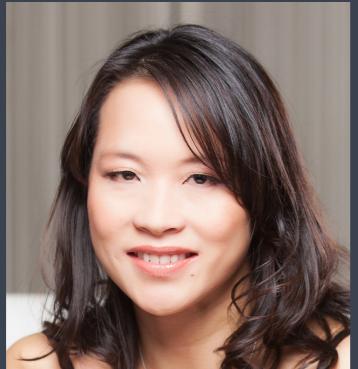


# Introduction to Variant Analysis

Harvard Chan Bioinformatics Core

<https://tinyurl.com/Intro-to-variant-analysis>

Sponsored by DF/HCC and HMS Foundry



Shannan Ho Sui  
*Director*



Victor Barrera



Amelie Jule



Zhu Zhuo



James Billingsley



Radhika Khetani  
*Director of Education*



Meeta Mistry



Will Gammerdinger



Emma Berdan



Sergey Naumenko



Maria Simoneau



We are hiring!  
Email [bioinformatics@hspf.harvard.edu](mailto:bioinformatics@hspf.harvard.edu)  
for more information



Peter Kraft  
*Faculty Advisor*

# Consulting

- Experimental design help
- 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/>



NIEHS



THE HARVARD CLINICAL  
AND TRANSLATIONAL  
SCIENCE CENTER



# 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 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 support researchers at Harvard by providing training.

HBC's training team is made up of experts in training and community building who work on research projects to ensure our training is effective.

Our hands-on workshops focus on **bioinformatics**, with an emphasis on **experimental design** and **data analysis**. We also provide training for **wet-lab biologists** aiming to learn how to analyse 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 workshop is designed for researchers with no prior experience in bioinformatics.



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

**DF/HCC**  
DANA-FARBER / HARVARD CANCER CENTER



THE HARVARD CLINICAL  
AND TRANSLATIONAL  
SCIENCE CENTER



Our dedicated training team holds workshops to help researchers learn how to analyse their data or 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 to teach researchers how to perform sequencing-based experiments and analysing the resulting data.

**bioinformatics**)

**basic data skills** (e.g. NGS) data)\*\*

and **advanced topics** (e.g. 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 support researchers at Harvard by providing training.

HBC's training team is made up of experts in training and community building who work on research projects to ensure our training is effective.

Our hands-on workshops are designed with an emphasis on **experimental design** and **informatics**, for **wet-lab biologists** and **bioinformaticians** alike.

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



THE HARVARD CLINICAL  
AND TRANSLATIONAL  
SCIENCE CENTER



Our dedicated training team holds workshops to help researchers learn how to analyze high-throughput sequencing (NGS) data.

The training team also devote substantial time to material development, and the 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**, **informatics**, and **reproducibility**. Our workshops are designed for wet-lab biologists and bioinformaticians, covering topics such as designing experiments and analysing the resulting data.

(bioinformatics)

(high-throughput sequencing (NGS) data)\*\*

or the advanced workshops.

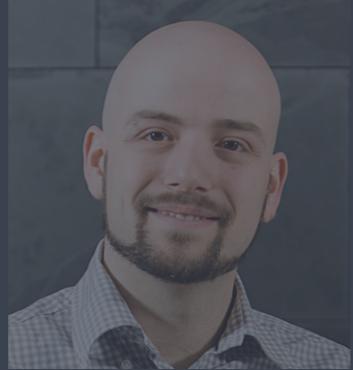
<http://bioinformatics.sph.harvard.edu/training/>

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

# Introductions!



Shannan Ho Sui  
*Director*



Victor Barrera



Amelie Jule



Zhu Zhuo



James Billingsley



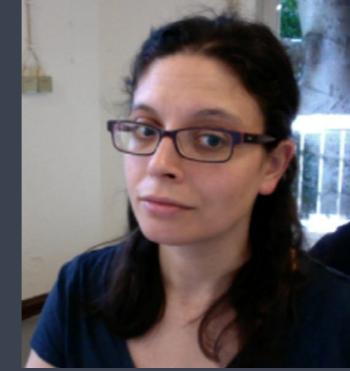
Radhika Khetani  
*Director of Education*



Meeta Mistry



Will Gammerdinger



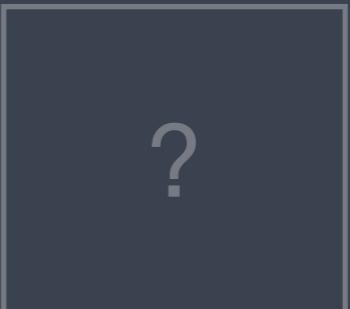
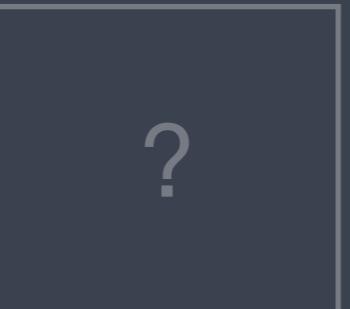
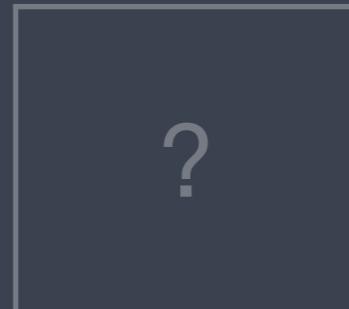
Emma Berdan



Sergey Naumenko



Maria Simoneau

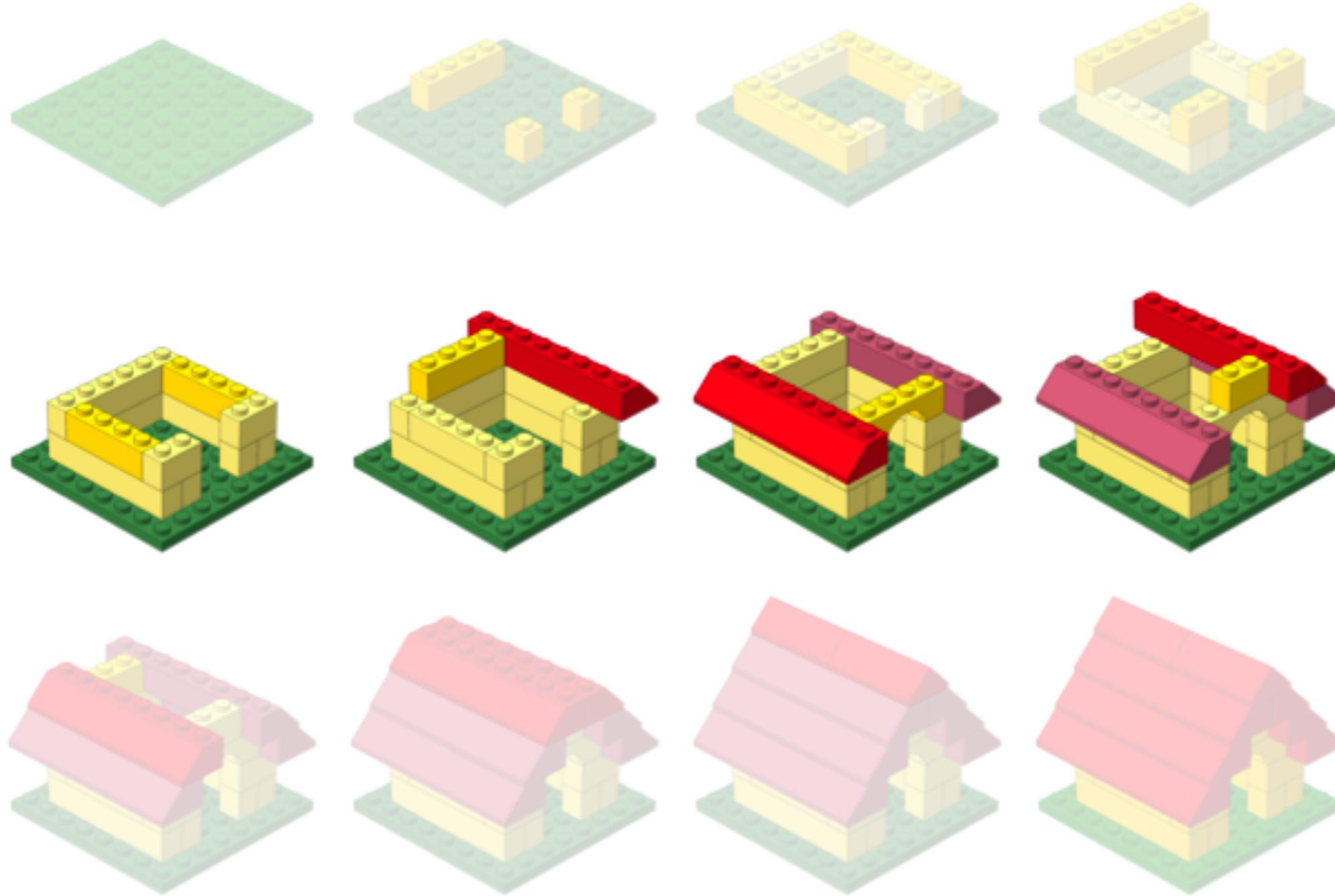


We are hiring!  
Email [bioinformatics@hspf.harvard.edu](mailto:bioinformatics@hspf.harvard.edu)  
for more information



Peter Kraft  
*Faculty Advisor*

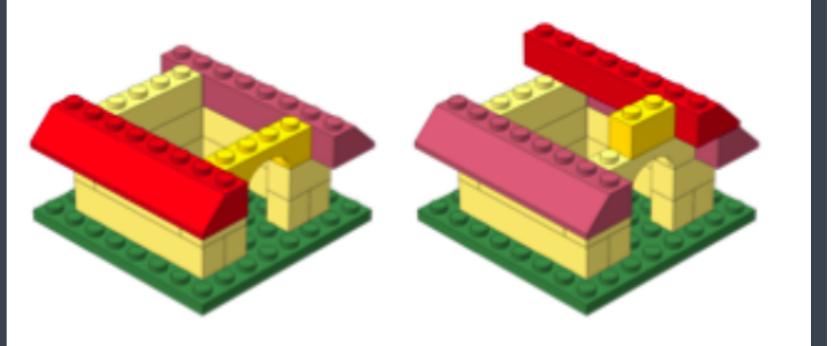
# Workshop Scope...



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

Bioinformatics data analysis

# Learning Objectives



- ✓ Evaluate QC metrics for variant calling
- ✓ Call variants using GATK
- ✓ Filter variants to retain only high-quality variant calls
- ✓ Annotate variants using SnpEff and dbSNP
- ✓ Prioritize variants by their impact
- ✓ Visualize variants in IGV

# Logistics

# Course webpage

<https://tinyurl.com/Intro-to-variant-analysis>

# Course schedule online



## Workshop Schedule

### Pre-reading

#### Before the workshop:

Please read the following page to learn about the dataset we will be using:

[ICGC-TCGA DREAM Mutation Calling Challenge Synthetic Dataset](#)

### Day 1

Time	Topic	Instructor
9:30 - 10:10	Workshop Introduction	Will
10:00 - 11:30	Introduction to Variant Calling	Sergey
11:30 - 11:50	Project Organization	Will

# Course materials online



## Project Organization

[View on GitHub](#)

### Project Organization

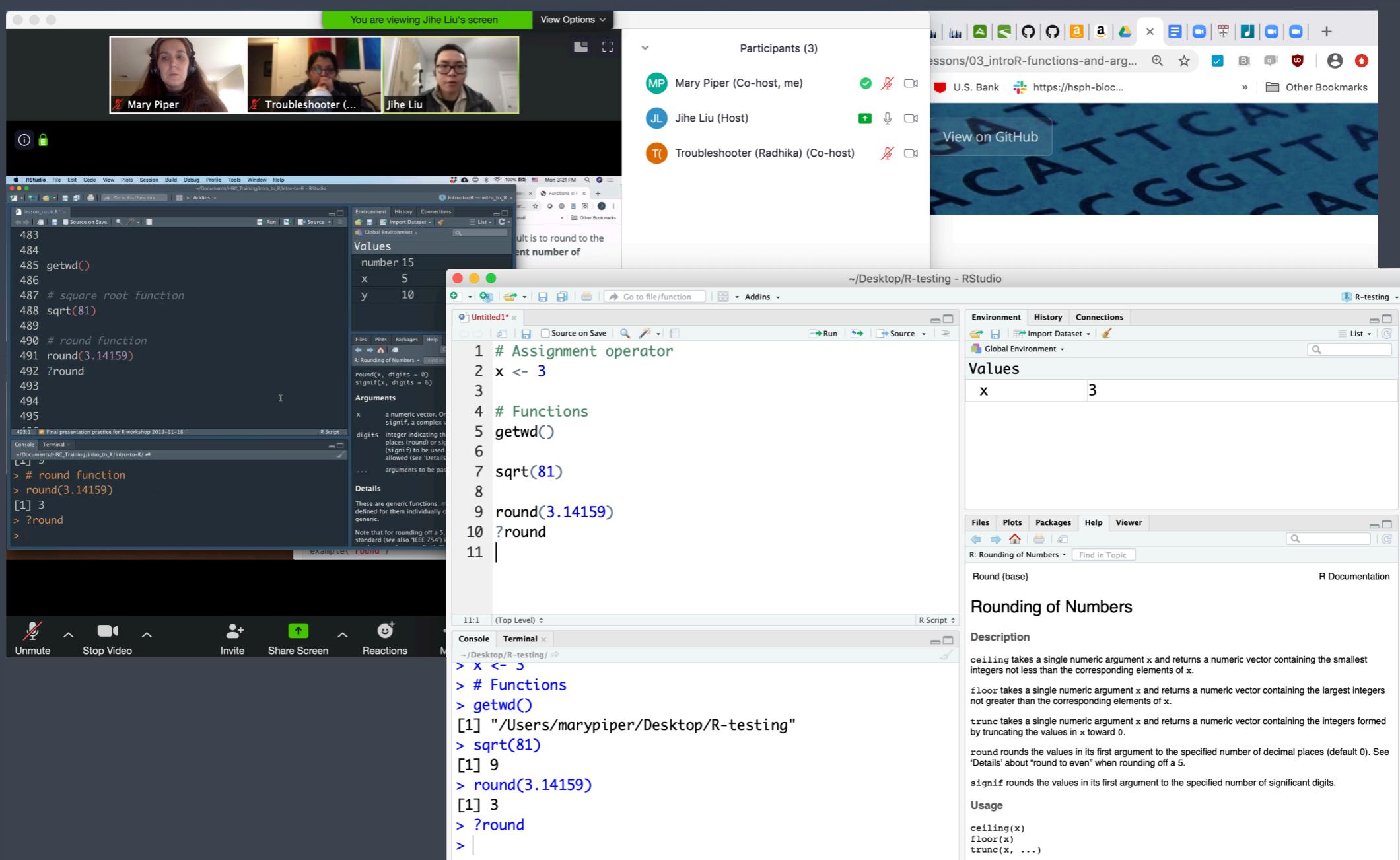
#### Learning Objectives

- Configure a workspace on the `scratch` drive
- Organize dataset for analysis
- Differentiate between using `home` and `scratch` drives

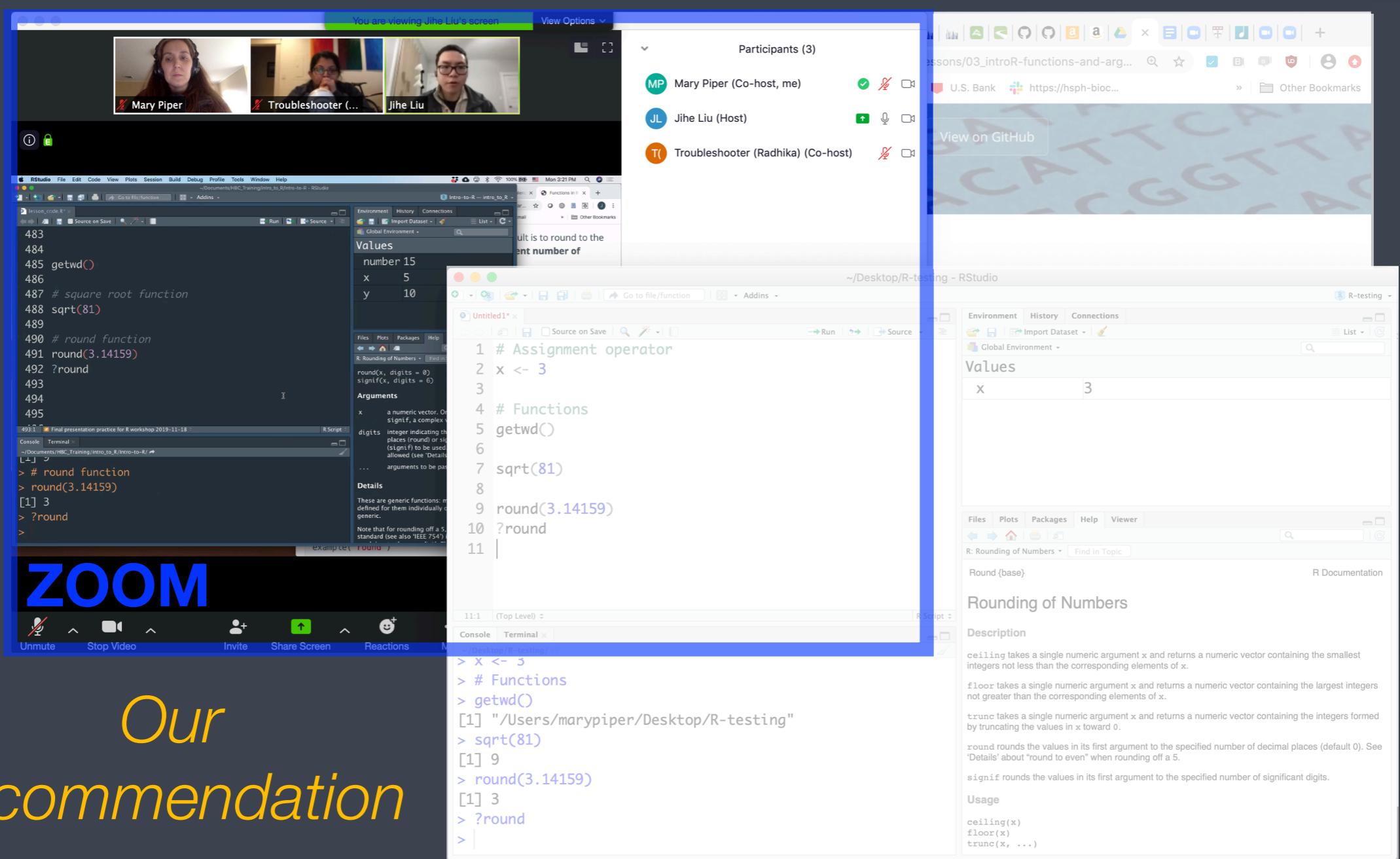
#### Logging into O2

For this workshop we will be using training accounts to log in. These have been created for us by the [HMS Research Computing \(HMS-RC\) team](#), they are the folks that manage the O2 cluster. We will be providing each of you with your own training account associated with a password for the duration of this workshop. Your training account and password can be found [here](#).

# Single screen & 3 windows?

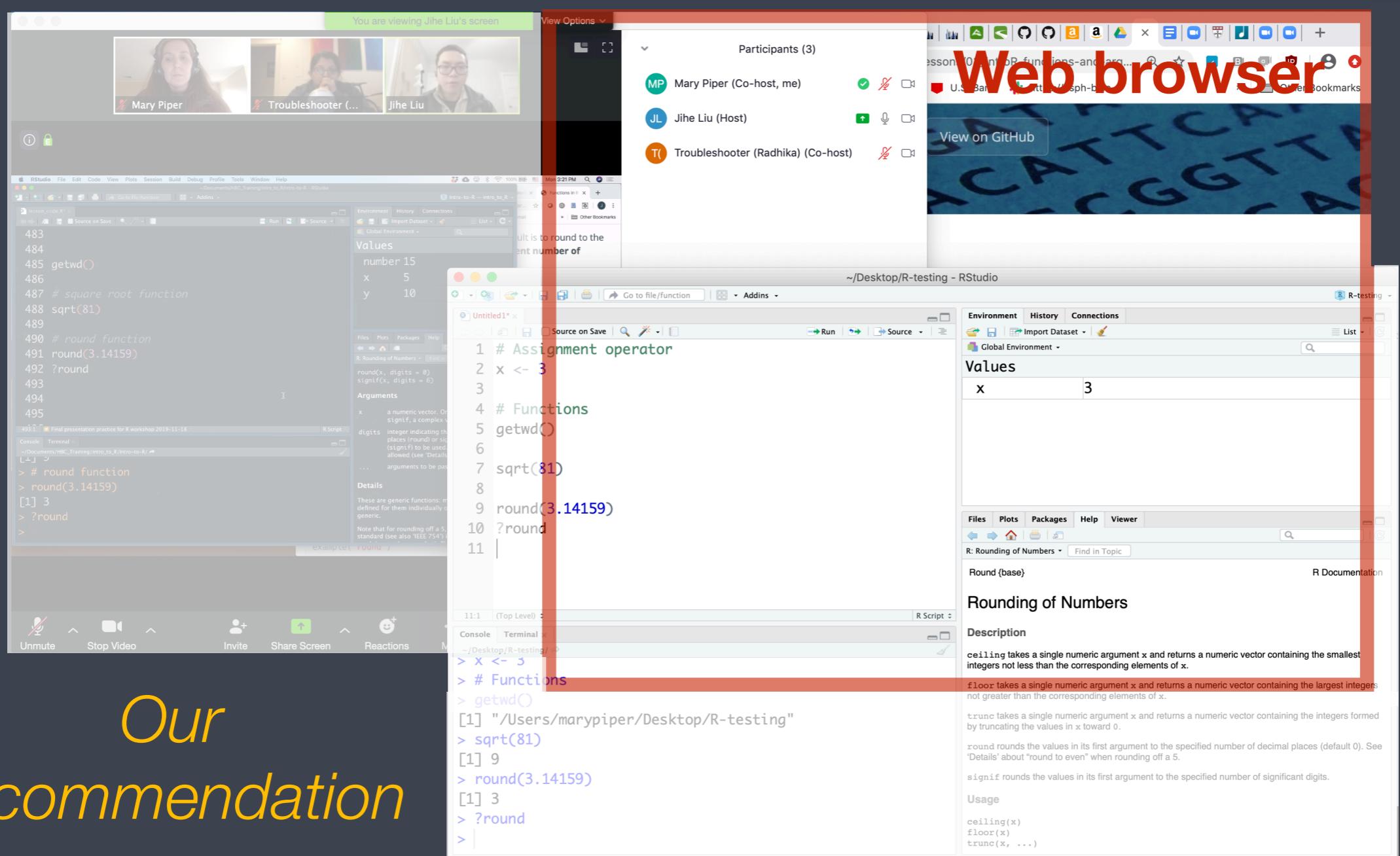


# Single screen & 3 windows?



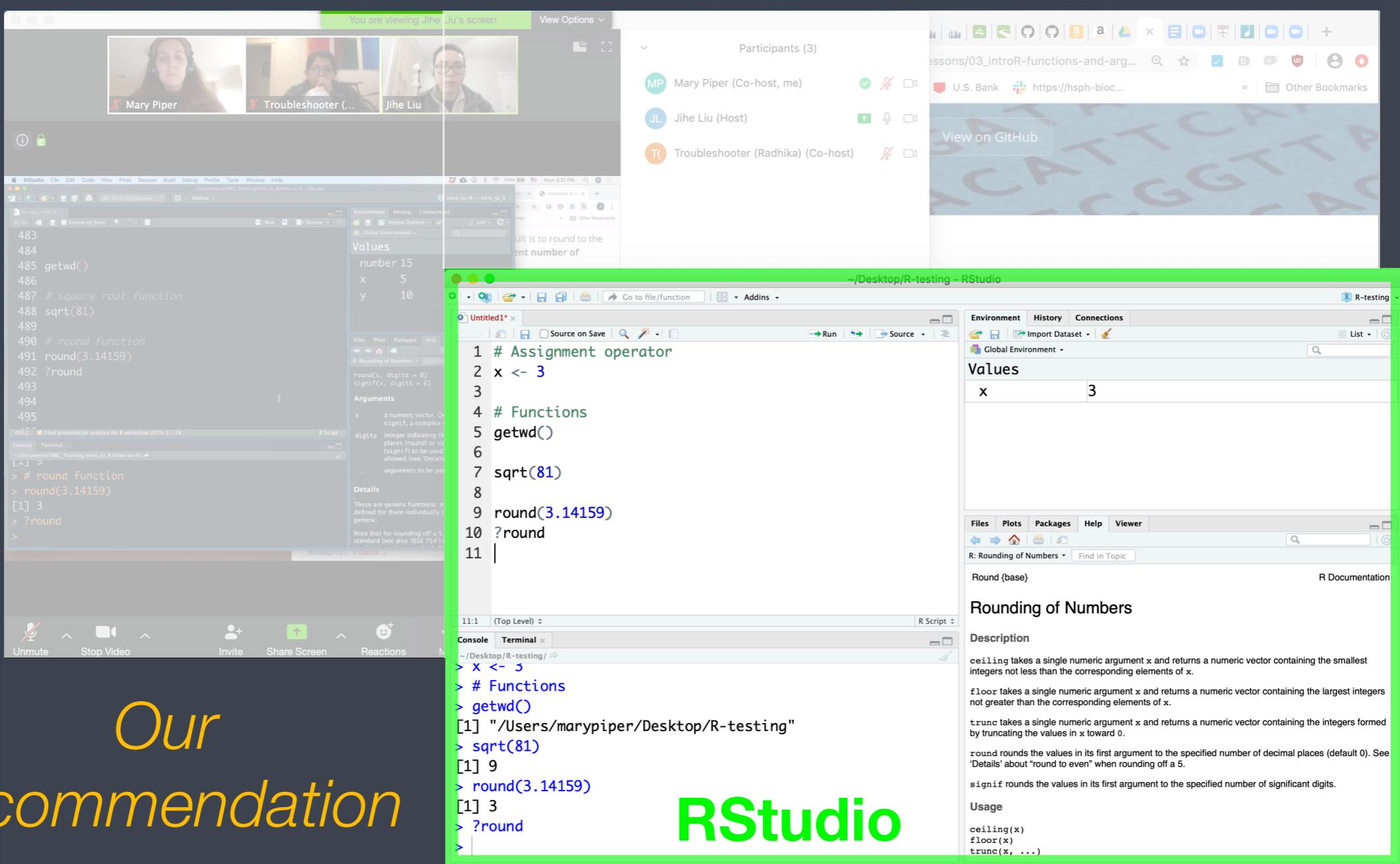
*Our  
recommendation*

# Single screen & 3 windows?

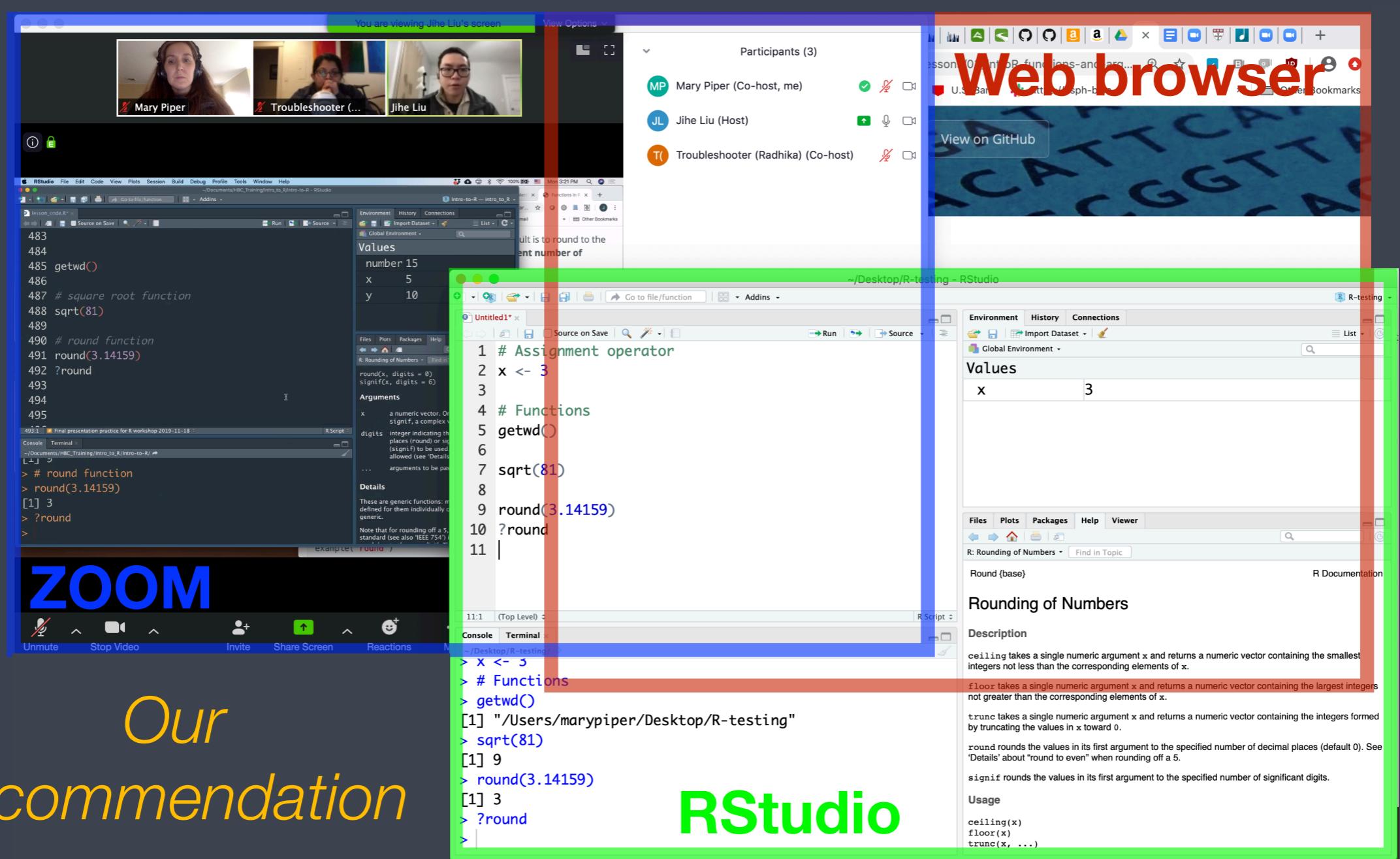


Our  
recommendation

# Single screen & 3 windows?



# 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
- ❖ Homework load is heavier in the beginning of this workshop series and tapers off

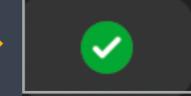
# Odds and Ends (1/2)

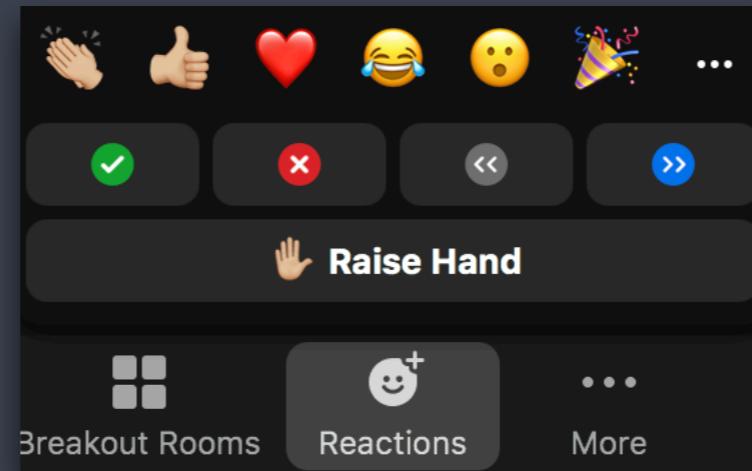
- ❖ Quit/minimize all applications that are not required for class

# Odds and Ends (1/2)

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

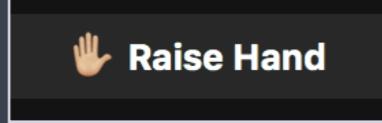
# Odds and Ends (1/2)

- ❖ 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**)



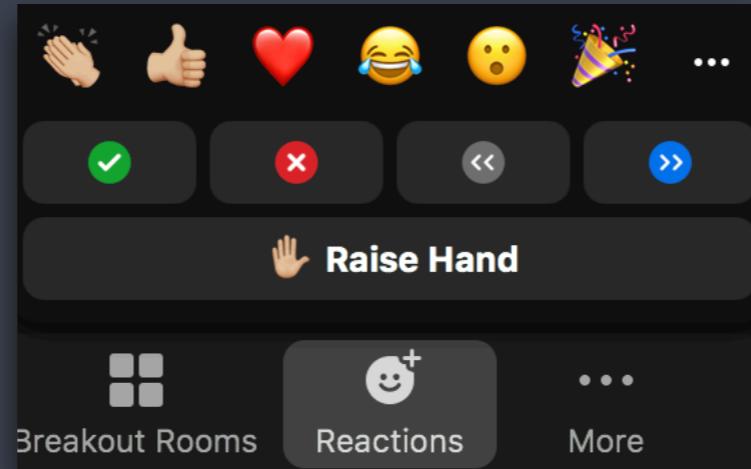
# Odds and Ends (2/2)

## ❖ 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](mailto:hbctraining@hsph.harvard.edu)

*HBC consulting:* [bioinformatics@hsph.harvard.edu](mailto:bioinformatics@hsph.harvard.edu)

Twitter

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