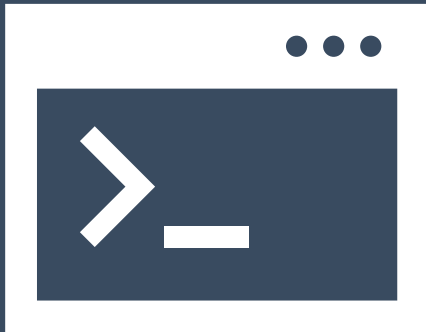


Peak Analysis

<https://tinyurl.com/Peak-analysis>



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



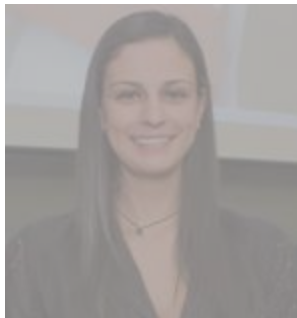
Will Gammerdinger



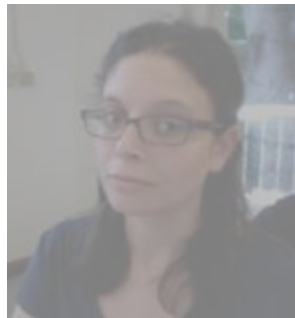
Noor Sohail



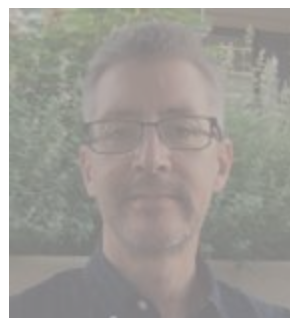
Elizabeth
Partan



Alex Bartlett



Emma Berdan



James Billingsley



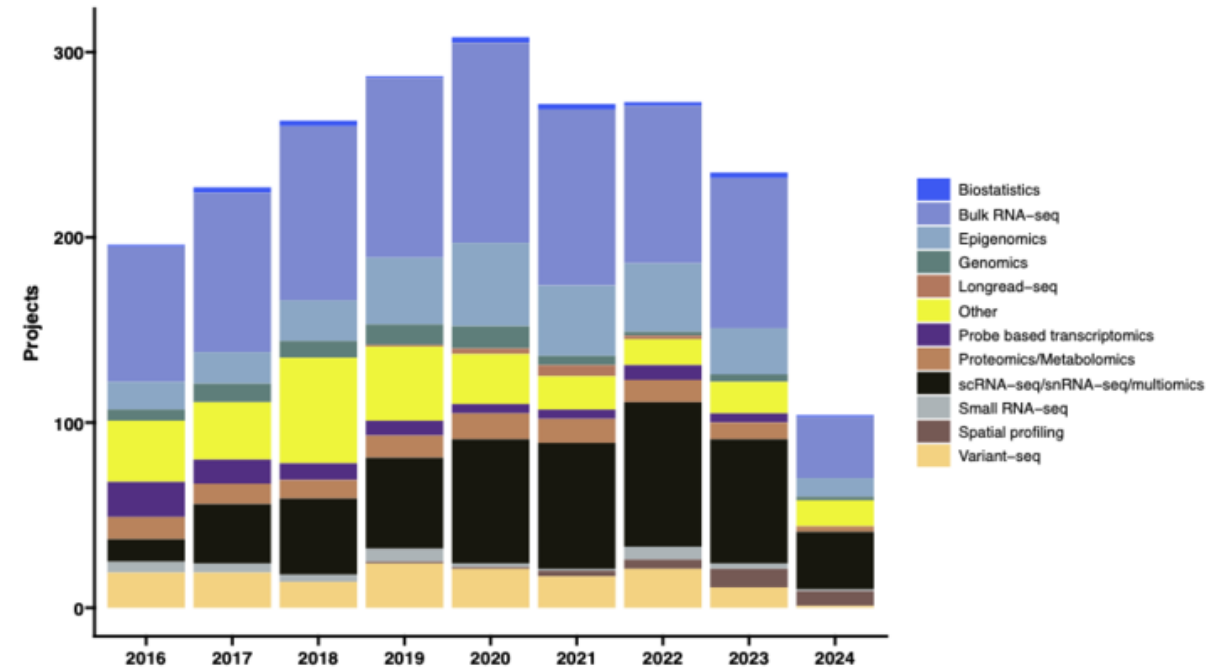
Zhu Zhuo



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



NIEHS



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 & Peak Analysis

- ❖ Bulk RNA-seq

- ❖ Differential Gene Expression (Bulk & scRNA)

- ❖ 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 & Peak Analysis
 - ❖ Bulk RNA-seq
 - ❖ Differential Gene Expression (Bulk & scRNA)
 - ❖ scRNA-seq
 - ❖ Variant Calling
 - ❖ Current Topics in Bioinformatics

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



HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

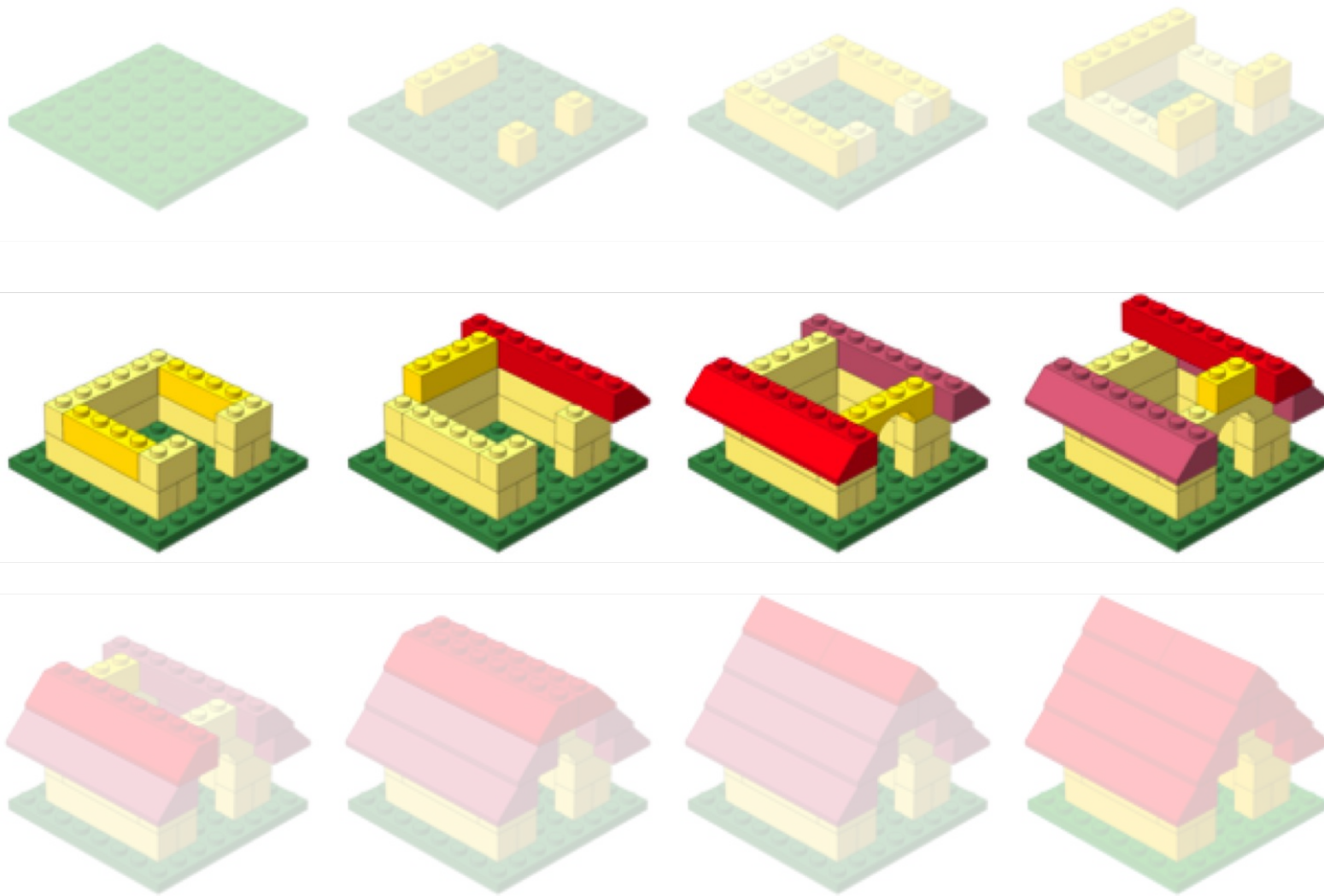
DF/HCC
DANA-FARBER / HARVARD CANCER CENTER



THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

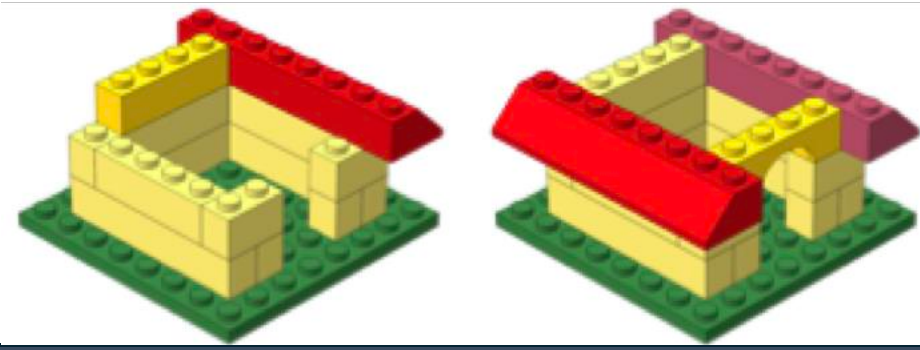


Workshop scope



Bioinformatic Data Analysis

Workshop Scope



- ❖ Describe peak data and different file formats generated from peak calling algorithms
- ❖ Assess various metrics used to assess the quality of peak calls
- ❖ Compare peak calls across samples within a dataset
- ❖ Create visualizations to evaluate peak annotations
- ❖ Evaluate differentially enriched regions between two sample groups

Logistics



Course schedule

Day 1

Time	Topic	Instructor
09:30 - 09:45	Workshop Introduction	Meeta
09:45 - 10:15	Pre-reading discussion	Meeta
10:15 - 11:00	Understanding peaks and peak file formats	Meeta
11:00- 11:05	Break	
11:05 - 12:00	Assessing peak quality metrics	Will

Before the next class:

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

1. [Assessing sample similarity and identifying potential outliers](#)

Click here for a preview of this lesson

2. [Concordance across replicates using peak overlaps](#)

Click here for a preview of this lesson

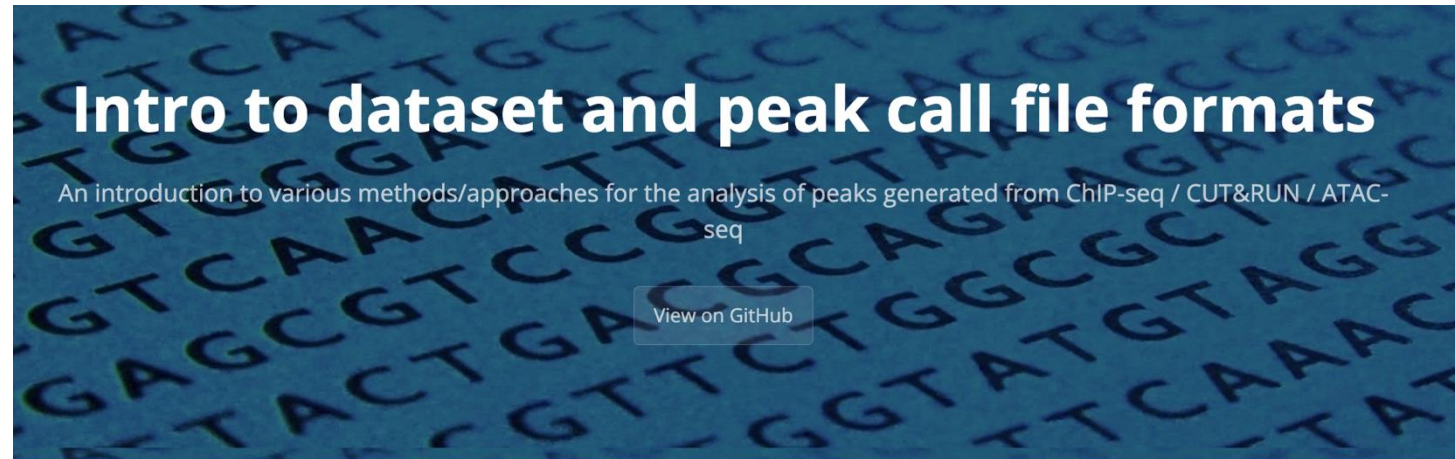
3. **Complete the exercises:**

- Each lesson above contains exercises; please go through each of them.
- Copy over your solutions into the [Google Forms](#) the **day before the next class**.

<https://tinyurl.com/Peak-analysis>

Course materials

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



Contributors: Heather Wick, Upendra Bhattarai, Meeta Mistry, Will Gammerdinger

Approximate time: 40 minutes

Learning Objectives

In this lesson, we will:

- Explain the dataset and the biological context
- Define peaks as genomic coordinate data
- Describe file formats for peak data

Introduction to the dataset

For this workshop we will be working with ChIP-seq data from a publication in Neuron by *Baizabal et al., 2018* ¹.

<https://tinyurl.com/Peak-analysis>

Single Screen & 3 Windows

The screenshot shows a video conference with three participants: Mary Piper (Co-host, me), Jihe Liu (Host), and Troubleshooter (Radhika) (Co-host). The main window displays the RStudio interface, which is split into three panes. The left pane shows the R console with the following 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  
[1] 3  
> round(3.14159)  
[1] 3  
> ?round  
>
```

The middle pane shows the R script editor with the following 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 right pane shows the RStudio Environment pane with the following values:

```
Global Environment  
x 3
```

The bottom pane shows the RStudio Console with the following output:

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

The bottom right pane shows the R Documentation for the 'round' function, titled 'Rounding of Numbers'. It includes a description of the function and its usage.

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

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: Mary Piper (Co-host, me), Jihe Liu (Host), and Troubleshooter (Radhika) (Co-host). The RStudio IDE window in the middle shows R code being executed, with the output '3' displayed. The web browser window on the right shows a GitHub page for 'Rounding of Numbers'.

```
# Assignment operator
x <- 3

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

Environment History Connections

Global Environment

Values

x 3

Files Plots Packages Help Viewer

R: Rounding of Numbers

Round (base)

R Documentation

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 R code being executed, with the console output showing the results of the code. 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 Console 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
```

RStudio Environment:

```
Values  
number 15  
x 5  
y 10
```

RStudio Script Editor:

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

RStudio Console:

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

RStudio Environment:

```
Values  
x 3
```

RStudio Help:

Rounding of Numbers

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 from a video conference. The top-left window shows a video call with three participants: Mary Piper, Troubleshooter (Radhika), and Jihe Liu. The bottom-left window shows an R console with the following 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  
[1] 3  
> round(3.14159)  
[1] 3  
> ?round  
>
```

The bottom-right window shows the R Studio interface with the following code in the script editor:

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

The R Studio window also shows the Environment pane with the variable x set to 3, and the Help pane displaying the documentation for the round function.

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:

```
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 R 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

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

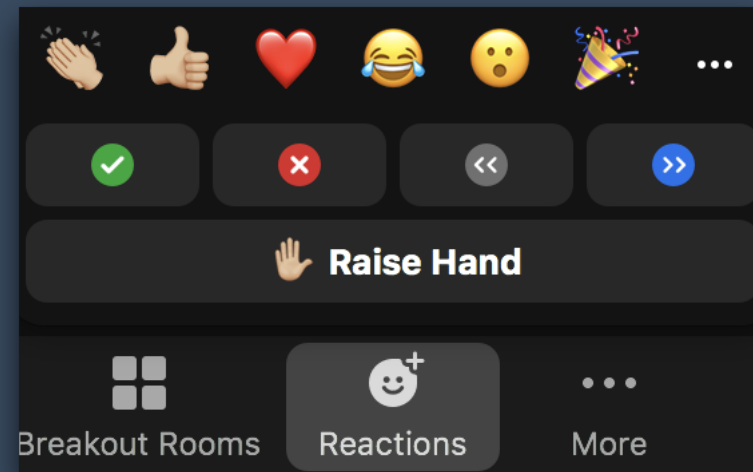
Odds & Ends

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

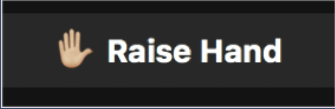
❖ Are you all set?

❖  = "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
- ❖ *HBC consulting:* bioinformatics@hsph.harvard.edu