

Short Read Workshop Day 7

Counting Reads and Differential Expression

Rutendo Sigauke
2023

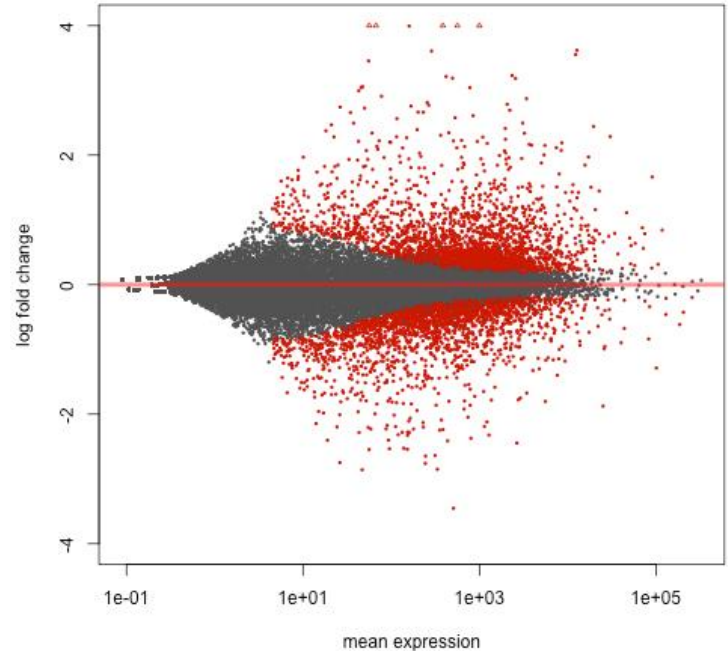
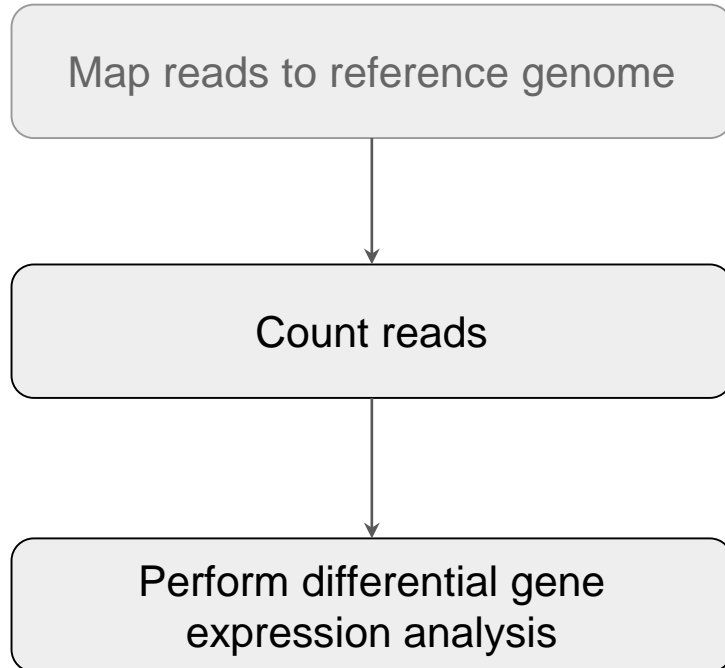
Day 7 Overview

- M&Ms
- featureCounts
- DESeq2



Goal of the Day

Find genes that are different between samples



featureCounts counts reads over features in R

There are several options in featureCounts

```
fc <- featureCounts(files=bam_file_list,
  annot.ext=gtf,
  isGTFAnnotationFile=TRUE,
  GTF.featureType="exon",
  GTF.attrType="gene_id",
  useMetaFeatures=TRUE,
  allowMultiOverlap=TRUE,
  largestOverlap=TRUE,
  countMultiMappingReads=TRUE,
  isPairedEnd=TRUE,
  strandSpecific=1,
  nthreads=N)
```

	union	intersection_strict	intersection_nonempty
	gene_A	gene_A	gene_A
	gene_A	no_feature	gene_A
	gene_A	no_feature	gene_A
	gene_A	gene_A	gene_A
	gene_A	gene_A	gene_A
	ambiguous (both genes with --nonunique all)	gene_A	gene_A
	ambiguous (both genes with --nonunique all)		
	alignment_not_unique (both genes with --nonunique all)		

aligned read:
start: 113217600 end: 113217650



GTF

```
chr1 unknown exon 113217048 113217252 . + . gene_id "MOV10";p_id "P5535";transcript_id "NM_001130079"
chr1 unknown exon 113217048 113217351 . + . gene_id "MOV10";p_id "P5535";transcript_id "NM_020963"
chr1 unknown exon 113217470 113217671 . + . gene_id "MOV10";p_id "P5535";transcript_id "NM_001130079"
chr1 unknown CDS 113217535 113217671 . + 0 gene_id "MOV10";p_id "P5535";transcript_id "NM_001130079"
chr1 unknown start_codon 113217535 113217537 . + gene_id "MOV10";p_id "P5535";transcript_id "NM_001130079"
```

↑
feature type

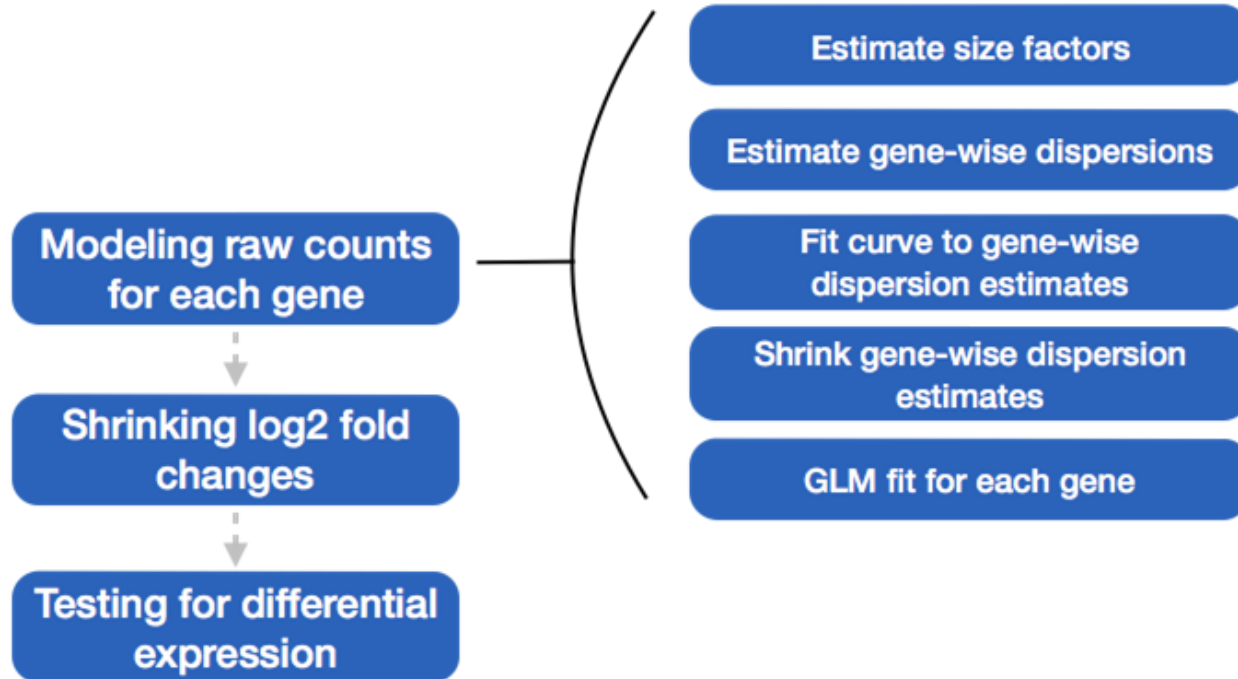
↑
feature

The are other tools for counting reads but featureCounts is more efficient

Method	Number of reads	Number of fragments	Time (min)	Memory (MB)
<i>featureCounts</i>	4 385 354	4 796 948	1.0	16
<i>SummarizeOverlaps</i> (whole genome at once)	4 385 354	3 942 439	12.1	3400
<i>SummarizeOverlaps</i> (by chromosome)	4 385 354	3 942 439	41.7	661
<i>htseq-count</i>	4 385 207	4 769 913	22.7	101

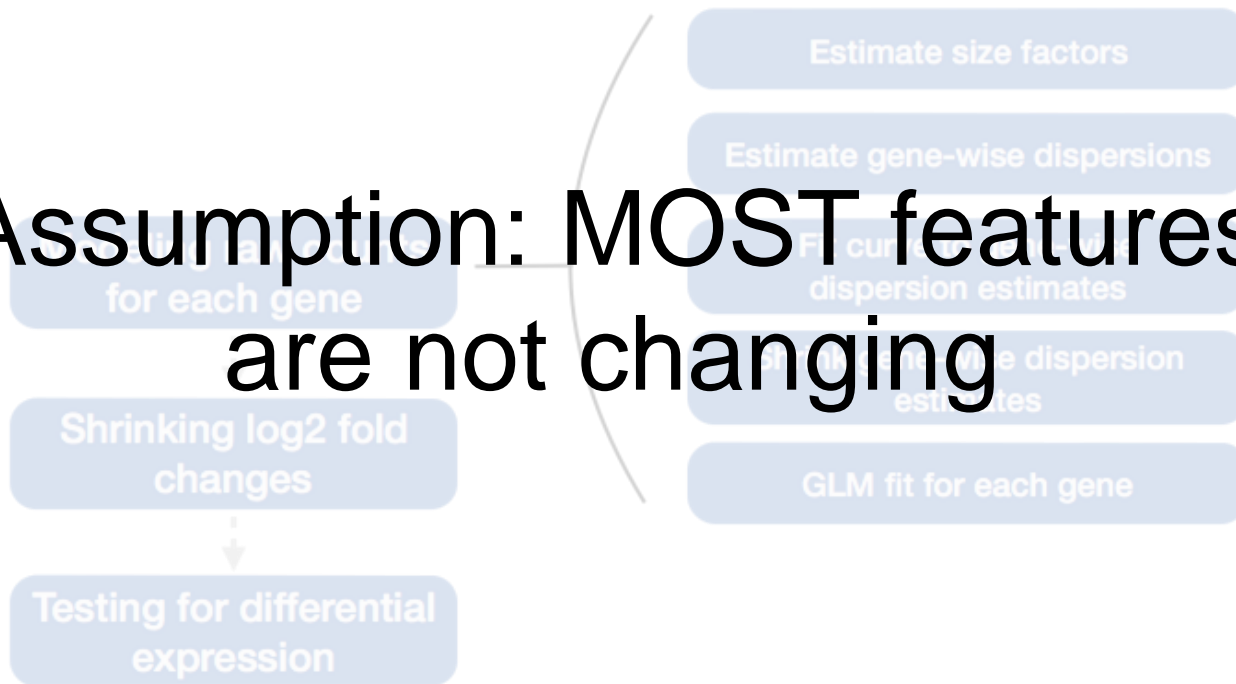
featureCounts is faster and more efficient.

DESeq2 Recap



DESeq2 Recap

Assumption: MOST features
are not changing



Counting reads with featureCounts

- Follow [featureCounts](#) worksheet:
 - Open R and install Rsubread
 - Get `d7_featureCounts.R` and `d7_featureCounts.sbatch` scripts
 - Edit both scripts and execute the sbatch script

Challenge Question

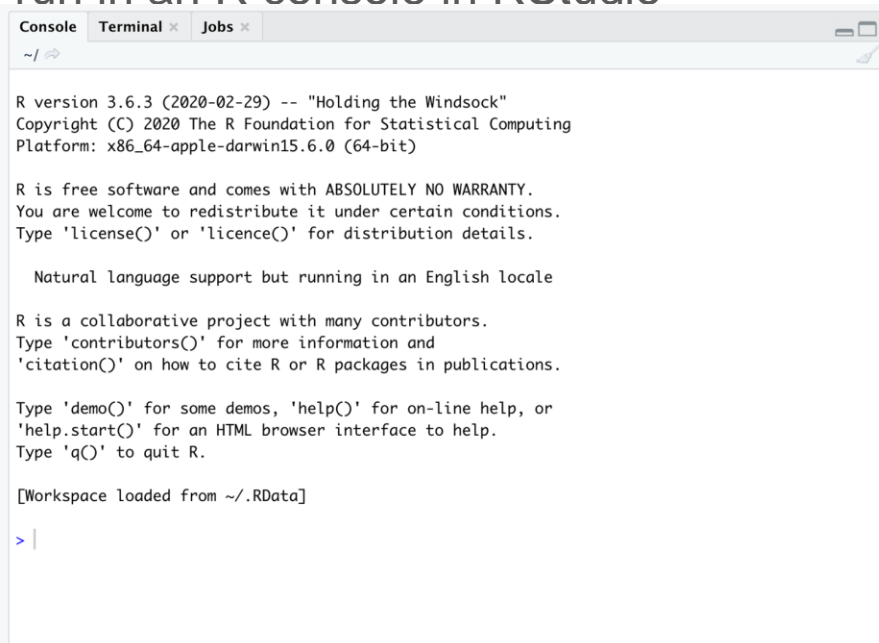
- What feature would you use to count reads for RNA-seq?
 - A. Gene
 - B. Exon
 - C. Transcripts

Challenge Question

- What feature would you use to count reads for RNA-seq?
 - A. Gene
 - B. Exon ✓
 - C. Transcript

Run DESeq2...

- Follow [DESeq2](#) worksheet
 - This will be run in an R console in RStudio



```
Console Terminal x Jobs x
~/
R version 3.6.3 (2020-02-29) -- "Holding the Windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

> |
```

Challenge Question

- How would you run DESeq2 on the supercomputer?

Challenge Question

- How would you run DESeq2 on the supercomputer?
 - Install DESeq2 in your R packages directory
 - Make a conditions table that matches your count table
 - Run the R script through an sbatch script

Homework

- Explore DE genes with a heatmap
- Run DESeq2 to explore differential expression with nicotine and caffeine treatments