



```
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")
```

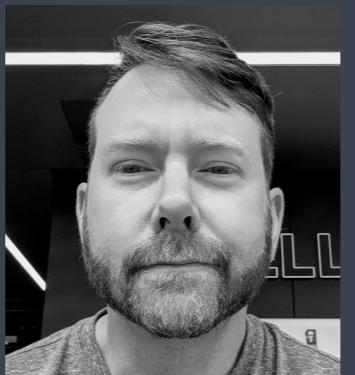
Introduction to Differential Gene Expression Analysis

Harvard Chan Bioinformatics Core

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



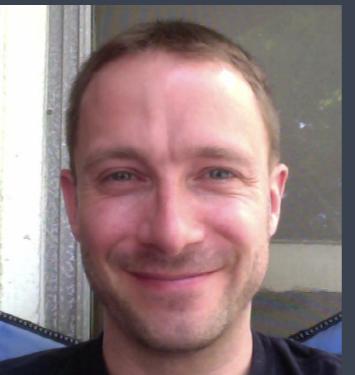
Shannan Ho Sui
Director



John Hutchinson
Associate Director



Victor Barrera



Rory Kirchner



Zhu Zhuo



Preetida Bhetariya



Meeta Mistry



Mary Piper



Jihe Liu



Radhika Khetani
Training Director



Ilya Sytchev



James Billingsley



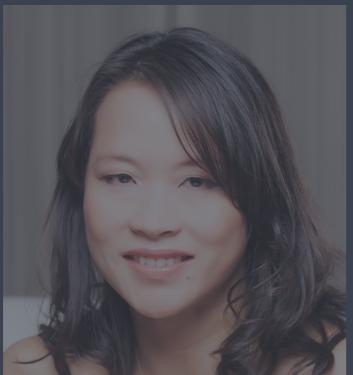
Sergey Naumenko



Joon Yoon



Peter Kraft
Faculty Advisor



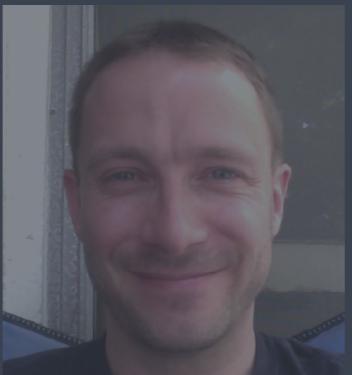
Shannan Ho Sui
Director



John Hutchinson
Associate Director



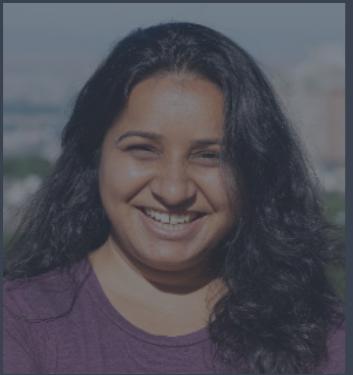
Victor Barrera



Rory Kirchner



Zhu Zhuo



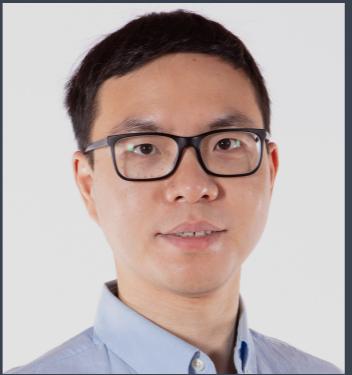
Preetida Bhetariya



Meeta Mistry



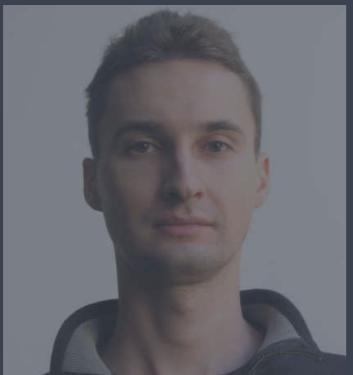
Mary Piper



Jihe Liu



Radhika Khetani
Training Director



Ilya Sytchev



James Billingsley



Sergey Naumenko



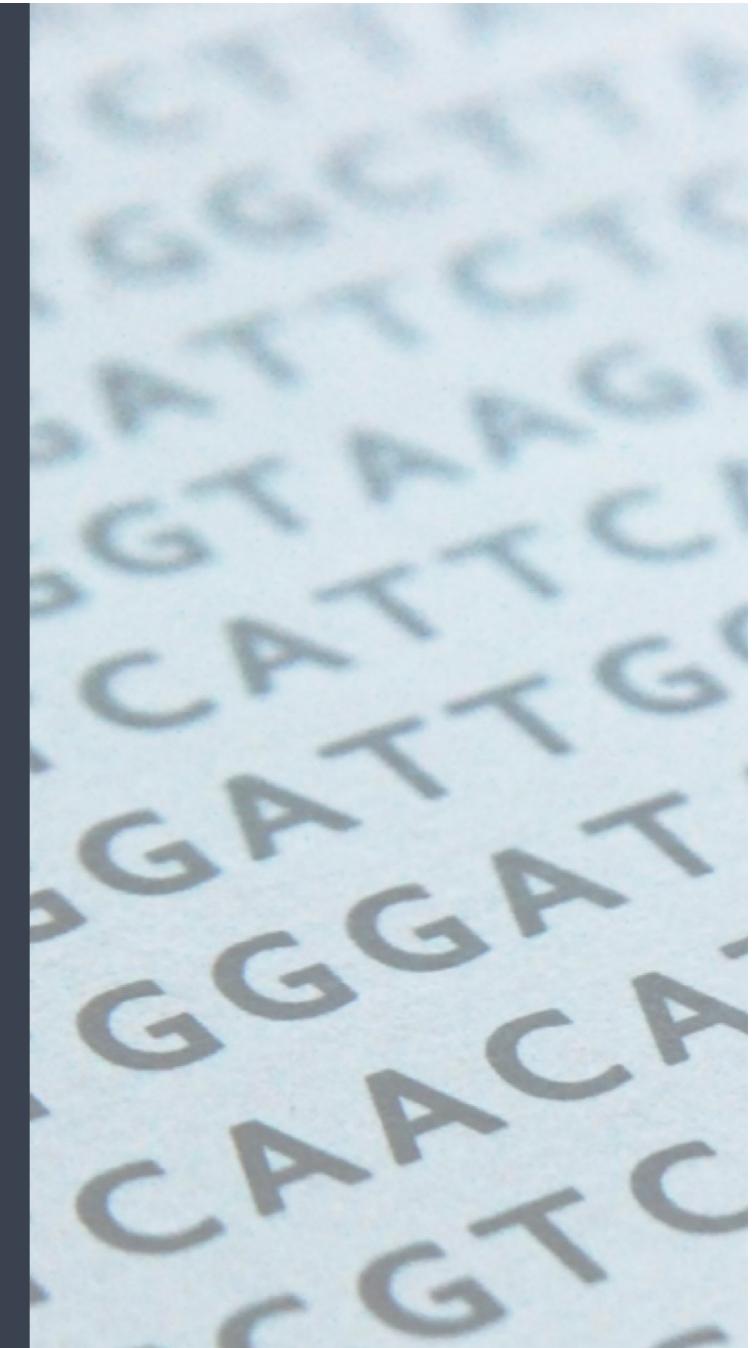
Joon Yoon



Peter Kraft
Faculty Advisor

Consulting

- RNA-seq, small RNA-seq and ChIP-seq analysis
- Genome-wide methylation
- WGS, resequencing, exome-seq and CNV studies
- Quality assurance and analysis of gene expression arrays
- Functional enrichment analysis
- Grant support



Training

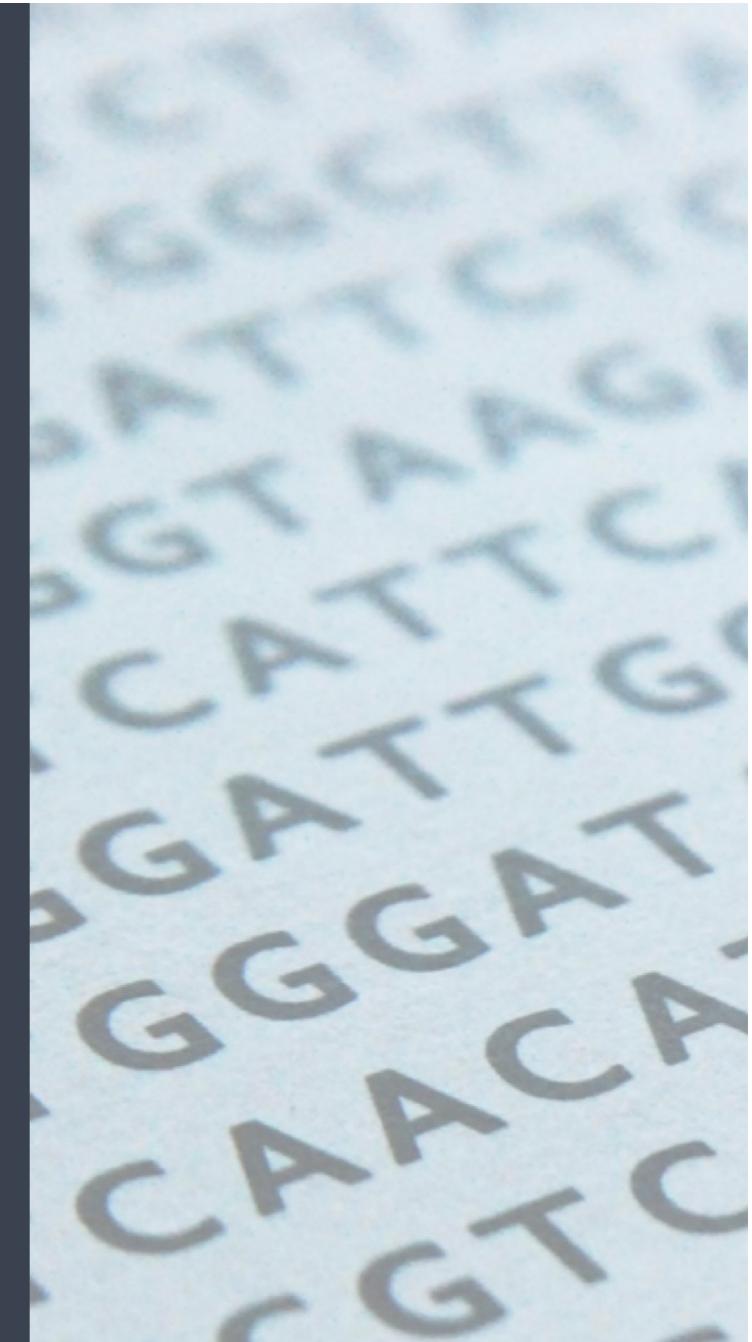
Our short workshops are divided into 2 categories:

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/>





HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

NIEHS / CFAR
Bioinformatics
Core

HSCI
HARVARD STEM CELL
INSTITUTE

Center for Stem
Cell
Bioinformatics

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

 **HARVARD**
MEDICAL SCHOOL

Harvard
Catalyst
Bioinformatics
Consulting

HMS
Foundry



HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

NIEHS / CFAR
Bioinformatics
Core

HSCI
HARVARD STEM CELL
INSTITUTE

Center for Stem
Cell
Bioinformatics

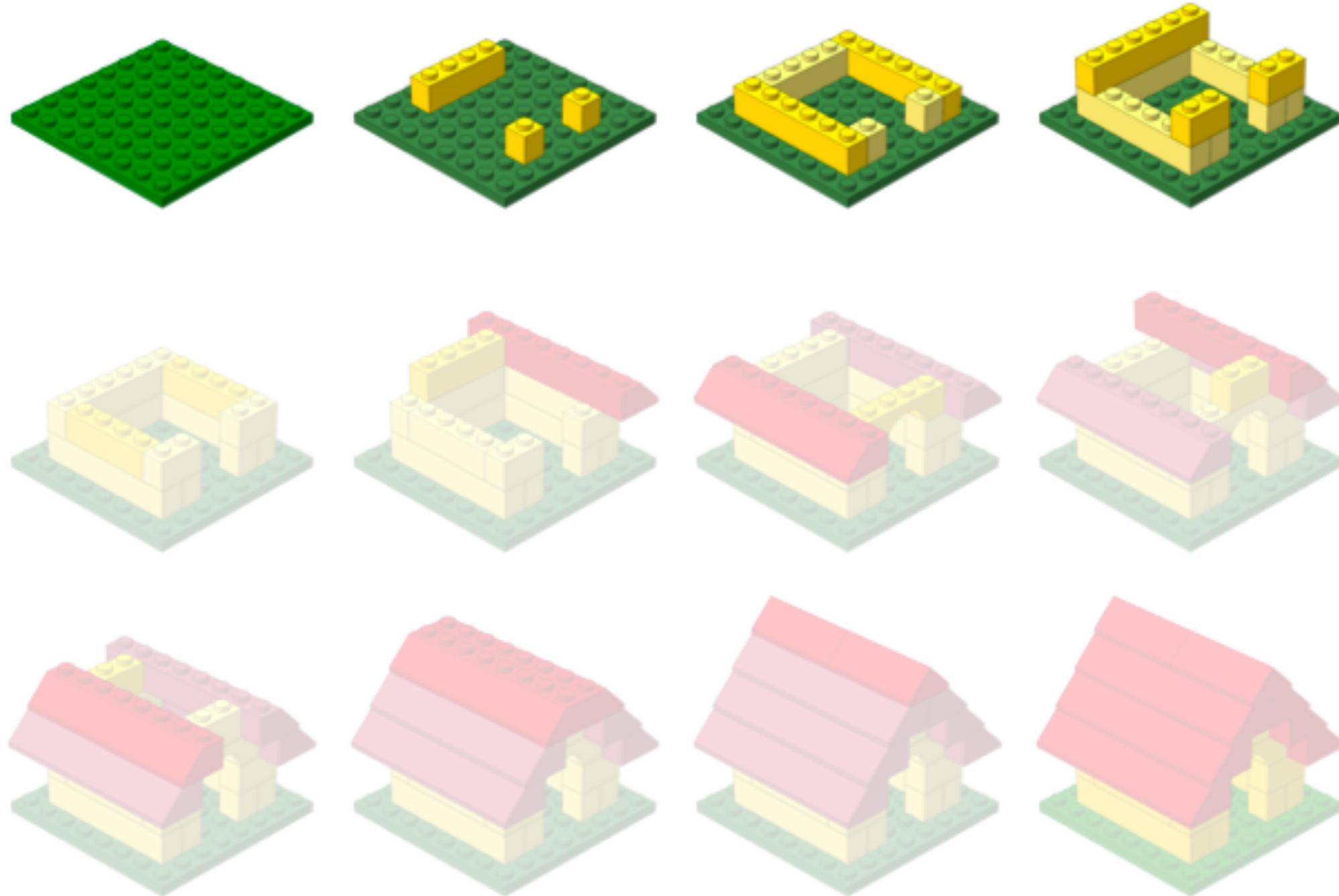
 **HARVARD CATALYST**
THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

 **HARVARD**
MEDICAL SCHOOL

Harvard
Catalyst
Bioinformatics
Consulting

HMS
Foundry

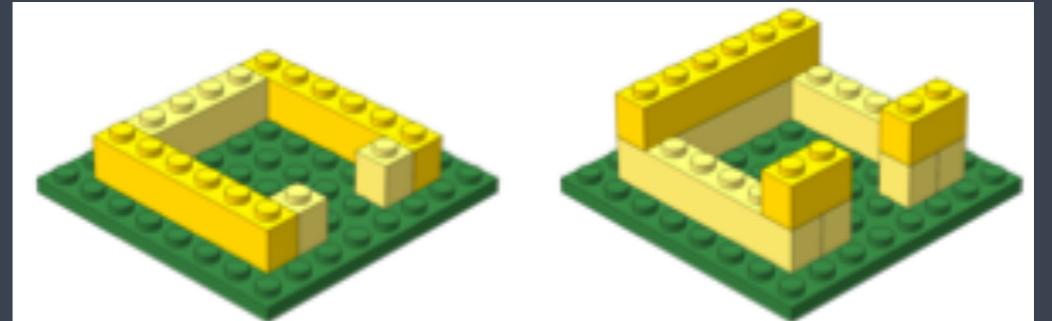
Workshop Scope...



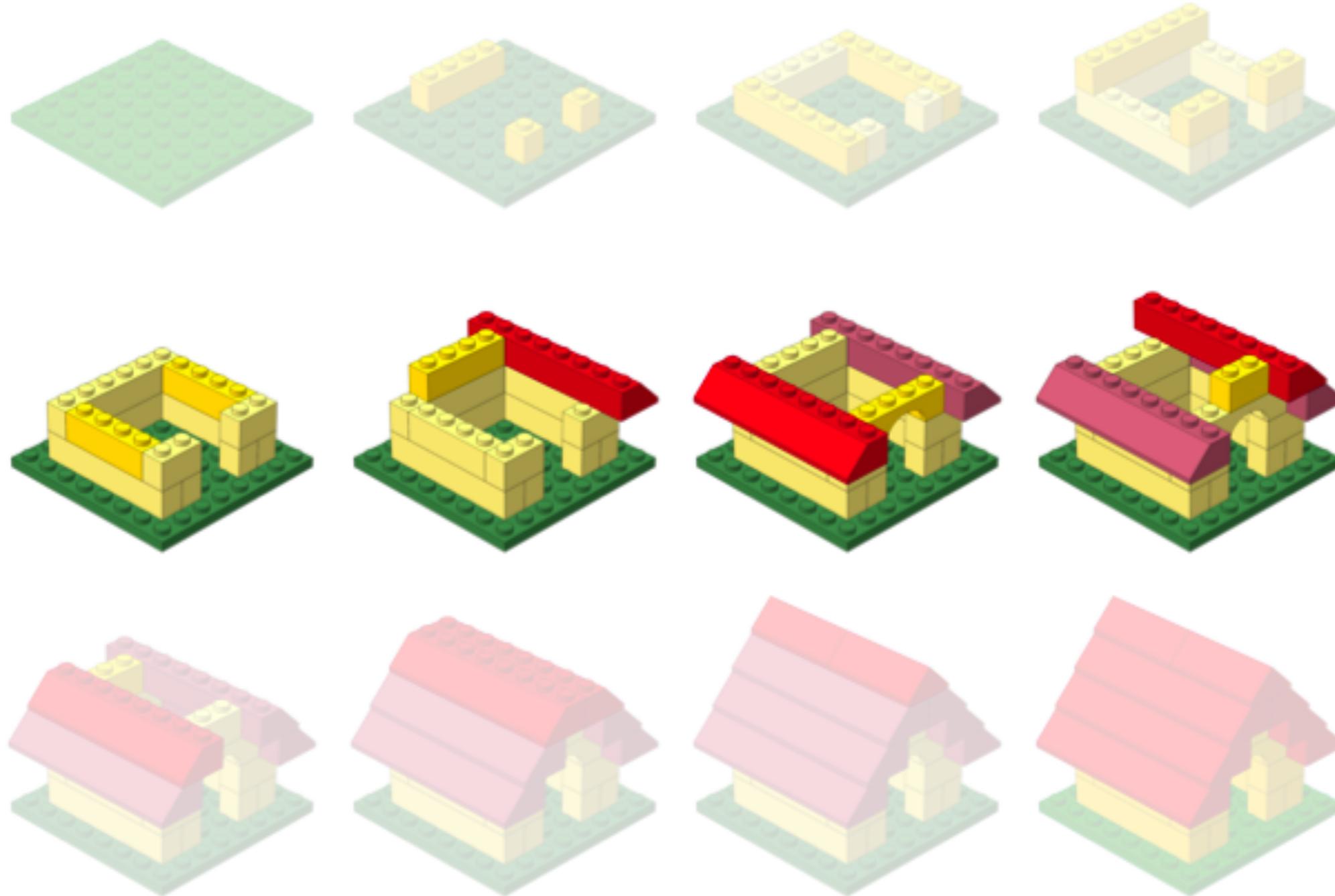
<http://anoved.net/tag/lego/page/3/>

Setting up to perform Bioinformatics analysis

Setting up...



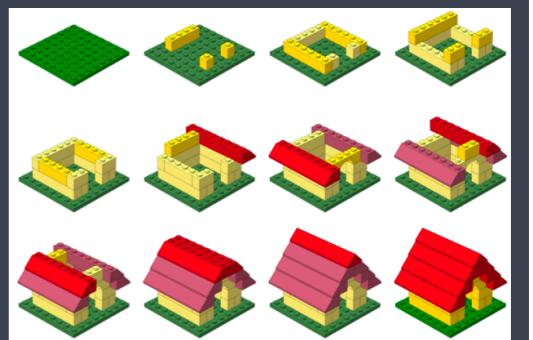
- ✓ Introduction to the command-line interface (shell, Unix, Linux)
 - Dealing with large data files
 - Performing bioinformatics analysis
 - Using tools
 - Accessing and using compute clusters
- ✓ R
 - Parsing and working with smaller results text files
 - Statistical analysis, e.g. differential expression 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	Radhika
10:30 - 10:45	R refresher Q & A	Mary
10:45 - 11:15	RNA-seq pre-reading discussion	Meeta/Mary
11:15 - 12:00	Intro to DGE / setting up DGE analysis	Meeta

Self-Learning Part 1

1. [RNA-seq counts distribution](#)
2. [Count normalization](#)
3. Sample-level QC (PCA and hierarchical clustering)

Assignment #1

- R script
- Dropbox link
- [Email us](#) about questions regarding the homework that you need answered before the next class.
- Post 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

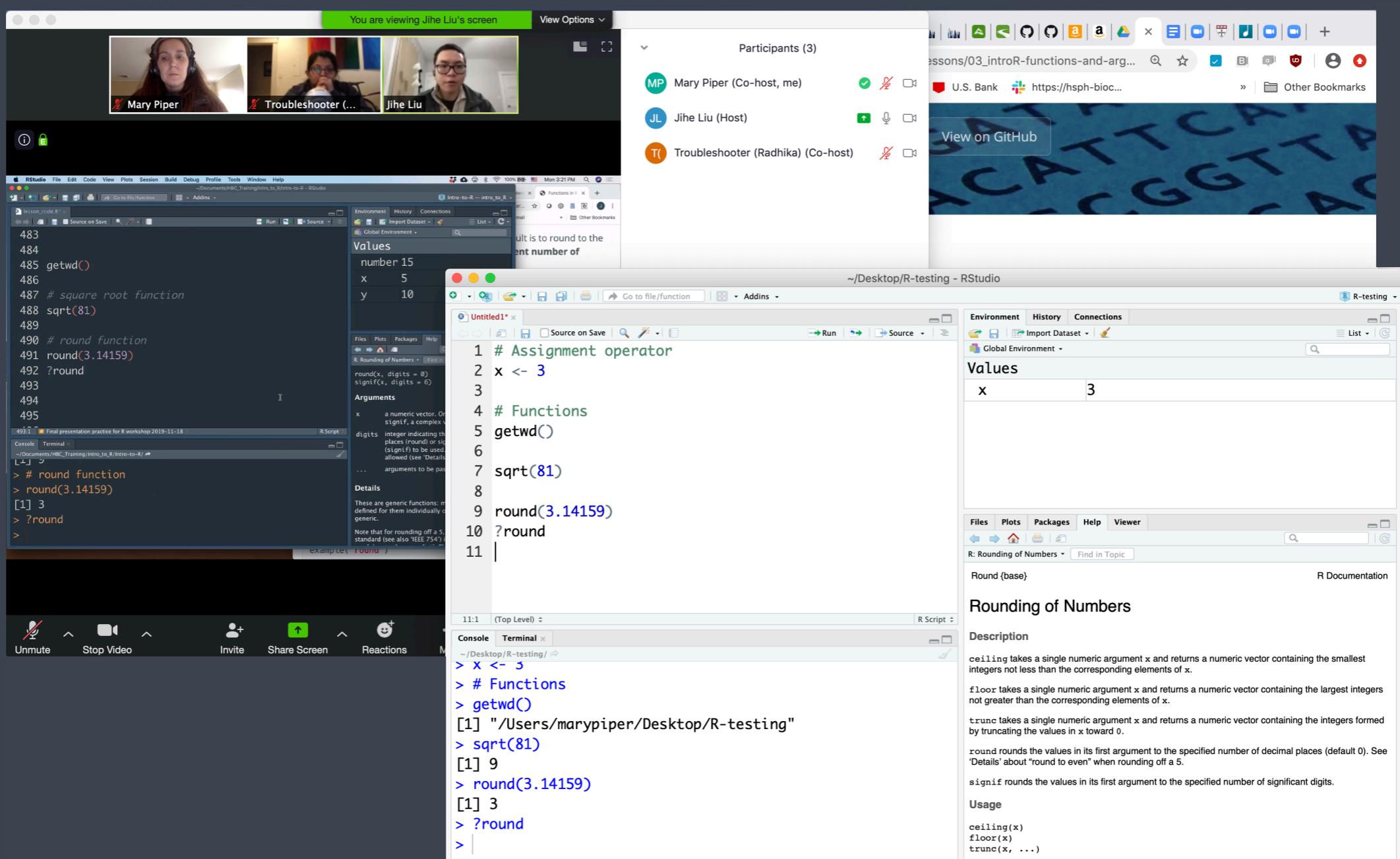
Normalization

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

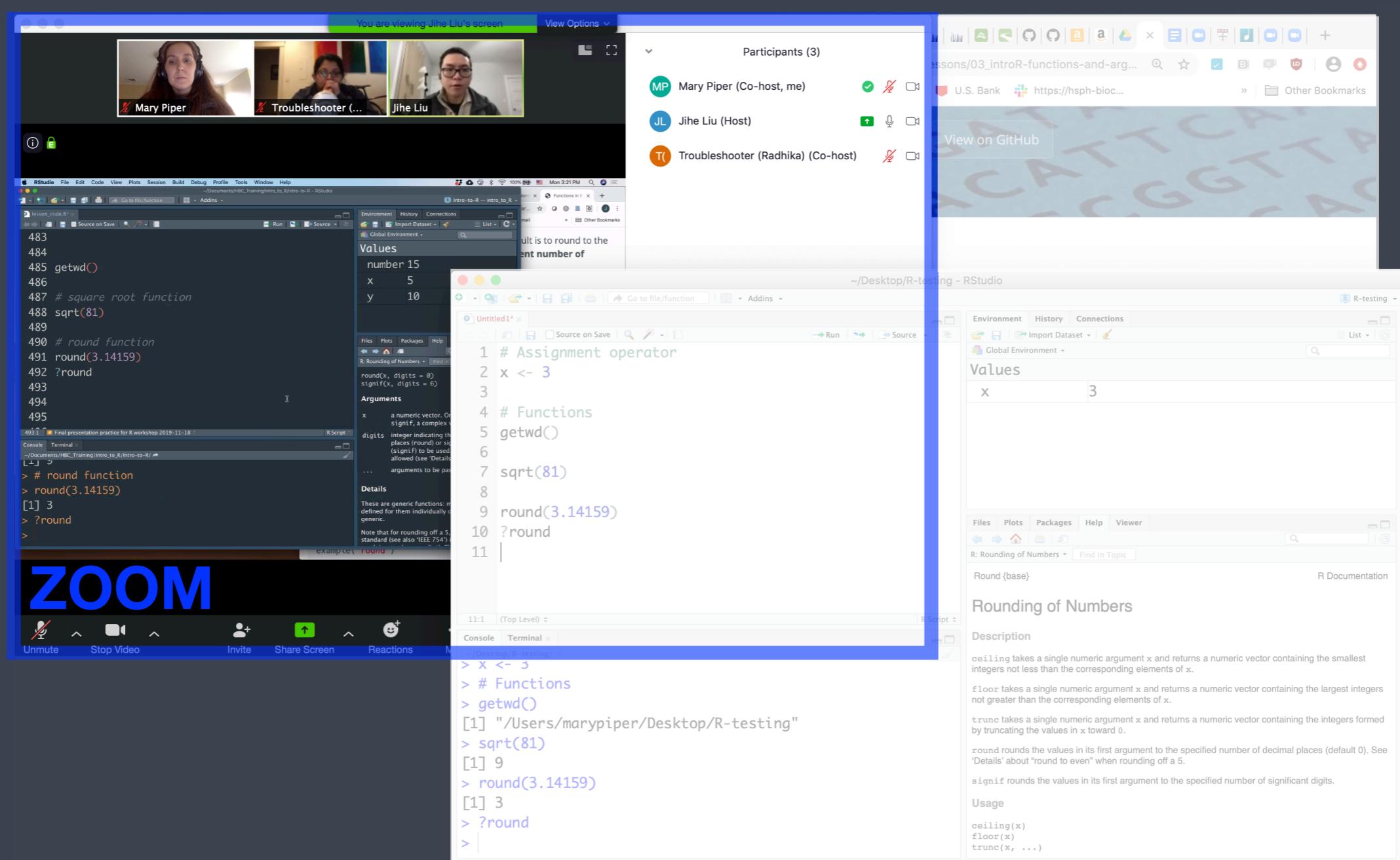
```
graph TD; A["Pseudocounts with  
Kallisto, Sailfish, Salmon"] --> B["Read counts  
associated with genes"]; B --> C["Normalization"]; C --> D["Unsupervised clustering analyses"]; C -.-> E["Quality control"]
```

The flowchart illustrates the DE analysis workflow. It begins with 'Pseudocounts with Kallisto, Sailfish, Salmon', followed by 'Read counts associated with genes'. This leads to 'Normalization', which then leads to 'Unsupervised clustering analyses'. A bracket on the right side groups 'Normalization' and 'Unsupervised clustering analyses' under the heading 'Quality control'.

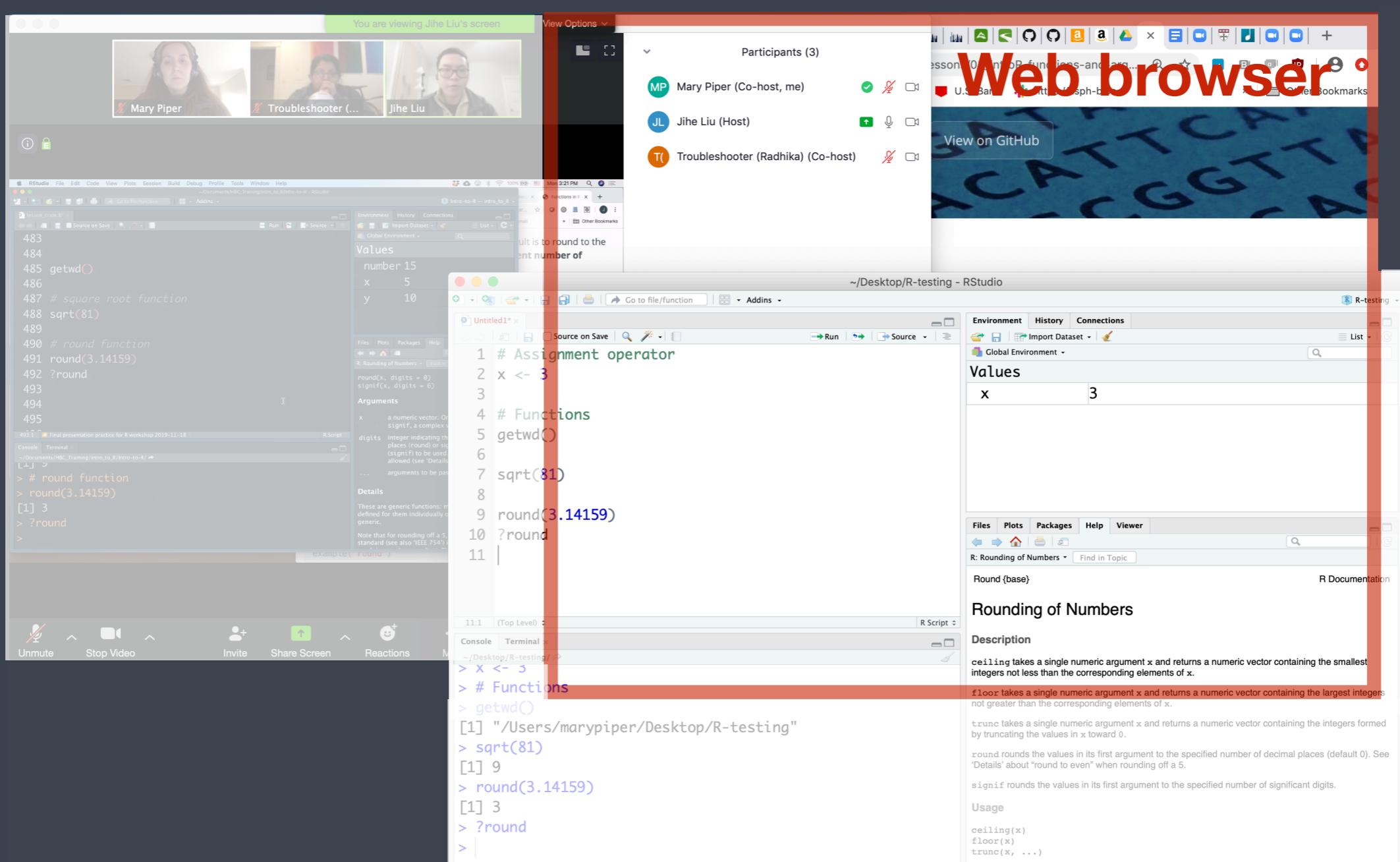
The 3 Window problem...



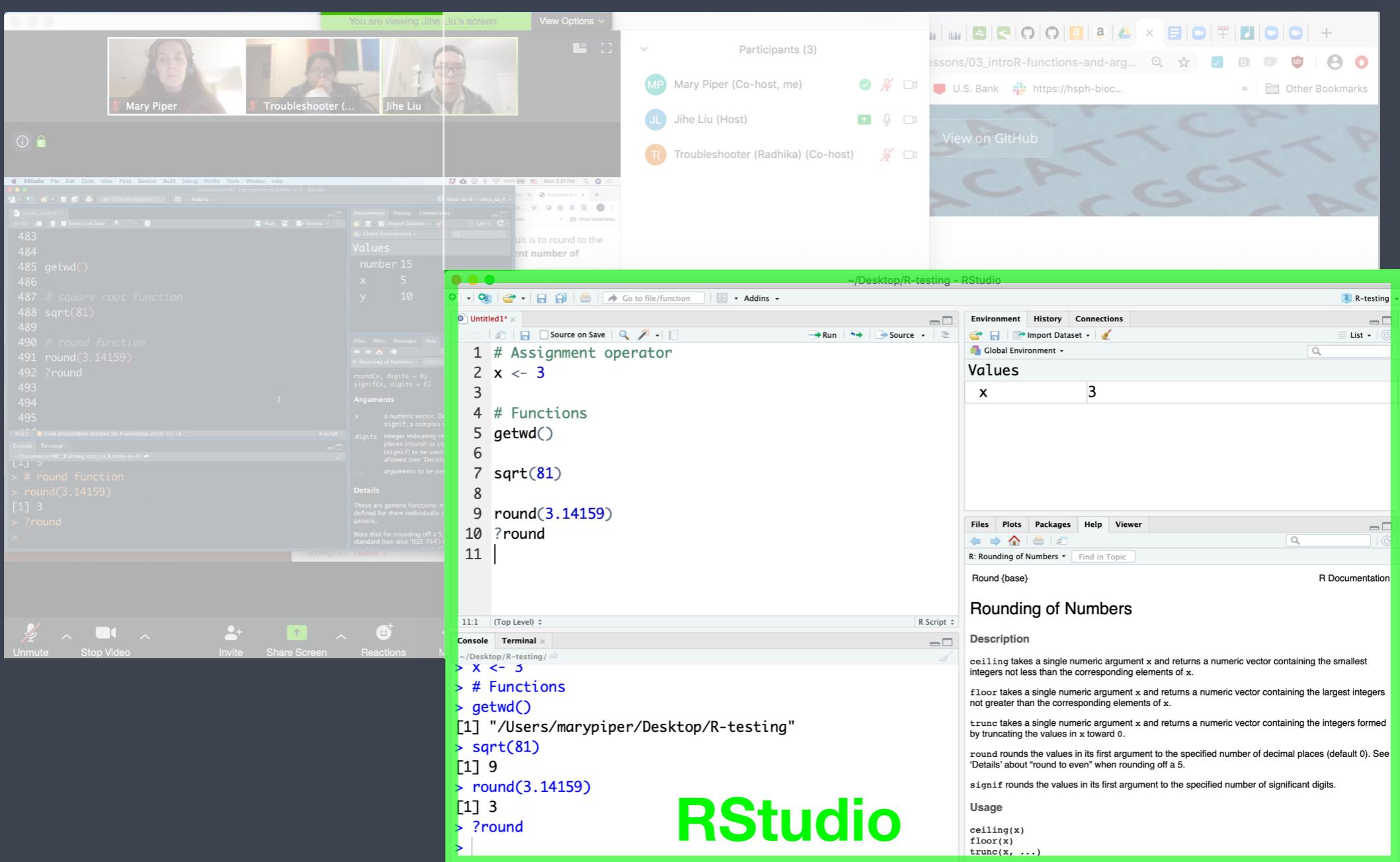
The 3 Window problem...



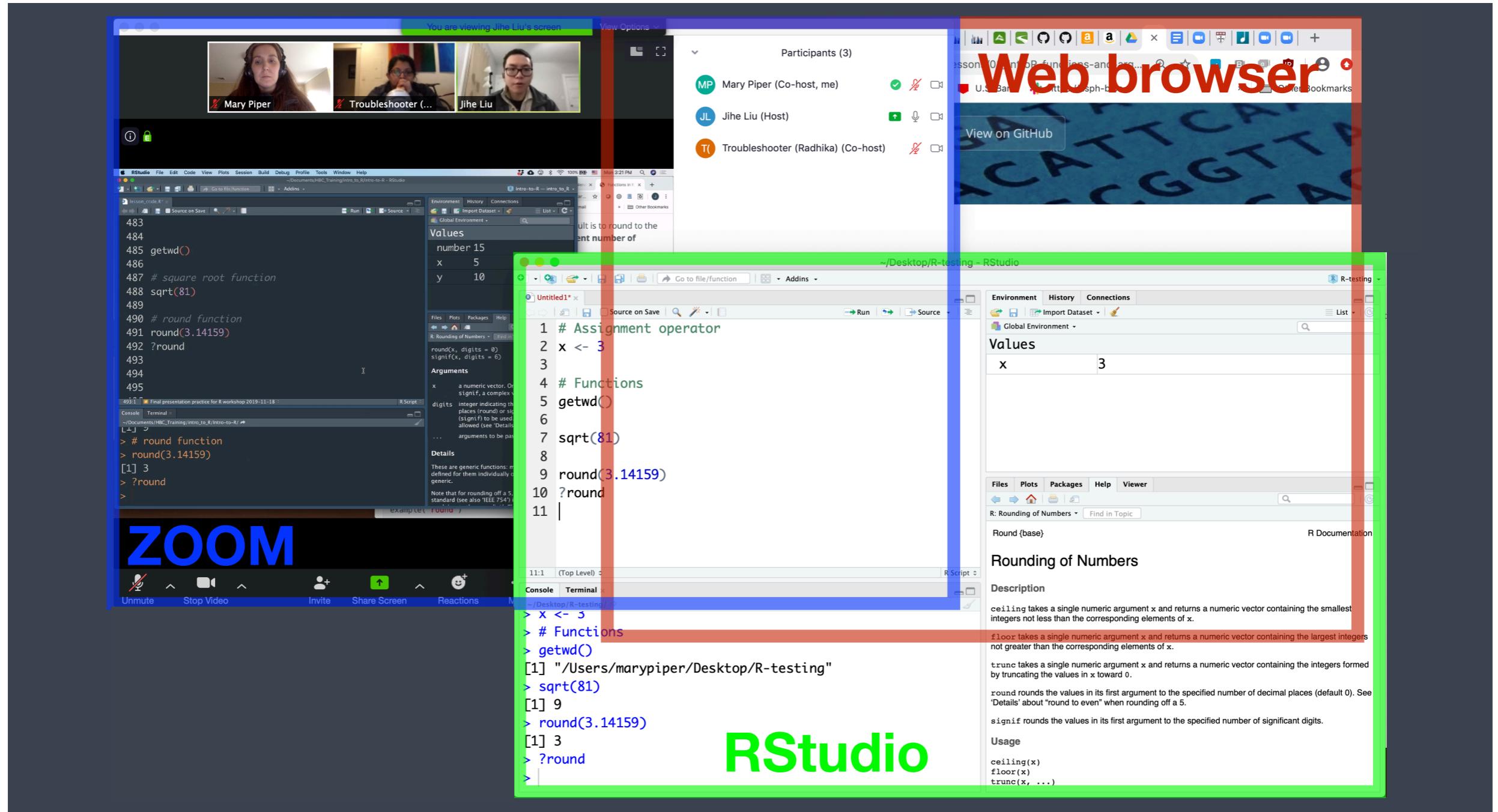
The 3 Window problem...



The 3 Window problem...

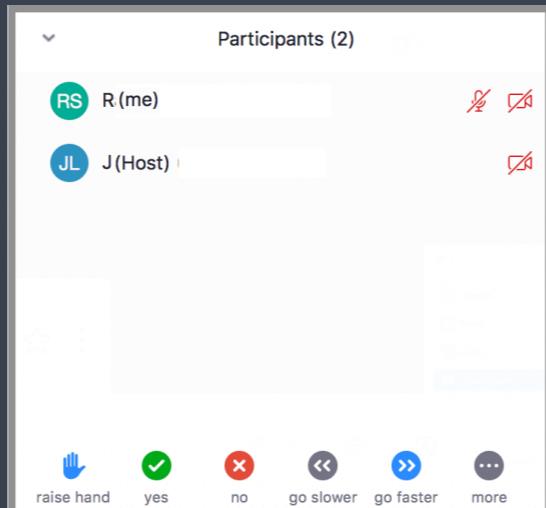


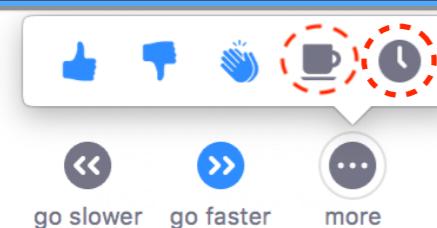
The 3 Window problem...



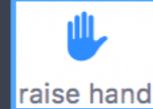
Odds and Ends (1/2)

- ❖ Quit/minimize all applications that are not required for class
- ❖ Click on “Participants” to open that panel in Zoom



- ▶  = "agree", "I'm all set" (equivalent to a **green post-it**)
- ▶  = "disagree", "I need help" (equivalent to a **red post-it**)
- ▶  If you are away from the computer use the coffee cup or clock icon

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)