**Pipeline overview**

The TaRGET II ATAC-seq pipeline is designed for standardized data processing of mouse ATAC-seq samples. It incorporates automatic quality control, generates user friendly files for computational analysis and outputs genome browser tracks for data visualization. To ensure consistent and reproducible data processing, the entire workflow, associated software and libraries are built into a singularity image, which can be run on computational clusters with job submission as well as on stand-alone machines. Pipeline installation requires minimal user input. All the software and genome references used for TaRGET II ATAC-seq data processing are included in the pipeline image. The pipeline supports both single- and paired-end data, it accepts FASTQ files, performs alignments, peak calling and signal track generation. The major outputs produced by the pipeline are:

1. JSON file with quality control measurement of user supplied dataset. Quality control measurement of ENCODE dataset is provided for references.
2. Log file with processing status of each step
3. Processed files after alignment (.bam) and peak calling (.narrowPeak)
4. Bigwig file for visualization purpose

**Usage of the pipeline**

An easy 2-step guideline for TaRGET II ATAC-seq pipeline usage:

1. Download the singularity image of the pipeline.
2. Run the code below from the same directory as your fastq.gz file:

singularity run -B ./:/process <path-to-image> -r <SE/PE> -g <mm10 > -o <read\_file1> -p <read\_file2>

**Software Used in the Pipeline**

cutadapt (v1.16) was used to find and remove adapter sequence, and other types of unwanted sequence from high-throughput sequencing reads.

fastqc (v0.11.7) was used to provide quality control checks on raw sequence data.

bwa (v0.7.16a) was used to map sequence reads against the mouse reference genome.

methylQA (v0.1.9) was used to generate the reads density and reads mapping statistics.

macs2 (v2.2.5) was used for peak calling process.

samtools (v1.9) and bedtools (v2.29) were used to process the sequence alignments and genome features.

**Genomic References Used in the Pipeline**

The following mouse genome and gene references were built into the singularity image and used for TaRGET ATAC-seq data processing:

1. BWA indexes of mouse genome (mm10)
2. Chromosome size file of mm10
3. Promoter region file of genes from GENCODE vM20 annotation file