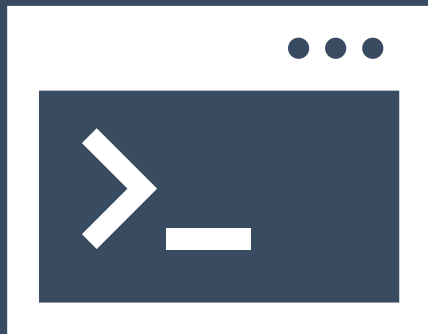


Introduction to differential gene expression (DGE) analysis

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



Harvard Chan Bioinformatics Core



Introductions!





Shannan Ho Sui
Director



Meeta Mistry
Associate Director



Lorena Pantano
*Director of Bioinformatics
Platform*



John Quackenbush
Faculty Advisor



Open 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



Open Bhattarai



Heather Wick



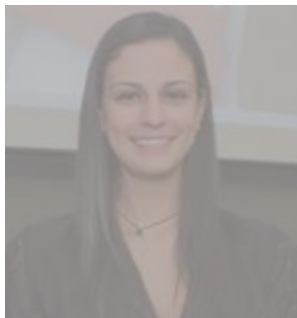
Will Gammerdinger



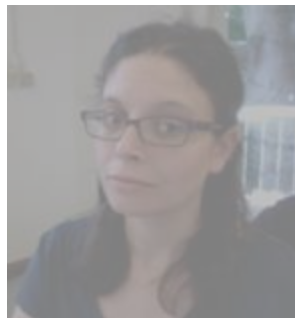
Noor Sohail



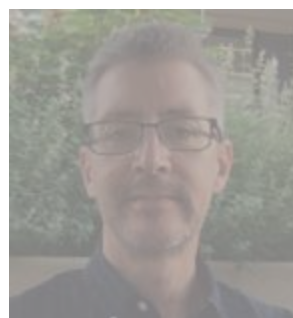
Elizabeth
Partan



Alex Bartlett



Emma Berdan



James Billingsley



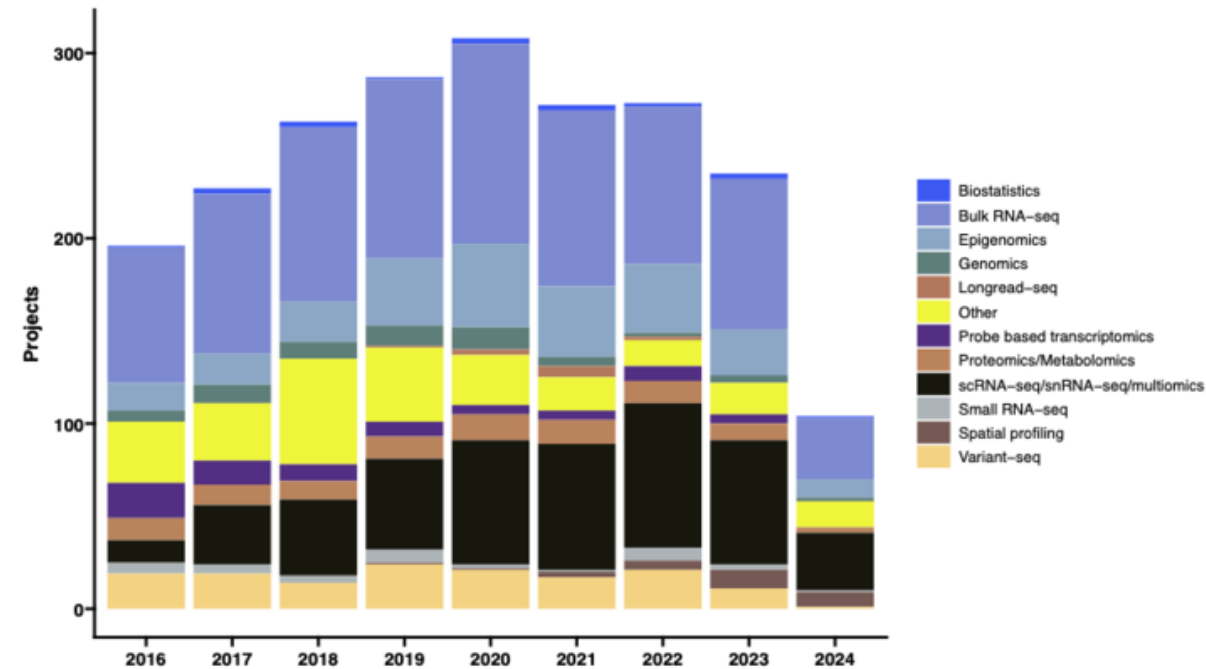
Zhu Zhuo



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

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

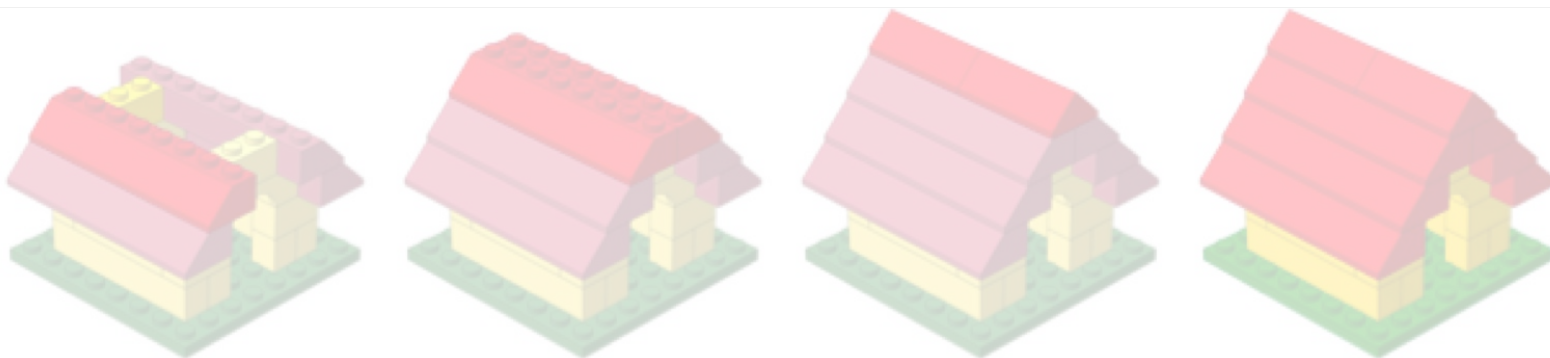
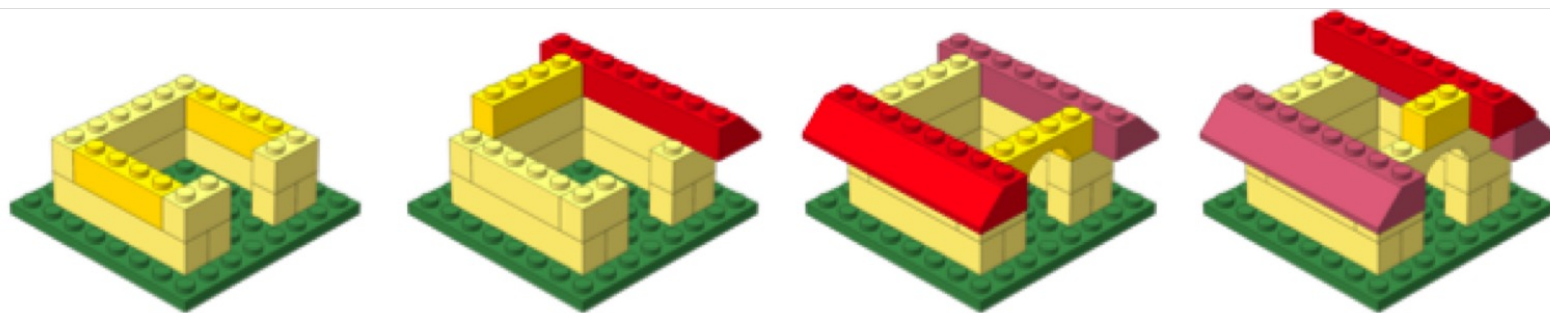
DF/HCC
DANA-FARBER / HARVARD CANCER CENTER



THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

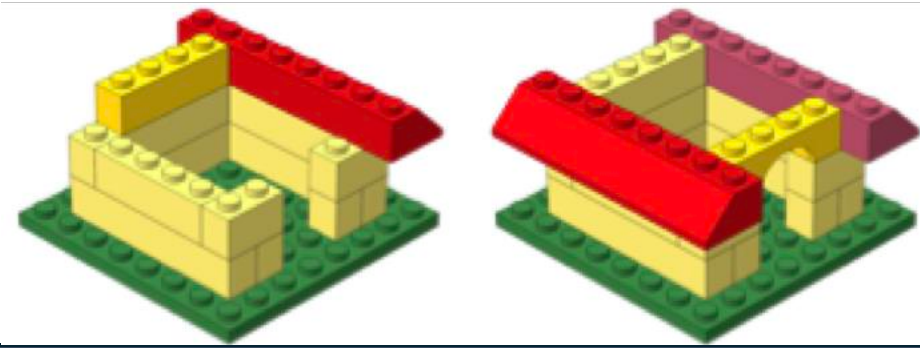


Workshop scope



Bioinformatic 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	Workshop Introduction	Meeta
10:30 - 11:00	RNA-seq pre-reading discussion	All
11:00 - 11:45	Intro to DGE / setting up DGE analysis	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:

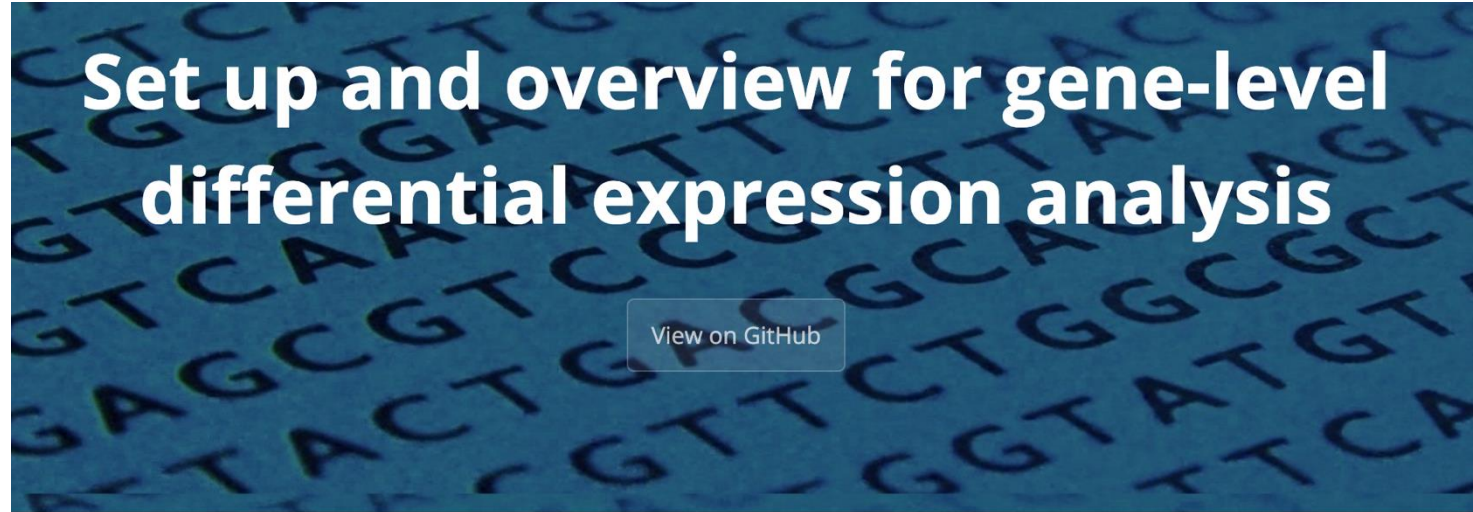
- [RNA-seq counts distribution](#)

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

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

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 Zoom meeting window with three participants: Mary Piper, Troubleshooter (Radhika), and Jihe Liu. The main screen displays three overlapping windows:

- Zoom Meeting:** Shows the video call interface with three participants and a 'Participants (3)' list.
- RStudio IDE:** Displays R code and output. The console shows the execution of several R functions:

```
> # round function
> round(3.14159)
[1] 3
> ?round
>
```
- Web Browser:** Shows a page titled 'Rounding of Numbers' with a description of the function: `round(x, digits = 0)`. The page content includes:

```
round(x, digits = 0)
signif(x, digits = 5)

Arguments
x a numeric vector, or
signif, a complex
digits integer indicating the
places (round) or the
(signif) to be used
allowed (see 'Details'
... arguments to be pa

Details
These are generic functions: #
defined for them individually
generic.
Note that for rounding off a 5,
standard (see also IEEE 754).
```

Single Screen & 3 Windows

The screenshot shows a Zoom meeting in progress. The main window displays a shared RStudio environment. The RStudio interface includes a source editor with R code, a console, and a documentation pane. The code in the source editor 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
```

The console shows the following output:

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

The documentation pane on the right shows the 'Rounding of Numbers' section, including a description of the 'round' function and its usage.

The Zoom interface shows three participants: Mary Piper (Co-host, me), Jihe Liu (Host), and Troubleshooter (Radhika) (Co-host). A 'Zoom' watermark is visible in the center of the screen.

Our Recommendation

Single Screen & 3 Windows

Participants (3)

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

Values

x	3
---	---

Rounding of Numbers

Description

Values 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, ...)
```

Web Browser

*Our
Recommendation*

Single Screen & 3 Windows

The image is a composite of three windows. The top-left window is a Zoom meeting interface showing three participants: Mary Piper, Troubleshooter (Radhika), and Jihe Liu. The top-right window is a browser showing a search for 'Rounding of Numbers' in R Documentation. The bottom window is an RStudio window showing R code being executed, with the console output showing the results of the code.

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

Console Terminal

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

Values

Variable	Value
x	3

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*

R Studio

Single Screen & 3 Windows

Zoom

Web Browser

R Studio

```
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/mariyaper/Desktop/R-testing"
> sqrt(81)
[1] 9
> round(3.14159)
[1] 3
> ?round
>
```

Participants (3)

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

Environment History Connections

Global Environment

Values

x	3
---	---

Files Plots Packages Help Viewer

R: Rounding of Numbers

Round (base)

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

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

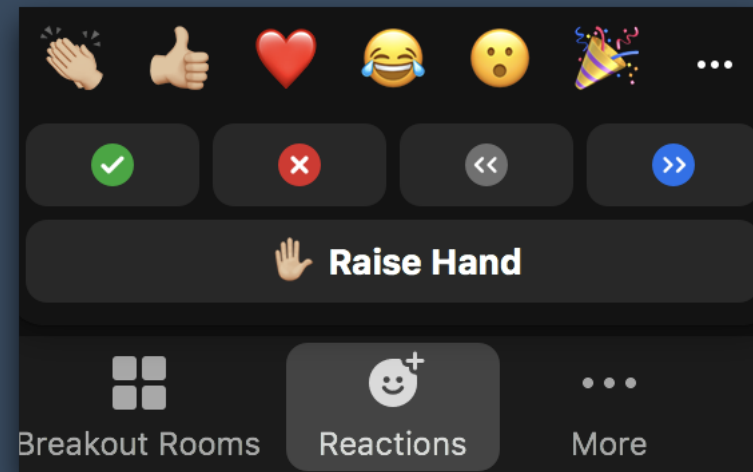
Odds & Ends

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

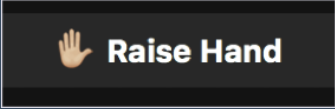
❖ Are you all set?

❖  = "agree", "I'm all set"

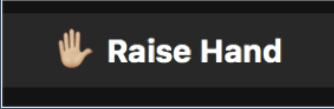
❖  = "disagree", "I need help"



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