You are provided with two files (A VCF and a SAM file). The files can be downloaded from here (https://drive.google.com/drive/folders/11UD52i99CaCSBEJFNb8Y1afo9p3hL8cL?usp=sharing). Use BCFtools, SAMtools and bash utilities to answer the questions that follow.

Make your submissions on your GitHub and send the link to the repo via email to ibra.lujumba@gmail.com. Your submissions should include commands/scripts used to obtain answers to the questions as well as answers to the questions.

Deadline: 26th January 2023

Manipulating VCF files

1. Describe the format of the file and the data stored
2. What does the header section of the file contain
3. How many samples are in the file
4. How many variants are in the file
5. How would you extract the chromosome, position, QualByDepth and RMSMappingQuality fields? Save the output to a tab-delimited file
6. Extract data that belongs to chromosomes 2,4 and MT
7. Print out variants that do not belong to chr20:1-30000000
8. Extract variants that belong to SRR13107019
9. Filter out variants with a QualByDepth above 7
10. How many contigs are referred to in the file. Check the header section
11. Comment on the eighth and ninth columns of the file
12. Extract data on the read depth of called variants for sample SRR13107018
13. Extract data on the allele frequency of alternate alleles. Combine this data with the chromosome and position of the alternate allele

Manipulating SAM files

1. Describe the format of the file and the data stored
2. What does the header section of the file contain
3. How many samples are in the file
4. How many alignments are in the file
5. Get summary statistics for the alignments in the file
6. Count the number of fields in the file
7. Print all lines in the file that have @SQ and sequence name tag beginning with NT_
8. Print all lines in the file that have @RG and LB tag beginning with Solexa
9. Extract primarily aligned sequences and save them in another file
10. Extract alignments that map to chromosomes 1 and 3. Save the output in BAM format
11. How would you obtain unmapped reads from the file
12. How many reads are aligned to chromosome 4
13. Comment of the second and sixth column of the file
14. Extract all optional fields of the file and save them in "*optional_fields.txt*"