



```
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



John Hutchinson
Associate Director



Victor Barrera



Zhu Zhuo



Preetida Bhetariya



Radhika Khetani
Training Director



Meeta Mistry



Jihe Liu



Will Gammerdinger



Maria Simoneau



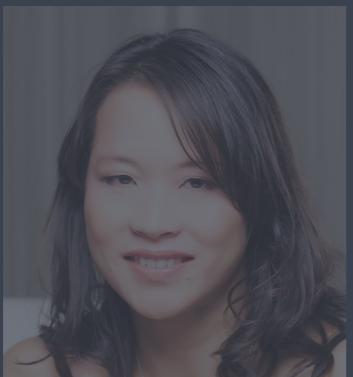
James Billingsley



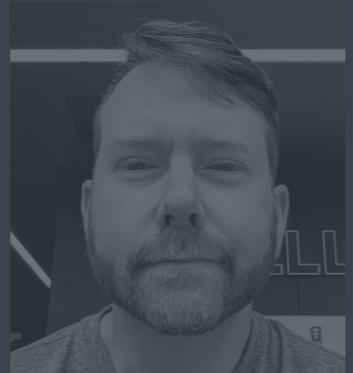
Sergey Naumenko



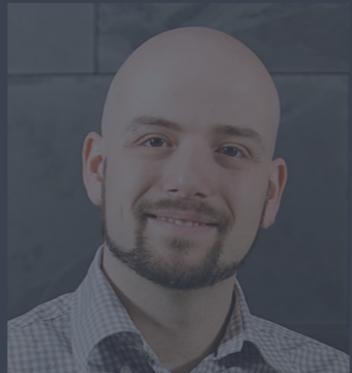
Peter Kraft
Faculty Advisor



Shannan Ho Sui
Director



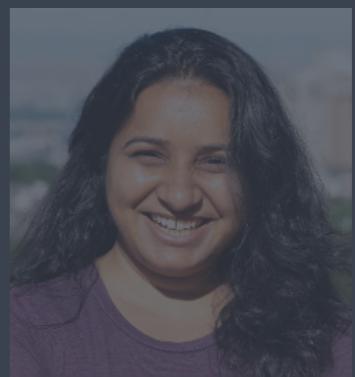
John Hutchinson
Associate Director



Victor Barrera



Zhu Zhuo



Preetida Bhetariya



Radhika Khetani
Training Director



Meeta Mistry



Jihe Liu



Will Gammerdinger



Maria Simoneau



James Billingsley



Sergey Naumenko



Peter Kraft
Faculty Advisor

Consulting

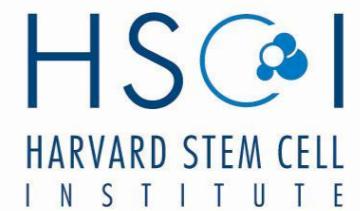
- RNA-seq analysis: bulk, single cell, small RNA
- ChIP-seq and ATAC-seq analysis
- Genome-wide methylation
- WGS, resequencing, exome-seq and CNV studies
- QC & analysis of gene expression arrays
- Functional enrichment analysis
- Grant support

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



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

NIEHS



Training



1. Basic Data Skills - No prior programming knowledge needed (no prerequisites)
2. Advanced Topics: Analysis of high-throughput sequencing (NGS) data - Certain “Basic” workshops are prerequisites.
3. Current Topics in Bioinformatics Series - Free, monthly, 3-hour workshops (often hands-on)

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

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



Training



HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

1. Basic Data Skills - No prior programming knowledge needed (no prerequisites)
2. Advanced Topics: Analysis of high-throughput sequencing (NGS) data - Certain “Basic” workshops are prerequisites.
3. Current Topics in Bioinformatics Series - Free, monthly, 3-hour workshops (often hands-on)

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

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

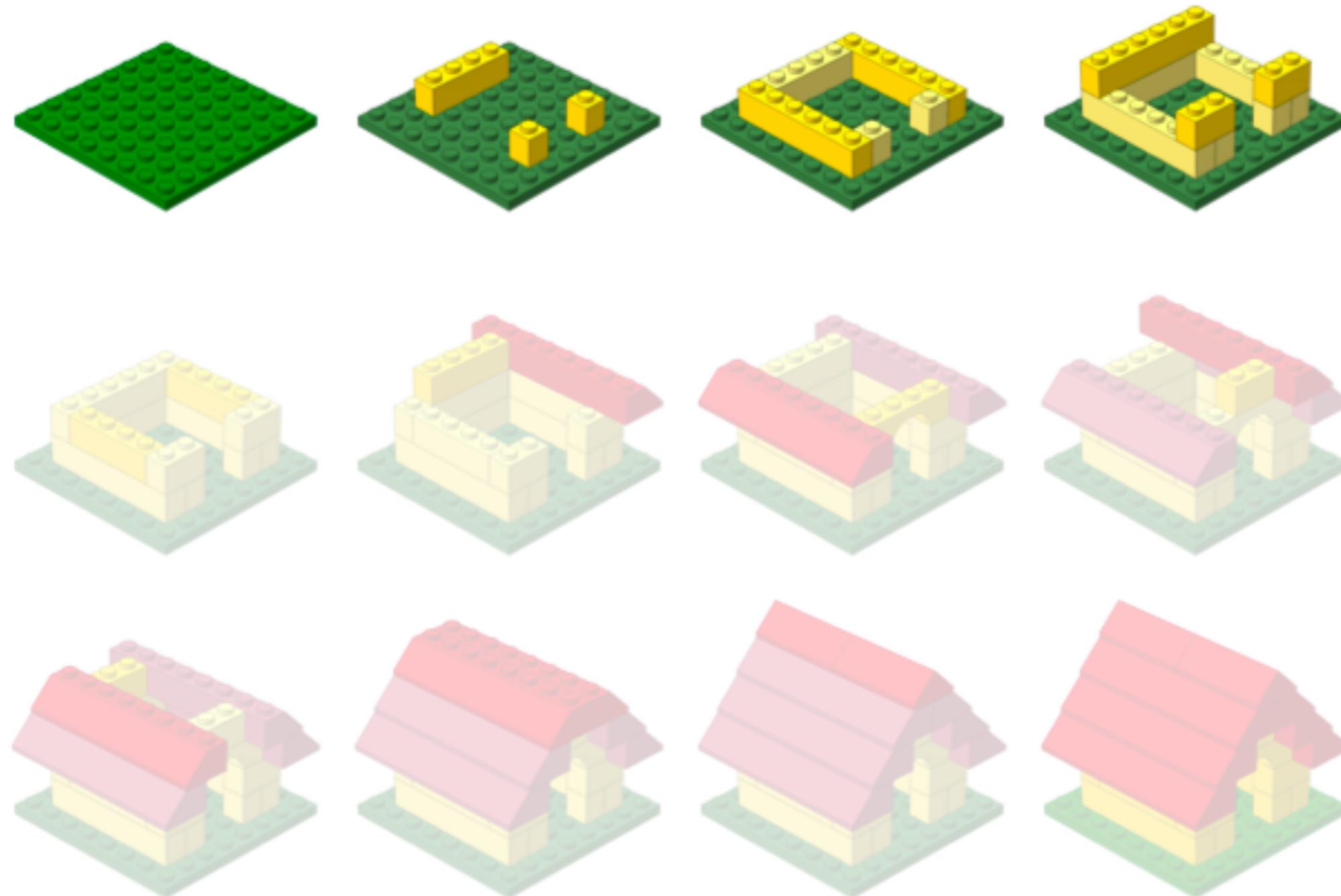
DF/HCC
DANA-FARBER / HARVARD CANCER CENTER

HSCI
HARVARD STEM CELL
INSTITUTE

 HARVARD
CATALYST
THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

 HARVARD
MEDICAL SCHOOL

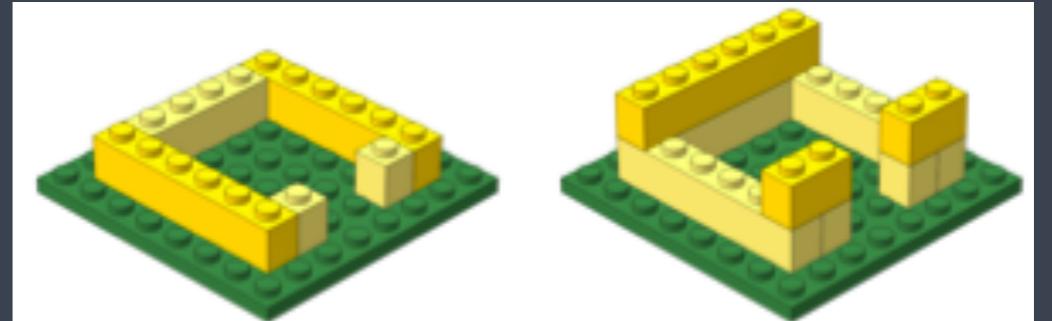
Workshop Scope...



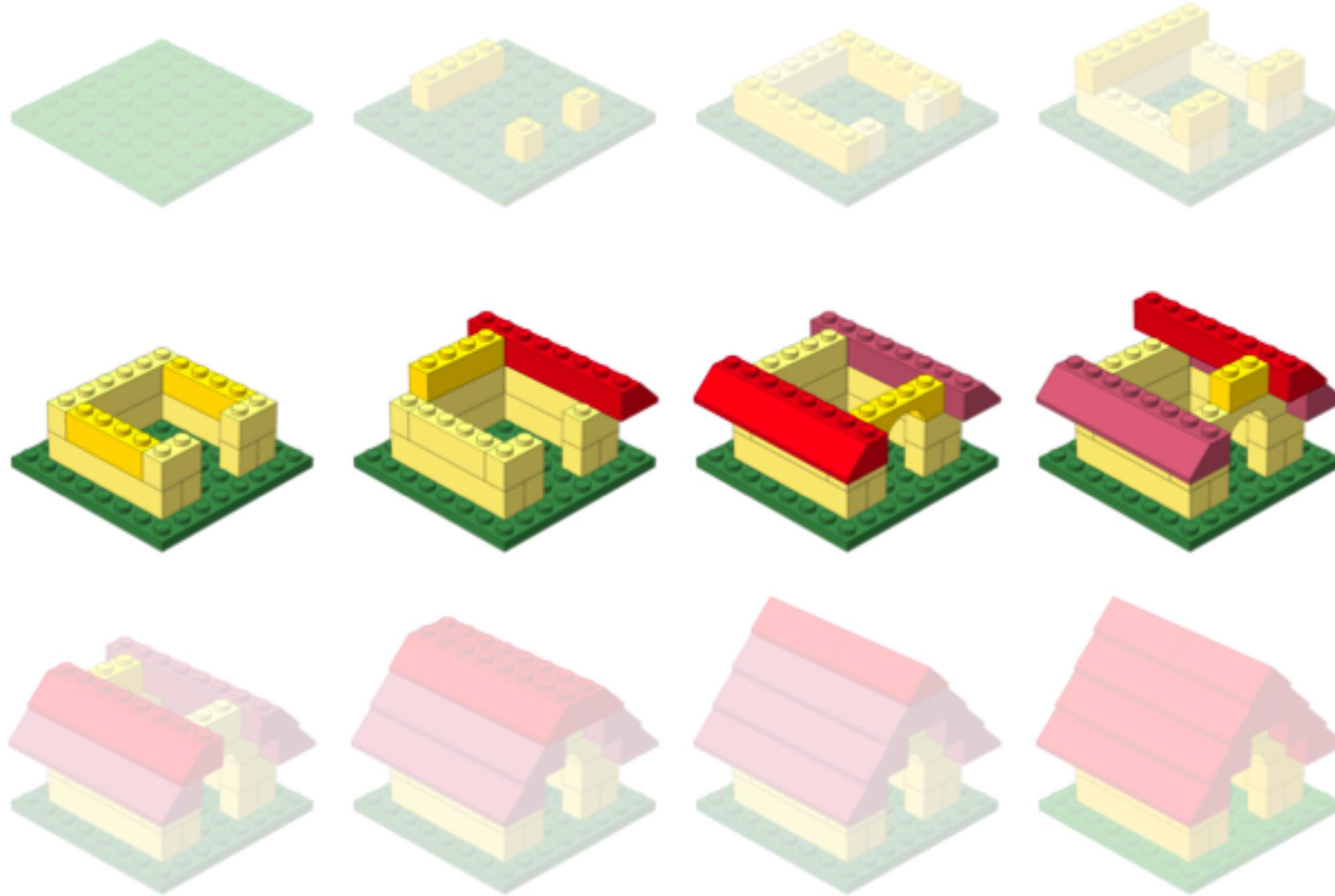
<http://anoved.net/tag/lego/page/3/>

Setting up to perform Bioinformatics analysis

Setting up...



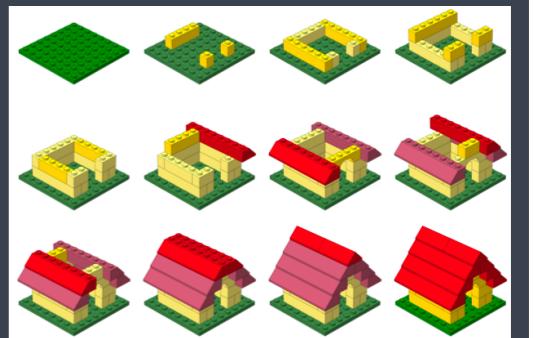
- ✓ Introduction to the command-line interface (shell, Unix, Linux)
 - Dealing with large data files
 - Performing bioinformatics analysis
 - Using tools
 - Accessing and using compute clusters
- ✓ R
 - Parsing and working with smaller results text files
 - Statistical analysis, e.g. differential expression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

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

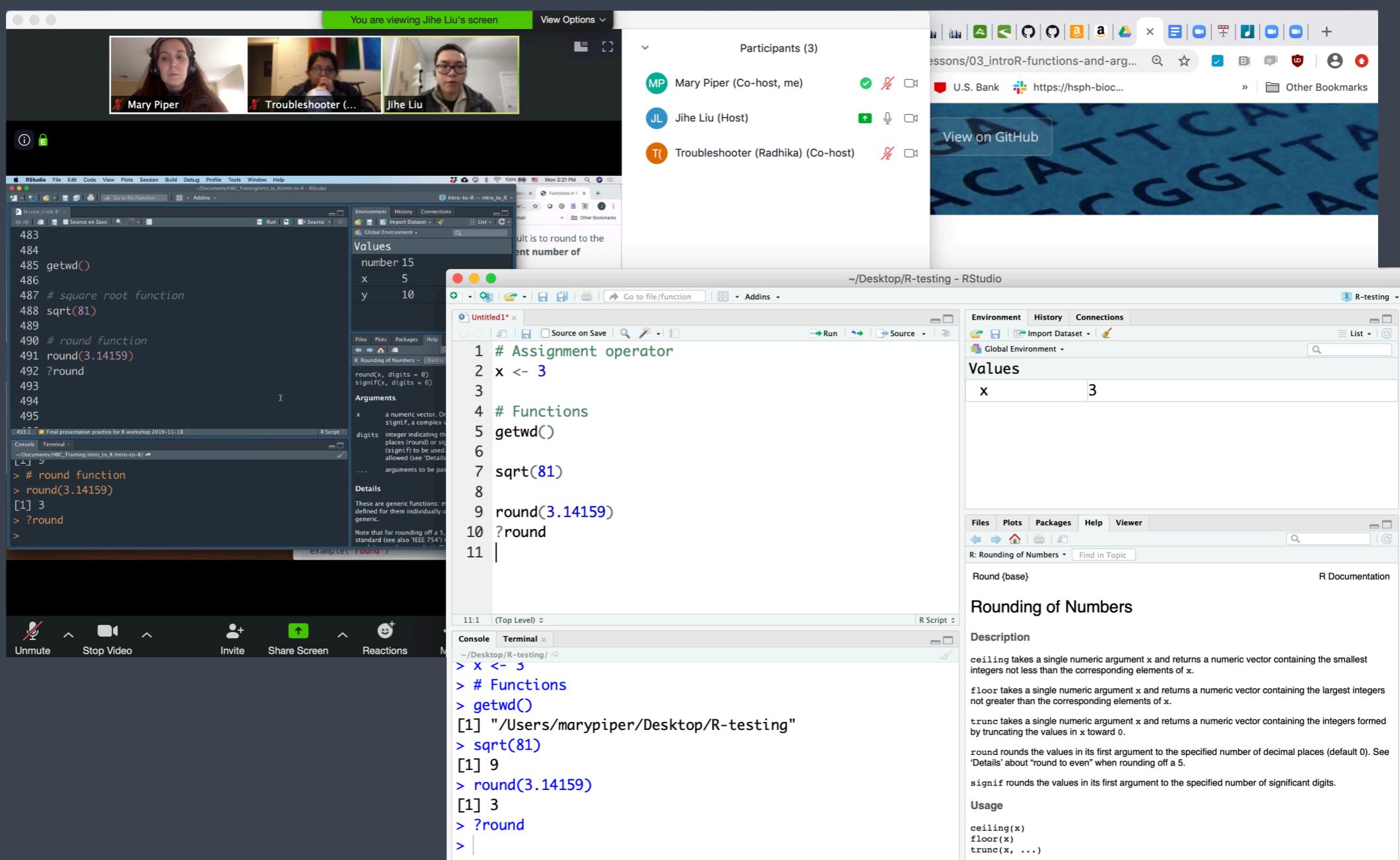
Normalization

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

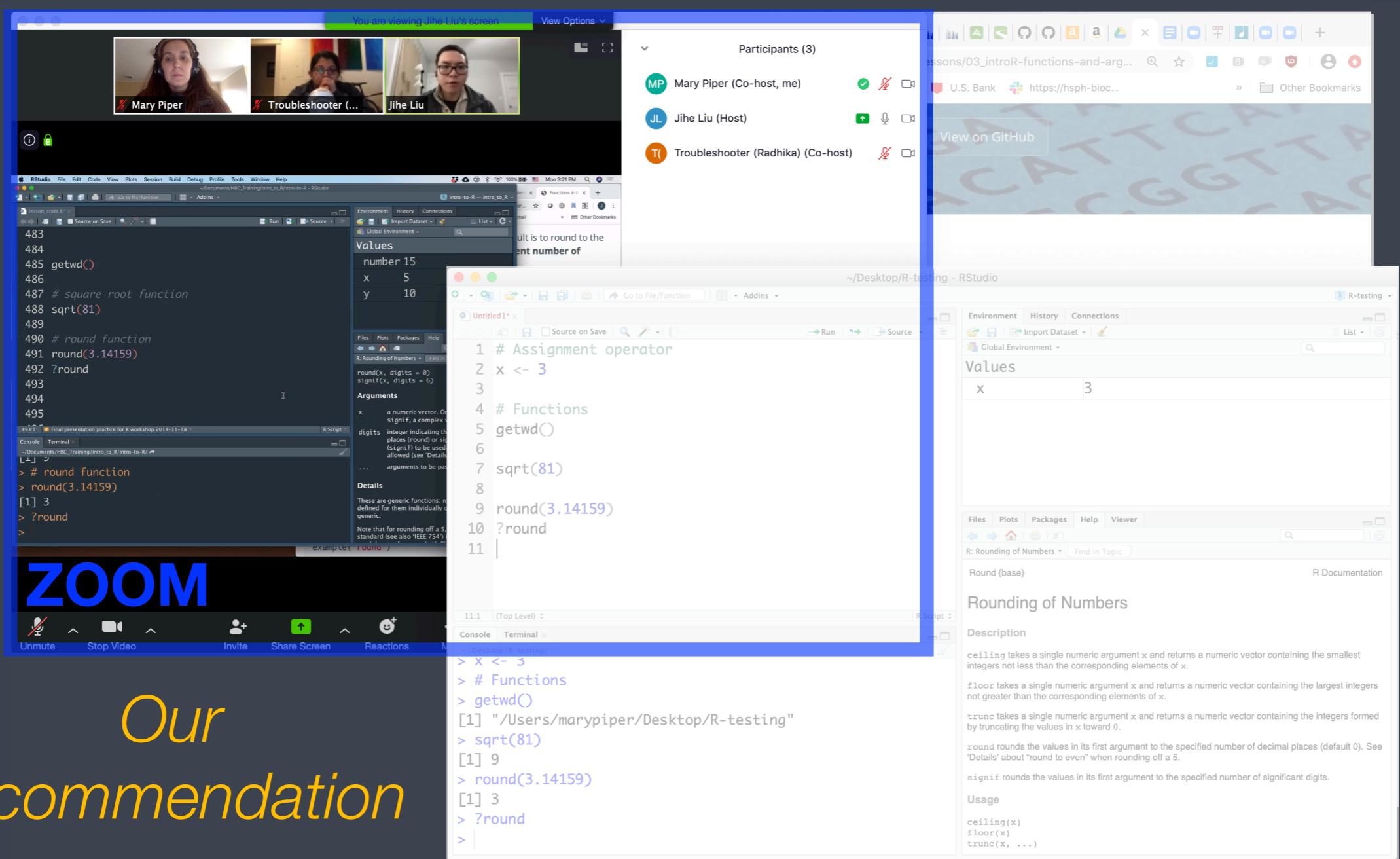
```
graph TD; A["Pseudocounts with  
Kallisto, Sailfish, Salmon"] --> B["Read counts  
associated with genes"]; B --> C["Normalization"]; C --> D["Unsupervised clustering analyses"]; C -.-> E["Quality control"]
```

The flowchart illustrates the DE analysis workflow. It begins with a purple box labeled "Pseudocounts with Kallisto, Sailfish, Salmon". An arrow points down to a pink box labeled "Read counts associated with genes". Another arrow points down to a green box labeled "Normalization". From "Normalization", an arrow points down to a light green box labeled "Unsupervised clustering analyses". A bracket on the right side of the flowchart groups the "Normalization" and "Unsupervised clustering analyses" boxes under the label "Quality control".

Single screen & 3 windows?

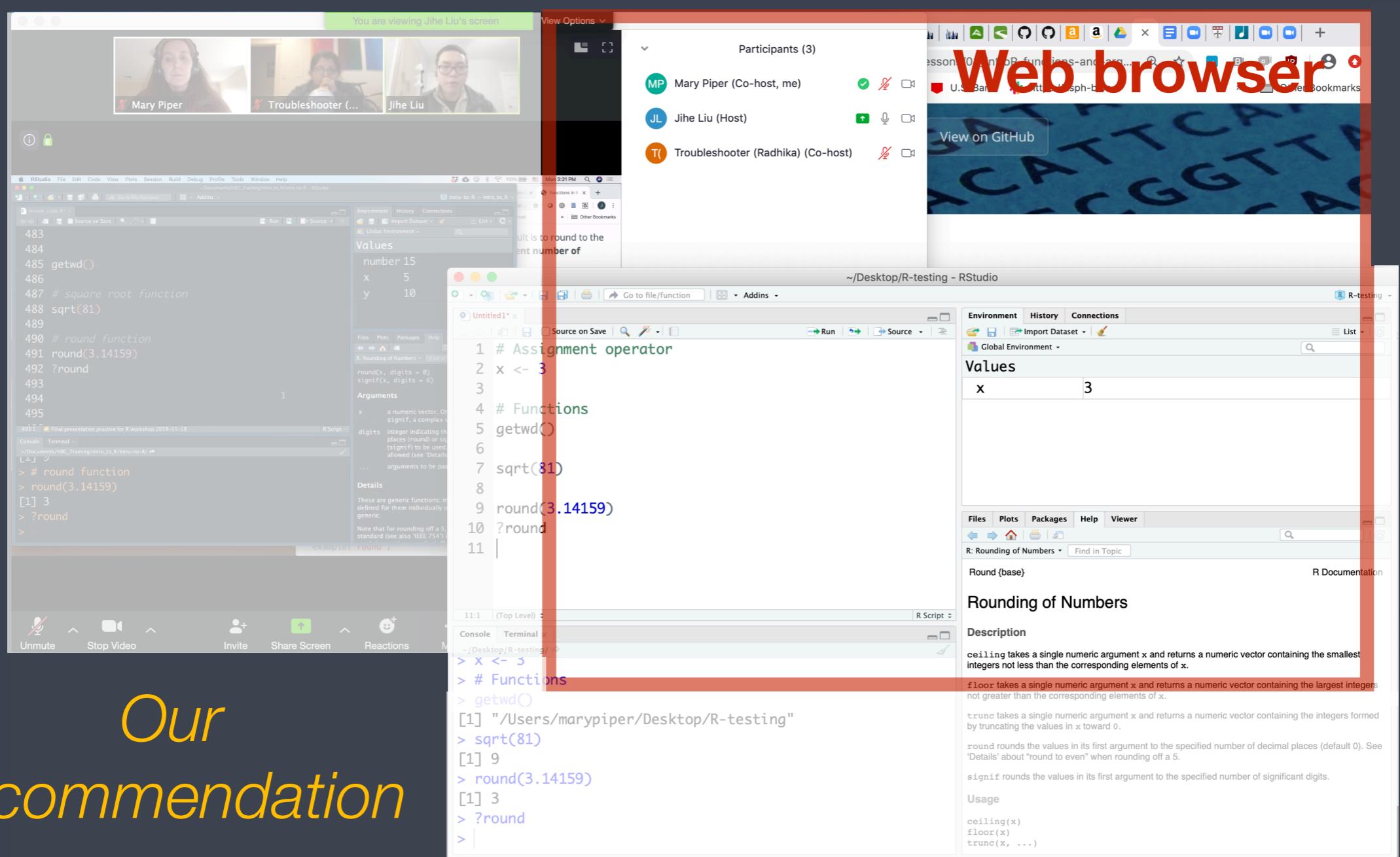


Single screen & 3 windows?



*Our
recommendation*

Single screen & 3 windows?



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". It contains the following R code:

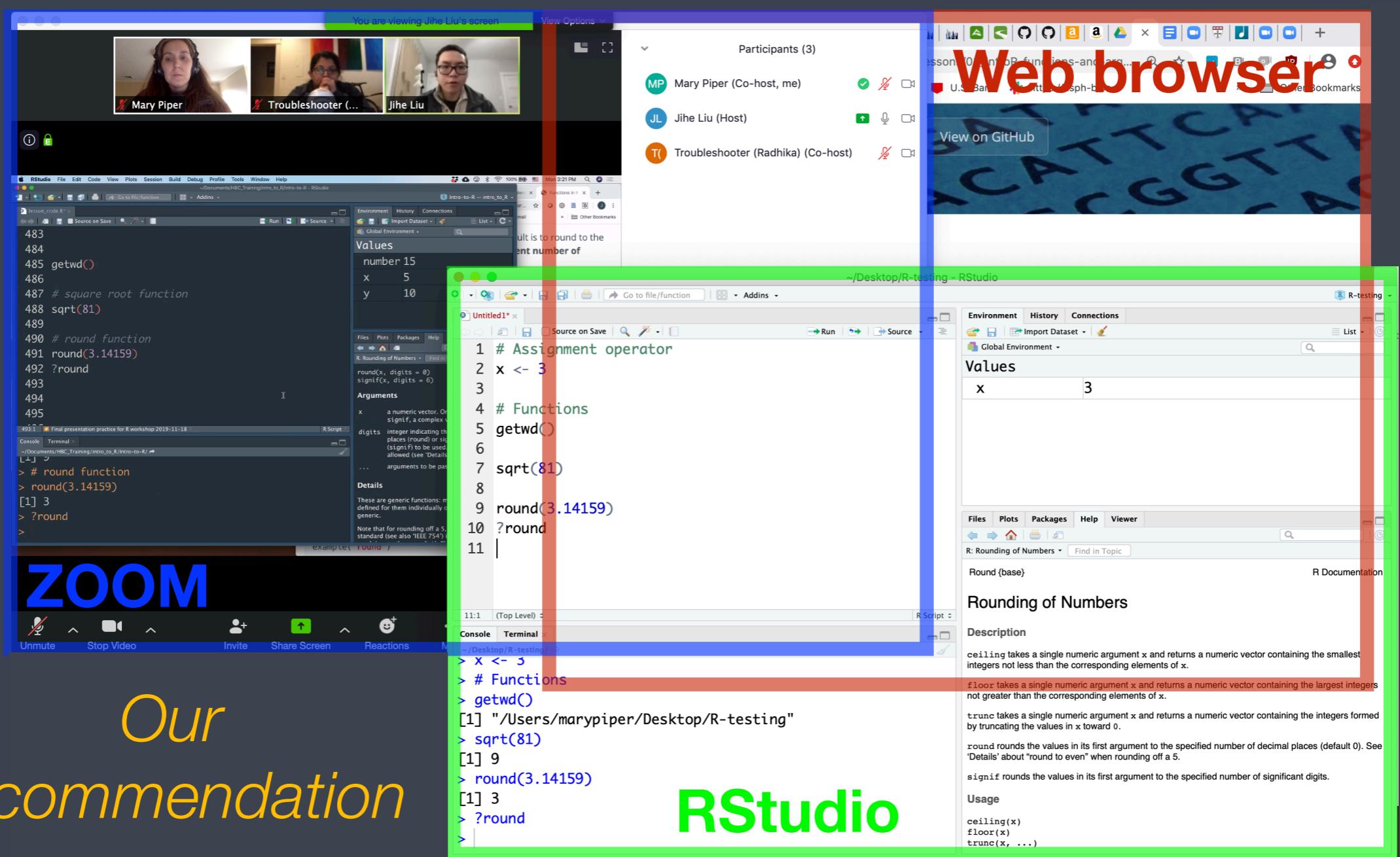
```
483  
484  
485 getwd()  
486  
487 # square root function  
488 sqrt(81)  
489  
490 # round function  
491 round(3.14159)  
492 ?round  
493  
494  
495
```
- Bottom Left Window:** An RStudio session titled "Untitled1*". It contains the following R code:

```
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 GitHub page for "introR-functions-and-args".
- Middle Right Window:** An RStudio session titled "~/Desktop/R-testing - RStudio". It contains the following R code:

```
11:1 (Top Level) >  
Console Terminal x  
~/Desktop/R-testing/ >  
> x <- 3  
> # Functions  
> getwd()  
[1] "/Users/marypiper/Desktop/R-testing"  
> sqrt(81)  
[1] 9  
> round(3.14159)  
[1] 3  
> ?round  
>
```
- Bottom Right Window:** A detailed view of the "round" function documentation from the R documentation website.

Our recommendation: RStudio

Single screen & 3 windows?

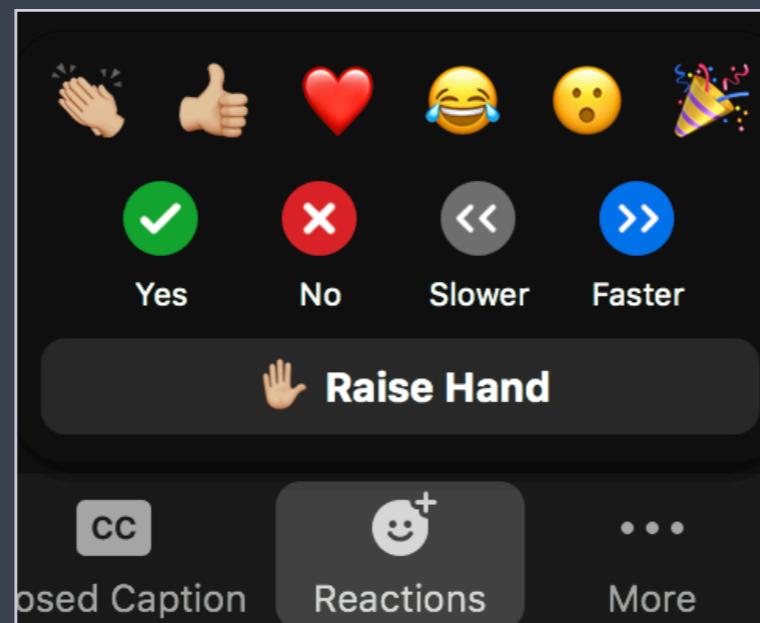


Odds and Ends

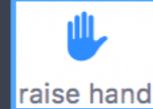
- ❖ Quit/minimize all applications that are not required for class

Odds and Ends

- ❖ 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**)
 - Type “X” in chat = "disagree", “I’m not ready” (equivalent to a **red post-it**)



Odds and Ends (2/2)

- ❖ Questions for the presenter?
 - Post the question in the Chat window OR
 - Raise your hand  when the presenter asks for questions
- ❖ Technical difficulties with R or RStudio?
 - 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

Twitter

[@bioinfocore](https://twitter.com/bioinfocore)