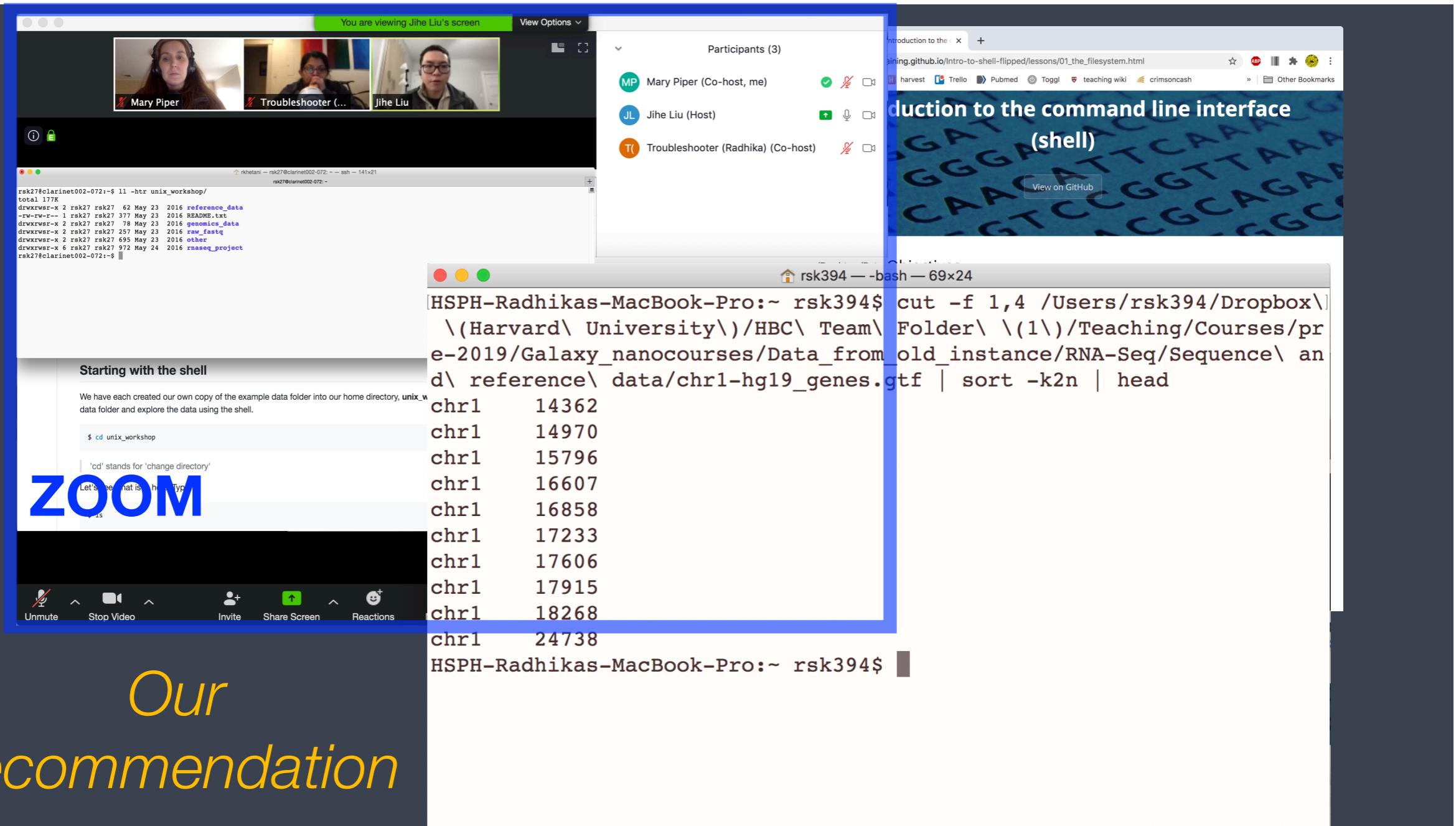
Quality control of sequence reads



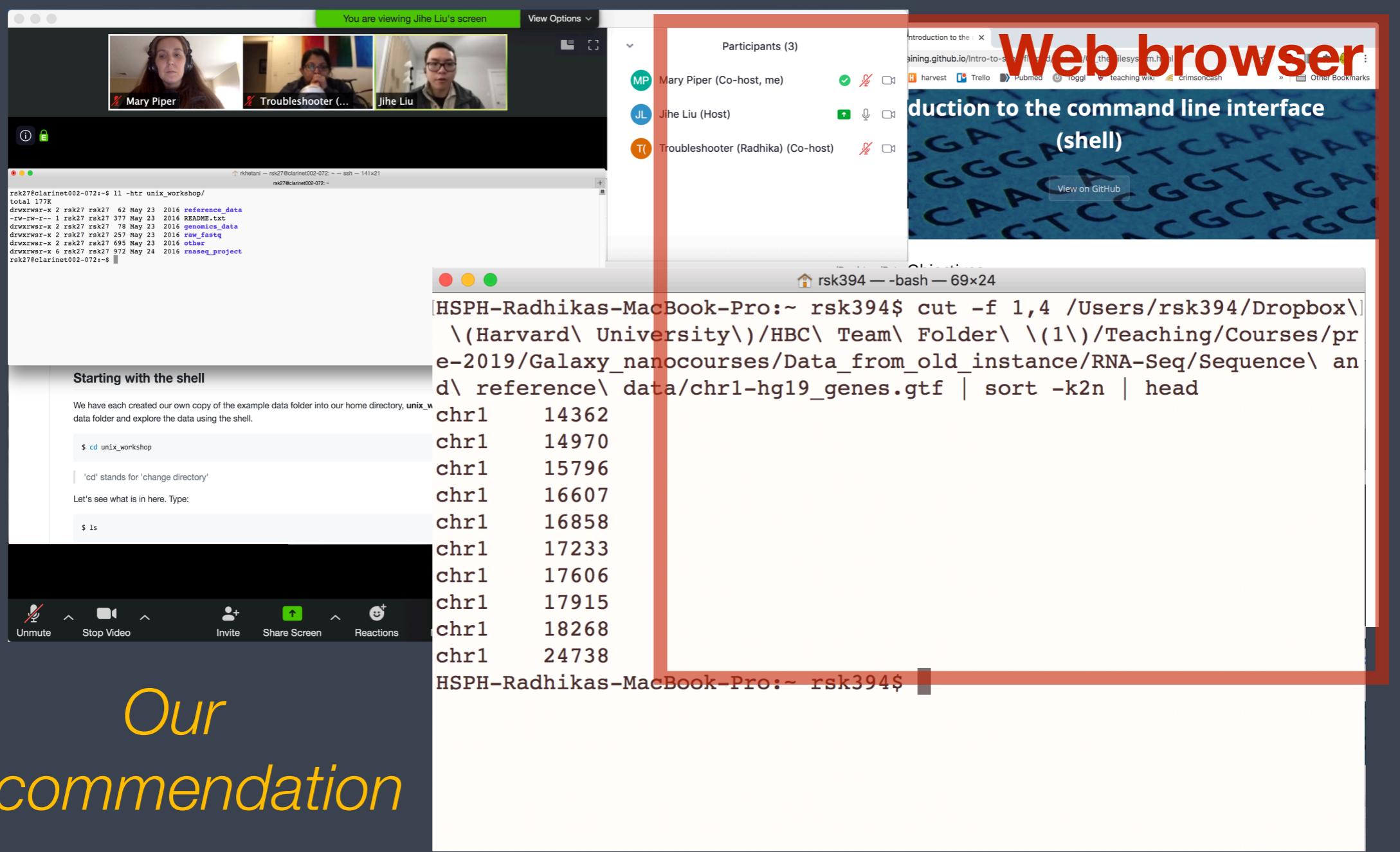
Single screen & 3 windows?



Single screen & 3 windows?



Single screen & 3 windows?



Single screen & 3 windows?

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

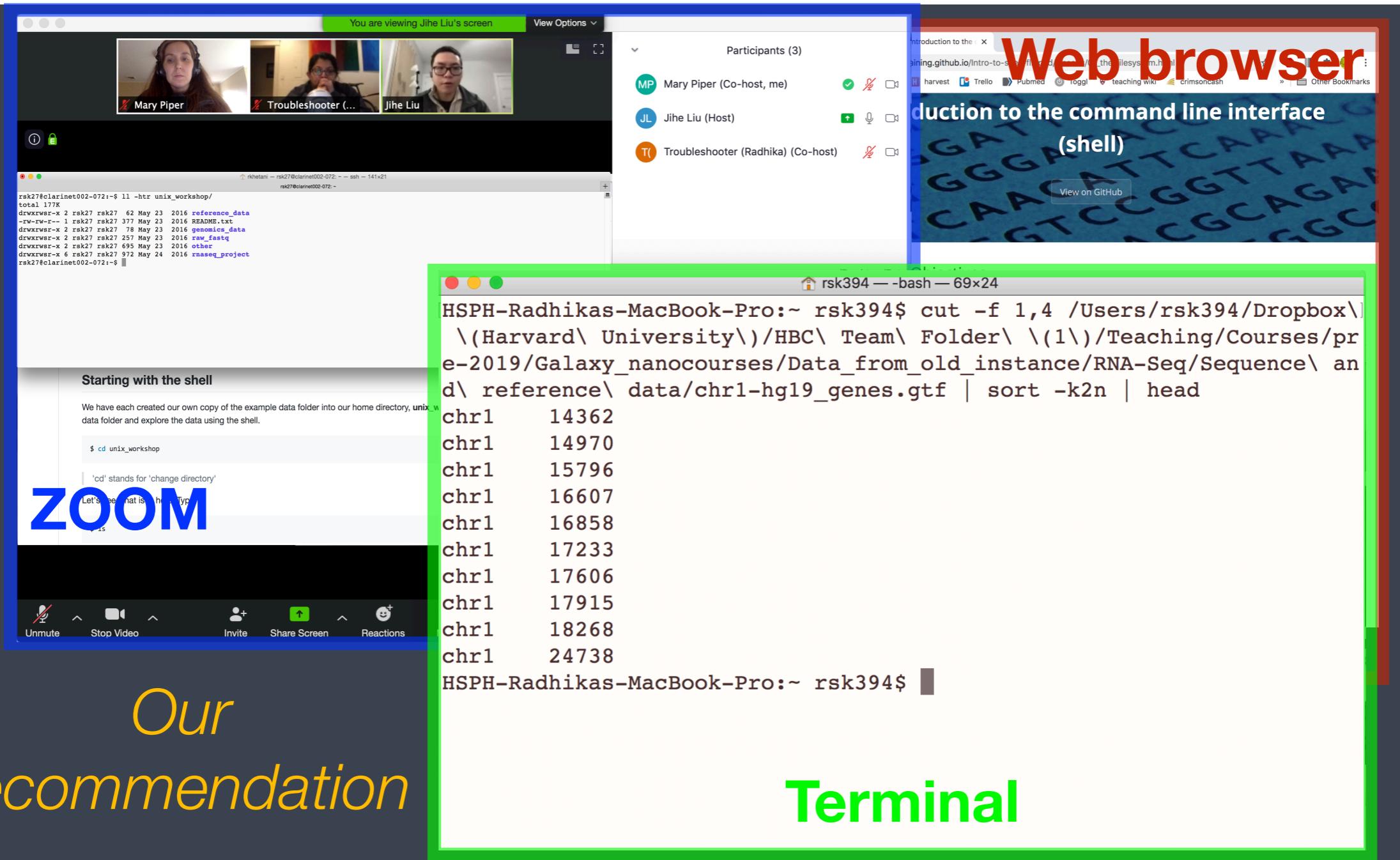
- Video Feed:** Shows three participants: Mary Piper, Troubleshooter (Radhika), and Jihe Liu.
- Participants List:** Shows three participants: Mary Piper (Co-host, me), Jihe Liu (Host), and Troubleshooter (Radhika) (Co-host).
- Terminal Session:** A green-highlighted window showing a command-line interface.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



Course participation

- ▶ Please keep your videos on, we would love to see your faces!
- ▶ 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!



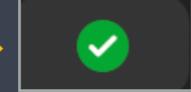
Homework and Expectations

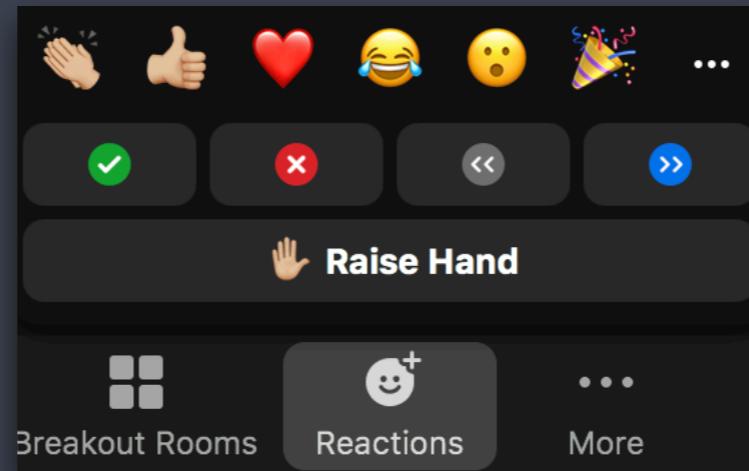
- ❖ At-home lessons and exercises after each session
- ❖ Cover material not previously discussed
- ❖ Provides us feedback to help pace the course appropriately
- ❖ 3-5 hours to complete

Odds and Ends (1/2)

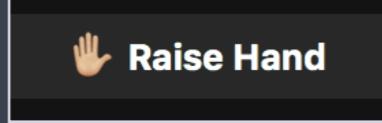
- ❖ Quit/minimize all applications that are not required for class

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**)

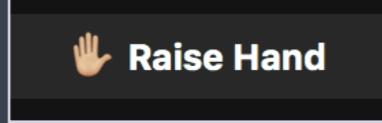


Odds and Ends (2/2)

- ❖ 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