



```
dds <- DESeqDataSetFromMatrix(countData = cts,
                                colData = coldata,
                                design= ~ batch + condition)

dds <- DESeq(dds)
resultsNames(dds) # lists the coefficients
res <- results(dds, name="condition_trt_vs_untrt")
# or to shrink log fold changes association with condition:
res <- lfcShrink(dds, coef="condition_trt_vs_untrt", type="apeglm")
```

Bulk RNA-seq Analysis Part II

Differential Gene Expression

Harvard Chan Bioinformatics Core

<https://tinyurl.com/hbc-dge-online>



Shannan Ho Sui
Director



Victor Barrera



Amelie Jule



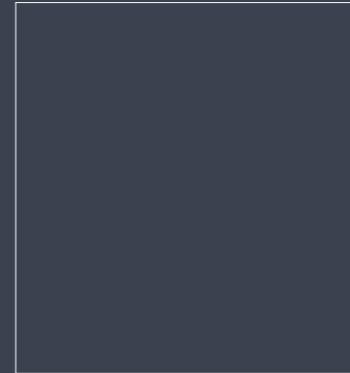
Zhu Zhuo



Radhika Khetani
Director of Education



Meeta Mistry



Heather Wick
Starts on June 19th



Will Gammerdinger



Emma Berdan



Sergey Naumenko



Maria Simoneau



Noor Sohail



James Billingsley

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



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.*

<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 NGS 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

CFAR
HARVARD UNIVERSITY
CENTER FOR AIDS RESEARCH

 **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 and interpret NGS data.

The training team also devote substantial time to material development, consulting, and teaching. Our training team also participate in consultations on best practices in NGS analysis.

Workshops focus on the analysis of high-throughput sequencing data, with an emphasis on **experimental design**, **data quality**, and **reproducibility**. Our workshops are designed for both **wet-lab biologists** and **computational biologists** performing experiments and analysing the resulting NGS 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 **computational biologists** training in NGS 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



**HARVARD
MEDICAL SCHOOL**

Our dedicated training team holds workshops to help researchers learn how to analyze and interpret NGS data.

The training team also devote substantial time to material development, consulting, and teaching. Our training team also participate in consultations on best practices in NGS analysis.

Workshops focus on the analysis of high-throughput sequencing data, with an emphasis on **experimental design**, **data quality**, and **reproducibility**. Our workshops are designed for both **wet-lab biologists** and **computational biologists** performing experiments and analysing the resulting NGS data.

bioinformatics)

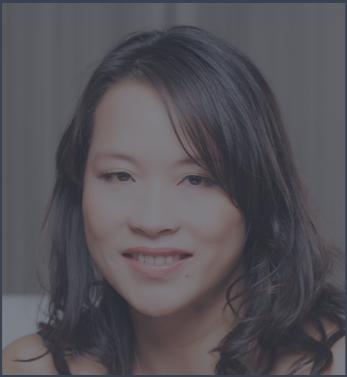
(NGS) data**

or the advanced workshops.

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

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

Introductions!



Shannan Ho Sui
Director



Victor Barrera



Amelie Jule



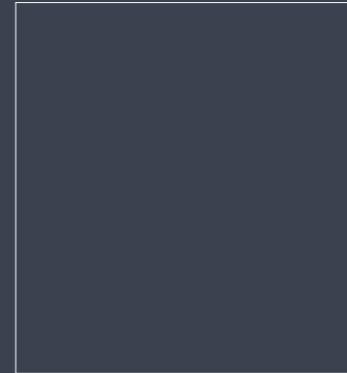
Zhu Zhuo



Radhika Khetani
Director of Education



Meeta Mistry



Heather Wick
Starts on June 19th



Will Gammerdinger



Emma Berdan



Sergey Naumenko



Maria Simoneau

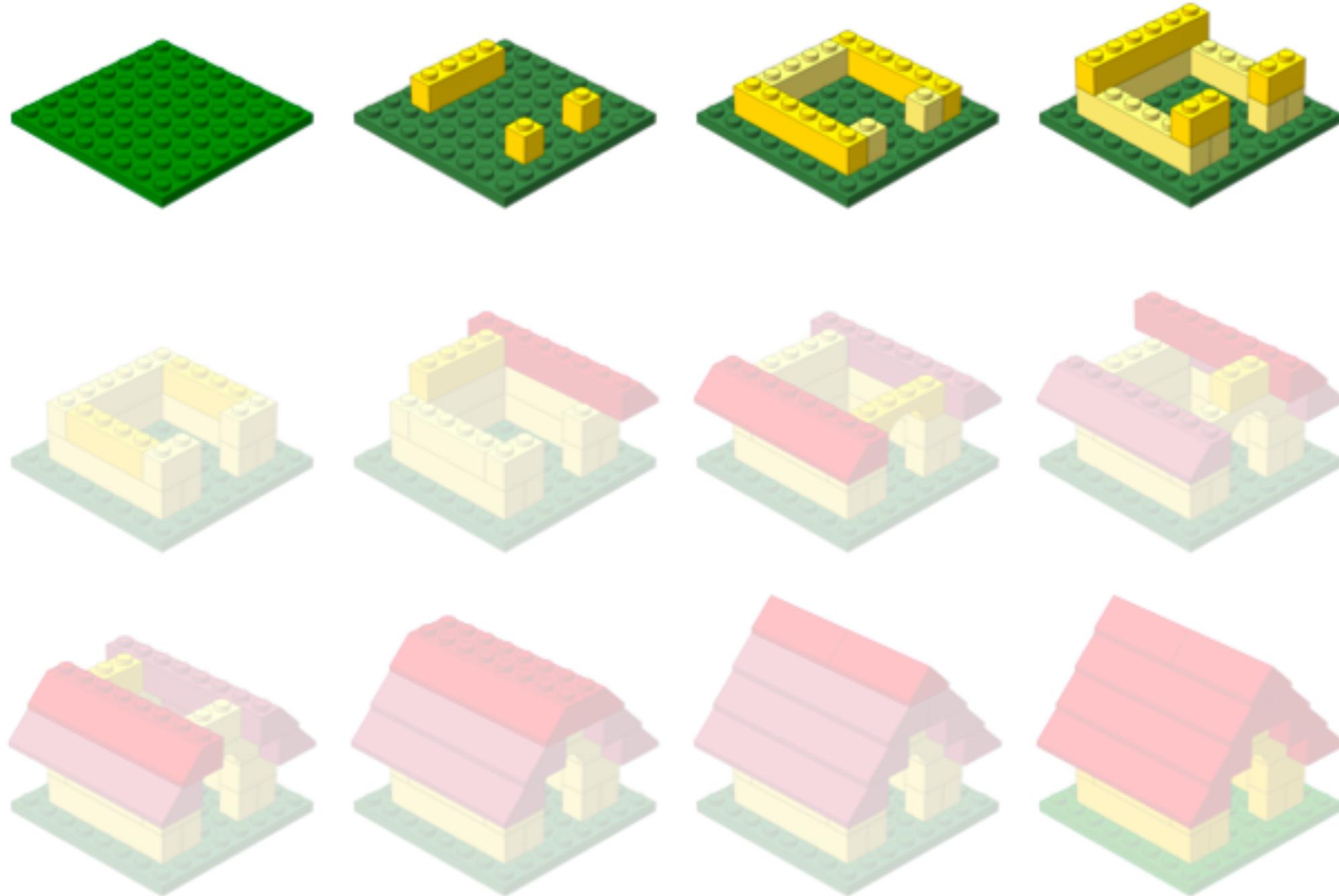


Noor Sohail



James Billingsley

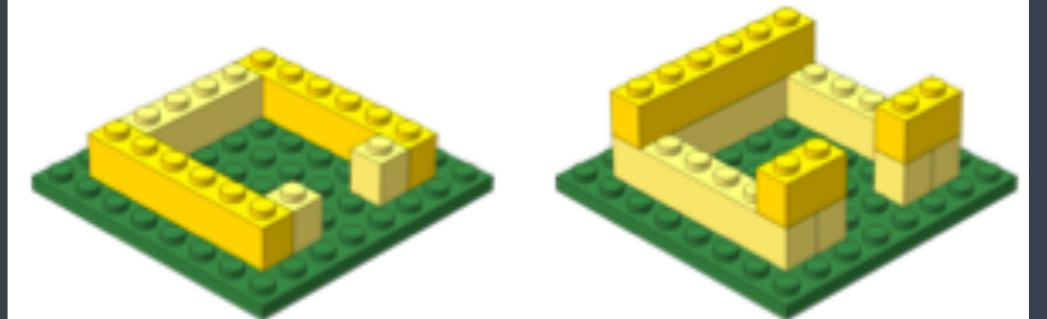
Workshop Scope...



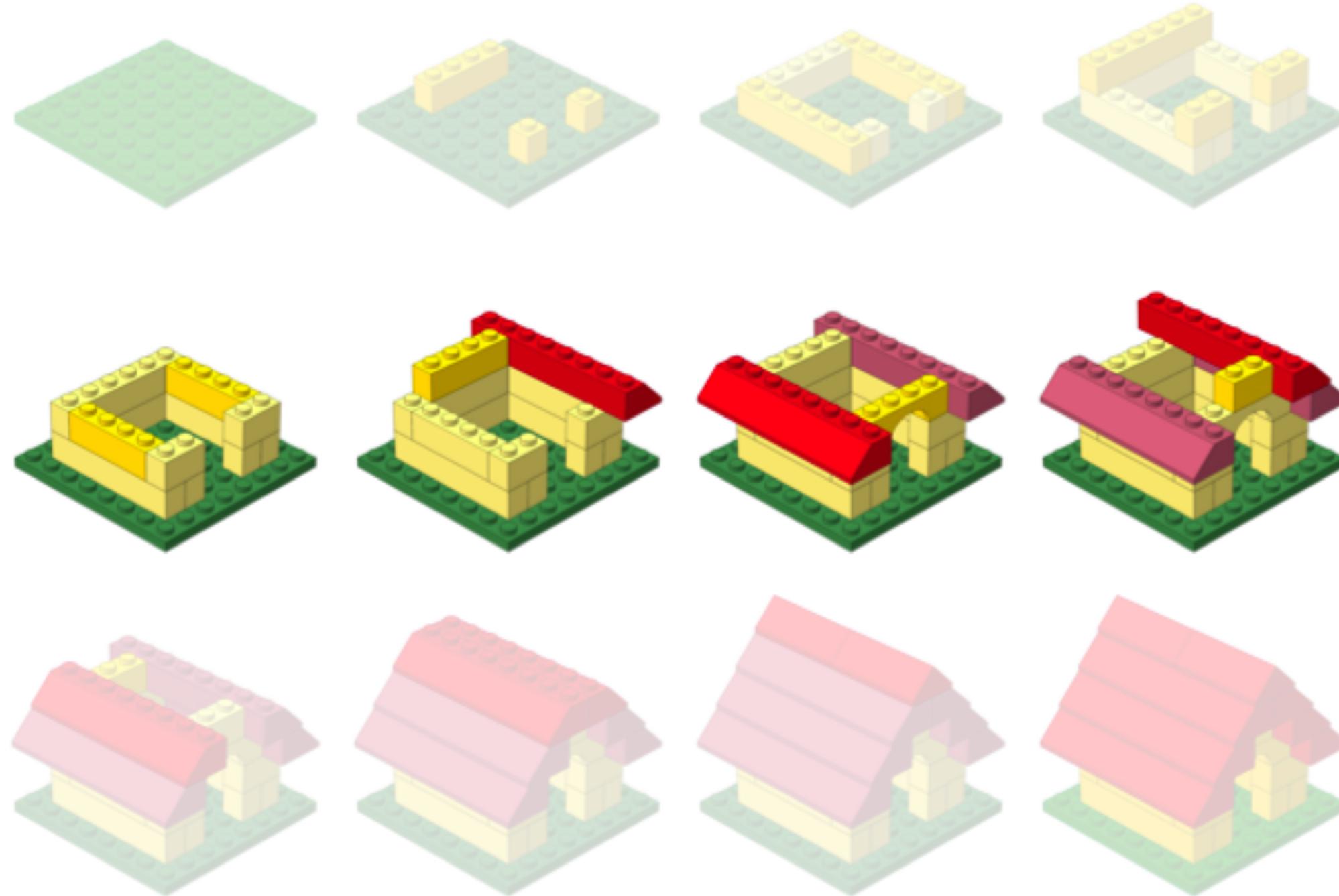
<http://anoved.net/tag/lego/page/3/>

Setting up to perform Bioinformatics analysis

Setting up...



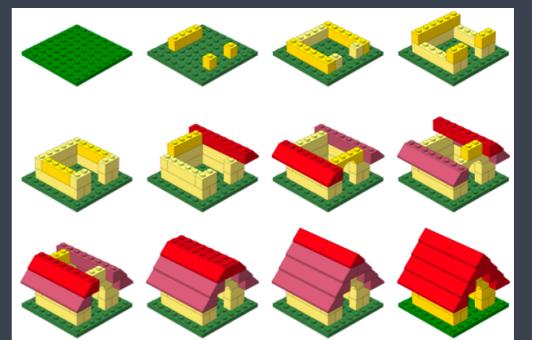
- ✓ Introduction to the command-line interface (shell, Unix, Linux)
 - Dealing with large data files
 - Performing bioinformatics analysis
 - Using tools
 - Accessing and using compute clusters
- ✓ R
 - Parsing and working with smaller results text files
 - Statistical analysis, e.g. differential expression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



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

Bioinformatics data analysis

Workshop Scope



Differential Gene Expression analysis

- ✓ Understand the considerations for performing statistical analysis on RNA-seq data
- ✓ Start with gene counts (after alignment and counting)
- ✓ Perform QC on count data
- ✓ Use DESeq2 to perform differential expression analysis on the count data and obtain a list of significantly different genes
- ✓ Visualize results of the analysis
- ✓ Perform functional analysis on the lists of differentially expressed genes

Logistics

Course webpage

<https://tinyurl.com/hbc-dge-online>

Course schedule online

Workshop Schedule

Pre-reading

1. [Workflow \(raw data to counts\)](#)
2. [Experimental design considerations](#)

Day 1

Time	Topic	Instructor
10:00 - 10:30	Workshop Introduction	Jihe
10:30 - 10:45	R refresher Q & A	Radhika
10:45 - 11:15	RNA-seq pre-reading discussion	Radhika
11:15 - 12:00	Intro to DGE / setting up DGE analysis	Meeta

Before the next class:

1. Please **study the contents** and **work through all the code** within the following lessons:
 - [RNA-seq counts distribution](#)
 - [Count normalization](#)
 - [Sample-level QC \(PCA and hierarchical clustering\)](#)
2. **Complete the exercises:**
 - Each lesson above contain exercises; please go through each of them.
 - **Copy over** your code from the exercises into a text file.
 - **Upload the saved text file to Dropbox** the **day before the next class**.

Questions?

- **If you get stuck due to an error** while running code in the lesson, [email us](#)
- Post any **conceptual questions** that you would like to have **reviewed in class here**.

Course materials

Available online for the
foreseeable future



Introduction to DGE

[View on GitHub](#)

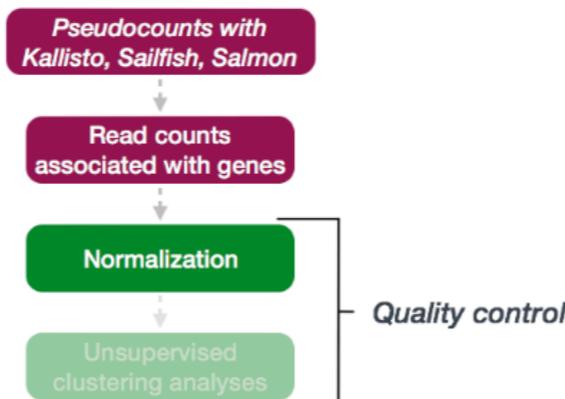
Approximate time: 60 minutes

Learning Objectives

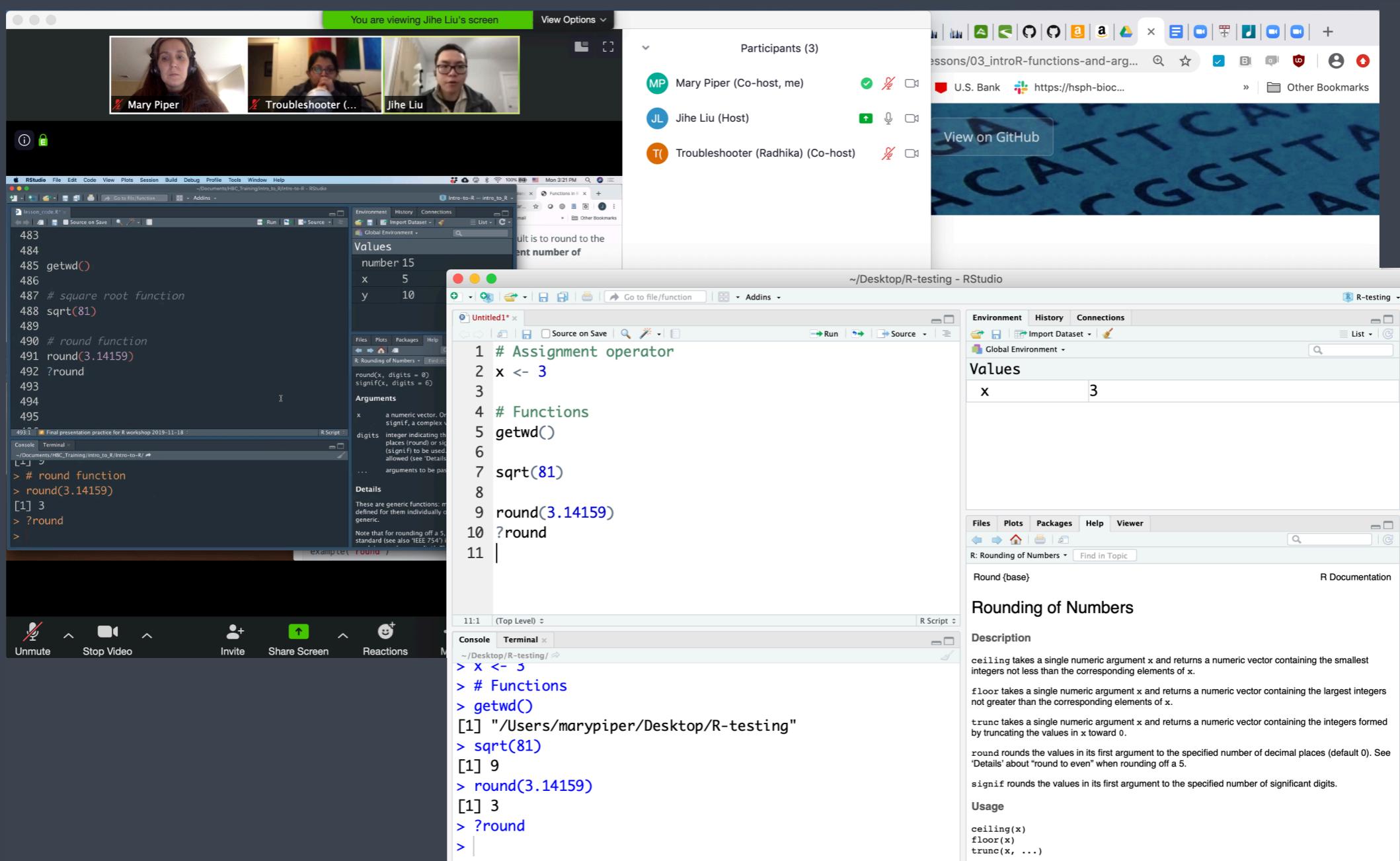
- Explore different types of normalization methods
- Become familiar with the `DESeqDataSet` object
- Understand how to normalize counts using DESeq2

Normalization

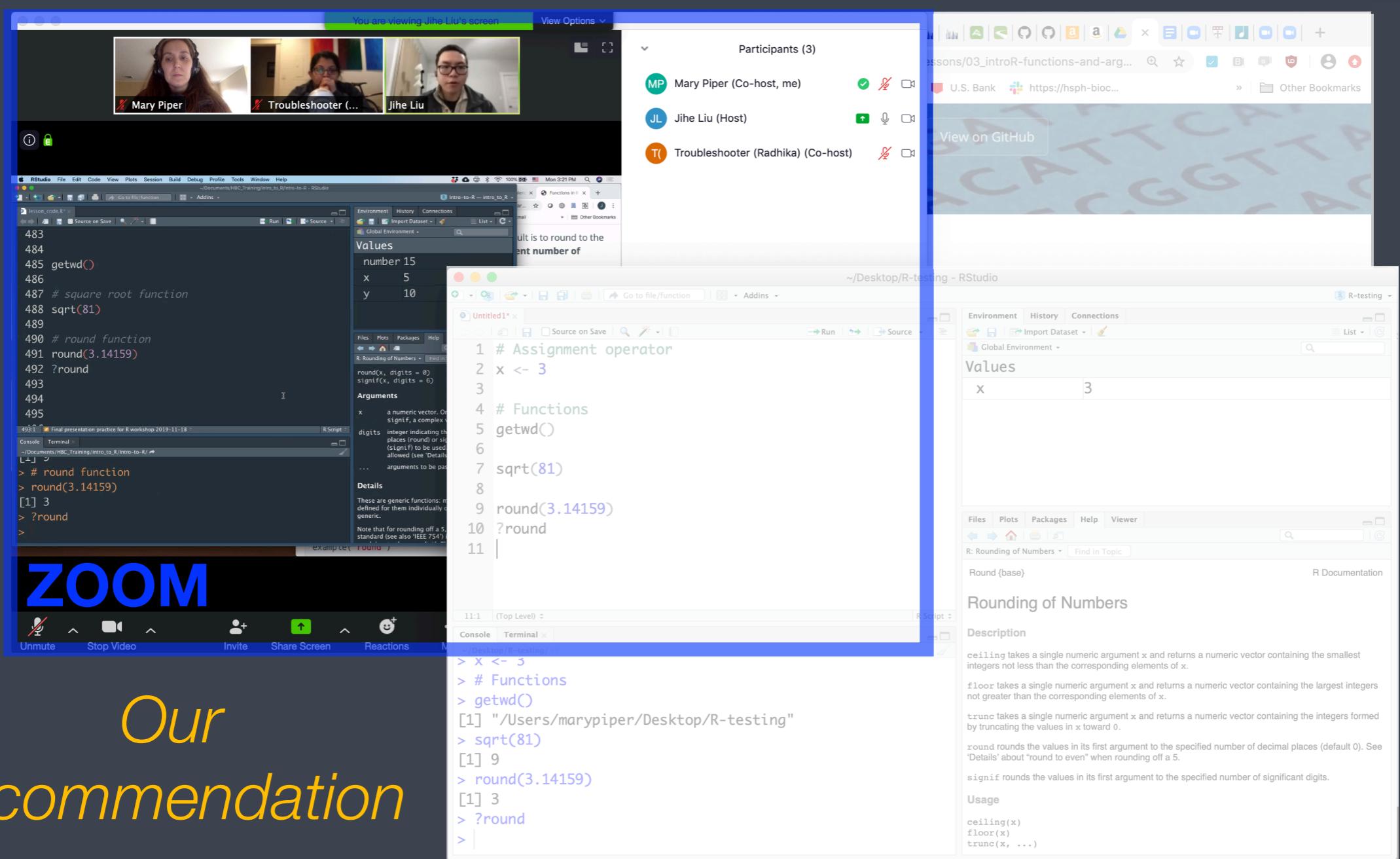
The first step in the DE analysis workflow is count normalization, which is necessary to make accurate comparisons of gene expression between samples.



Single screen & 3 windows?

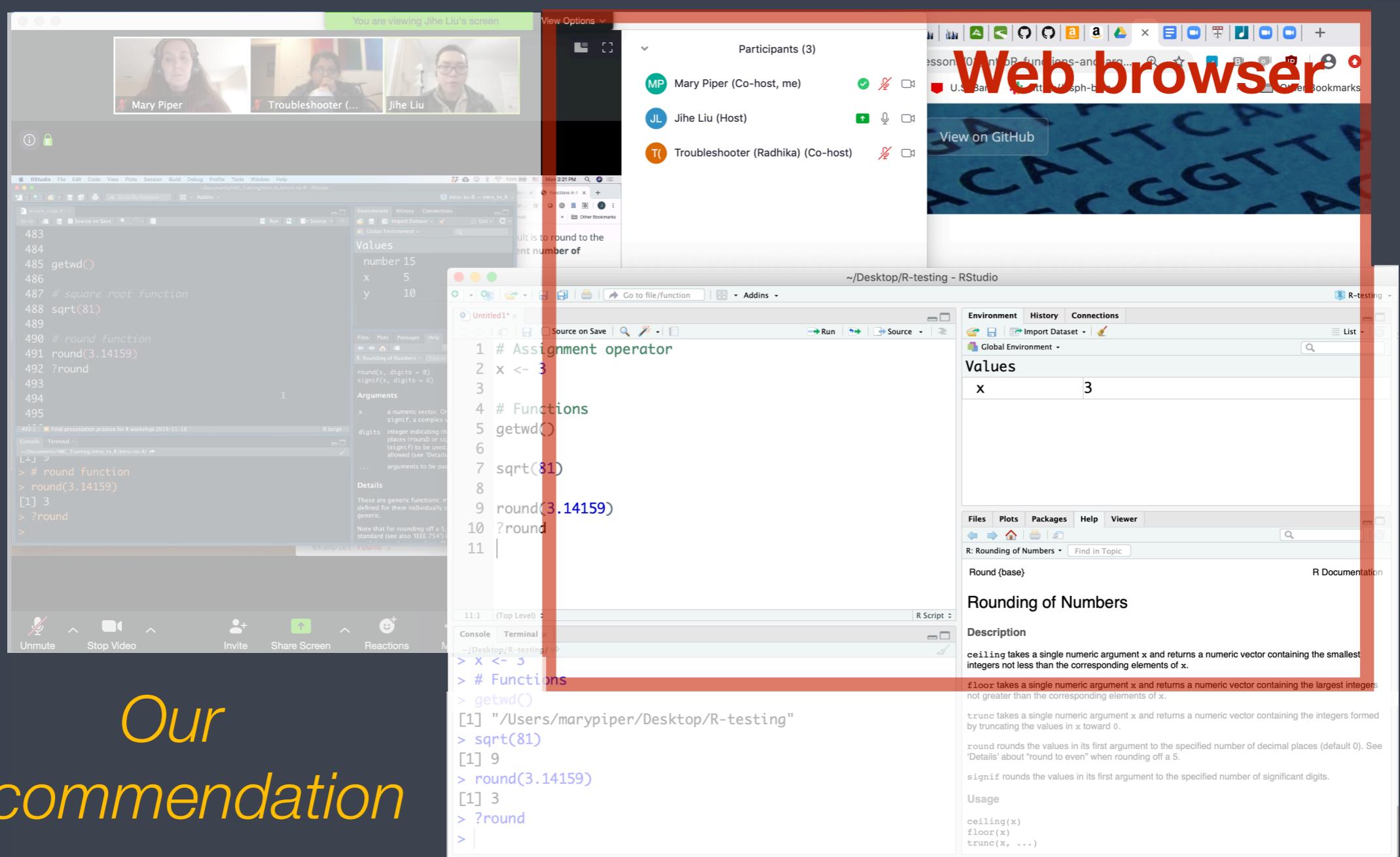


Single screen & 3 windows?



*Our
recommendation*

Single screen & 3 windows?



Our
recommendation

Single screen & 3 windows?

The screenshot shows a video conference interface with three windows:

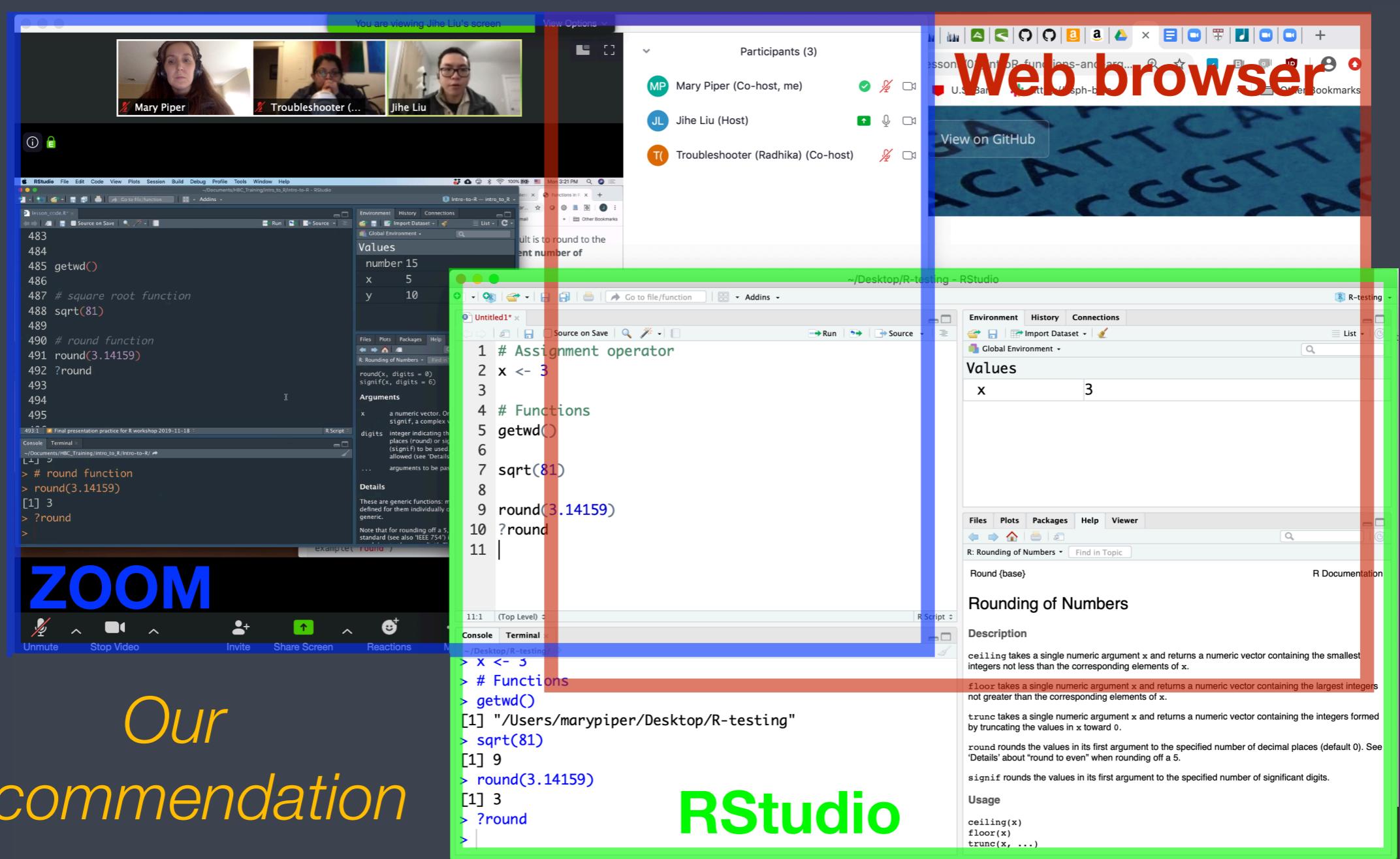
- Top Left Window:** A video feed showing three participants: Mary Piper, Troubleshooter (Radhika), and Jihe Liu.
- Middle Left Window:** An RStudio session titled "intro_to_R -- intro_to_R.R" containing the following R code:

```
483  
484  
485 getwd()  
486  
487 # square root function  
488 sqrt(81)  
489  
490 # round function  
491 round(3.14159)  
492 ?round  
493  
494  
495
```
- Bottom Left Window:** An RStudio session titled "Untitled1*" containing the following R code:

```
1 # Assignment operator  
2 x <- 3  
3  
4 # Functions  
5 getwd()  
6  
7 sqrt(81)  
8  
9 round(3.14159)  
10 ?round  
11
```
- Top Right Window:** A web browser window showing a presentation slide with the title "R: Rounding of Numbers". The slide includes sections for "Description", "Details", and "Usage". It also contains a screenshot of the RStudio session from the bottom left window.
- Bottom Right Window:** A large green box highlighting the RStudio session from the bottom left window, specifically the "Assignment operator" section.

Our recommendation: RStudio

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!



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
- ❖ Homework load can be heavier in the beginning of this workshop series, but it tapers off

Odds and Ends (1/2)

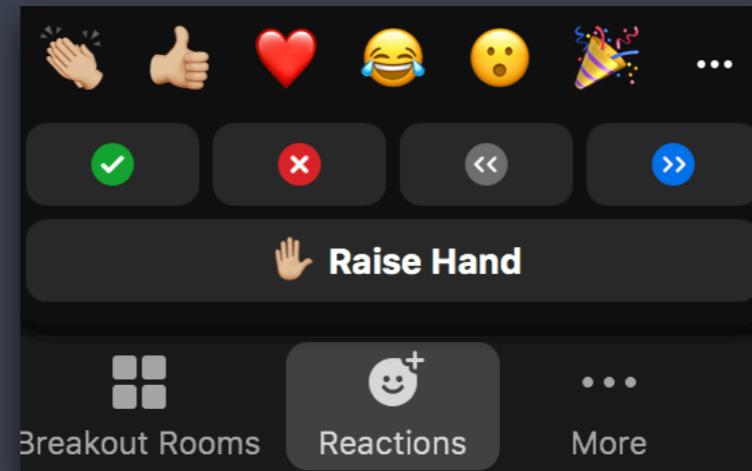
- ❖ Quit/minimize all applications that are not required for class

Odds and Ends (1/2)

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

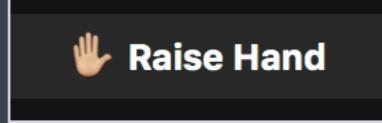
Odds and Ends (1/2)

- ❖ 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**)



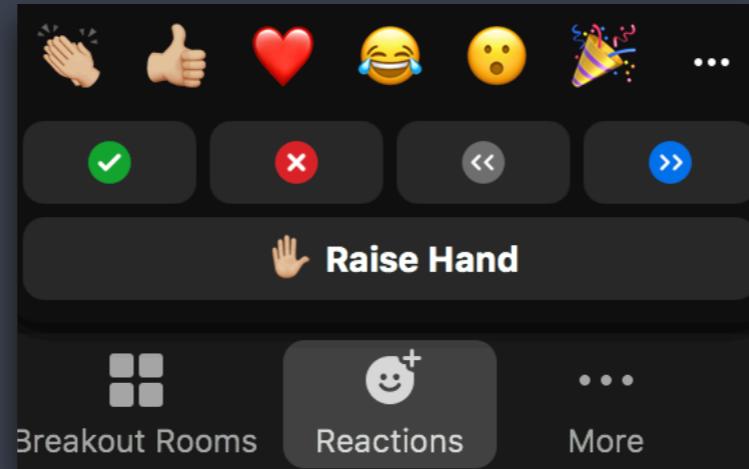
Odds and Ends (2/2)

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



Contact us!

HBC training team: hbctraining@hsph.harvard.edu

HBC consulting: bioinformatics@hsph.harvard.edu