Step-by-Step Guide: Variant Calling

Author: Itunuoluwa Isewon PhD

Email: itunu.isewon@covenantuniversity.edu.ng

You have been provided with a text file which contains six sample ID, upload by clicking on the upload button and drop the file.

Quality control

Obtain Fastq files:

1. in the tool Bar, click on Get Data
2. choose “Download and Extract Reads in FASTA/Q format from NCBI SRA”
3. Change the select input type to “List of SRA accession, then chose your sample id file and run tool
4. In this tutorial we’ll use six datasets.

sample	condition
SRR15044361	test
SRR15044360	test
SRR15044359	test
SRR15044358	control
SRR15044357	control
SRR15044356	control

Perform QC:

1. on the search bar, type fastqc
2. choose the desired fastq files (paired end) in the raw read tab.
3. leave all other tabs unchanged
4. Once it runs, two files are generated, a raw data file and a Webpage file
5. View the result by clicking on the webpage file produced
6. repeat for the second data and compare their results.

Multiqc

Why: it helps us to obtain a more intuitive comparison

1. on the search bar, type multiqc
2. On the “Which tool was used generate logs?” tab, choose Fastqc
3. Then click on “Insert FastQC output”
4. Type of output is raw data
5. Add the raw data files generated earlier
6. Leave all other parameters at default
7. Run tool
8. View the result by clicking on the webpage file produced

Variant calling

Mapping

1. search for **Map with BWA-MEM** in the tool search bar, choose the options for longer reads
2. We would be using a built-in genome
3. Choose **Aspergillus flavus NRRL3357** as the reference genome
4. Leave other parameters as default

Descriptive statistics

1. search for **Samtools flagstat** in the tool search bar, choose the options for longer reads
2. select the file generated from the BWA-MEM and leave the output format as txt
3. run tool
4. view results

Generate genotype likelihoods

1. search for **bcftools mpileup** in the tool search bar
2. we are using single Bam alignment input
3. select the file generated from the BWA-MEM
4. Reference genome is **Aspergillus flavus NRRL3357**
5. Output format is uncompressed VCF
6. run tool

Variant calling

1. search for **bcftools call** in the tool search bar, choose the options for longer reads
2. select the file generated from the **bcftools mpileup**
3. leave all other parameters default
4. Output format is uncompressed VCF
5. run tool
6. View result

Remove homologous variants and variants with missing phenotype

1. search for **Filter data on any column using simple expressions** in the tool search bar
2. select the file generated from the **bcftools call**
3. supply the condition c10 != '0/0' : sample genotype information are on the tenth column, != means not equal to, '0/0' represents homologous variants (portions of the genome not different from the refence)
4. run tool
5. View result
6. search for **Filter data on any column using simple expressions** in the tool search bar
7. select the file generated from the last step
8. supply the condition c10 != './.' : './.' denotes missing data
9. run tool
10. View result

Sorting

1. Find sort in the search bar, choose “Sort data in ascending or descending order”
2. Sort on column 6: Quality
3. Keep every other parameter as default.
4. Variants with high quality are now on top.