Lecture 6 Practice 1

- Generate a SAM file by mapping fastq(/2_disk/Bioinf2018/EcoTestRead2.fq) to E.coli genome.
  1. How many reads are in the data?
  2. How many reads are unmapped?
  3. How may different types of quality scores can you observe? (hint: cut, sort, uniq)?
  4. How many different CIGAR string do you see? What are the 5 most common CIGAR strings?
  5. How many reads align to the reverse strand?
  6. How many reads have a MAPQ (mapping quality) of 0 and what does that value mean in a SAM file from BWA?

Lecture 6 Practice 2

- Generate a sorted and indexed BAM file based on the fastq (/2_disk/Bioinf2018/EcoTestRead2.fq).
  1. Find the number of uniquely mapped reads
  2. Find the number of high quality alignments (MAPQ>30) for each strand separately
  3. A genomic feature has its start site on the forward strand on Genome I-2Kb.
     1. How many reads fall within this location?
     2. Print the position of each read (hint: there are not that many)
     3. Report the number of reads in this region for each strand separately.