



mRNA-Seq data processing workflow

The paired end short read sequence data (illumina) was processed using in house scripting (bash and R) on University of Edinburgh high performance computing facility (Eddie3). The statistical analysis was carried out in **R v ≥ 3.6** ¹ using **tximport v3.11**². Briefly the trimming was carried out using **Trimmomatic v 0.39**³ using the following criteria: `java -jar trimmomatic/0.39/trimmomatic-0.36.jar PE -threads ${vCPU} -basein ${sra_id}_1.fastq.gz baseout ${sra_id}_trimmed.fq.gz ILLUMINACLIP:/trimmomatic/0.39/adapters/TruSeq3-PE-2.fa:2:30:10:1:TRUE SLIDINGWINDOW:5:20 MINLEN:50`.

The trimmed reads were then mapped to the Ensembl's pig genome (Sscrofa 11.1 v100 GCA_000003025.6) using **Kallisto v0.46.0**⁴. The indexes were build using the following gene model fasta files `Sus_scrofa.Sscrofa11.1.cdna.all.fa.gz`. The quantification was carried out using (default criteria): `kallisto quant --bias -l index -o output < (zcat ${sample}_1_trimmed.fq.gz) < (zcat ${sample}_2_trimmed.fq.gz)`

References:

1. R Core Team. R: A language and environment for statistical computing. *R Found. Stat. Comput.* (2017).
2. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences [version 2; referees: 2 approved]. *F1000Research* (2016). doi:10.12688/F1000RESEARCH.7563.2
3. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–20 (2014).
4. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).