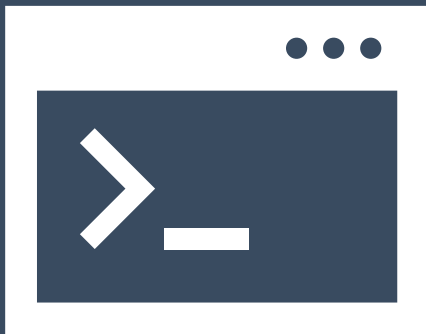


# Introduction to differential gene expression (DGE) analysis

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



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



Alex Bartlett



Elizabeth  
Partan



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



Alex Bartlett



Elizabeth  
Partan



Emma Berdan



James Billingsley



Zhu Zhuo

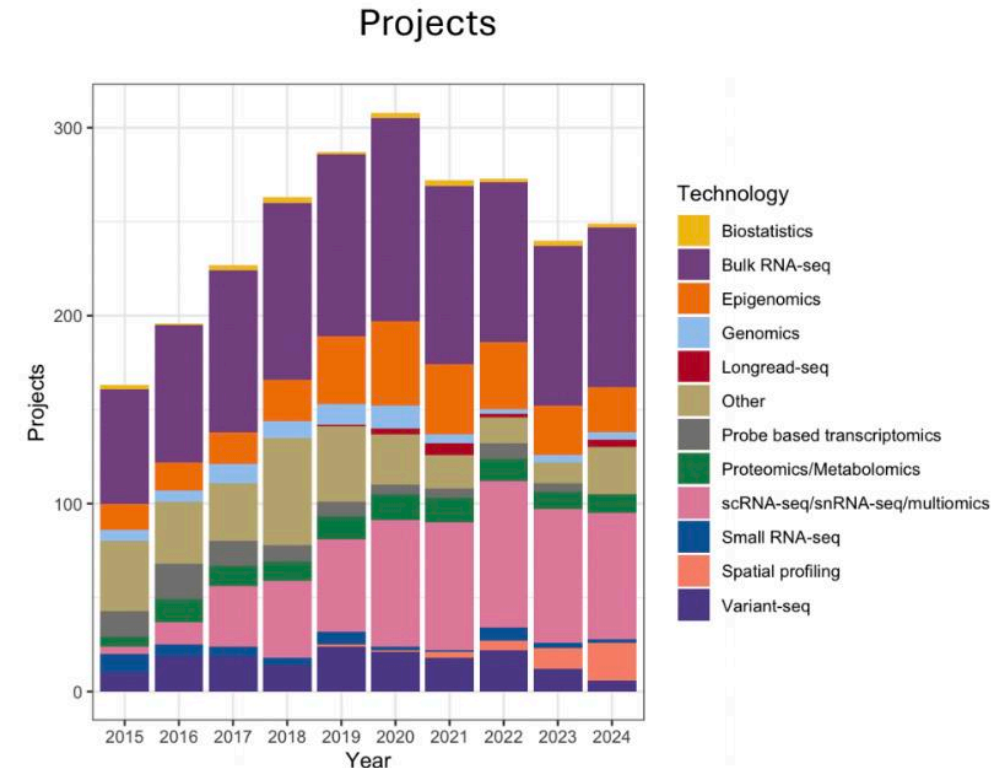


Maria Simoneau



# 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>



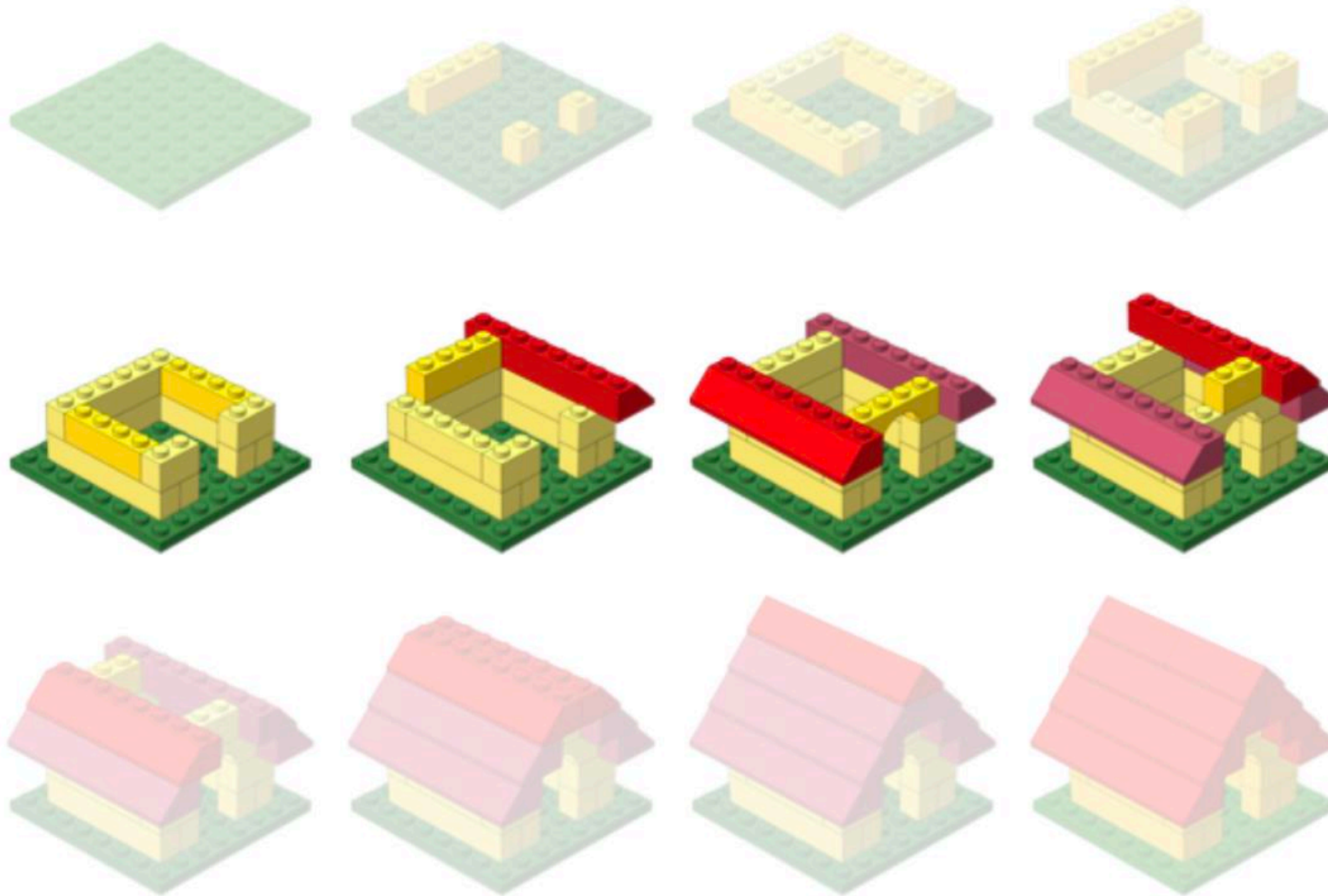
**HARVARD**  
**T.H. CHAN**  
SCHOOL OF PUBLIC HEALTH

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



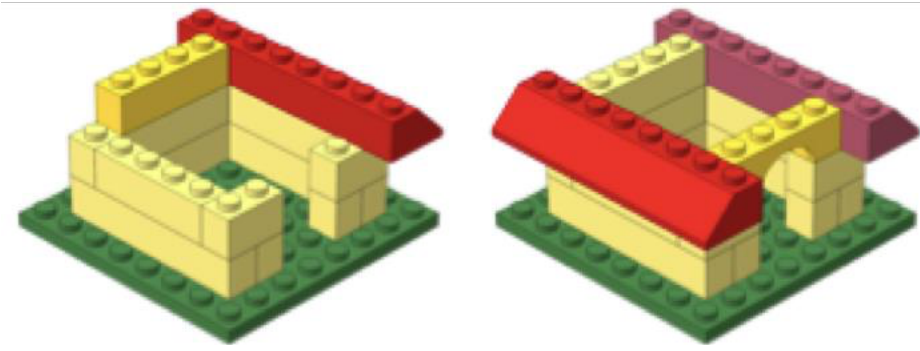


# **Workshop scope**



Bioinformatics Data Analysis

# Workshop Scope



- ❖ 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 schedule

## Workshop Schedule

### Pre-reading

1. [Workflow \(raw data to counts\)](#)
2. [Experimental design considerations](#)

### Day 1

Time	Topic	Instructor
10:00 - 10:30	<a href="#">Workshop Introduction</a>	Meeta
10:30 - 11:00	RNA-seq pre-reading discussion	All
11:00 - 11:45	<a href="#">Intro to DGE / setting up DGE analysis</a>	Noor
11:45 - 12:00	Overview of self-learning materials and homework submission	Meeta

### Before the next class:

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

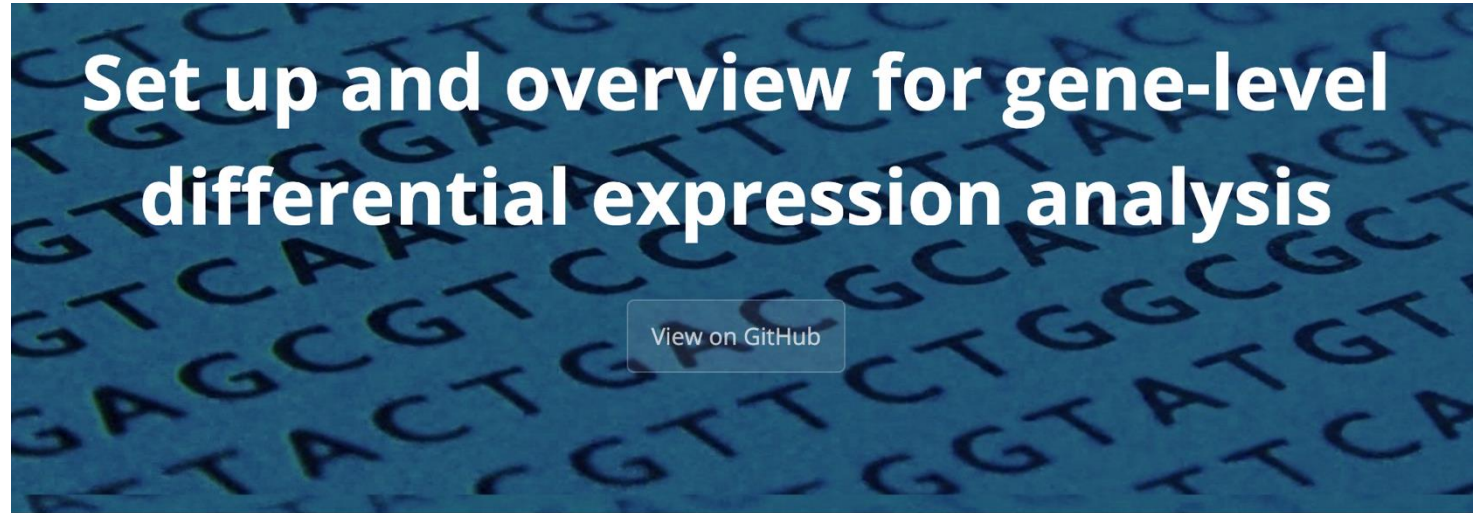
- [RNA-seq counts distribution](#)

*[Click here for a preview of this lesson](#)*

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

# Course materials

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



Approximate time: 60 minutes

## Learning Objectives

- Describe the RNA-seq and the differential gene expression analysis workflow
- Explain the experiment and its objectives
- Create a project in R
- Setup for the analysis of RNA-seq data

## Differential gene expression analysis



# 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. The top-left pane shows the R script editor with the following 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
```

The bottom-left pane shows the console output:

```
> # round function  
> round(3.14159)  
[1] 3  
> ?round  
>
```

The right pane shows the Environment window with the following values:

Variable	Value
x	3

The bottom-right pane shows the Documentation window for the 'Rounding of Numbers' topic, which includes a description of the functions ceiling, floor, trunc, round, and signif, along with their usage.

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

The console output shows the execution of the code:

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

# 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 with three windows. The top window is a video conference showing three participants: Mary Piper, Troubleshooter, and Jihe Liu. The bottom-left window is an RStudio IDE showing a script with R code and a console with output. The bottom-right window is a web browser showing a page titled "Web Browser" with a green outline.

**Participants (3)**

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

**Web Browser**

**RStudio Script:**

```
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
```

**RStudio Console:**

```
> # round function  
> round(3.14159)  
[1] 3  
> ?round  
>
```

**RStudio Environment:**

Values

x
3

**RStudio Console Output:**

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

**RStudio Environment - Rounding of Numbers**

Values

x
3

Description

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 bottom-right window is the R Studio IDE, showing a script editor, environment pane, and documentation pane.

**Video Conference Window:** Shows three participants in a video call. The status bar at the top indicates "You are viewing Jihe Liu's screen".

**R Console Window:** Displays the following R code and output:

```
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
```

**R Studio IDE Window:** Shows the R Studio interface with the following components:

- Script Editor:** 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
```
- Environment Pane:** Shows the variable `x` with the value `3`.
- Documentation Pane:** Displays the documentation for the `round` function, including a description and usage examples.

**Our Recommendation**

**R Studio**



# Single Screen & 3 Windows

The image is a composite of three overlapping windows, each with a colored border: blue for Zoom, green for a web browser, and pink for R Studio.

- Zoom (Blue border):** Shows a video call with three participants: Mary Piper, Troubleshooter (...), and Jihe Liu. Below the video feed is a code editor with 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
```
- Web Browser (Green border):** Shows a page titled "Web Browser" with a URL bar displaying "https://hspb-bioc...".
- R Studio (Pink border):** Shows the R Studio interface with a script editor, console, and environment pane. The script editor contains 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
```

The console shows the output of the code:

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

The environment pane shows the variable `x` with value `3`. The help pane shows the documentation for the `round` function.

**Our Recommendation**

**R Studio**

# 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

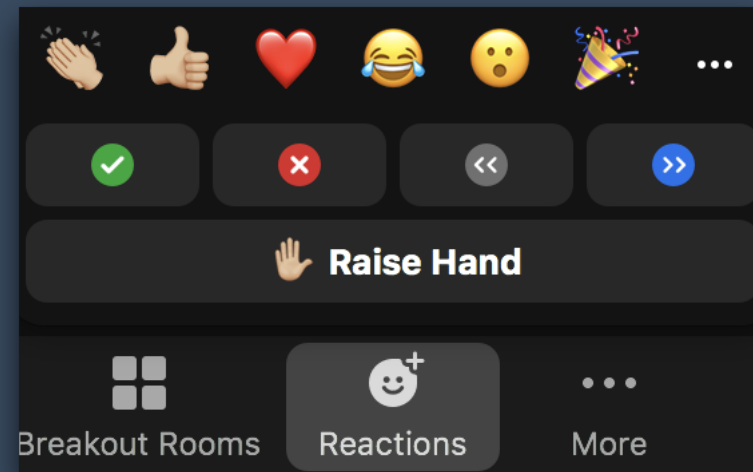
# Odds & Ends

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

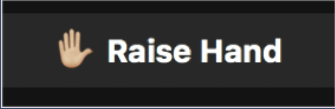
❖ Are you all set?

❖  = "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)