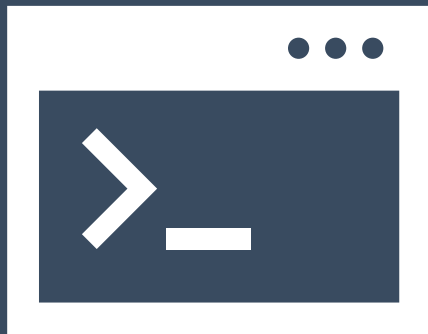


Pseudobulk and related approaches for scRNA-seq analysis

<https://tinyurl.com/hbc-pseudobulk-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



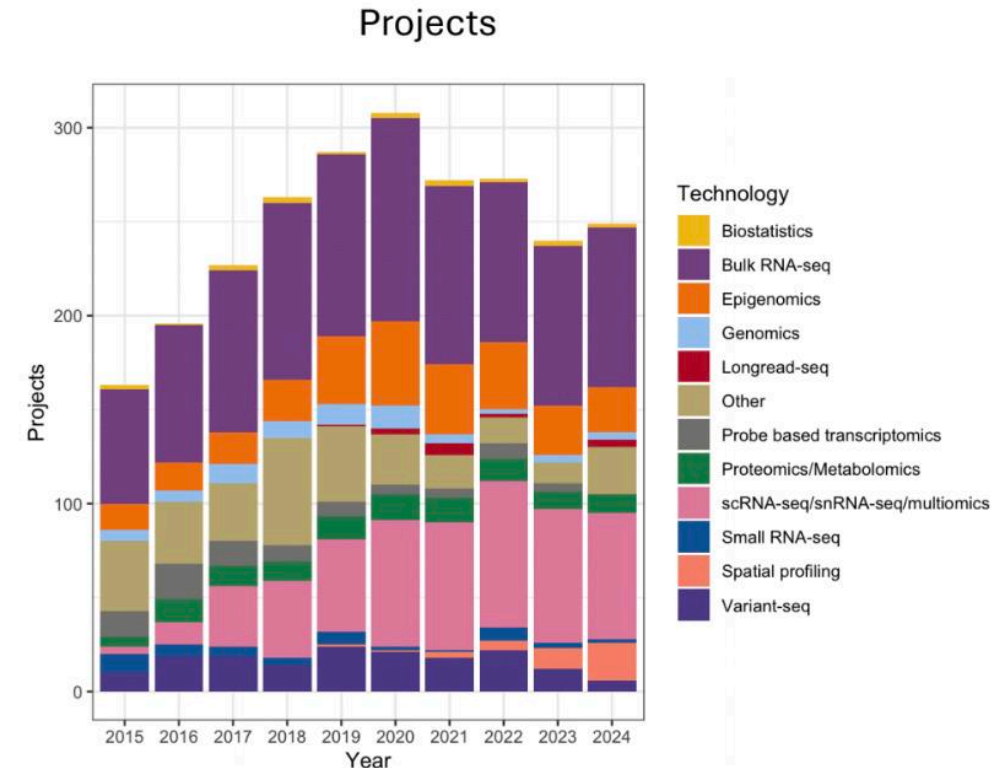
Zhu Zhuo



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



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

- ❖ 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>

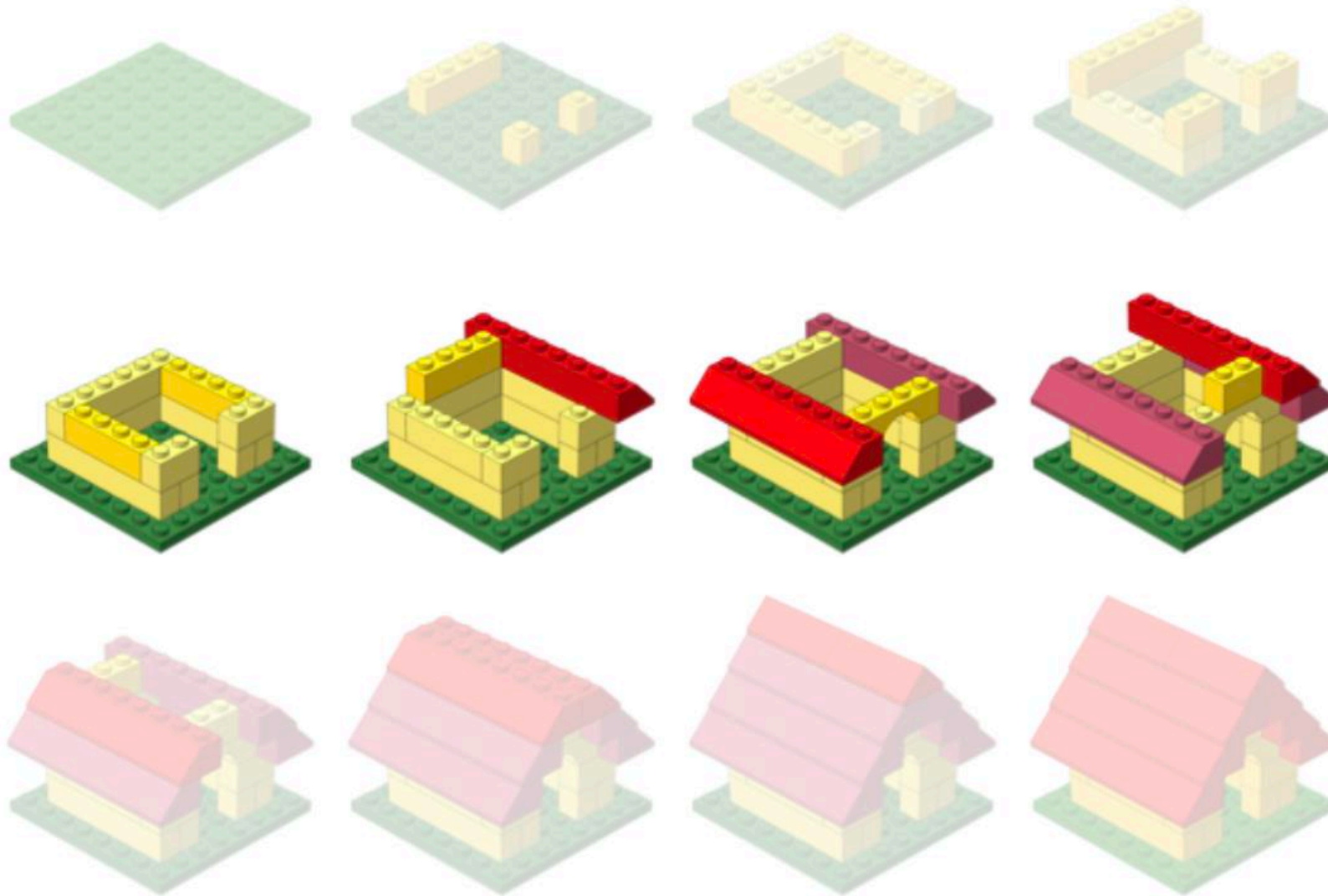


HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

DF/HCC
DANA-FARBER / HARVARD CANCER CENTER

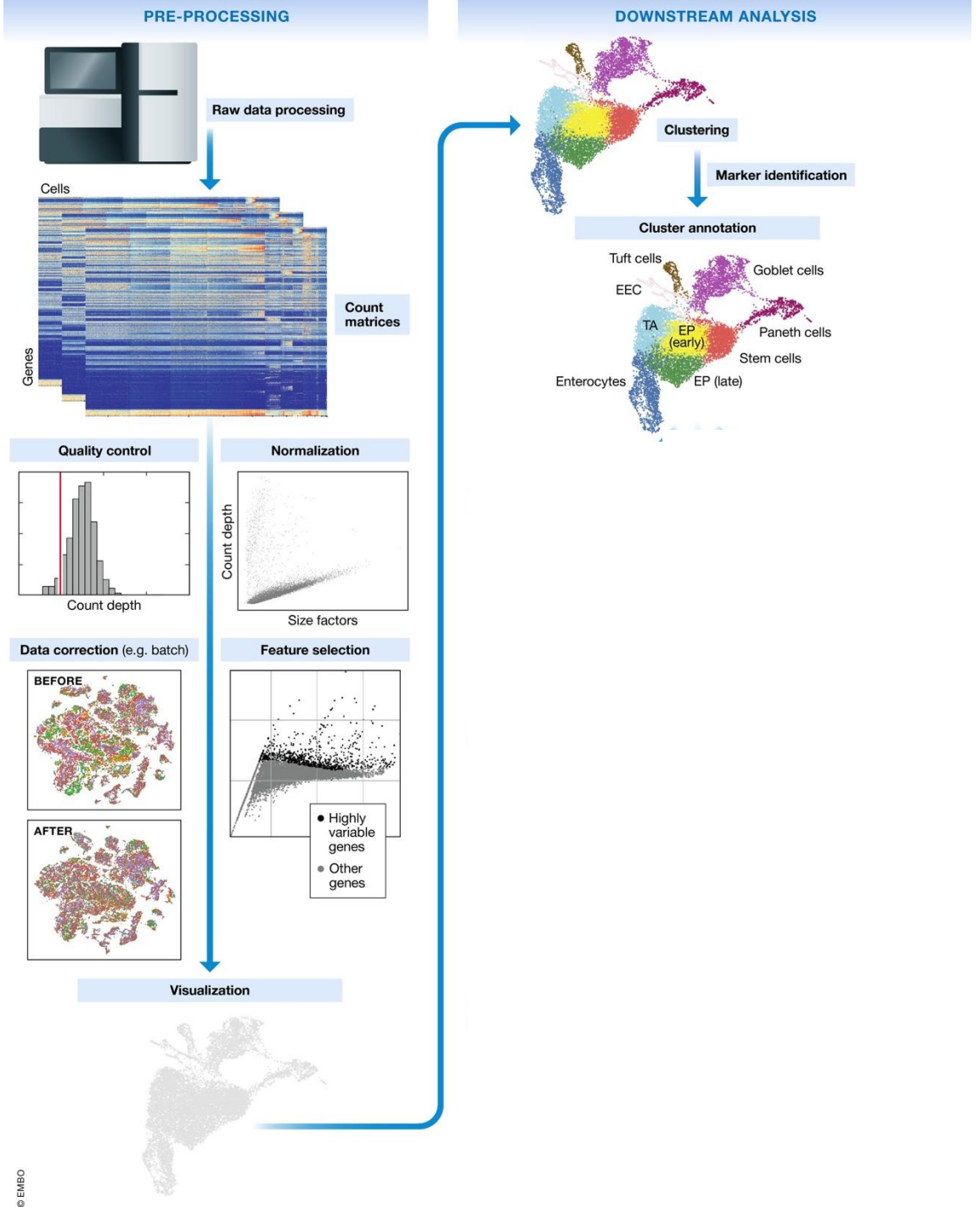


Workshop scope



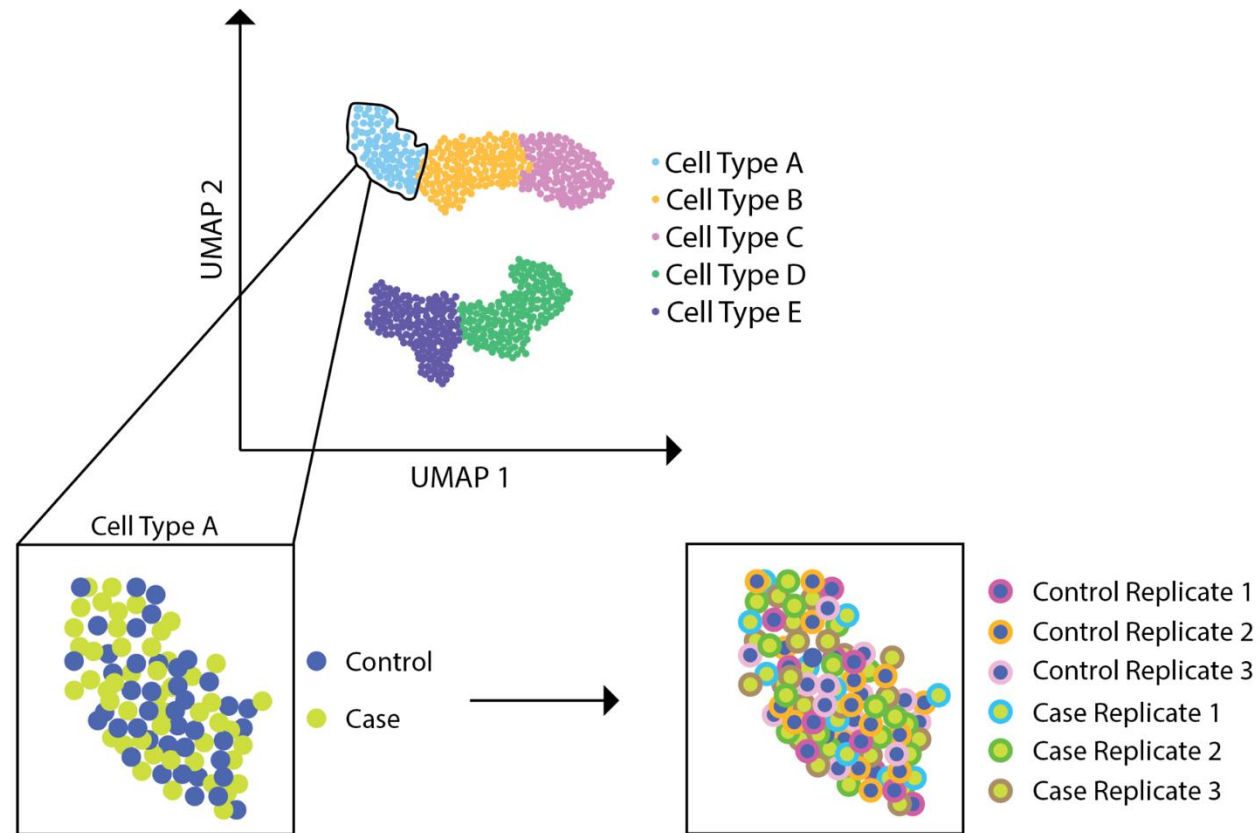
Bioinformatics Data Analysis

What is NOT covered in this workshop

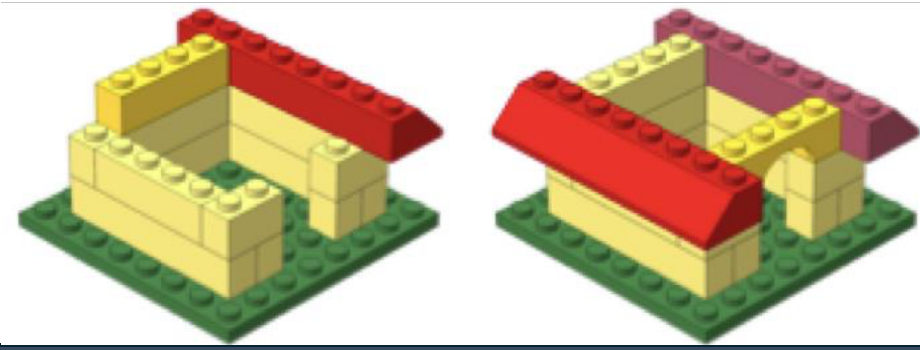


Luecken, MD and Theis, FJ. Current best practices in single-cell RNA-seq analysis: a tutorial, Mol Syst Biol 2019
(doi: <https://doi.org/10.15252/msb.20188746>)

What IS covered in this workshop



Workshop Scope



- ❖ Understanding considerations for when to use different DGE algorithms on scRNA-seq data
- ❖ Using FindMarkers to evaluate significantly DE genes
- ❖ Aggregating single cell expression data into a pseudobulk counts matrix to run a DESeq2 workflow
- ❖ Evaluating expression patterns of differentially expressed genes at the pseudobulk and single cell level
- ❖ Application of methods for evaluating differential proportions of cells between conditions

Logistics



Course schedule

Day 1

Time	Topic	Instructor
09:30 - 09:45	Workshop introduction	Will
09:45 - 10:15	Pre-reading review and Q&A	All
10:15 - 10:25	Break	
10:25 - 11:00	Project setup and data exploration	Will
11:00 - 11:50	Differential expression analysis using <code>FindMarkers()</code>	Noor
11:50 - 12:00	Overview of self-learning materials and homework submission	Will

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](#)

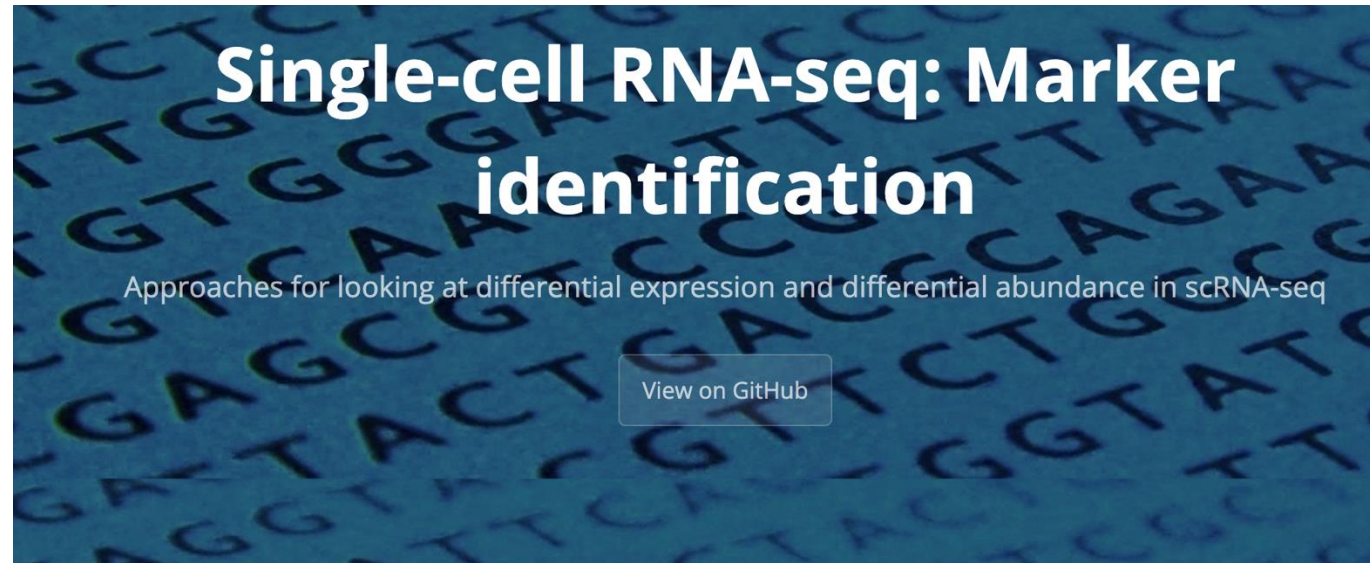
2. [DE analysis of pseudobulk data using DESeq2](#)

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

<https://tinyurl.com/hbc-pseudobulk-DGE>

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/hbc-pseudobulk-DGE>

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 console 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 R console output:

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

The top-right pane shows the RStudio Environment pane with the following values:

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

The bottom-right pane shows the RStudio Console pane 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-right pane also shows the RStudio Console pane with the following code:

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

The bottom-right pane also shows the RStudio Documentation pane for 'Rounding of Numbers'.

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

```
# Assignment operator
x <- 3

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

11.1 (Top Level) Console Terminal

```
> x <- 3
> # Functions
> getwd()
[1] "/Users/mariyaper/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

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 workflow on a single screen:

- Top Left (Video Conference):** A Zoom window titled "You are viewing Jihe Liu's screen". It shows three participants: Mary Piper, Troubleshooter (...), and Jihe Liu. The interface includes video thumbnails, a chat area, and controls like "Unmute", "Stop Video", "Invite", "Share Screen", and "Reactions".
- Bottom Left (R Console):** A terminal window showing R code and output. The code includes comments and function calls like `getwd()`, `sqrt(81)`, `round(3.14159)`, and `?round`. The output shows the current directory and the result of the `round` function.
- Right (R Studio IDE):** The R Studio interface, which is highlighted with a pink border. It consists of several panes:
 - Source:** Contains R code for assignment (`x <- 3`), functions (`getwd()`, `sqrt(81)`, `round(3.14159)`), and a help call (`?round`).
 - Environment:** Shows the current environment with a variable `x` of type `numeric` and value `3`.
 - Files:** Displays the file explorer for the project directory.
 - Plots:** A pane for visualizing data.
 - Packages:** A pane for managing installed and available packages.
 - Help:** Displays the documentation for the `round` function, including its description and usage.

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 this window.
- Web Browser Window (Top Right):** Shows a browser with a URL bar containing 'https://hsph-bioc...'. A large green 'Web Browser' label is overlaid on this window.
- R Studio Window (Bottom):** Shows the R Studio interface with a script editor containing R code for assignment operators, functions, and rounding. A large purple 'R Studio' label is overlaid on this window.

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

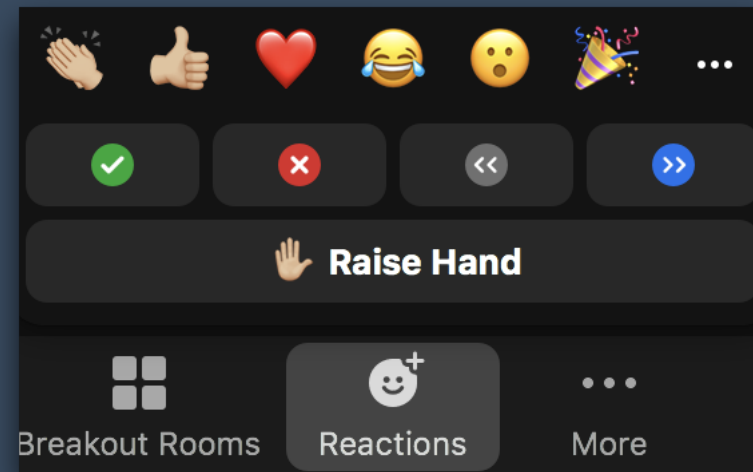
Odds & Ends

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

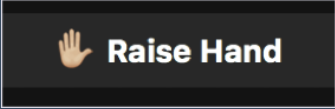
❖ Are you all set?

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