

QC the Alignment

After performing alignment, we get the BAM format, which further undergoes downstream analysis and we then, perform quality control using JAVA based tool, FastQC

FastQC: It provides a simple way to do some quality control checks on a raw sequence data coming from high throughput sequencing pipelines. It provides set of analysis, which can use to give a quick impression of whether your data has any problems of which should be aware before doing any further analysis.

FastQC has following main functions:

- ☐ Import of data from BAM/SAM/FastQC file format.
- ☐ Provide a quick overview to tell you in which areas there may be problems.
- ☐ Summary graphs and tables to quickly access your data.
- ☐ Export of results to an HTML based permanent report.
- ☐ Offline operation to allow automated generation of reports without running the interactive e application.

➔ Please follow the link below to see the FastQC reports and do not forget to rate the workflow,

[Link: <https://usegalaxy.org/u/ajay.ducs/h/genomic-data-science-capstone>]

BASE STATISTICS FOR OUR SIX SAMPLE ARE:

Measure	Value	Measure	Value
Filename	Bowtie2 on data 5_ alignments	Filename	Bowtie2 on data 6_ alignments
File type	Conventional base calls	File type	Conventional base calls
Encoding	Sanger / Illumina 1.9	Encoding	Sanger / Illumina 1.9
Total Sequences	91185969	Total Sequences	46971267
Sequences flagged as poor quality	0	Sequences flagged as poor quality	0
Sequence length	100	Sequence length	100
%GC	48	%GC	46

Measure	Value
Filename	Bowtie2 on data 7_ alignments
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	89061863
Sequences flagged as poor quality	0
Sequence length	100
%GC	52

Measure	Value
Filename	Bowtie2 on data 8_ alignments
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	162403466
Sequences flagged as poor quality	0
Sequence length	100
%GC	46

Measure	Value
Filename	Bowtie2 on data 9_ alignments
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	121992326
Sequences flagged as poor quality	0
Sequence length	100
%GC	48

Measure	Value
Filename	Bowtie2 on data 10_ alignments
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	136312421
Sequences flagged as poor quality	0
Sequence length	100
%GC	46

NOTE: DUE TO SPACE LIMIT, I HAVE SHOWN HERE ONLY BASIC STATISTICS. GO TO WORKFLOW LINK FOR FURTHER QUALITY CHECKS.