

# Differential gene expression analysis workshop

---



**Owen M. Wilkins, PhD**

Bioinformatics scientist

Center for Quantitative Biology, Geisel School of Medicine at Dartmouth

**Email:** [DataAnalyticsCore@groups.dartmouth.edu](mailto:DataAnalyticsCore@groups.dartmouth.edu)

**Website:** (<https://sites.dartmouth.edu/cqb/projects-and-cores/data-analytics-core/>)

---

08/23/21



# Who are we? The Data Analytics Core



## Mission statement

*Facilitate advanced genomic & bioinformatic data analysis solutions to CQB faculty & the Dartmouth research community*



**James O'Malley, PhD**  
Director



**Shannon Soucy, PhD**  
Senior Research Scientist



**Tim Sullivan, BA**  
Bioinformatics Research Scientist



**Owen Wilkins, PhD**  
Bioinformatics Research Scientist

### ➤ Genomic data analysis

- Analytical support for Dartmouth researchers
- Pipeline development & maintenance

### ➤ Bioinformatics consulting

### ➤ Publication & grant support

- Writing for methods & results sections
- Letters of support

### ➤ Training

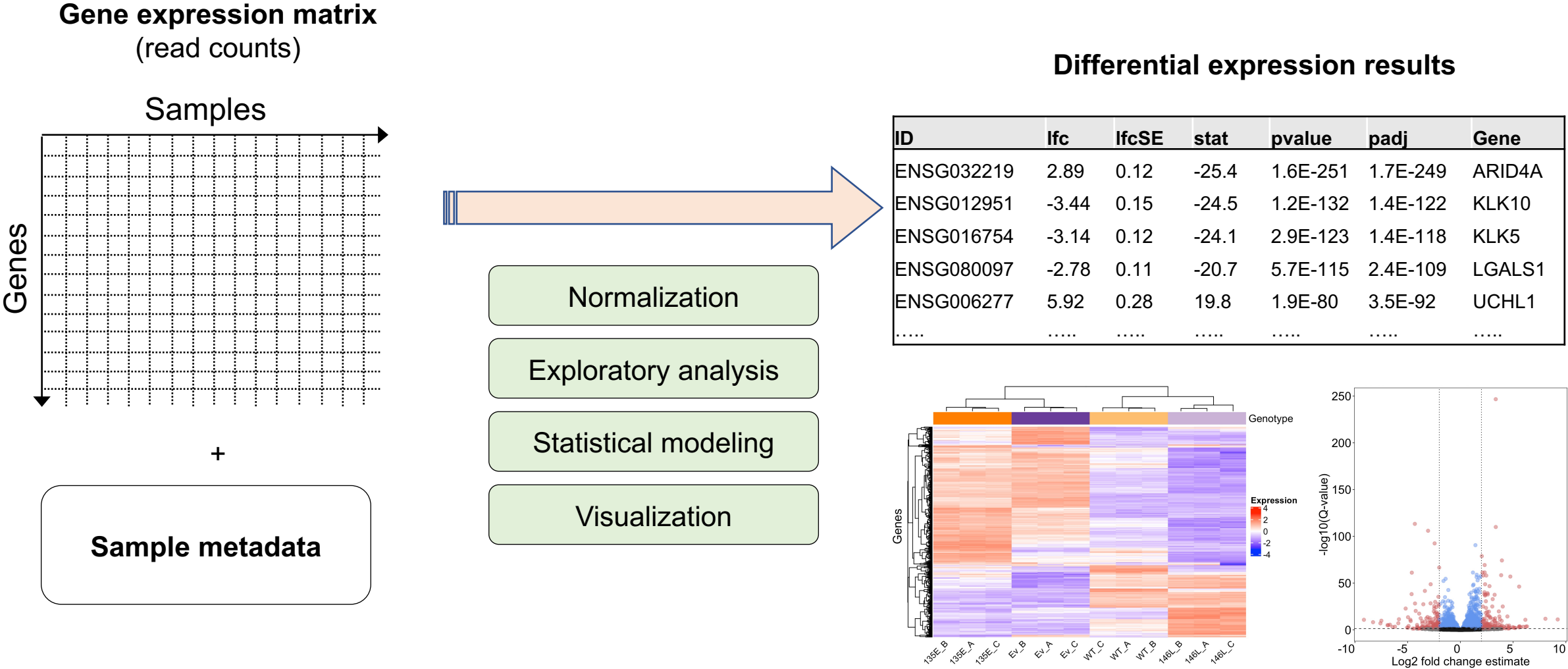
- Group workshops

# Goals of the workshop

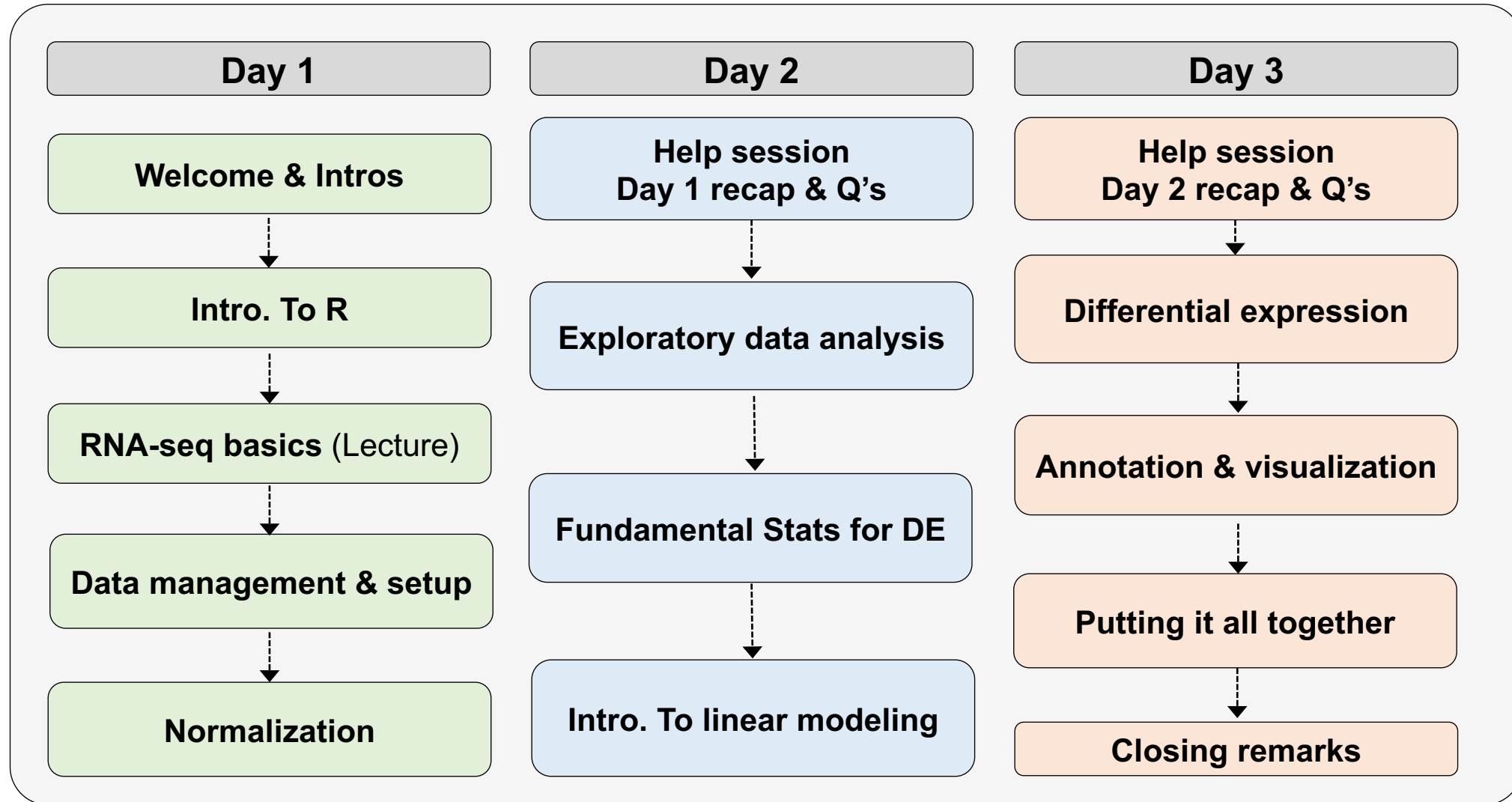


- Understand the basic principles of a differential expression analysis using RNA-seq data
- Develop a working understanding of the fundamental statistics behind a typical differential expression analysis using R/Bioconductor packages
- Perform a differential expression analysis using R/Bioconductor packages
- Learn how to explore the results and make robust insights from your data

# Workshop outline



# Workshop outline



# Schedule

- Can be found at: <https://github.com/Dartmouth-Data-Analytics-Core/RNA-seq-Differential-Expression-workshop-June-2021/blob/master/schedule.md>
- 12pm-5pm each day
- Schedule is best guess, and we may deviate from it based on time
- If you will be absent for a session, just let us know

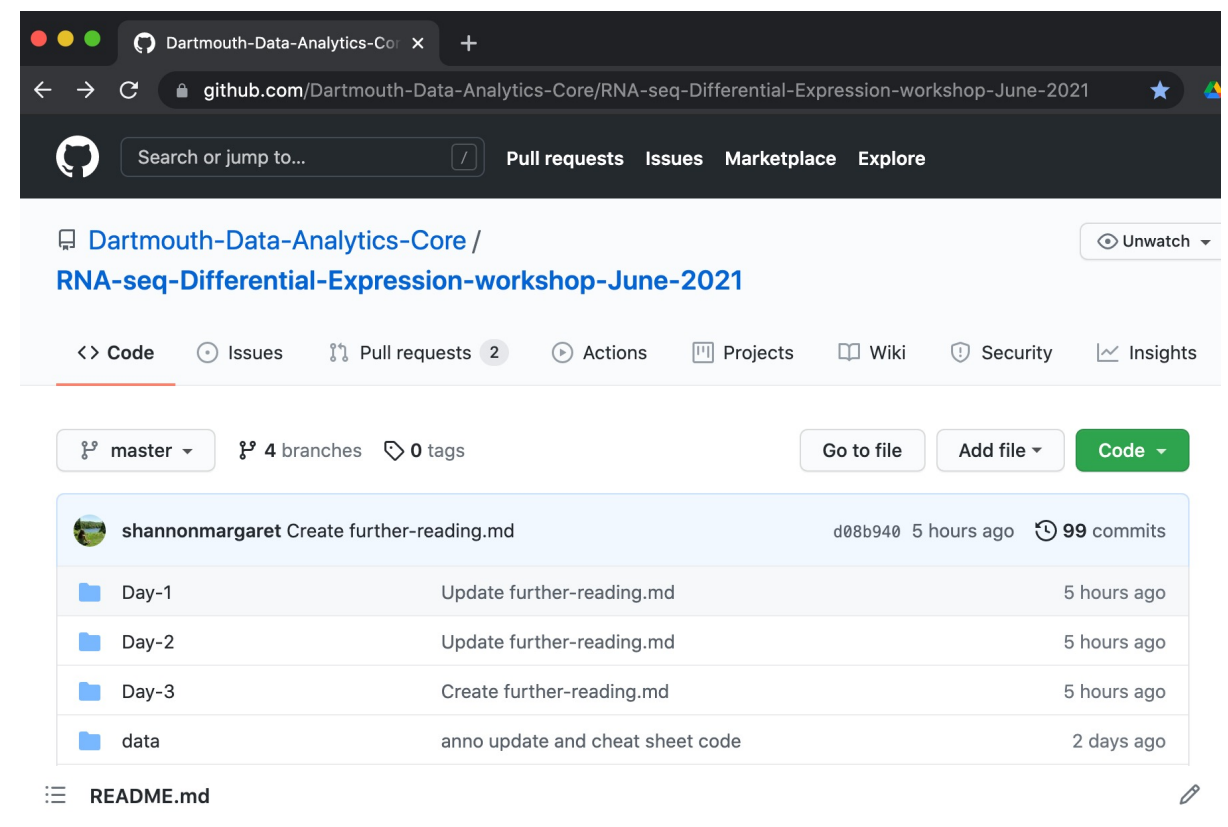
# Logistics I

- Course materials are all online (and will stay there):

<https://github.com/Dartmouth-Data-Analytics-Core/RNA-seq-Differential-Expression-workshop-June-2021>

- You will download the materials on your local machine during the workshop

- We will be copying R code from markdowns (.md) into the RStudio



The screenshot shows a GitHub repository page. The repository name is 'Dartmouth-Data-Analytics-Core / RNA-seq-Differential-Expression-workshop-June-2021'. The page includes tabs for 'Code', 'Issues', 'Pull requests' (with a count of 2), 'Actions', 'Projects', 'Wiki', 'Security', and 'Insights'. Below the repository name, there are buttons for 'Go to file', 'Add file', and 'Code'. A commit history table is visible, showing commits by 'shannonmargaret' with details like 'Create further-reading.md', 'Update further-reading.md', and 'anno update and cheat sheet code'. The commit hash 'd08b940' and '5 hours ago' are also shown. At the bottom, there is a 'README.md' file icon.

| Commit Hash | Author          | Message                          | Time        |
|-------------|-----------------|----------------------------------|-------------|
| d08b940     | shannonmargaret | Create further-reading.md        | 5 hours ago |
|             |                 | Update further-reading.md        | 5 hours ago |
|             |                 | Update further-reading.md        | 5 hours ago |
|             |                 | Create further-reading.md        | 5 hours ago |
|             |                 | anno update and cheat sheet code | 2 days ago  |

## RNA-seq differential expression workshop, August 2021

This workshop will be delivered on August 23, 25, & 27 by the Data Analytics Core (DAC) of the [Center for Quantitative Biology at Dartmouth](#).

The DAC aims to facilitate advanced bioinformatic, computational, and statistical analysis of complex genomics data for the Dartmouth research community.

If you have questions about this workshop, or would like to discuss data analysis services available from the Data Analytics Core, please visit our [website](#), or email us at: [DataAnalyticsCore@groups.dartmouth.edu](mailto:DataAnalyticsCore@groups.dartmouth.edu)



# Logistics II

- Multiple tabs open:
  - Web browser
  - RStudio
- Copy & paste code from Markdowns to Rstudio
- If you finish: edit the code, try different options, generate scripts
- Use the *cheat sheets for R!*

The screenshot displays a web browser window on the left and an RStudio interface on the right. The browser window shows a GitHub page for 'RNA-seq-Differential-Expression' with a code editor containing R code for filtering and saving results. The RStudio window shows the 'Global Environment' pane with various objects like 'colData\_sub', 'cts', 'dds', 'ha', 'ht1', 'mat\_scaled', 'mat1', 'p', and 'n2'. The 'Plots' pane shows a 'Hierarchical clustering of DEGs (padj<0.05)' heatmap. The heatmap has 'Genes' on the y-axis and 'Group' on the x-axis. The color scale for 'Expression' ranges from -4 (blue) to 4 (red). The 'Group' legend indicates 'Dex' (blue) and 'untreated' (red). The heatmap shows a clear separation of genes between the two groups.

377 lines (279 sloc) | 16.7 KB

```
head(res_ord)
```

We've now added a lot of useful information to our results that will help us interpret them in more detail. We may also wish to restrict the table to only those results that were statistically significant (at a threshold of 5%).

```
# subset @ 5% adjusted pval sig. level
res_order_FDR_05 <- res_ord[res_ord$padj<0.05,]
nrow(res_order_FDR_05)
```

Now write the table to a .csv file so that you can view it in other software (e.g. Excel) or share with others.

```
# write csv file for complete results
write.csv(as.data.frame(res_ord), file="DE_results.csv")

# write csv for significant results
write.csv(as.data.frame(res_order_FDR_05), file="DE_results.FDR.0.05.csv")
```

### Part 2: Visualization of Differential Expression

Several specific plot types exist that are useful for visualizing the results of a differential expression analysis, each providing insight on complimentary aspects of the results.

Below we will explore the major plot types useful for visualization of RNA-seq differential expression results, including:

- Volcano plots
- MA plots
- Heatmaps (hierarchical clustering)

#### Volcano plot

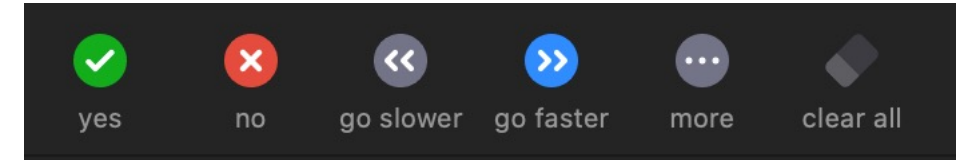
Volcano plots contrast the **log2 fold change** (effect size) against the **-log10 P-value** (statistical significance). The  $-\log_{10}()$  of a really small number is a very large value, therefore any gene that has a very small  $P$ -value will appear higher up along the y-axis. In contrast, the  $-\log_{10}$  of 1 is equal to 0, therefore genes with low statistical significance ( $P$ -values approaching 1) will appear lower down on the y-axis.

```
> draw(ht1, row_title = "Genes", column_title = "Hierarchical clustering of DEGs (padj<0.05)")
> |
```

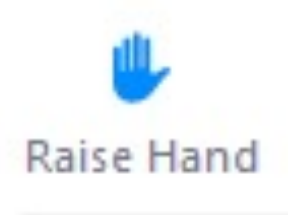


# Logistics III

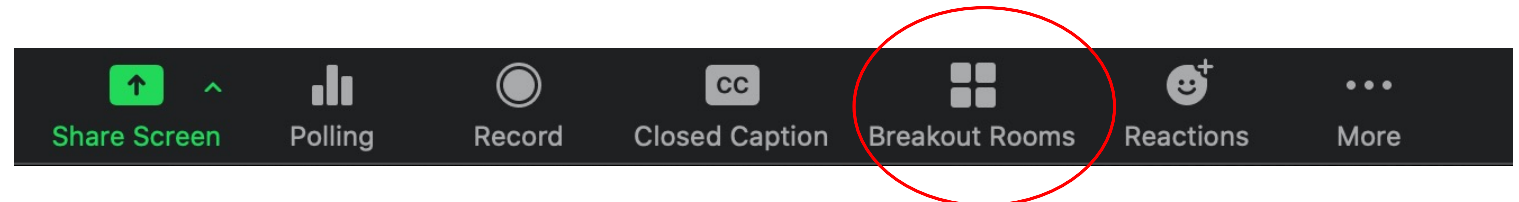
- Use buttons in Participants tab in zoom



- You'll be muted, but if you want to ask a question, just raise your hand



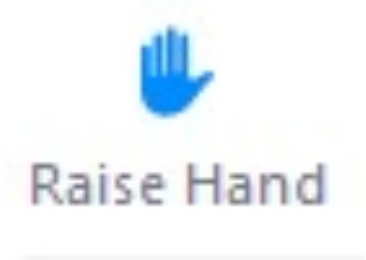
- We'll be using *breakout rooms (BRs)*
  - We will use these when we split up to run code independently
  - We've tried to pair everyone based on experience
  - If your stuck, message us, and we will come help you in your (BR)
  - When we are going to move on, breakout rooms will close



- Please be courteous on zoom..

# How to get help?

- **Raise your hand in zoom** (bottom right, participants tab)



- **Use the slack channel to message one of us**
  - Use the ***general*** channel if it might benefit everyone
  - Message us directly if its specific



- **If all else fails, email us:**
  - DAC: [DataAnalyticsCore@groups.dartmouth.edu](mailto:DataAnalyticsCore@groups.dartmouth.edu)
  - Shannon Soucy ([Shannon.Margaret.Soucy@Dartmouth.edu](mailto:Shannon.Margaret.Soucy@Dartmouth.edu))
  - Tim Sullivan ([Timothy.J.Sullivan@dartmouth.edu](mailto:Timothy.J.Sullivan@dartmouth.edu))
  - Owen Wilkins ([omw@Dartmouth.edu](mailto:omw@Dartmouth.edu))

# Questions after the workshop?

## Bioinformatics office hours

Friday 1-2pm (every other week, check calendar):  
<https://sites.dartmouth.edu/cqb/upcoming-events/calendar/>

Zoom link: <https://dartmouth.zoom.us/s/96998379866>

Passcode: bioinfo

- At the end, please give us feedback about this workshop, there will be a survey!
- And please ask lots of questions!

# Questions?

## **...then.. Introductions!**

Name, department/program, research interests, why are you taking the workshop?