**Introduction**

This document describes a bioinformatics pipeline for a RNA sequencing data obtained from any organism with a reference genome and annotation. The input to this pipeline are a data samplesheet and fastq files, followed by the deployment of a 6 sub-workflows, as shown in the pipeline metro schematic. These sub-workflows are for pre-processing/quality control (QC), alignment and quantification, visualization, and differential gene analysis. The final sub-workflow – summary/final QC results in the generation of an extensive report based and log file.

**Pipeline overview**

This pipeline is built to process input fastq data using the following steps. The final output of this pipeline can be read out in different formats, that are usable for various downstream platforms.

Sub-workflow 1 : Pre-processing

cat Merge sequenced FastQ file lanes (for paired end data) into Reads 1 and 2

Input: fastq

Output: fastq

fastp QC on raw FastQ data and Adapter and quality trimming

Input: fastq

Output: html, json, trimmed fastq and txt

FastQC QC on trimmed fastq reads

Input: trimmed fastq

Output: html

Sub-workflow 2: Genome alignment and quantification

Salmon Quantification

Input: trimmed fastq

Output: json, quant files, aux info

[STAR](https://nf-co.re/rnaseq/3.11.2/output" \l "star-and-salmon)   Alignment of reads to reference genome and output a bam

Input: trimmed fastq

Output: SAM

Sub-workflow 3: Post alignment processing

[SAMtools](https://nf-co.re/rnaseq/3.11.2/output" \l "samtools) Convert sam to bam – SAMtools view

Sort bam in genomic order - SAMtools sort

Create index file – index

Confirm stats - idxstats

Input: SAM, BAM

Output: BAM, txt, BAI, idxstat

Picard Mark duplicates

Input: sorted BAM

Output: BAM with duplicates marked/flagged (de-duplexed)

BEDtools Convert BAM to Bed files

Input: BAM

Output: BED

bedGraphtoBigWig Generate bigwig coverage files

Input: BED

Output: bigWig

Sub-workflow 4: Visualization

IGV Visualize sorted BAM files

Input: BAM, BAI

Output: Images (PDF, PNG)

UCSC Visualize bigWig format

Input: bigWig

Output: Images (PDF, PNG)

Sub-workflow 5: Disease signature

Tximport Import pseudocount data into DESeq2

Input: json, quant files, aux info

Output: count data (TPM)

DESeq2 To create signature matrix

Input: count data (TPM)

Output: signature matrix (csv)

CIBERSORTx Immune cell deconvolution

Input: signature matrix (csv)

Output: txt, heatmap

Sub-workflow 6: Final QC and exit report

MultiQC Show QC for raw reads, trimming, alignment and quantification

Input: FastQC reports and QC from other tools

Output: html and additional info files