Assay for Transposase-Accessible Chromatin (ATAC-Seq) data processing workflow

The paired end Illumina ATAC-Seq raw data was processed using in house scripting (bash and R) on University of Edinburgh high performance computing facility (Eddie3). The ATAC-Seq data statistical analysis was carried out in **R v ≥ 3.6** [1] using **ChIPseeker v1.26.0** [2] package for annotation and metrics. Briefly quality control of the raw sequence data was performed using **FastQC v0.11.9**[3] and **multiQC v1.9**[4] (Ewels et al., 2016) programmes. The paired end reads were trimmed using **Trimmomatic v0.39**[5] (Bolger et al., 2014) and the following trimming criteria*:*

trimmomatic-0.39.jar ILLUMINACLIP:/Trimmomatic-0.39/adapters/NexteraPE-PE.fa:2:30:10:1:true SLIDINGWINDOW:5:20 MINLEN:30.

The trimmed reads were then mapped to the Sscrofa11.1 pig reference genome available from Ensembl v100 (GCA_000003025.6) using **Bowtie2 v2.3.5.1**[6,7] and the default flags of the *--very-sensitive* mode. The BAM files that were generated were then sorted and indexed using **samtools v1.6**[8]. ATAC-Seq peak calling was performed using **Genrich v0.5**[9] (John M. Gaspar, 2020) under ATAC-Seq mode using the following flags:

Genrich -t input.bam -o ouput_peak.bed -f output_peak.log -z -j -r -e MT -v.

Intersectionality of the peaks called was explored using **bedtools' v2.26.0**[10] intersect and subtract functions:

bedtools intersect -sortout -u -a A.bed -b B.bed C.bed D.bed > ABCD.bed

bedtools subtract -A -a A.bed -b ABCD.bed > A_specific.bed

References:

1. R Core Team. R: A language and environment for statistical computing. *R Found. Stat. Comput.* (2017).
2. G Yu, LG Wang, QY He. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 2015, 31(14):2382-2383.
3. Andrews S.; S Bittencourt a, 2010. FastQC: a quality control tool for high throughput sequence data – ScienceOpen. Babraham Institute.

4. Ewels, P., Magnusson, M., ... Käller, M., 2016. MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048. doi:10.1093/bioinformatics/btw354
5. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–20 (2014).
6. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–9 (2012).
7. Langmead, B., Wilks, C., Antonescu, V. & Charles, R. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35,** 421–432 (2019).
8. Li, H., Handsaker, B., ... Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. doi:10.1093/bioinformatics/btp352
9. John M. Gaspar (2020). Genrich: detecting sites of genomic enrichment. Available at: https://github.com/jsh58/Genrich.
10. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).