Stage 2

Project 3: Run a simple NGS analysis pipeline

In this section, you will implement a simple NGS analysis on a simple dataset

Starting Datasets:

- Forward Strand
- Reverse Strand
- Reference Genome

Proposed Pipeline:

Download dataset (**wget**) => Quality Control (**FastQC**) => Trimming (**FastP**) => Genome Mapping (**bwa**) => Variant Calling (**bcftools/freebayes**)

Feel free to add software as you prefer.

Submission:

- We look forward to receiving your final pipeline **script.sh** (you can use bash, snakemake, nextflow or any pipeline tool you know how to use).
- Alongside, create a **setup.sh** file that anyone can use to install all the tools needed for making the pipeline work.
- Make a **requirement.txt** file that simply lists all the tools you used
- Upload the 3 files to your team's github repo. Each team member should have a folder and their folder should contain their 3 scripts.
- Copy the link to the team's repo and paste it on HackBio Submission platform
- Finally, be ready to discuss your pipeline with everyone

Resources

- Introduction to Whole Genome Sequencing and Variant Calling
- Raw Sequence to Variant Calling Pipeline with FreeBayes (Hands-On)
- Galaxy Tutorial for Variant Calling (with Code)
- Using For loops in BASh