Project 6 Baking bad

1. Download all data

```
# SRR941816: fermentation 0 minutes replicate 1
wget ftp.sra.ebi.ac.uk/vol1/fastq/SRR941/SRR941816/SRR941816.fastq.gz (413 Mb)
# SRR941817: fermentation 0 minutes replicate 2
wget ftp.sra.ebi.ac.uk/vol1/fastq/SRR941/SRR941817/SRR941817.fastq.gz (455 Mb)
# SRR941818: fermentation 30 minutes replicate 1
wget ftp.sra.ebi.ac.uk/vol1/fastq/SRR941/SRR941818/SRR941818.fastq.gz (79.3 Mb)
# SRR941819: fermentation 30 minutes replicate 2
wget ftp.sra.ebi.ac.uk/vol1/fastq/SRR941/SRR941819/SRR941819.fastq.gz (282 Mb)
```

```
# Reference genome + annotation
wget ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/146/045/GCF_000146045.2_R64/GCF_000146045.2_R64_genomic.gff.gz
wget ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/146/045/GCF_000146045.2_R64/GCF_000146045.2_R64_genomic.gff.gz
```

2. Analysis Pipeline

```
conda install hisat2
hisat2-build GCF_000146045.2_R64_genomic.fna GCF_000146045.2_R64_genomic.gff
```

```
hisat2 -p 2 -x '/hisat_index/GCF_000146045.2_R64_genomic.gff' -U 'after/SRR941818.fastq' | samtools sort > out_after_ferment1.bam
```

```
hisat2 -p 2 -x 'hisat_index/GCF_000146045.2_R64_genomic.gff' -U 'after/SRR941819.fastq' | samtools sort > out_after_ferment19.bam
```

```
hisat2 -p 2 -x '/hisat_index/GCF_000146045.2_R64_genomic.gff' -U '/before/SRR941816.fastq' | samtools sort > out_before_ferment16.bam
```

```
featureCounts -g gene_id -a '/home/anya/Desktop/IB/workshop/project_6/GCF_000146045.2_R64_genomic.gff' -o feature_counts_out.txt out_after_
```

```
featureCounts -g gene_id -a gff_to_gft_output.gft -o feature_counts_out.txt out_after_ferment18.bam out_after_ferment19.bam out_before_ferm
```

```
cat feature_counts_out.txt | cut -f 1,7-10>simple_counts.txt
```

Find differentially expressed genes with Deseq2

```
cat simple_counts.txt | R -f deseq2.r
```

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/5805e853-6ea6-4673-ac83-de3bfcc738e0/result.txt

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/101c7477-0243-402c-9103-383164022b0b/norm-matrix-deseq2.txt

Heatmap code:

```
library(ComplexHeatmap)
library(circlize)
library(readxl)


ndata <- read_table("norm-matrix-deseq2.txt")

data <- read.table(file = 'norm-matrix-deseq2.tcv', sep = '\t', header = TRUE)

gene = data[,1]
vals = as.matrix(data[,2:ncol(data)])
vals = jitter(vals, factor = 1, amount=0.00001)

column_names = c('Aerobic 1', 'Aerobic 2',
                 'Fermentation 1', 'Fermentation 2')

score = NULL

for (i in 1:nrow(vals)) {
  row=vals[i,]
  zscore=(row-mean(row))/sd(row)
  score =rbind(score,zscore)
}
row.names(score) = gene
zscore=score
mat = as.matrix(zscore)


col_fun = colorRamp2(c(-1.5, 0, 1.5), c("#1CD500", "#F7F7F7", "#E21E00"))

Heatmap(mat, col = col_fun, show_row_names = FALSE,
        row_title = 'Genes', row_title_side = 'left',
        column_labels = column_names, # column_title = 'Samples'
        column_title_side = 'top', column_dend_side = 'bottom',
        column_names_side = "top", column_names_rot = 45,
        name = 'z-score')
```

Volcano plot code:

```
library(EnhancedVolcano)

deseq <- read.table(file='result.txt')
deseq_50 <- deseq[c(1:50),1]

EnhancedVolcano(deseq,
                lab = rownames(deseq),
                selectLab = c(deseq_50),
                x = 'log2FoldChange',
                y = 'padj',
                col=c('grey', '#CCA9E6', 'blue', '#EB6060'),
                hline = c(5e-30),
                hlineCol = c('red'),
                hlineType = c('longdash'),
                hlineWidth = c(1.0),
                colAlpha = 0.5)
```

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/3e17e447-8689-4685-8364-54982fb15897/mapper_genes_1456757_slimTerms.html