```
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

Invitation to chat:
What questions are you interested in answering using DGE?

Differential Gene Expression

Harvard Chan Bioinformatics Core

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

Consulting

- Transcriptomics: bulk, single cell, small RNA
- Epigenomics: ChIP-seq, CUT&RUN, ATAC-seq, DNA methylation
- Variant discovery: WGS, resequencing, exome-seq and CNV
- Multiomics integration
- Spatial biology
- Experimental design and grant support

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

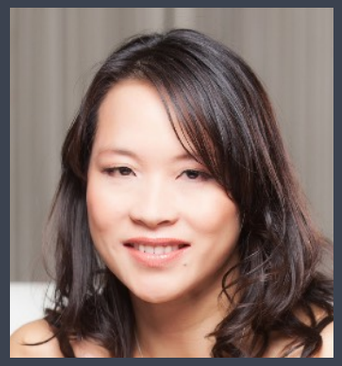
CFAR
HARVARD UNIVERSITY
CENTER FOR AIDS RESEARCH

HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

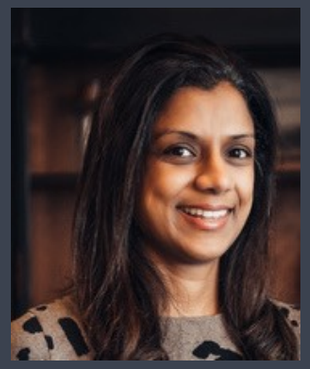
NIEHS

HARVARD
CATALYST
THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

HARVARD
MEDICAL SCHOOL



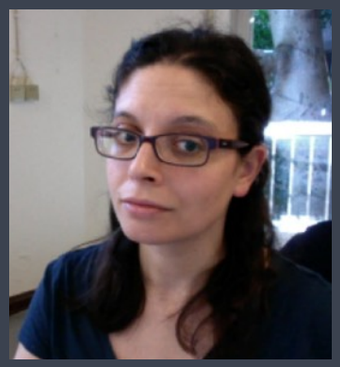
Shannan Ho Sui
Director



Meeta Mistry
Associate Director

Lorena Pantano
Director of Bioinformatics Platform

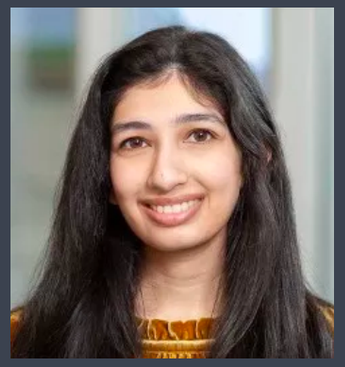
John Quackenbush
Faculty Advisor



Emma Berdan

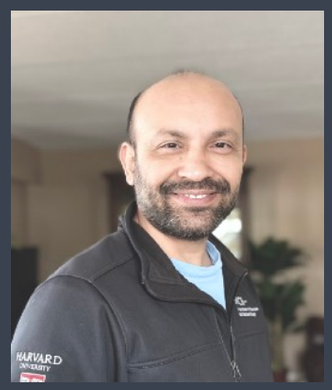
Heather Wick

Will Gammerdinger



Noor Sohail

Alex Bartlett



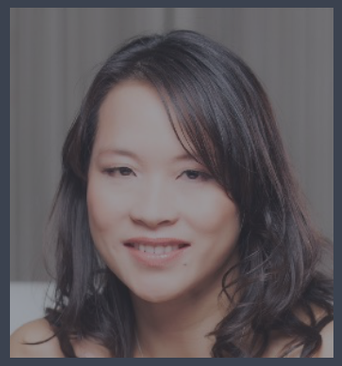
Upen Bhattarai

Elizabeth Partan

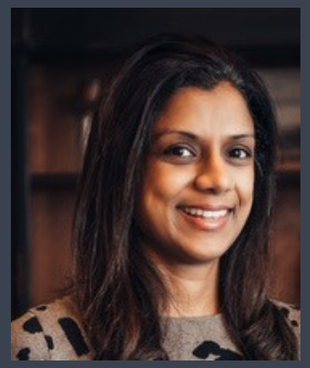
James Billingsley

Zhu Zhuo

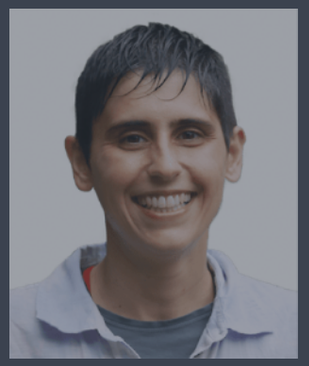
Maria Simoneau



Shannan Ho Sui
Director



Meeta Mistry
Associate Director



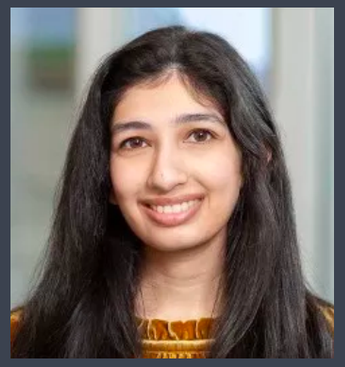
Lorena Pantano
Director of Bioinformatics Platform

John Quackenbush
Faculty Advisor

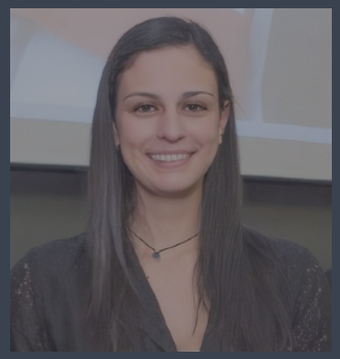
Emma Berdan

Heather Wick

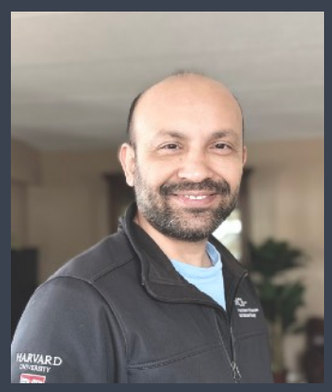
Will Gammerdinger



Noor Sohail



Alex Bartlett



Upen Bhattarai

Elizabeth Partan

James Billingsley

Zhu Zhuo

Maria Simoneau

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 H
researchers at Harvard b

HBC's training team is m
training and community b
research projects to ensu

Our hands-on workshops
an emphasis on **experim
for **wet-lab biologists** ai
data.

We offer three types of w

1. Short, 3-hour monthly
2. Basic Data Skills**
3. Advanced Topics: Ana

**The basic data skills

Our dedicated training team holds workshop to help
or NGS data.

o devote substantial time to material development,
raining team also participate in consultations on
best practices in NGS analysis.

ysis of high-throughput sequencing data, with
and **reproducibility**. Our workshops are designed
cing-based experiments and analysing the resulting

matics)

NGS) data**

r the advanced workshops.

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

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

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

Training

A key component of the H
researchers at Harvard b

HBC's training team is m
training and community b
research projects to ensu

Our hands-on workshops
an emphasis on **experim
for **wet-lab biologists** ai
data.

We offer three types of w

1. Short, 3-hour monthly
2. Basic Data Skills**
3. Advanced Topics: Ana

**The basic data skills

Our dedicated training team holds workshop to help
or NGS data.

o devote substantial time to material development,
raining team also participate in consultations on
best practices in NGS analysis.

ysis of high-throughput sequencing data, with
and **reproducibility**. Our workshops are designed
cing-based experiments and analysing the resulting

matics)

NGS) data**

r the advanced workshops.

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

DF/HCC DANA-FARBER / HARVARD CANCER CENTER

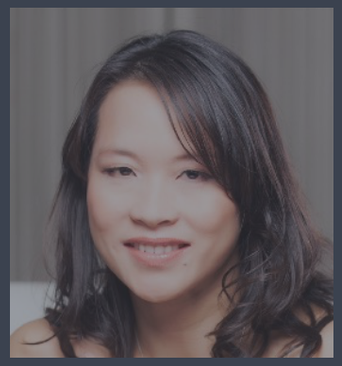
HARVARD CATALYST
THE HARVARD CLINICAL AND TRANSLATIONAL SCIENCE CENTER

HARVARD MEDICAL SCHOOL

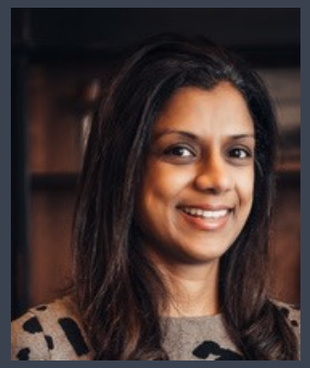
http://bioinformatics.sph.harvard.edu/training/

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

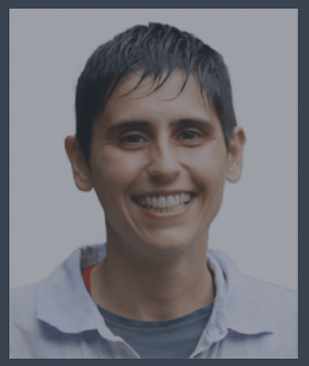
Introductions!



Shannan Ho Sui
Director



Meeta Mistry
Associate Director



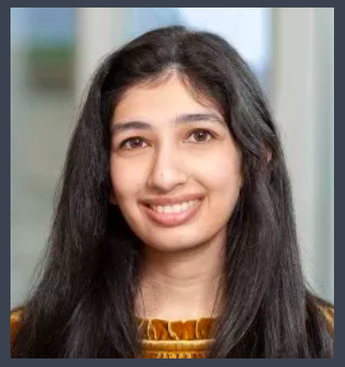
Lorena Pantano
Director of Bioinformatics Platform

John Quackenbush
Faculty Advisor

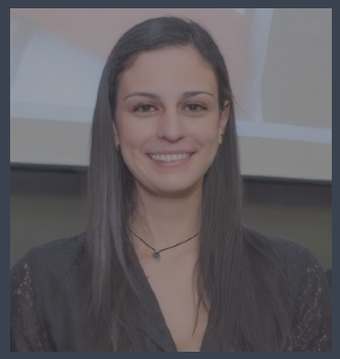
Emma Berdan

Heather Wick

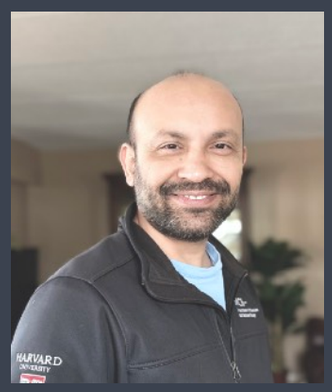
Will Gammerdinger



Noor Sohail



Alex Bartlett



Upen Bhattarai

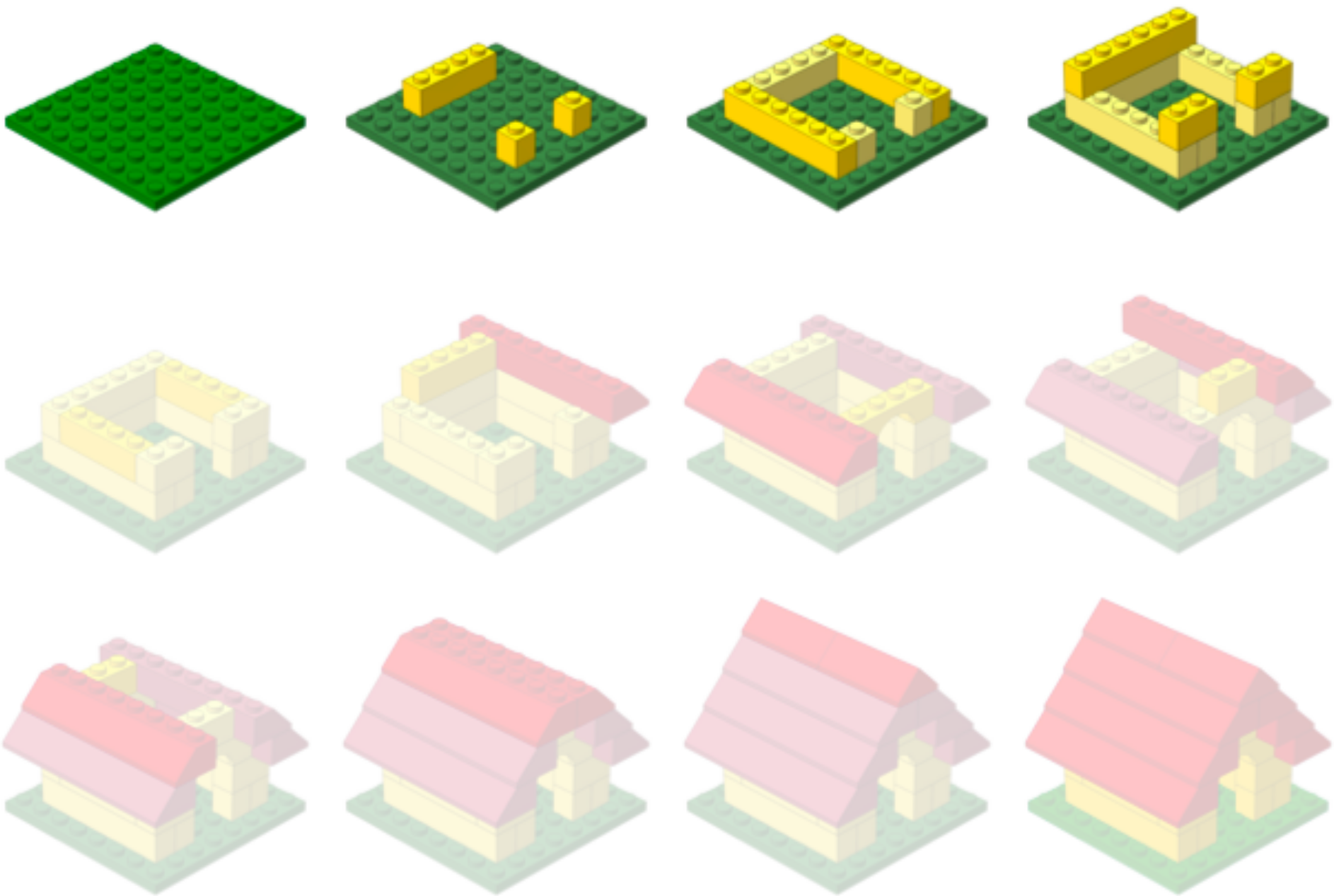
Elizabeth Partan

James Billingsley

Zhu Zhuo

Maria Simoneau

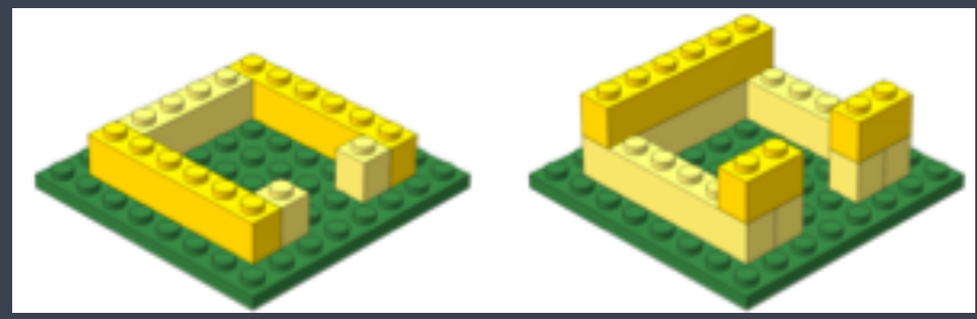
Workshop Scope…



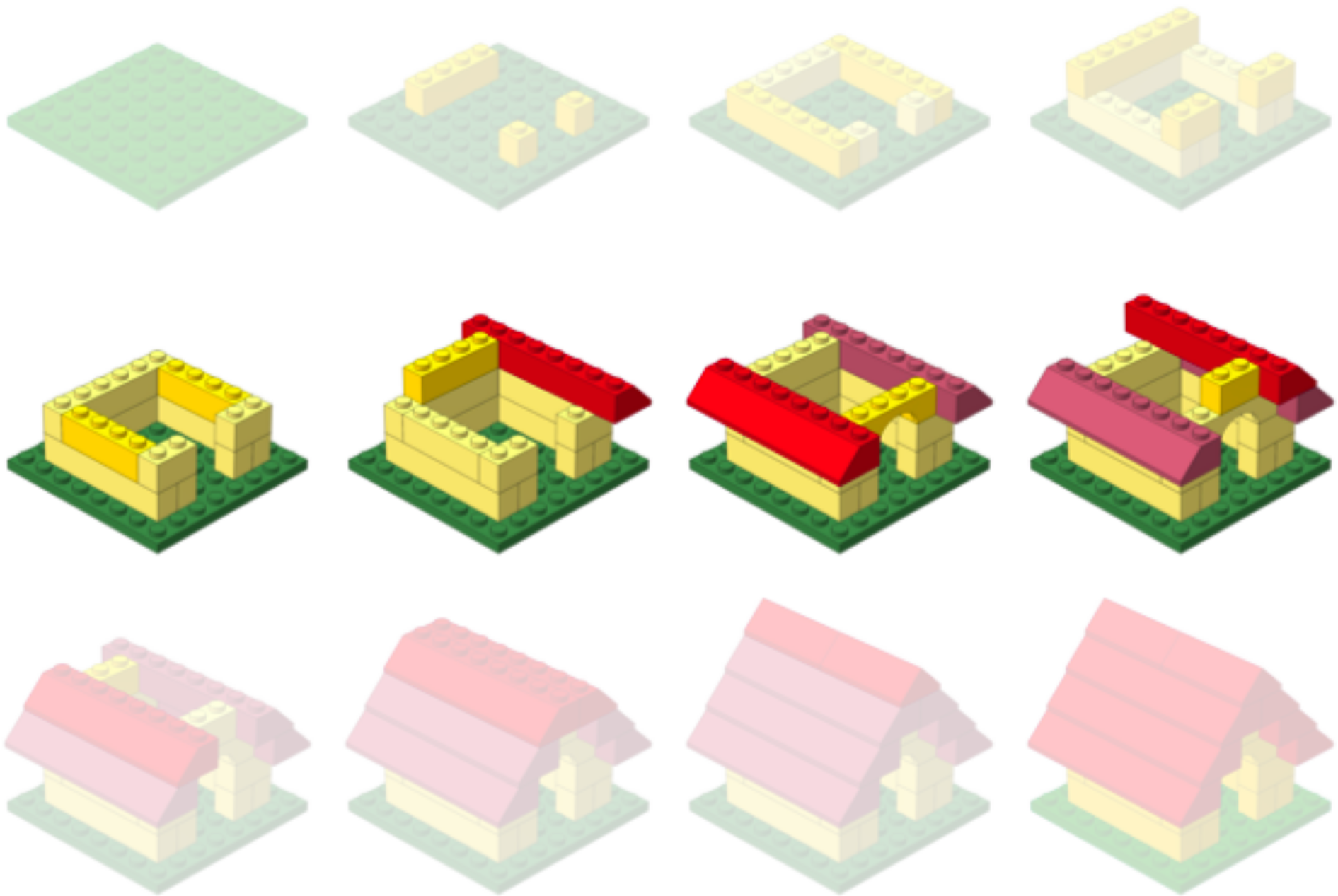
http://anoved.net/tag/lego/page/3/

Setting up to perform Bioinformatics analysis

Setting up…



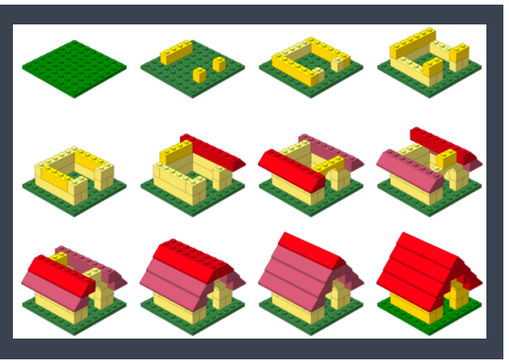
- ✓ Introduction to the command-line interface (shell, Unix, Linux)
  - Dealing with large data files
  - Performing bioinformatics analysis
  - Using tools
  - Accessing and using compute clusters
- ✓ R
  - Parsing and working with smaller results text files
  - Statistical analysis, e.g. differential 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 runnning code in the lesson, email us
- Post any **conceptual questions** that you would like to have **reviewed in class** here.

Course webpage

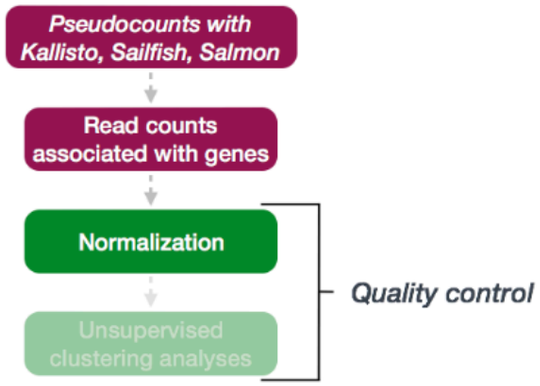
Approximate time: 60 minutes

Learning Objectives

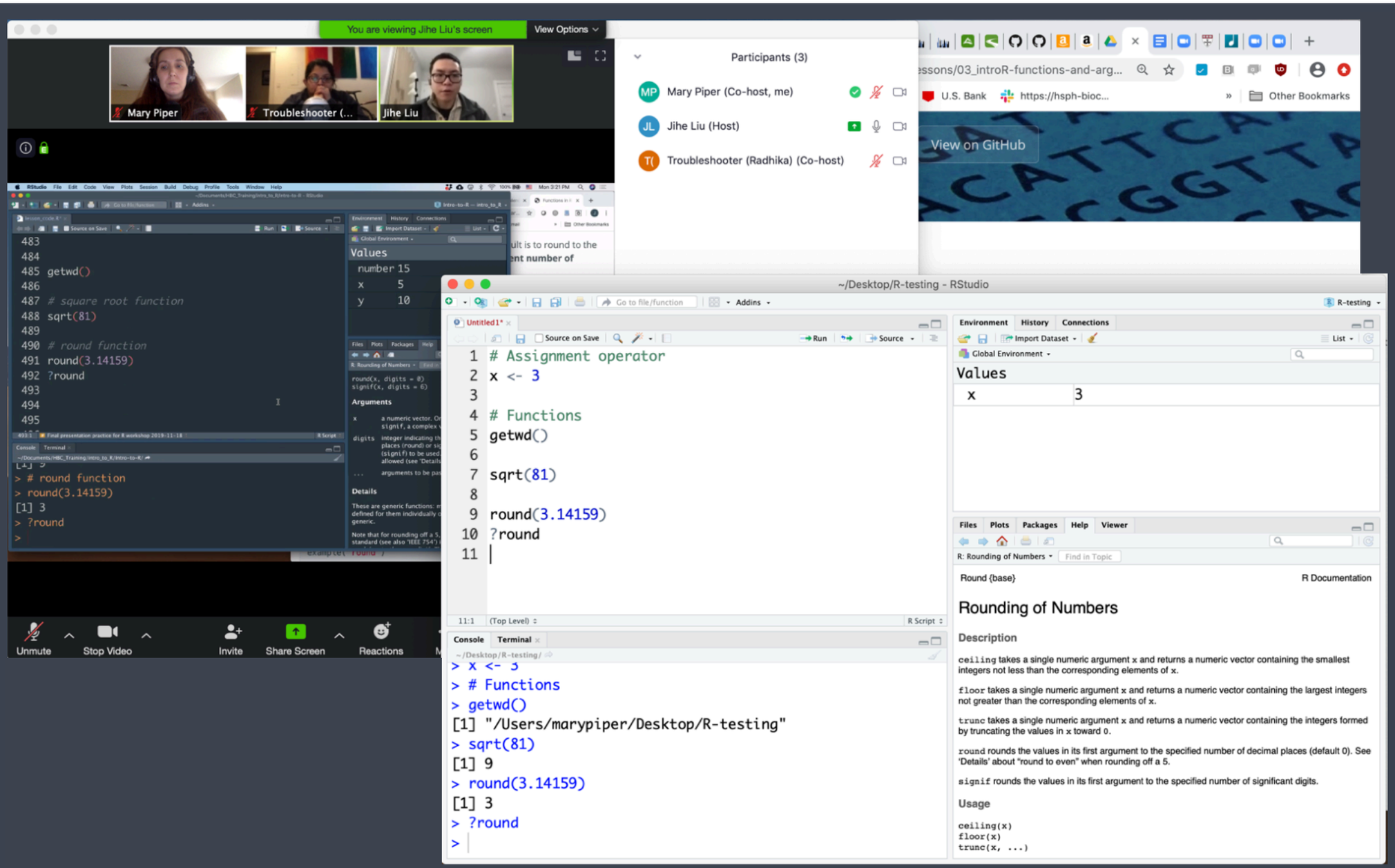
- Explore different types of normalization methods
- Become familiar with the `DESeqDataSet` object
- Understand how to normalize counts using DESeq2

Normalization

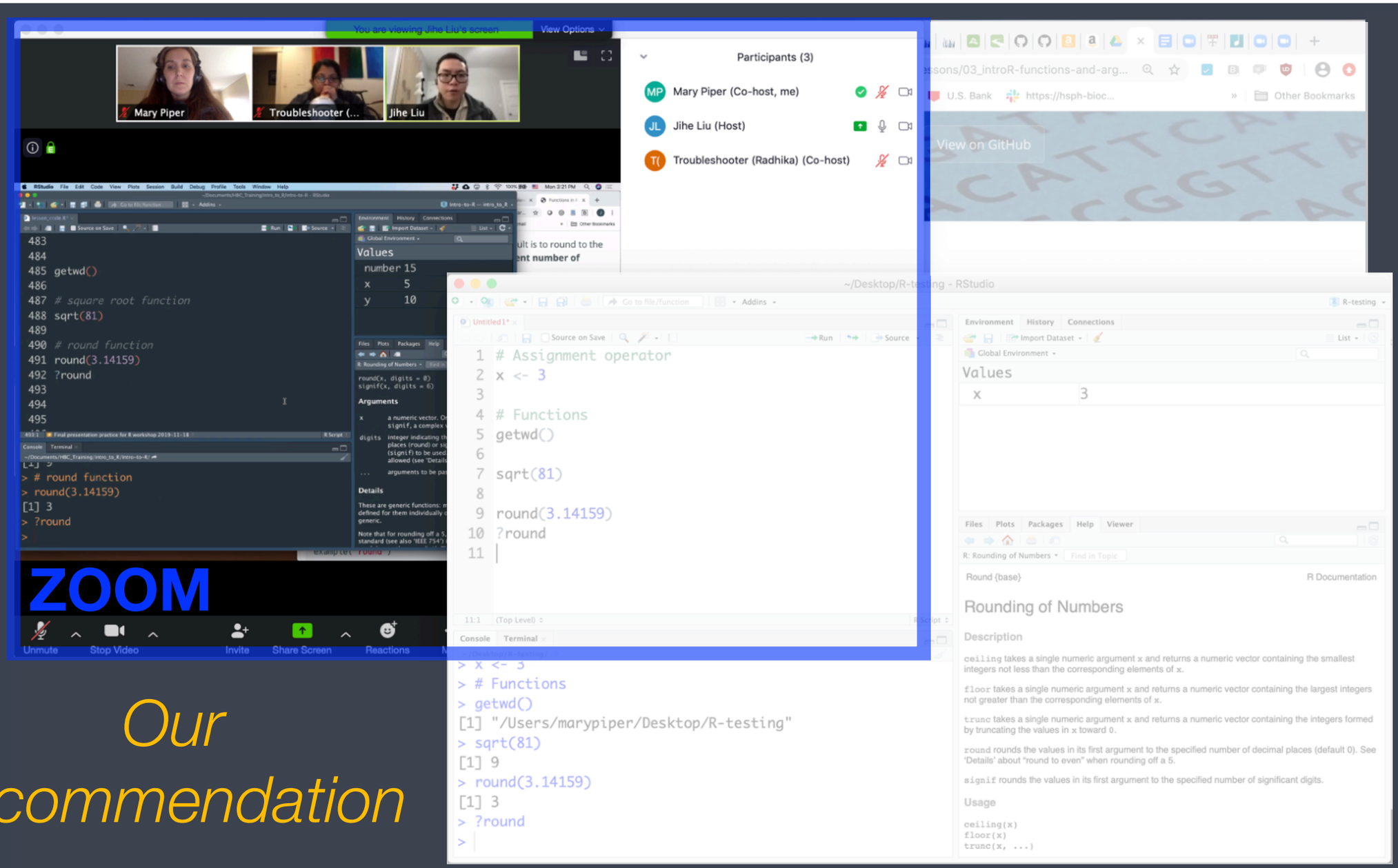
The first step in the DE analysis workflow is count normalization, which is necessary to make accurate comparisons of gene expression between samples.



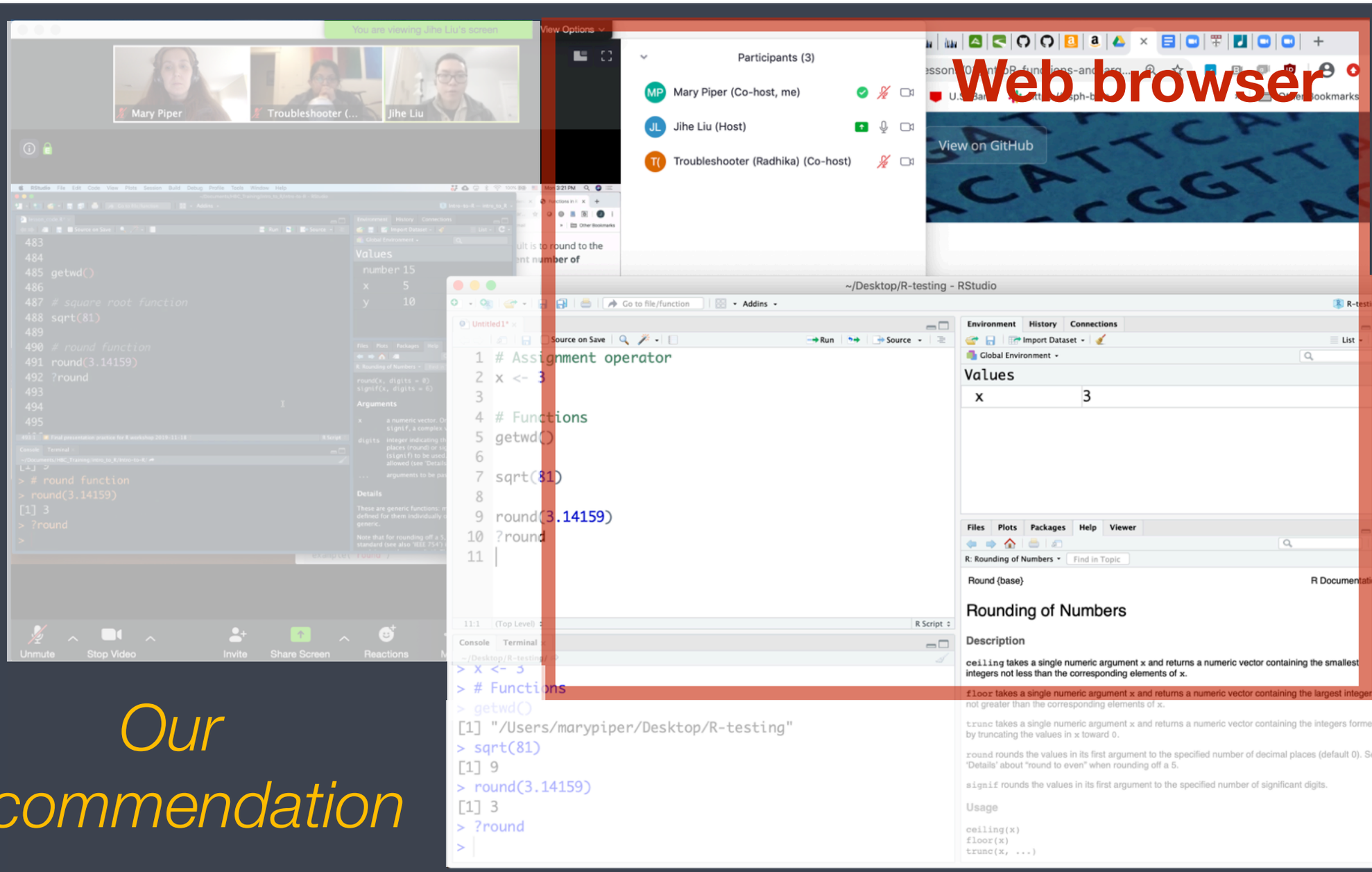
Single screen & 3 windows?



Single screen & 3 windows?

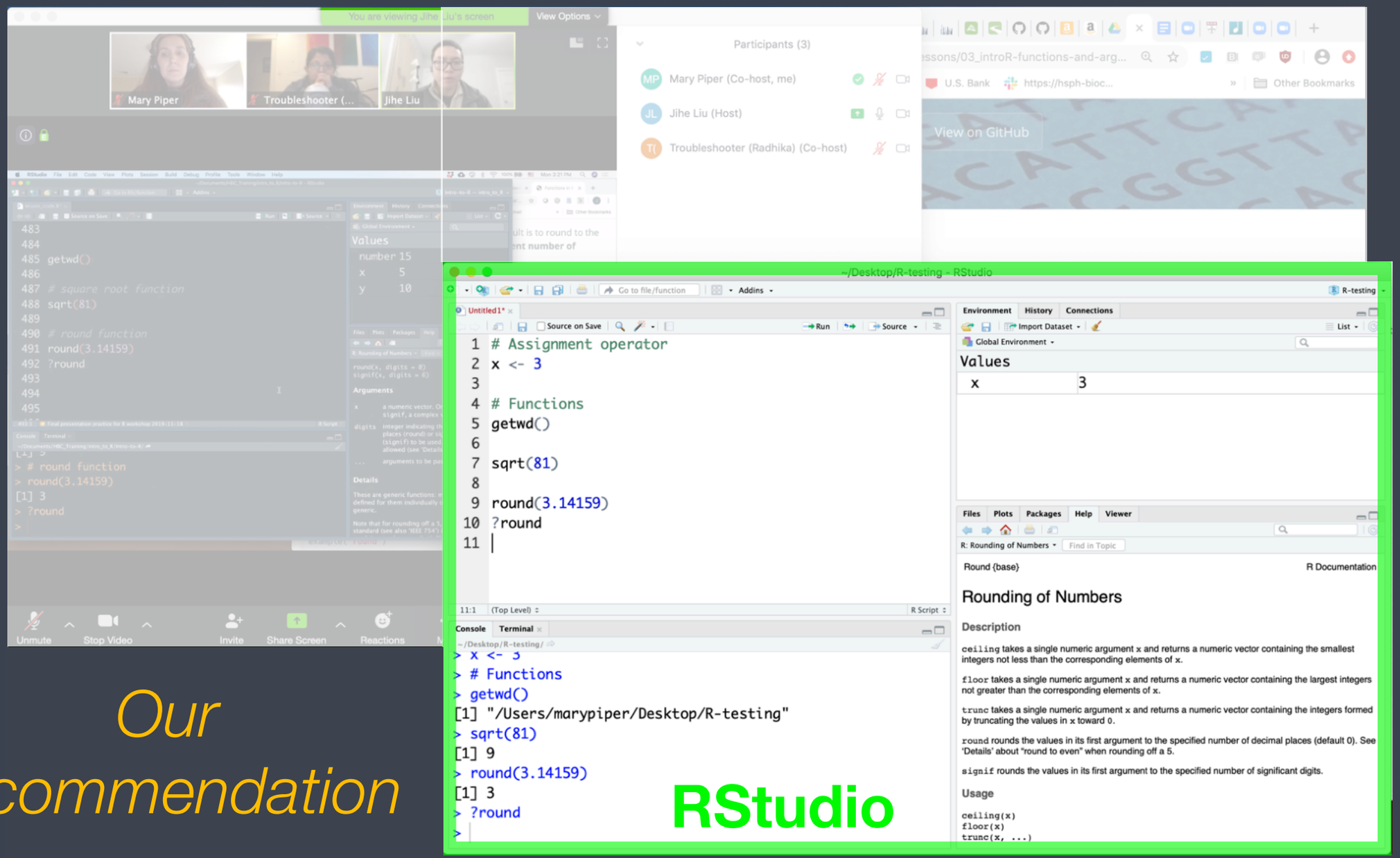


Single screen & 3 windows?



Our recommendation

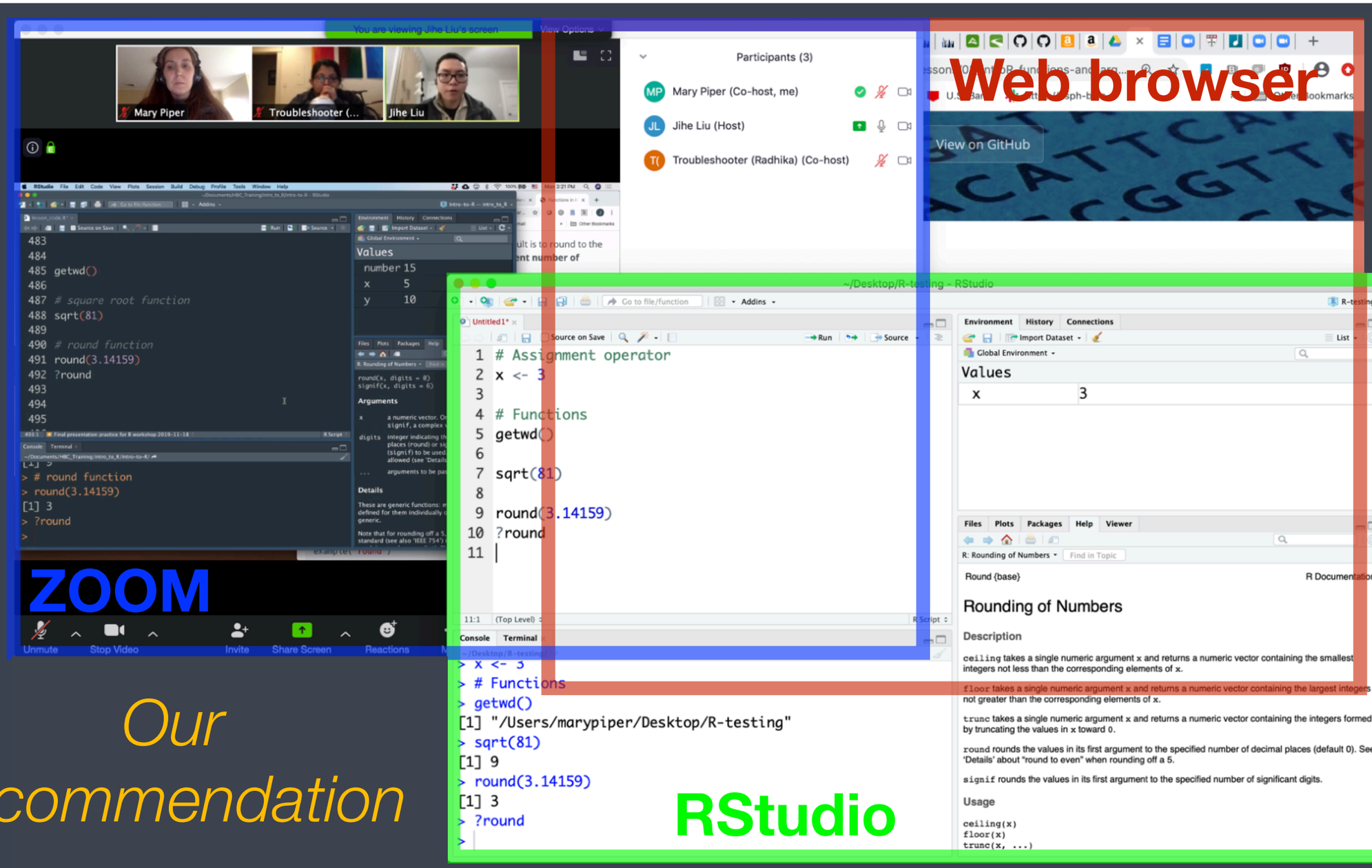
Single screen & 3 windows?



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

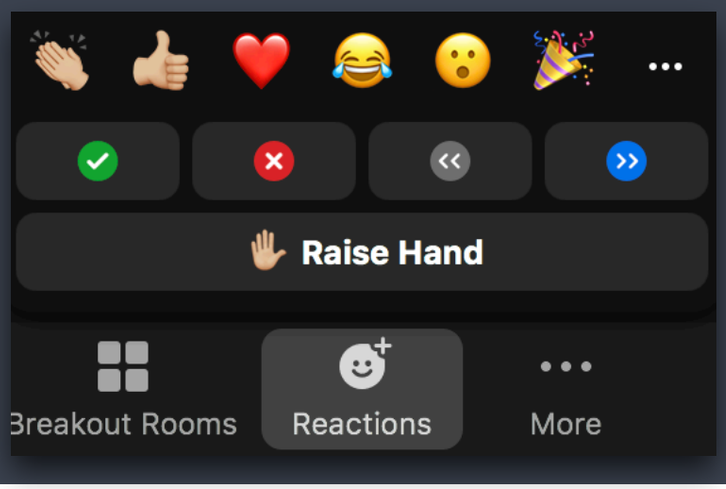
- At-home lessons and exercises after each session
- Cover material not previously discussed
- Provides us feedback to help pace the course appropriately
- 3-5 hours to complete
- Homework load can be heavier in the beginning of this workshop series, but it tapers off

Odds and Ends

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

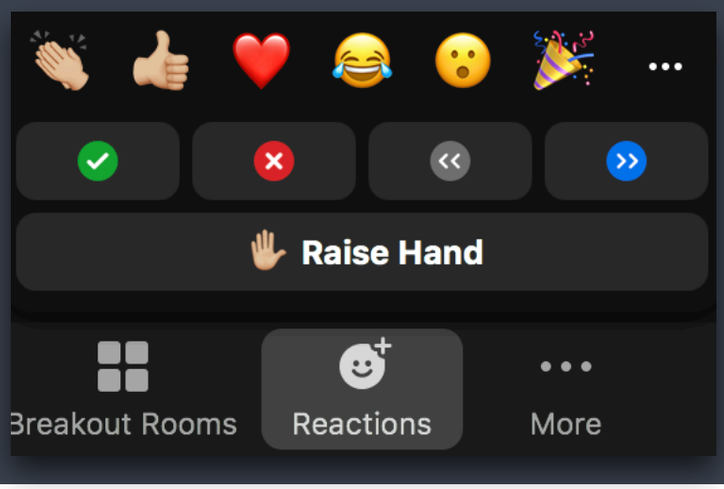
Odds and Ends (1/2)

- 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)

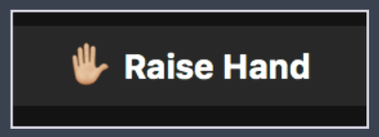


Odds and Ends (2/2)

- Questions for the presenter?
  - Post the question in the Chat window OR
  - ✋ **Raise Hand** 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 RStudio or R?
  - 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