

# DNA methylation analysis

**fastq quality control with fastqc v0.11.9**



**trim reads with trim-galore v0.6.5**

- trim 5 bp from the 5' end of read 2
- filter reads with phred < 33
- reads with Ns from either side of the read
- trim adapter sequences anywhere in the read
- remove reads shorter than 20bp after trimming



**Bismark v0.22.3 for alignment**

- align to genome GRCm38
- bowtie2-2.3.5



**filter for uniquely mapped reads only**

- is default in bismark v0.22.3 when run with bowtie2



**Deduplicate reads**

- using deduplicate\_bismark Bismark v0.22.3 with default parameters



Group 50 CpGs together in one probe and filter for probes > 10x coverage using seqmonk v1.45.4



treat the 3 replicates as replicate set and use the mean methylation level for following analysis



**Generate featurecounts lists in seqmonk overlapping**

- genes
- CpG islands
- promoters (2kb upstream of annotated transcript)
- intergenic regions, excluding repeats
- repeats excluding overlapping genes: RepeatMasker database from UCSC
  - filter for L1s > 5kb
  - filter for LAPEz > 6kb
  - filter for MMERV10C > 4.5kb



**Plot in R v3.3.1**