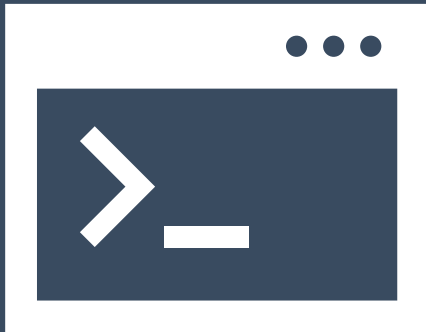


# Differential expression analysis of Single Cell RNA-seq

<https://tinyurl.com/DGE-analysis-scRNAseq>



Harvard Chan Bioinformatics Core



# Introductions!





Shannan Ho Sui  
*Director*



Meeta Mistry  
*Associate Director*



Lorena Pantano  
*Director of Bioinformatics  
Platform*



John Quackenbush  
*Faculty Advisor*



Upen Bhattarai



Heather Wick



Will Gammerdinger



Noor Sohail



Elizabeth  
Partan



Alex Bartlett



Emma Berdan



James Billingsley



Zhu Zhuo



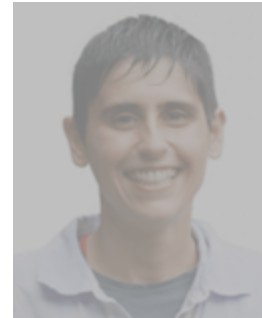
Maria Simoneau



Shannan Ho Sui  
*Director*



Meeta Mistry  
*Associate Director*



Lorena Pantano  
*Director of Bioinformatics  
Platform*



John Quackenbush  
*Faculty Advisor*



Upen Bhattarai



Heather Wick



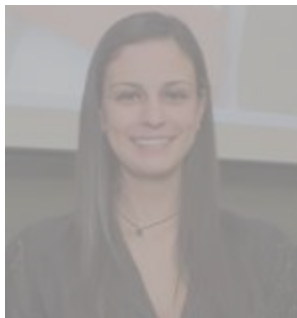
Will Gammerdinger



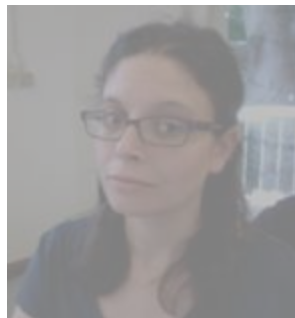
Noor Sohail



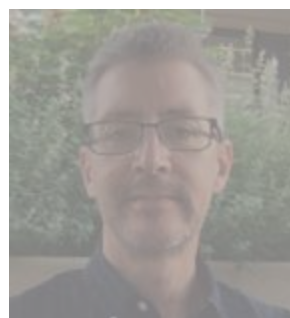
Elizabeth  
Partan



Alex Bartlett



Emma Berdan



James Billingsley



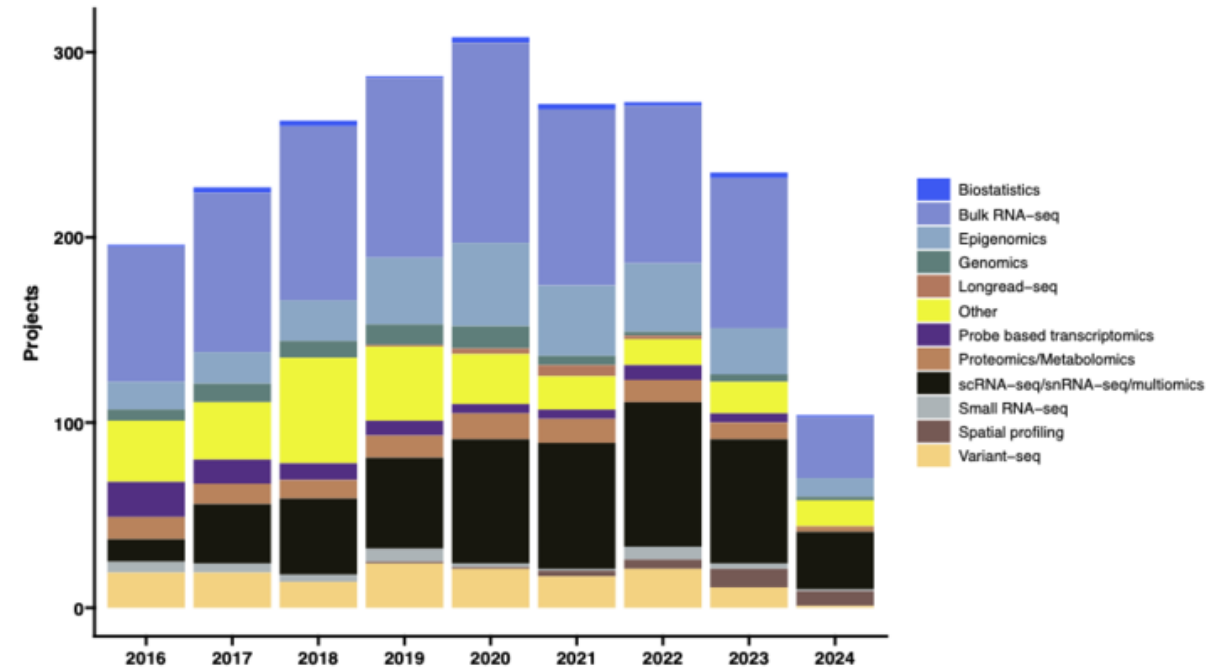
Zhu Zhuo



Maria Simoneau

# 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





# 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



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

NIEHS

---



THE HARVARD CLINICAL  
AND TRANSLATIONAL  
SCIENCE CENTER

---



**HARVARD**  
MEDICAL SCHOOL

# Training

- ❖ Hands-on workshops design to reflect best practices, reproducibility and an emphasis on experimental design

- ❖ Basic Data Skills

- ❖ Shell

- ❖ R

- ❖ Advanced Topics: Analysis of high-throughput sequencing data

- ❖ Chromatin Biology

- ❖ Bulk RNA-seq

- ❖ Differential Gene Expression

- ❖ scRNA-seq

- ❖ Variant Calling

- ❖ Current Topics in Bioinformatics

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

# Training

- ❖ Hands-on workshops design to reflect best practices, reproducibility and an emphasis on experimental design
  - ❖ Basic Data Skills
    - ❖ Shell
    - ❖ R
  - ❖ Advanced Topics: Analysis of high-throughput sequencing data
    - ❖ Chromatin Biology
    - ❖ Bulk RNA-seq
    - ❖ Differential Gene Expression
    - ❖ scRNA-seq
    - ❖ Variant Calling
  - ❖ Current Topics in Bioinformatics

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



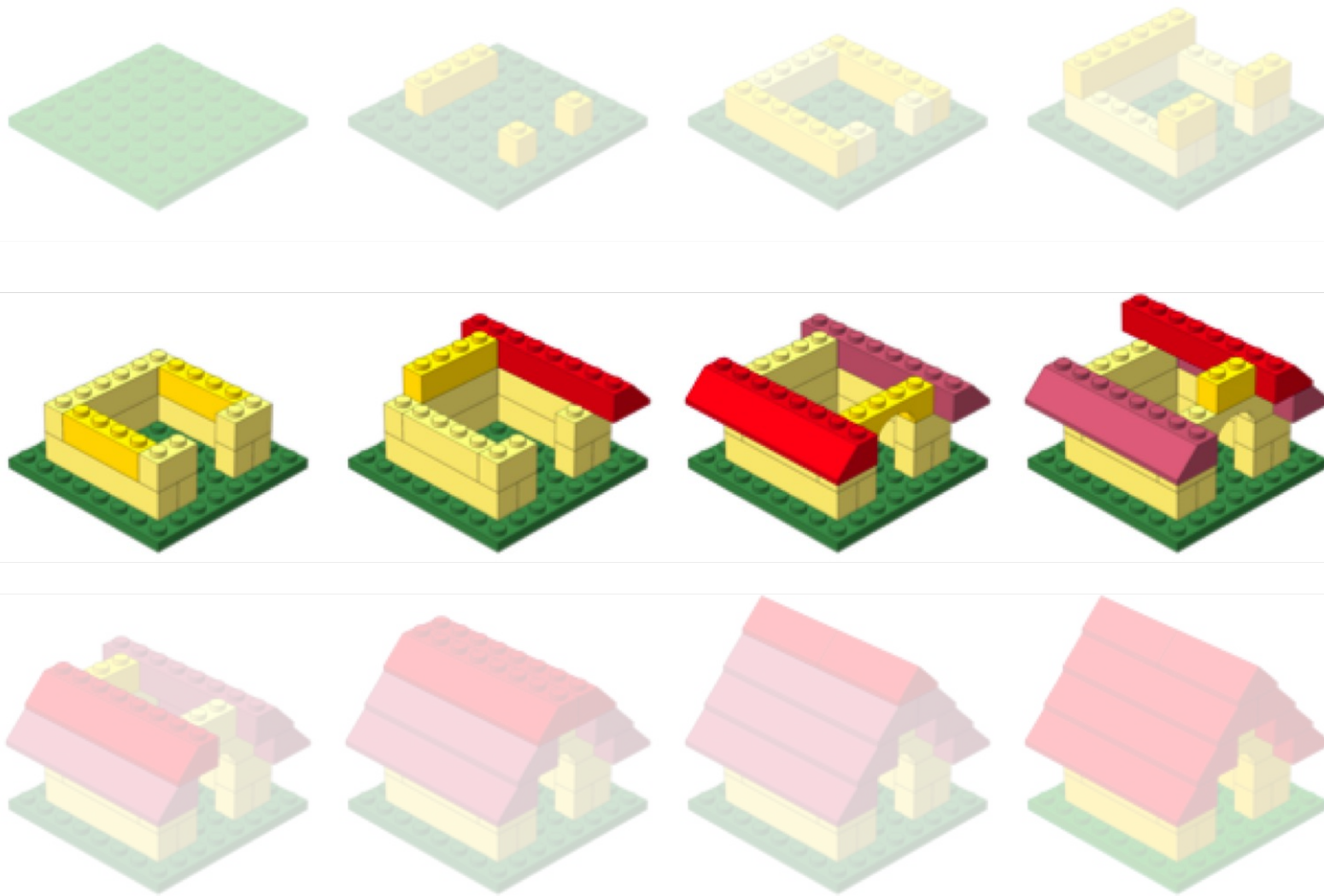
**HARVARD**  
**T.H. CHAN**  
SCHOOL OF PUBLIC HEALTH

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



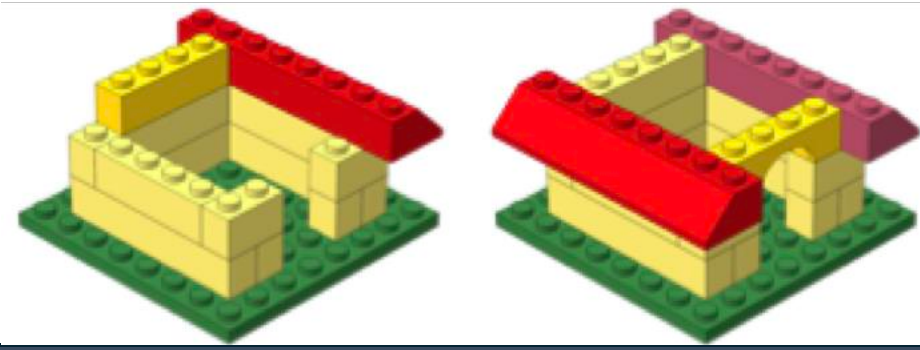


# **Workshop scope**



Bioinformatic Data Analysis

# Workshop Scope



- ❖ Understanding considerations for when to use different DGE algorithms on scRNA-seq data
- ❖ Using FindMarkers to evaluate significantly different genes
- ❖ Pseudobulking a counts matrix in order to run DESeq2 for a DGE analysis
- ❖ Visualizing and evaluating expression patterns of differentially expressed genes
- ❖ Calculating differential abundance with MiloR

# Logistics



# Course schedule

## Day 1

Time	Topic	Instructor
09:30 - 09:45	Workshop introduction	Noor
09:45 - 10:30	Project setup and data exploration	Noor
10:30 - 10:40	Break	
10:40 - 11:45	Differential expression analysis using <code>FindMarkers()</code>	Meeta
11:45 - 12:00	Overview of self-learning materials and homework submission	Meeta

### Before the next class:

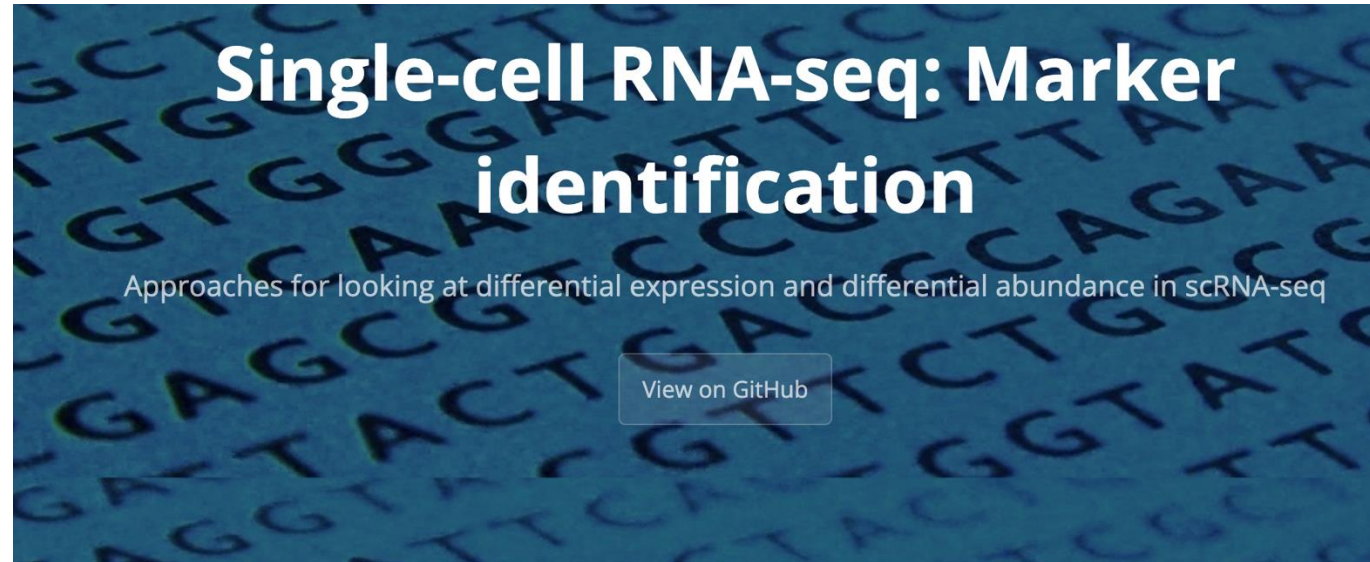
I. Please **study the contents** and **work through all the code** within the following lessons:

1. [Aggregating counts by celltype using pseudobulk approach](#)  
*Click here for a preview of this lesson*
2. [DE analysis of pseudobulk data using DESeq2](#)  
*Click here for a preview of this lesson*

<https://tinyurl.com/DGE-analysis-scRNAseq>

# Course materials

- ❖ We continuously update our materials to reflect changes in the field/software



Approximate time: 75 minutes

## Learning Objectives:

- Evaluate differential gene expression between conditions using a Wilcoxon rank sum test
- Create visualizations for differentially expressed genes
- Discuss other statistical tests for differential expression analysis

## Differential expression between conditions using `FindMarkers()`

In our current UMAP, we have merged samples across the different conditions and used integration to align cells of the same celltype across samples. Now, what if we were interested in a particular celltype and **understanding how gene expression changes across the different conditions?**

<https://tinyurl.com/DGE-analysis-scRNAseq>



# Single Screen & 3 Windows

The image shows a video conference interface with three participants: Mary Piper (Co-host, me), Jihe Liu (Host), and Troubleshooter (Radhika) (Co-host). The main window displays the RStudio environment, which is split into several panes:

- Source Editor:** Contains R code for assignment operators, functions, and rounding. The code is as follows:

```
1 # Assignment operator
2 x <- 3
3
4 # Functions
5 getwd()
6
7 sqrt(81)
8
9 round(3.14159)
10 ?round
11
```
- Console:** Shows the output of the R code:

```
> x <- 3
> # Functions
> getwd()
[1] "/Users/mariyaper/Desktop/R-testing"
> sqrt(81)
[1] 9
> round(3.14159)
[1] 3
> ?round
>
```
- Environment:** Displays the current environment, showing the variable `x` with the value `3`.
- Documentation:** Shows the R documentation for the `round` function, titled "Rounding of Numbers".

The video conference interface also includes a top bar with the text "You are viewing Jihe Liu's screen" and a bottom bar with controls for Unmute, Stop Video, Invite, Share Screen, and Reactions.

# Single Screen & 3 Windows

**Zoom**

**Our Recommendation**

The image illustrates a workflow for a single-screen setup with three windows. The Zoom window at the top shows a video call with three participants. The RStudio IDE window at the bottom left shows a script with R code and a console output. The web browser window at the bottom right shows a GitHub page for 'Rounding of Numbers'.

```
# Assignment operator
x <- 3

# Functions
getwd()
sqrt(81)
round(3.14159)
?round
```

Console output:

```
> x <- 3
> # Functions
> getwd()
[1] "/Users/mariyaper/Desktop/R-testing"
> sqrt(81)
[1] 9
> round(3.14159)
[1] 3
> ?round
>
```

Web browser content:

### Rounding of Numbers

Description

ceiling takes a single numeric argument x and returns a numeric vector containing the smallest integers not less than the corresponding elements of x.

floor takes a single numeric argument x and returns a numeric vector containing the largest integers not greater than the corresponding elements of x.

trunc takes a single numeric argument x and returns a numeric vector containing the integers formed by truncating the values in x toward 0.

round rounds the values in its first argument to the specified number of decimal places (default 0). See 'Details' about "round to even" when rounding off a 5.

signif rounds the values in its first argument to the specified number of significant digits.

Usage

```
ceiling(x)
floor(x)
trunc(x, ...)
```

# Single Screen & 3 Windows

The image illustrates a single-screen setup for a video conference, RStudio, and a web browser. The video conference window at the top left shows three participants: Mary Piper, Troubleshooter, and Jihe Liu. The RStudio window at the bottom left shows a script editor with R code, a console with output, and a viewer pane showing the 'Rounding of Numbers' documentation. The web browser window at the top right shows a page titled 'Web Browser' with a large green text overlay.

**Participants (3)**

- Mary Piper (Co-host, me)
- Jihe Liu (Host)
- Troubleshooter (Radhika) (Co-host)

**Web Browser**

**RStudio**

```
1 # Assignment operator
2 x <- 3
3
4 # Functions
5 getwd()
6
7 sqrt(81)
8
9 round(3.14159)
10 ?round
11
```

**Console**

```
> x <- 3
> # Functions
> getwd()
[1] "/Users/mariyapiper/Desktop/R-testing"
> sqrt(81)
[1] 9
> round(3.14159)
[1] 3
> ?round
>
```

**Viewer**

**Rounding of Numbers**

integers not less than the corresponding elements of x.

floor: takes a single numeric argument x and returns a numeric vector containing the largest integers not greater than the corresponding elements of x.

trunc: takes a single numeric argument x and returns a numeric vector containing the integers formed by truncating the values in x toward 0.

round: rounds the values in its first argument to the specified number of decimal places (default 0). See 'Details' about "round to even" when rounding off a 5.

signif: rounds the values in its first argument to the specified number of significant digits.

**Usage**

```
ceiling(x)
floor(x)
trunc(x, ...)
```

*Our  
Recommendation*

# Single Screen & 3 Windows

The image is a composite of three overlapping windows, demonstrating a single-screen workflow for data science. The top-left window is a video conference interface showing three participants: Mary Piper, Troubleshooter (Radhika), and Jihe Liu. The bottom-left window is an R console window showing a script with R code and its output. The right window is the R Studio IDE, which is highlighted with a pink border. It displays a script with R code, a console window with the output of the code, and a sidebar with the Environment, History, and Connections panels. The R Studio window also shows a 'Values' panel with the variable 'x' set to 3, and a 'Rounding of Numbers' section with a description of the 'round' function.

*Our Recommendation*

**R Studio**



# Single Screen & 3 Windows

The image is a composite of three overlapping windows on a dark blue background, demonstrating a 'Single Screen & 3 Windows' setup.

- Zoom Window (Left):** Shows a video call with three participants: Mary Piper, Troubleshooter (...), and Jihe Liu. Below the video feed is the R Studio interface with a script editor containing R code for rounding and square root functions. A large blue 'Zoom' label is overlaid on the left side of this window.
- Web Browser Window (Top Right):** Shows a browser window with a URL bar and a search bar. A large green 'Web Browser' label is overlaid on the right side of this window.
- R Studio Window (Bottom):** Shows the R Studio interface with a script editor, console, and environment pane. A large purple 'R Studio' label is overlaid on the right side of this window.

At the bottom of the image, the text 'Our Recommendation' is written in a large, yellow, serif font.

# 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!**





# Course participation

- ❖ 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

# Using AI for Assignments

## ❖ Do

- ❖ Try to resolve error messages with it
- ❖ Test code written by AI on a dataset where you have expected results
- ❖ Take the time to review the generated code line-by-line

## ❖ Don't

- ❖ Implement it in replacement to learning
- ❖ Write code that you don't understand
- ❖ Assume the output from an AI process is correct

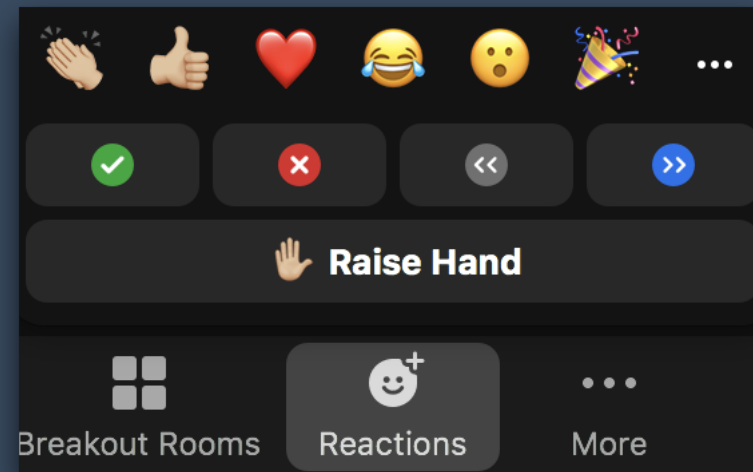
# Odds & Ends

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

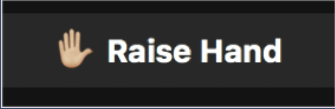
❖ Are you all set?

❖  = "agree", "I'm all set"

❖  = "disagree", "I need help"



# Odds & Ends

- ❖ Questions for the presenter?
  - ❖ Post the question in the Chat window OR
  - ❖  when the presenter asks for questions
  - ❖ Let the Moderator know

# Odds & Ends

## ❖ 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)