

# Post-pipeline RNA-seq

Daniel Fulop, PhD  
Gladstone Institutes

# Scope / Objectives

- Differential expression analysis using linear models in edgeR

Raw data



Demultiplex + QC



Map reads



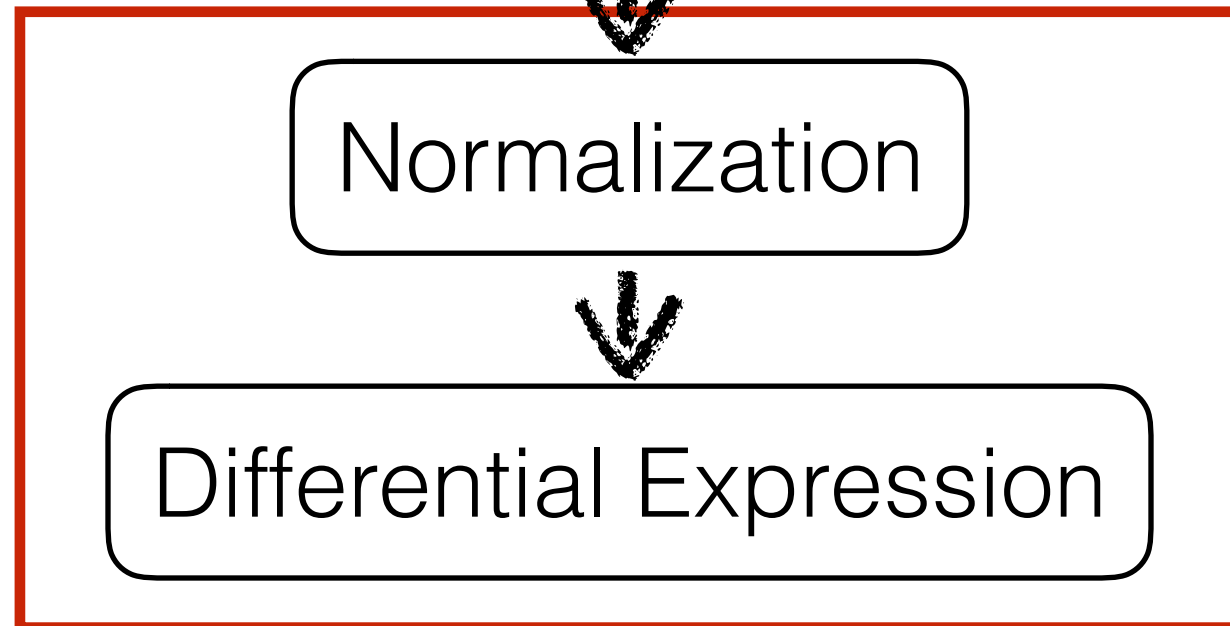
Count reads per feature



Normalization



Differential Expression



# EdgeR **DGEList** object

- R objects: data frame, vector, list
- DGEList is a custom object used by edgeR
  - raw & norm. counts, linear model, diff. exp.
- It's a container to which that we'll be adding data and results

# Dataset & biological question

Table 1

	GEO	SRA	CellType	Status
<b>MCL1.DG</b>	GSM1480297	SRR1552450	B	virgin
<b>MCL1.DH</b>	GSM1480298	SRR1552451	B	virgin
<b>MCL1.DI</b>	GSM1480299	SRR1552452	B	pregnant
<b>MCL1.DJ</b>	GSM1480300	SRR1552453	B	pregnant
<b>MCL1.DK</b>	GSM1480301	SRR1552454	B	lactating
<b>MCL1.DL</b>	GSM1480302	SRR1552455	B	lactating
<b>MCL1.LA</b>	GSM1480291	SRR1552444	L	virgin
<b>MCL1.LB</b>	GSM1480292	SRR1552445	L	virgin
<b>MCL1.LC</b>	GSM1480293	SRR1552446	L	pregnant
<b>MCL1.LD</b>	GSM1480294	SRR1552447	L	pregnant
<b>MCL1.LE</b>	GSM1480295	SRR1552448	L	lactating
<b>MCL1.LF</b>	GSM1480296	SRR1552449	L	lactating

Question: what are the genes involved in milk production

# Programming Pipes

- command-1's output becomes command-2's input
  - allows you to string together several operations
- %>% regular R pipe
- %<>% assignment pipe
- For more info: [magrittr.tidyverse.org](https://magrittr.tidyverse.org)

# What are we testing?

- That there's **no change in mean expression** between the conditions in question
- The simple 2-condition case is for a single variable, e.g. basal vs luminal cells (just Cell Type)
- These conditions can result from the combination of multiple experimental variables, e.g. Cell Type & Status