Analysis of RNAseq counts

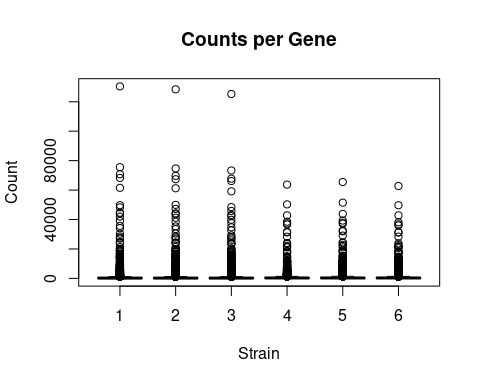
Murray Cadzow

20/04/2021

## Geneid Chr Start End Strand Length ...STAR.SRR014335.Aligned.out.sam  
## 1 YDL248W IV 1802 2953 + 1152 52  
## 2 YDL247W-A IV 3762 3836 + 75 0  
## 3 YDL247W IV 5985 7814 + 1830 2  
## 4 YDL246C IV 8683 9756 - 1074 0  
## 5 YDL245C IV 11657 13360 - 1704 0  
## 6 YDL244W IV 16204 17226 + 1023 6  
## ...STAR.SRR014336.Aligned.out.sam ...STAR.SRR014337.Aligned.out.sam  
## 1 46 36  
## 2 0 0  
## 3 4 2  
## 4 0 1  
## 5 3 0  
## 6 6 5  
## ...STAR.SRR014339.Aligned.out.sam ...STAR.SRR014340.Aligned.out.sam  
## 1 65 70  
## 2 0 1  
## 3 6 8  
## 4 1 2  
## 5 5 7  
## 6 20 30  
## ...STAR.SRR014341.Aligned.out.sam  
## 1 78  
## 2 0  
## 3 5  
## 4 0  
## 5 4  
## 6 19

This analysis will start to look at differential expression between the wild-type and mutant yeast strains.

There are 7120 genes (rows) that will be compared in order to establish if there has been differential expression. First, we shall take a look at the counts per gene.



The data is highly skewed which suggests a log transform would be useful. This will be done by adding 0.5 to the counts to prevent a log2(0) issue.

