# 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

**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
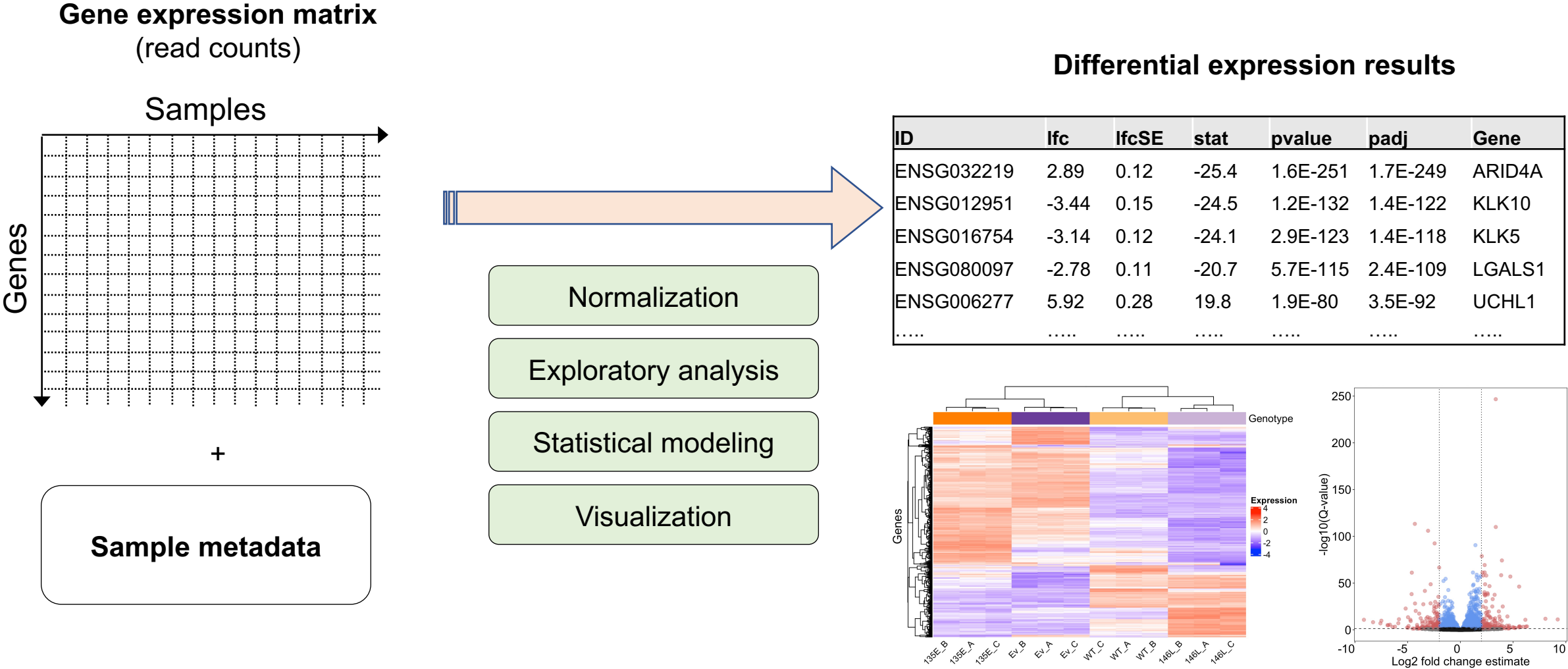  - Letters of support

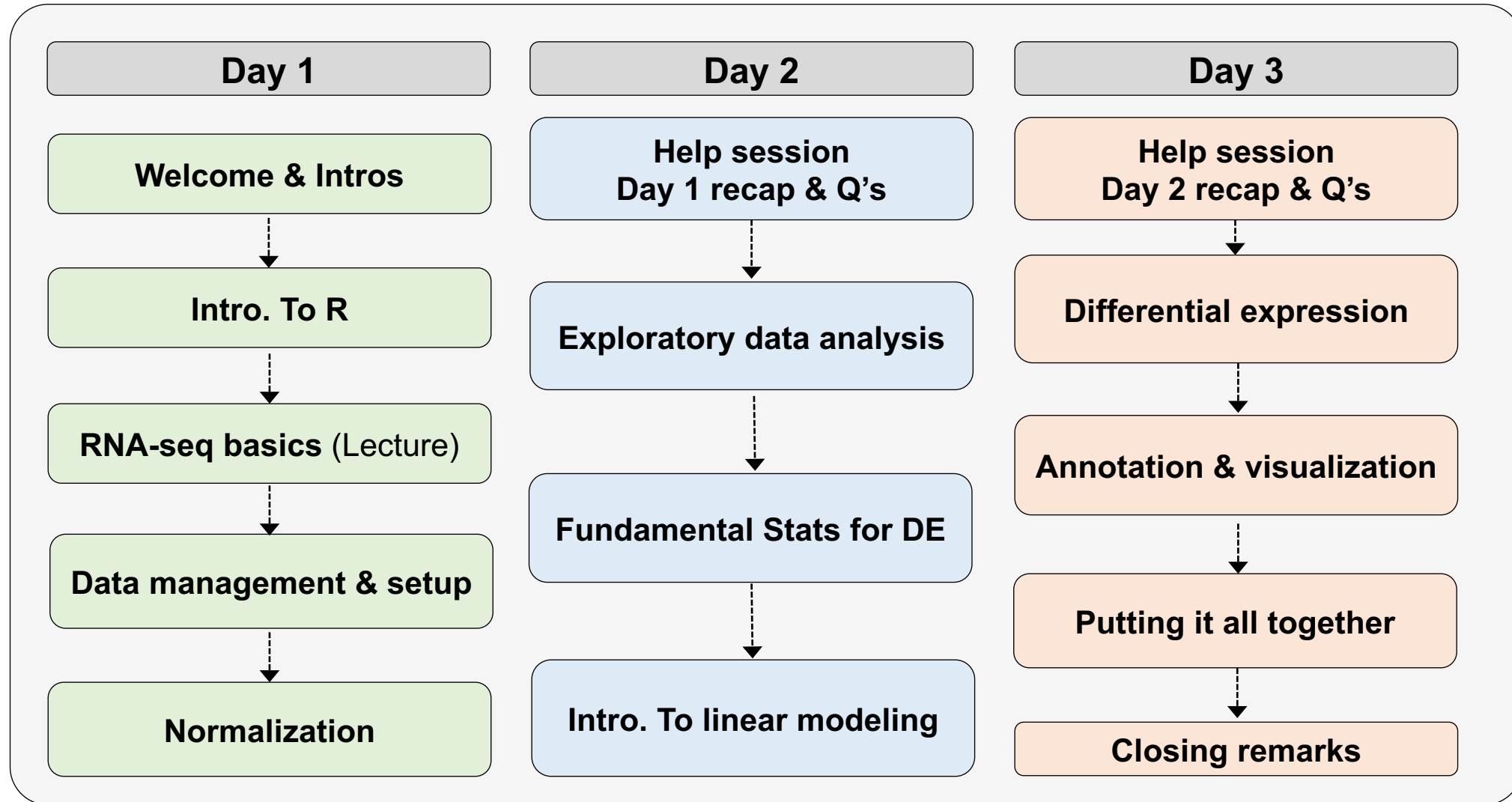➤ **Training**
  - Group workshops

# 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

**Gene expression matrix**
(read counts)

Samples

Genes

+

**Sample metadata**

Normalization

Exploratory analysis

Statistical modeling

Visualization

**Differential expression results**

| ID | lfc | lfcSE | stat | pvalue | padj | Gene |
|----|-----|-------|------|--------|------|------|
| ENSG032219 | 2.89 | 0.12 | -25.4 | 1.6E-251 | 1.7E-249 | ARID4A |
| ENSG012951 | -3.44 | 0.15 | -24.5 | 1.2E-132 | 1.4E-122 | KLK10 |
| ENSG016754 | -3.14 | 0.12 | -24.1 | 2.9E-123 | 1.4E-118 | KLK5 |
| ENSG080097 | -2.78 | 0.11 | -20.7 | 5.7E-115 | 2.4E-109 | LGALS1 |
| ENSG006277 | 5.92 | 0.28 | 19.8 | 1.9E-80 | 3.5E-92 | UCHL1 |
| ….. | ….. | ….. | ….. | ….. | ….. | ….. |

# Workshop outline

| Day 1 | Day 2 | Day 3 |
|---|---|---|
| Welcome & Intros | Help session Day 1 recap & Q's | Help session Day 2 recap & Q's |
| Intro. To R | Exploratory data analysis | Differential expression |
| RNA-seq basics (Lecture) | Fundamental Stats for DE | Annotation & visualization |
| Data management & setup | Intro. To linear modeling | Putting it all together |
| Normalization | | Closing remarks |

# 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
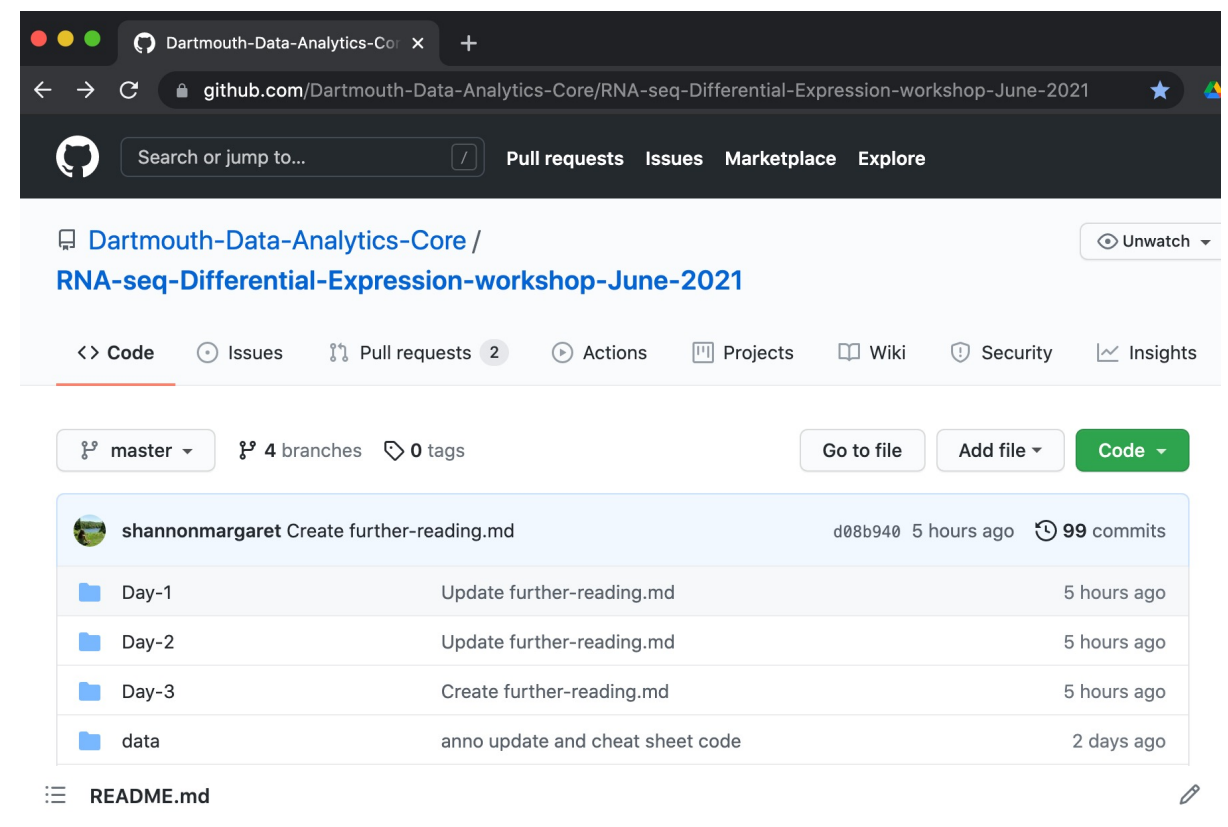
# Logistics I

➤ 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

# 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!*
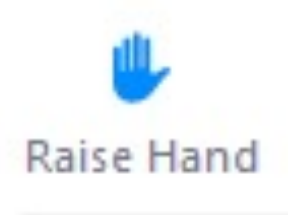
# Logistics III

➢ Use buttons in _Participants_ tab in zoom
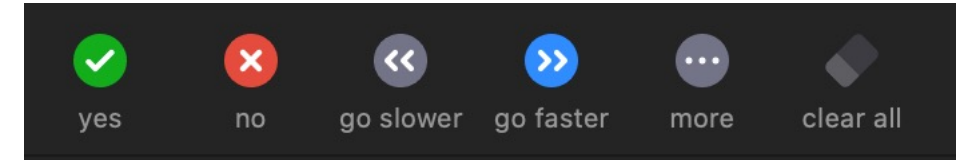
➢ You'll be muted, but if you want to ask a question, just raise your hand

➢ 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
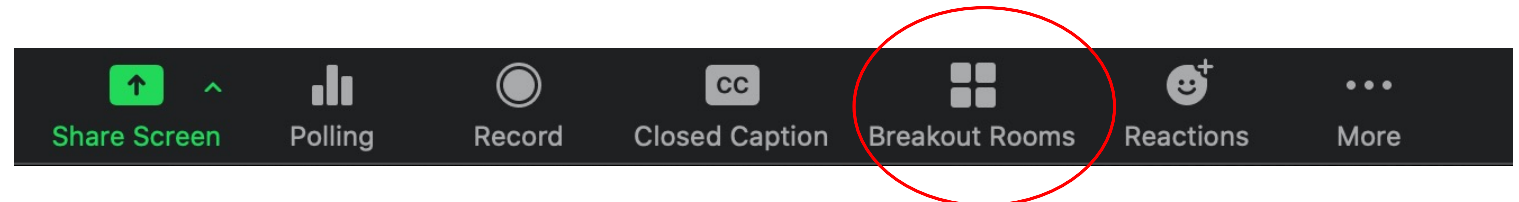
➢ Please be courteous on zoom..

# How to get help?

➤ **Raise your hand in zoom** (bottom right, participants tab)
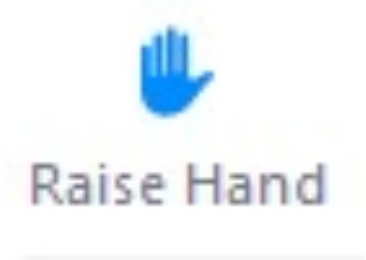
➤ **Use the slack channel to message one of us**

   - 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

   - Shannon Soucy (Shannon.Margaret.Soucy@Dartmouth.edu)

   - Tim Sullivan (Timothy.J.Sullivan@dartmouth.edu)

   - Owen Wilkins (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?