Alignment Pipeline

01. Alignment of raw reads in FASTQ format. Custom shell script for single-end and paired-end reads. **Bowtie2**

02. Bash-script for converting aligned reads in .sam format to binary alignment file (.bam). **Samtools**

03. Script for sorting aligned reads by leftmost coordinates. **Samtools**

04. Bash-script for locating and marking duplicate reads in alignment data. Where duplicate reads are defined as originating from a single fragment of DNA. **Picardtools**

05. Bash-script for indexing coordinates sorted reads. Samtools

06. Bash-script using collectalignmentsummarymetrics from **Picard**. Comparing the alignment file to the reference file (FASTA) and provides statistics to assess the alignment, such as the number of input reads and the percentage of reads that are mapped.

07. Bash-script using the flagstat function of **Samtools** package. The function is performing a full pass on input file to calculate and print statistics to stdout.

08. R-script for data manipulation of collected metrics (requires **dplyr** and **tidyverse** packages)

09. R-script for creating plots on alignment metric data (using **ggplot2** package in R)

10. Bash- script for performing **BAIT** analysis on aligned and sorted reads. Outputs ideograms and .bed files that will be used for further downstream analysis.