Module Planning template
NGS Bioinformatics - Remote classrooms

Module Title: NGS Data pre-processing and QC

➜ **Module Lead/Co-leads/Assistants** - Fatma and Shaun

➜ **List of learning outcomes specific for this module** <mark>(Please list 3 - 5 Learning Outcomes here; refer to How to writing learning outcomes</mark>)

➜ FastQ file format specification:
   ◆ Understand the FastQ file format specification and structure
   ◆ Describe the different quality scoring encodings e.g Illumina, Solexa, Phred score
   ◆ Differentiate between the long read and short read sequencing outputs

➜ Quality control (use two different datasets, 1 good and 1 bad):
   ◆ QC and interpretation of results
   ◆ Introduction to Quality Control and FastQC
   ◆ Execute FastQC using the workflow
   ◆ Understand the section in the configuration file for running FastQC (under development)
   ◆ Interpret different QC metrics, graphs and reports (good dataset and bad dataset)
   ◆ Provide general guidelines and best practices for improving data quality

➜ Trimming and filtering
   ◆ Differentiate between trimming and filtering
   ◆ Apply trimming and filtering using the workflow
   ◆ Discuss the impact of trimming and filtering
   ◆ Provide general guidelines and best practices for trimming and filtering

➜ Brief overview of NGS file formats generated during the process of variant calling (e.g FastQ, BAM / CRAM, bed, vcf mapped to the workflow)
   ◆ Describe different file formats (including encoding basis)
   ◆ Describe where specific file formats are generated and used in the workflow

➜ **Total number of hours/days to be spent on this content:** 6 hours

➜ **Tools/software and resources to be used**
   ◆ Bash terminal/FastQC/MultiQC/Samtools

- ◆ Description of datasets to be used : Small sample dataset to show good and bad data
- ◆ Other resources, or readings

→ **Overview of activities and exercises (**list specific practical activities which participants will do)
Introduction lecture:
Exercise/Activity 1:Run fastqc on some sample fastq files and examine and assess outputs.
Exercise 2: Read and extract specific information from a vcf file?
Exercise 3:

→ **Assessment**
Final Assessment

Appendix *(include any guideline documents related to content development such as breaking down content for remote classroom format, how to write LO's, designing assignments etc…)*
See links to other guidelines
-Overall trainer guideline and template documents will include information on lecture formats and recordings, contact sessions schedules, checklists, assignments and assessments, quizzes, feedback, audio-video conferencing, learning management platform, teaching assistants