



```
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

## *Differential Gene Expression*

Harvard Chan Bioinformatics Core

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



Shannan Ho Sui  
*Director*



Victor Barrera



James Billingsley



Zhu Zhuo



Meeta Mistry  
*Interim Director  
of Education*



Heather Wick



Will Gammerdinger



Emma Berdan



Sergey Naumenko



Maria Simoneau



Noor Sohail

# Consulting

- 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/>



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

---



# 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 training program is for researchers at Harvard to

[HBC's training team](#) is made up of experts in training and community based research projects to ensure

Our hands-on workshops place an emphasis on **experimentation** for **wet-lab biologists** analyzing their data.

We offer three types of workshops:

1. [Short, 3-hour monthly](#)
2. [Basic Data Skills](#)\*\*
3. [Advanced Topics: Analyzing](#)

\*\*The basic data skills



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

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



THE HARVARD CLINICAL  
AND TRANSLATIONAL  
SCIENCE CENTER



**HARVARD**  
**MEDICAL SCHOOL**

Our dedicated training team holds workshop to help researchers analyze their NGS data.

Participants do not have to 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 **reproducibility**. Our workshops are designed to help researchers design sequencing-based experiments and analysing the resulting

(informatics)

(NGS) data\*\*

for the advanced workshops.

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

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

# Training

A key component of the training program is for researchers at Harvard to

[HBC's training team](#) is made up of experts in training and community building. We support research projects to ensure

Our hands-on workshops have an emphasis on **experimentation** for **wet-lab biologists** analyzing their data.

We offer three types of workshops:

1. [Short, 3-hour monthly](#)
2. [Basic Data Skills](#)\*\*
3. [Advanced Topics: Analyzing](#)

\*\*The basic data skills



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

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



THE HARVARD CLINICAL  
AND TRANSLATIONAL  
SCIENCE CENTER



**HARVARD**  
MEDICAL SCHOOL

Our dedicated training team holds workshop to help researchers analyze their NGS data.

We do not devote substantial time to material development, but our training team also participate in consultations on best practices in NGS analysis.

Workshops focus on **analysis of high-throughput sequencing data**, with an emphasis on **reproducibility**. Our workshops are designed to help with planning-based experiments and analysing the resulting

(informatics)

(NGS) data\*\*

for the advanced workshops.

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

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

Introductions!



Shannan Ho Sui  
*Director*



Victor Barrera



James Billingsley



Zhu Zhuo



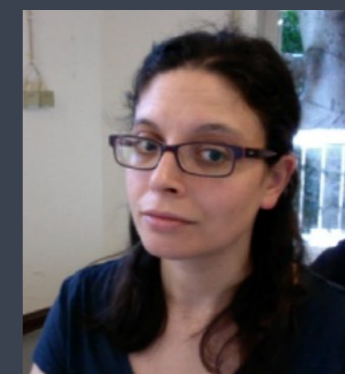
Meeta Mistry  
*Interim Director  
of Education*



Heather Wick



Will Gammerdinger



Emma Berdan



Sergey Naumenko



Maria Simoneau

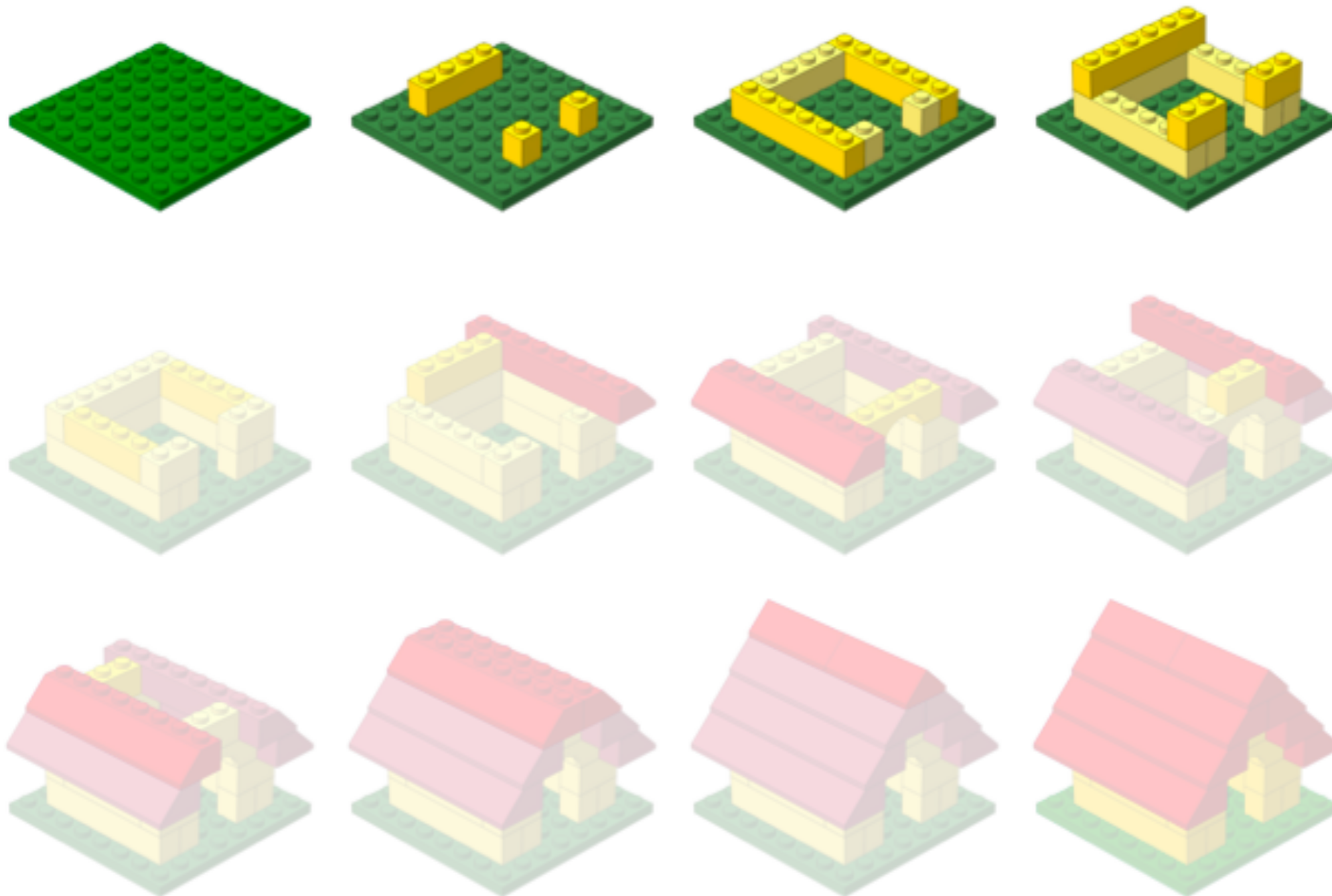


Noor Sohail



Introductions!

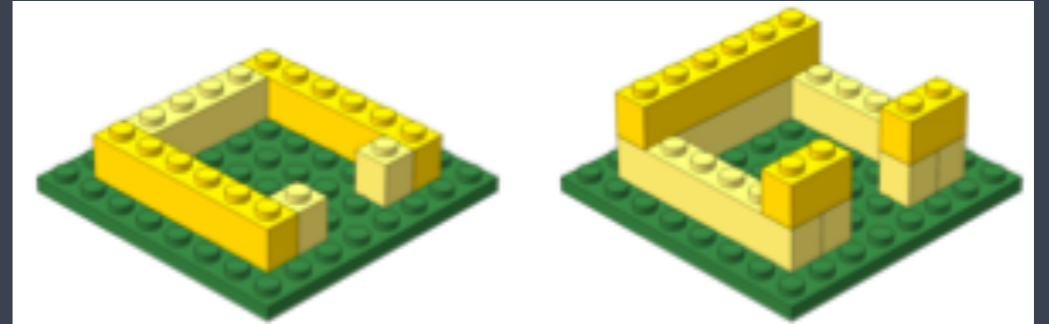
# Workshop Scope...



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

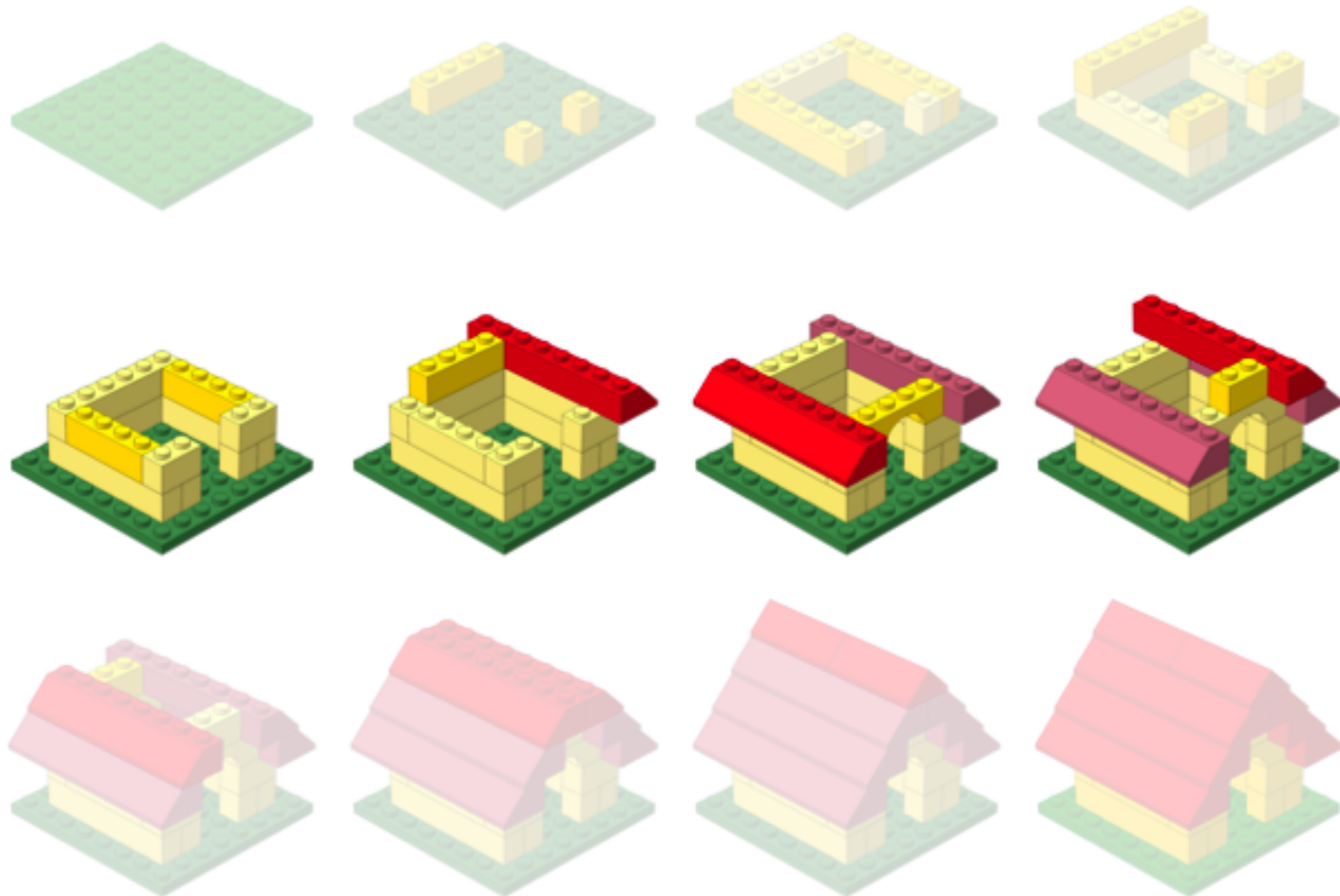
Setting up to perform Bioinformatics analysis

# 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

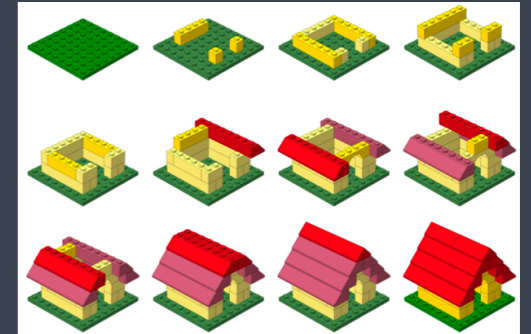




<http://anoved.net/tag/lego/page/3/>

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	<a href="#">Workshop Introduction</a>	Jihe
10:30 - 10:45	<a href="#">R refresher Q &amp; A</a>	Radhika
10:45 - 11:15	RNA-seq pre-reading discussion	Radhika
11:15 - 12:00	<a href="#">Intro to DGE / setting up DGE analysis</a>	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 running code in the lesson, [email us](#)
- Post any **conceptual questions** that you would like to have **reviewed in class** [here](#).

# Course webpage



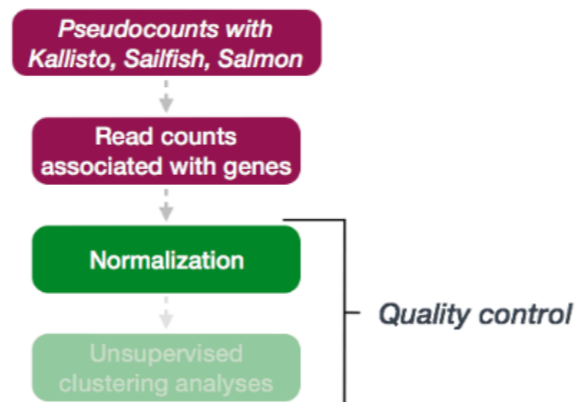
Approximate time: 60 minutes

## 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?

The image is a composite of three overlapping windows:

- Zoom Meeting:** Shows a video call with three participants: Mary Piper (Co-host, me), Troubleshooter (Radhika) (Co-host), and Jihe Liu (Host). The interface includes a 'Participants (3)' list and a 'View Options' menu.
- RStudio IDE:** Displays an R script with the following code:

```
483  
484 getwd()  
486  
487 # square root function  
488 sqrt(81)  
489  
490 # round function  
491 round(3.14159)  
492 ?round  
493  
494  
495
```

The 'Values' pane shows the results of the execution:

```
number 15  
x      5  
y      10
```
- Terminal:** Shows the execution of the R commands and their output:

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

# Single screen & 3 windows?

**ZOOM**

*Our recommendation*

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

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

**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?

**Web browser**

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

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

**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, ...)
```

*Our  
recommendation*

# Single screen & 3 windows?

The image is a composite of three overlapping windows. The top window is a Zoom meeting interface showing three participants: Mary Piper (Co-host, me), Jihe Liu (Host), and Troubleshooter (Radhika) (Co-host). The middle window is an RStudio IDE showing a script 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 bottom window is an R documentation page for the 'Rounding of Numbers' section. It includes a description of the 'round' function and its usage.

**Our recommendation**

**RStudio**

# Single screen & 3 windows?

**Web browser**

```
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
```

**ZOOM**

Participants (3)  
MP Mary Piper (Co-host, me)  
JL Jihe Liu (Host)  
T Troubleshooter (Radhika) (Co-host)

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

**RStudio**

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, ...)
```

**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

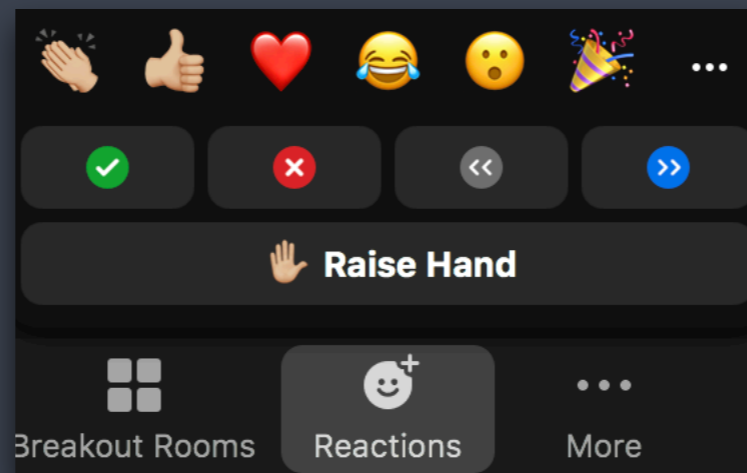
# Odds and Ends

- ❖ 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**)



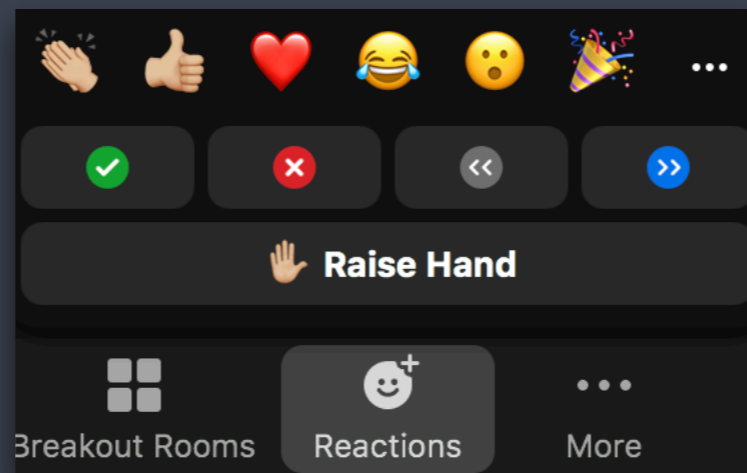
# Odds and Ends (2/2)

## ❖ Questions for the presenter?

- Post the question in the Chat window OR

-  when the presenter asks for questions

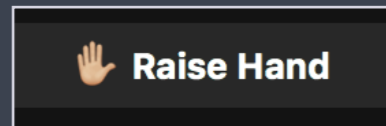
- Let the Moderator know



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

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