**Day 1 practicals with example code**

• Hardware (64-bit computer with 4 GB of RAM (8 GB preferred))

• FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/download.html#fastqc>)

• R (<https://www.r-project.org>, version 4.0.3 or later)

• Optional: QualiMap (<http://qualimap.bioinfo.cipf.es/>)

**Install packages**

install.packages("BiocManager")

BiocManager::install("NOISeq")

BiocManager::install("Repitools")

BiocManager::install("Rsamtools")

BiocManager::install("Rsubread")

BiocManager::install("rtracklayer")

**Experiment Design and Power Analysis**

Open [Scotty](http://scotty.genetics.utah.edu/) in your browser

Use *Yeast S. cerevisiae - Busby* as model dataset. For the cost section assume 1000$ per sample for library preparation. Sequencing costs are found for example [here](http://med.stanford.edu/gssc/rates.html).

How many replicates would you need to be able to detect at least 70% of differentially expressed genes with a 1.5 Fold Change between conditions at a confidence level of 0.01?

**Pre-alignment QC**

Run FASTQC on one of the fastq files you downloaded

pick file ERR188044\_chrX\_1.fastq.gz

How many total sequences are there? 1321477

What is the range (x - y) of read lengths observed? 76

What is the most common average sequence quality score? 38

**Run alignment**

Run alignment within R with *align* function from Rsubread.

Hint: need to index the reference file with *buildindex* function

library(Rsubread)

buildindex(basename="chrX\_data/genome/chrX\_index", reference="chrX\_data/genome/chrX.fa")

ERR188044\_align.stats <- align(index="chrX\_data/genome/chrX\_index",

readfile1="chrX\_data/samples/ERR188044\_chrX\_1.fastq.gz",

readfile2="chrX\_data/samples/ERR188044\_chrX\_2.fastq.gz",

output\_file="ERR188044\_alignResultsPE.BAM",phredOffset=64)

Count the reads with featureCount from Rsubread

ERR188044\_counts <- featureCounts("ERR188044\_alignResultsPE.BAM",

annot.ext = "chrX\_data/genes/chrX.gtf",

isGTFAnnotationFile = T,

isPairedEnd = T,

requireBothEndsMapped = T)

**Optional: post-alignment QC**

Index and sort the bam file

Use the asBam function from Rsamtools.

library(Rsamtools)

asBam("ERR188044\_alignResultsPE.BAM", "ERR188044\_alignResultsPE\_sorted")

Run QC with QualiMap