Stringtie resulted in 2 count files:

* **gene\_count\_matrix.csv (76,145 rows)**
* **transcript\_count\_matrix.csv (119,518)**

and in gtf files with gene annotation for Ballgown

Ballgown was used to extract FPKM data from Stringtie data and resulted in **FPKM\_full-table.csv** which has normalized reads (FPKM) for all biological replications of all transcripts (**119,600**)

The file **gene\_count\_matrix.csv** was used downstream for

* PCA analysis (Deseq2)
* MA plots

I am currently working on a file **FPKM\_full-table\_Fold\_change.csv w**hich will have log2FC values an p values for all transcripts; I will upload it here on our shared drive, soon.

Note: **FPKM\_full\_tables\_all\_rows** has unfortunately only the first 65,534 transcripts (out of 119,600); thus I will need to redo heatmaps and tables of up and downregulated transcripts

Note: this happened again when saving the file in excel as .xls; **only save as .csv !**

Adjusted/calculated FPKM table for heatmap and tables:

* + · Started with **FPKM \_full\_table.csv**
  + · Saved as **FPKM\_full-table\_remove\_some\_columns.csv**
    - · Removed some columns that do not include gene names, IDs or FPKM
  + · Saved as **FPKM\_full-table\_added\_value\_1FPKM.csv**

* + - · Added 1 FPKM to all to avoid dividing by 0
  + · Saved as **FPKM\_average+1\_fold\_change\_pvalue.csv**
    - Calculated fold changes and p-values for t20/control and t40/control

Sorted and filtered

UP-REGULATED

file name: **FPKM\_t20\_up\_1.5.xls**

**328 transcripts (rows)**

sorted for t20 (high to low)

filtered by FPKMt20 average >=3

log2FCt20 >=1.5

pvalue < 0.05

file name: **FPKM\_t40\_up\_1.5.xls**

**535 transcripts (rows)**

sorted for t40 (high to low)

filtered by FPKMt40 average >=3

log2FCt40 >=1.5

pvalue <0.05

file name: **FPKM\_t20\_t40\_up.1.5.xls**

**61 transcripts (rows)**

sorted for t20, then t40 (high to low)

filtered by FPKMt40 average >=3

log2FCt40 >=1

pvalue < 0.05

DOWN-REGULATED

file name: **FPKM\_t20\_down\_1.5.xls**

**124 transcripts (rows)**

sorted for t20 (low to high)

filtered by FPKMcontrol average >=3

log2FCt20 <= -1.5

pvalue < 0.05

file name: **FPKM\_t40\_down\_1.5.xls**

**494 transcripts (rows)**

sorted for t40 (low to high)

filtered by FPKMcontrol average >=3

log2FCt40 <= -1.5

pvalue <0.05

file name: **FPKM\_t20\_t40\_up.1.5.xls**

**61 transcripts (rows)**

sorted for t20, then t40 (low to high)

filtered by FPKMcontrol average >=3

log2FCt40 <=-1.5

pvalue < 0.05