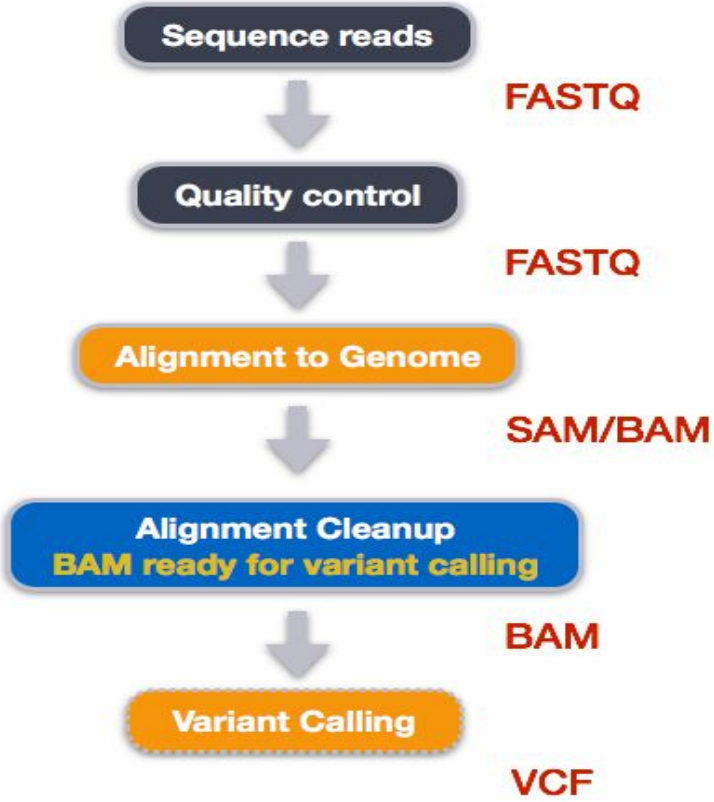
Variant calling

BY

Stephen Kanyerezi

AFRICAN
CENTERS
OF EXCELLENCE
IN BIOINFORMATICS

Variant calling workflow



AFRICAN CENTERS OF EXCELLENCE IN BIOINFORMATICS

Tools

All the steps involved are performed using specific tools

- How can I get these tools
- If they are different tools for a particular process, how do I choose the best tool
- Will I encounter challenges/problems/errors while using these tools. If so how do I go about it

Source of Data

- Sequencing lab
- Public Repository

How do I download the data from repositories

- Manual (Point and click)
- Programmatic

- What do I do after getting the data
- What is the format of my data
- Is the format standard
- Why do I need to understand the format

Fastq format

It contains sequenced fragments, each represented by 4 lines.

Below is the description of each line

1 => Always begins with ‘@’ and then information about the read

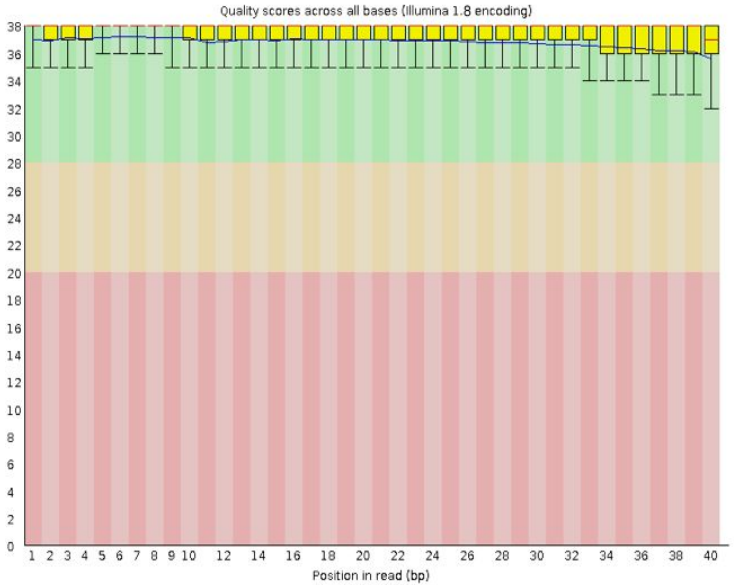
2 => The actual DNA sequence

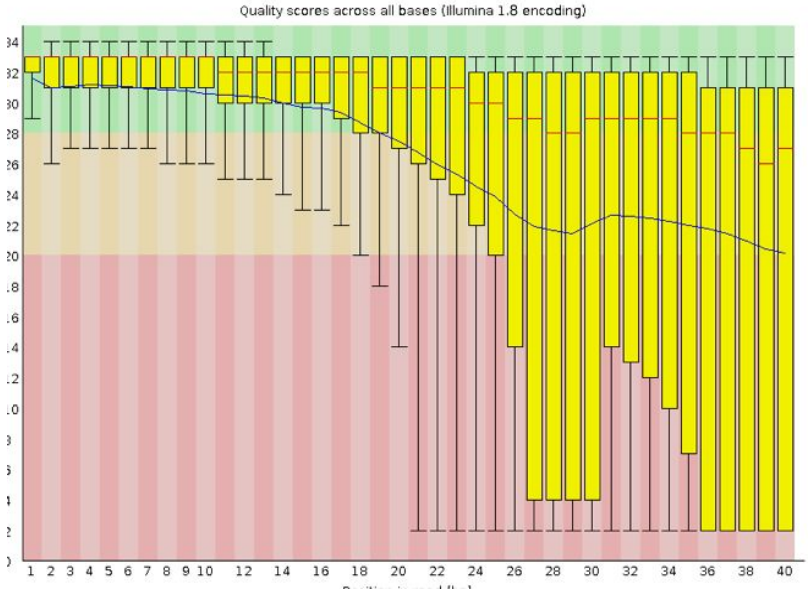
3 =>Always begins with a ‘+’

4 Has a string of characters which represent the quality scores; must have same number of characters as line 2

Quality Assessment

- Why is it necessary and how do I so it
- How do I deal with poor quality data
- Does this have an effect on my final results





AFRICAN CENTERS OF EXCELLENCE IN BIOINFORMATICS

Trimming and filtering

Various tools can do this

- Trimmomatic
- Trim galore, etc

Variant calling

We start with alignment

- Indexing the reference genome
- Aligning the reads to the reference genome

- ❖ Where do I get the reference genome
- ❖ After Alignment, what is the output

How do I deal with the alignment output

- Convert to bam (why)
- Sort
- index

Call variants

- Bcftools
- snpEff

Explore the output of calling variants

- Vcftools
- Bcftools
- Awk and other customized commands

I have multiple samples, should I run each sample independently

- Scripting

How do I improve the efficiency of my scripts

- Workflow managers
- Containerise
- Take note of the versions of programmes
- Good practice to use version control

AFRICAN CENTERS OF EXCELLENCE IN BIOINFORMATICS