

RNAseq and comparisons

Reanalyze data in this published paper [Baker et al 2014](#) “Slow growth of Mycobacterium tuberculosis at acidic pH is regulated by phoPR and host-associated carbon sources”

Data are downloaded to `/bigdata/gen220/shared/data/M_tuberculosis`

The Transcriptome file is also in the folder as `GCF_000008585.1_ASM858v1_cds_from_genomic.fna`

There is a `sra_info.tab` file which lists the sample accessions and their metadata.

Compare gene expression between two sets of conditions. - pH5.7 - pH7

And also growth carbon source - Glycerol - Pyruvate

1. Run Kallisto to get the gene expression calculated from each sample
2. Run pfam analysis to get the Protein domains found in each protein
3. Construct a tab delimited file which lists on each line
 - The Gene (LOCUS) name
 - The Protein length
 - An average TPM across replicates for each condition (eg there will be 4)
 - The Pfam Protein domains, separated by comma found in each Protein
 - GO Terms assigned to each domain