Next Generation Sequencing Data QC Report

| Sample Information

ID **Testing_trim**
Overall QC **Pass**

| Assay Information

Service Provider **Illumina**
Testing Item
Cardiomyopathy--Illumina
Version **3**
Created Date **02 Aug. 2022**

| Overview

M00168

Platform

SAMTOOLS

Alignment Tool

Version: 1.10-13-ga2916aa
Reference:

GATK(HAPLOT...

Calling Tool

Version: 3.8-0-ge9d806836
Reference: unknown

22 / 22

QC Score

FASTQ	BAM	VCF
14 / 14	4 / 4	4 / 4

Recommendation:

None

clean_read1.fastq	clean_read2.fastq	14 / 14
Read 1 File Name	Read 2 File Name	Testing Result

ITEM (Read 1)		CRITERIA	TESTING SAMPLE	QC
Total Reads[a]		≧ 50,000	79,987	✓
% of Duplicate Reads[b]		≦ 95%	92.08%	✓
Filtered Reads[c] (Condition: 100% of the filtered reads that the quality score > Q20)		≧ 40,000	63,622	✓
Quality Distribution	Median	≧ Q30 in each base	See Appendix	✓
	Lower Quartile	≧ Q20 in each base	See Appendix	
Sum of GC Content Deviations[d]		≦ 50%	45%	✓
% of N Content		≦ 20% in each base	See Appendix	✓
Sequences in Same Length		NO	See Appendix	✓
Sequences Length[e]		≧ 15 bp in all sequence	See Appendix	

ITEM (Read 2)		CRITERIA	TESTING SAMPLE	QC
Total Reads[a]		≧ 50,000	79,987	✓
% of Duplicate Reads[b]		≦ 95%	86.50%	✓
Filtered Reads[c] (Condition: 100% of the filtered reads that the quality score > Q20)		≧ 40,000	43,203	✓
Quality Distribution	Median	≧ Q20 in each base	See Appendix	✓
	Lower Quartile	≧ Q30 in each base	See Appendix	
Sum of GC Content Deviations[d]		≦ 50%	34%	✓
% of N Content		≦ 20% in each base	See Appendix	✓
Sequences in Same Length		NO	See Appendix	✓
Sequences Length[e]		≧ 15 bp in all sequence	See Appendix	

a. The count of the total number of sequences processed in this file.
b. The percentage of reads that start at the same position and is an indicator of library complexity. A low level of duplication may indicate a very high level of coverage of the target sequence.
c. The count of the total number of sequences passed the filtering criteria in this file.
d. The percentage of GC in the sample. The value is varies across the regions. For exome regions, the GC content is about 49–51%, and for whole-genome sequencing, the GC content is around 38–39%.
e. The range of fragment length in the sample.

trim_aligned.bam

File Name

4 / 4

Testing Result

ITEM	CRITERIA	TESTING SAMPLE	QC
Average Coverage Depth[a]	≧ 70X	72X	✓
% of the Bases Should in Target Region When the Minimum Coverage Depth ≧ 20X[b]	≧ 80%	84.86%	
% of Reads Mapped to Reference Genome[c]	≧ 85%	99.34%	✓
% of Bases in Target Region[d]	≧ 90%	90.31%	✓
% of the reads that the mapping quality ≧ Q30	≧ 50%	99.25%	✓

a. The average coverage depth of the bases in the target region. The number of reads covering a given base position is described as depth of coverage, and this parameter contributes to the accuracy, detection, sensitivity, and specificity of variant.
b. The percetage of the bases whose read depth is higher than 20X.
c. The percentage of the reads mapped to the reference genome.
d. The percentage of the bases covered by the sample in the target region.

trim_filtered_SNPs.vcf

File Name

4 / 4

Testing Result

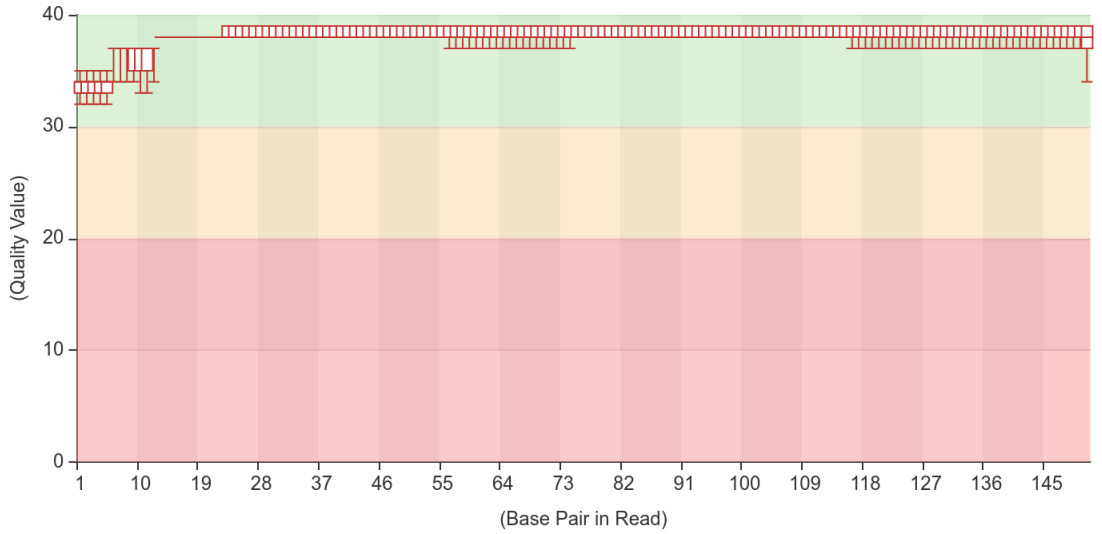
ITEM	CRITERIA	TESTING SAMPLE	QC
Ti/Tv Ratio	2 - 5	4.17	✓
% of the Variants that the Variant Quality by depth ≦ 10	≦ 30%	25.6%	✓
% of the Variants that the Genotype Quality ≦ Q10	≦ 30%	0%	✓
% of the Variants that the Depth of Variant ≦ 20X	≦ 20%	19.2%	✓

Appendix

| Base Quality

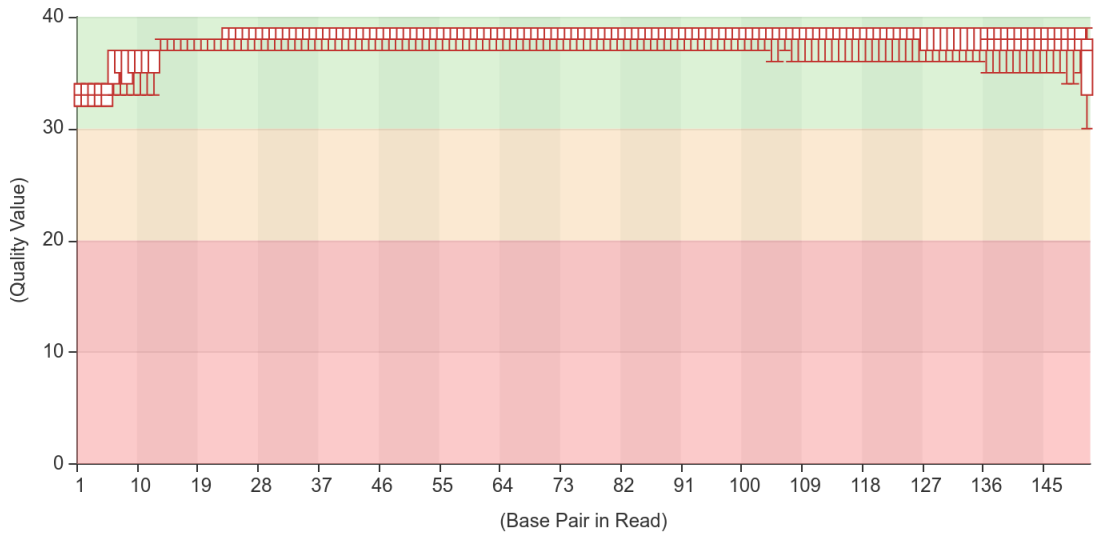
Quality Distribution / FASTQ-Read1

This chart shows the quality score across all bases at each position. The higher quality score means the base-calling results are more reliable and it's not unusual to see the quality falling towards the end of a read.



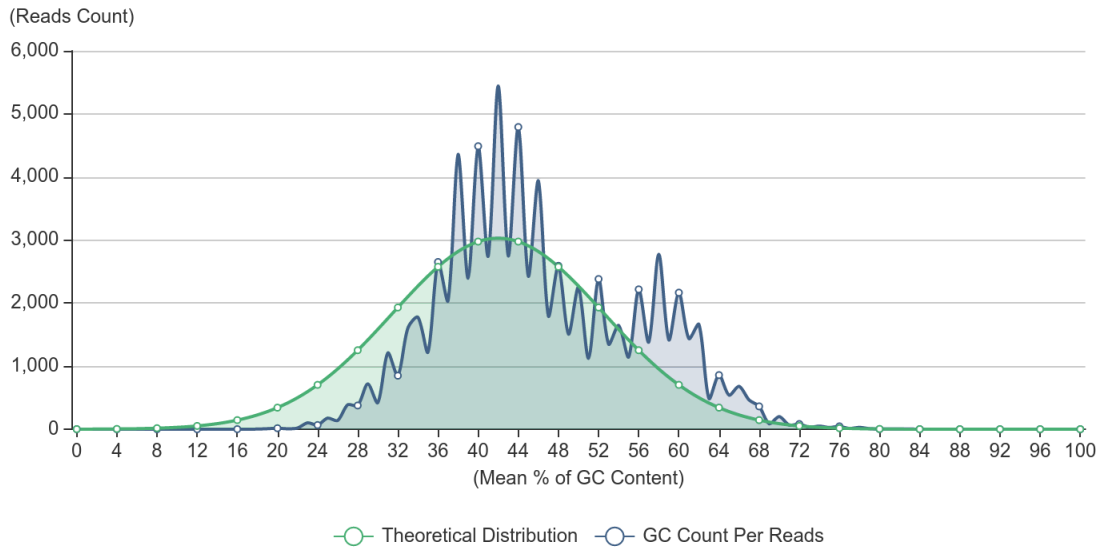
Quality Distribution / FASTQ-Read2

This chart shows the quality score across all bases at each position. The higher quality score means the base-calling results are more reliable and it's not unusual to see the quality falling towards the end of a read.



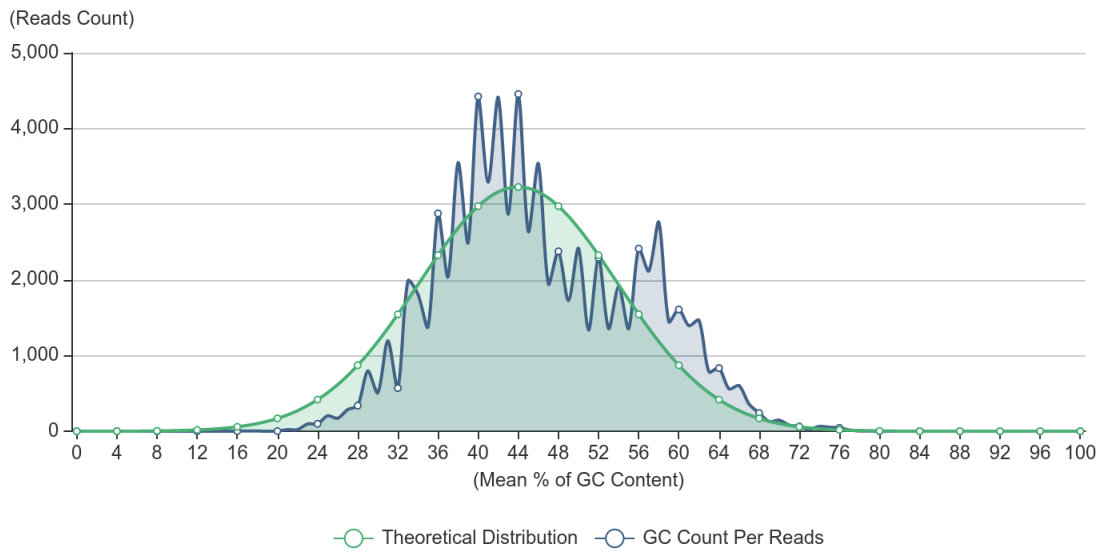
GC Content / FASTQ-Read1

This chart shows the GC distribution in the sample and the blue line is the theoretical normal distribution. The An unusually shaped distribution or high deviation value could indicate a contaminated library or some other kinds of biased subset.



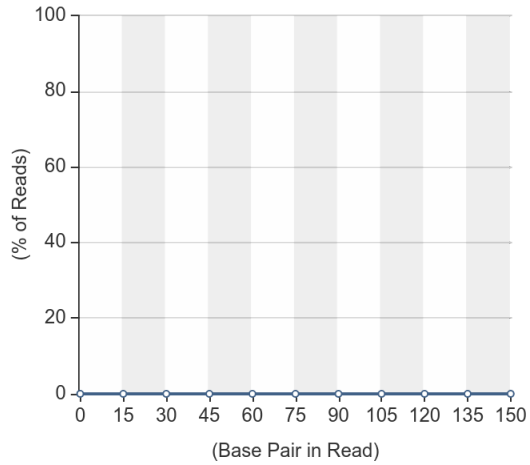
GC Content / FASTQ-Read2

This chart shows the GC distribution in the sample and the blue line is the theoretical normal distribution. The An unusually shaped distribution or high deviation value could indicate a contaminated library or some other kinds of biased subset.



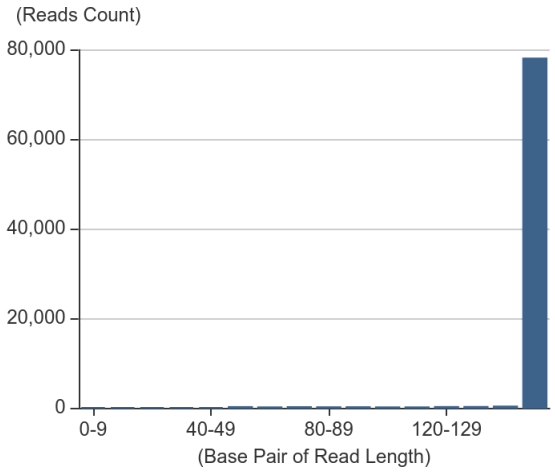
N Content / FASTQ-Read1

The charts shows the precentage of the un-called bases in each position. It's common to get few N bases, but if this proportion rises it represents that the base-calling pipeline could not interpret the sample data.



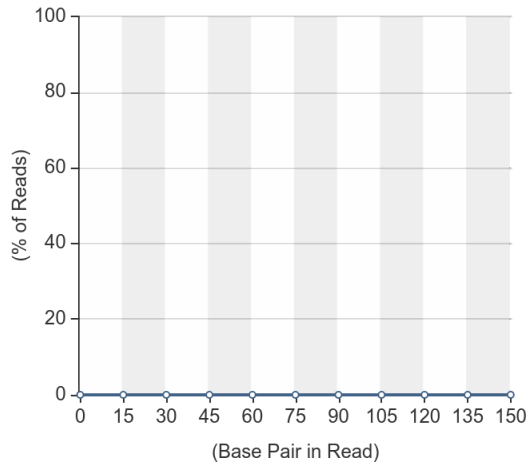
Length Distribution / FASTQ-Read1

This chart shows the distribution of fragment length in the sample, In many cases, the lenght should between 50 to 200. The length will be the same or not depends on your experimental setting of sequencing trimming pipeline.



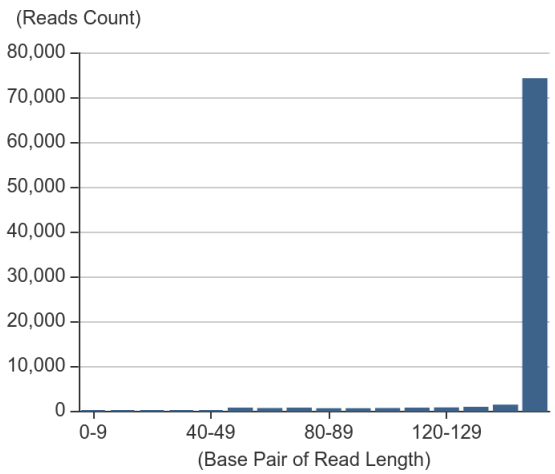
N Content / FASTQ-Read2

The charts shows the precentage of the un-called bases in each position. It's common to get few N bases, but if this proportion rises it represents that the base-calling pipeline could not interpret the sample data.



Length Distribution / FASTQ-Read2

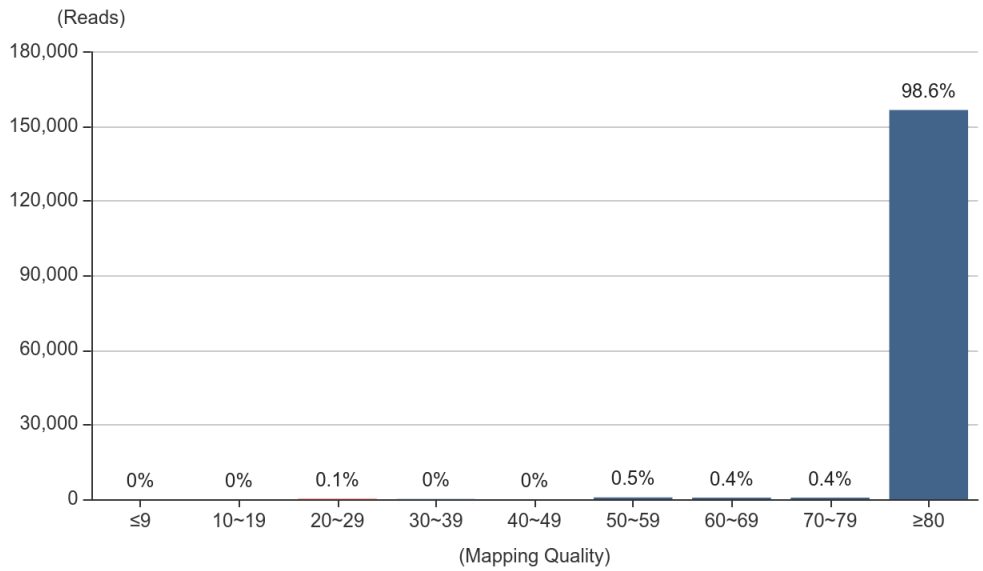
This chart shows the distribution of fragment length in the sample, In many cases, the lenght should between 50 to 200. The length will be the same or not depends on your experimental setting of sequencing trimming pipeline.



| Mapping Metrics

Mapping Quality Distribution

Mapping Quality Scores quantify the probability that a read is misplaced. The higher value reflects the mapping result is more reliable.



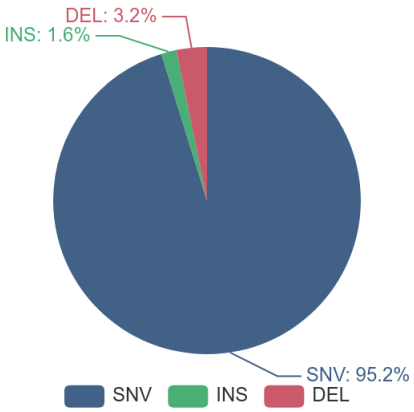
| Variant Quality

Ti/Tv

The ratio of transiton to transversion (Ti/Tv, also called Ts/Tv) could be an evaluation criteria of sequencing results. Transitions are mutations with the same type of nucleotide (T↔C, A↔G) and occur at higher frequencies than transversions (T↔A/G, C↔A/G).

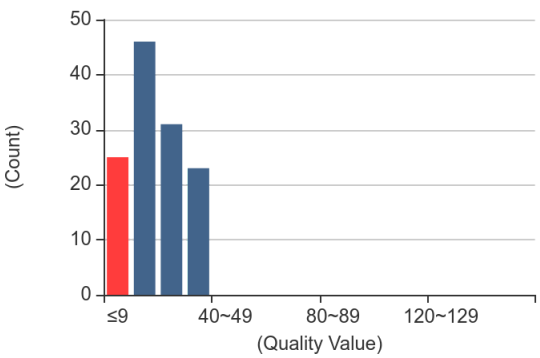
Variant Types

The variant type is based on the result of each allele with alignment to reference sequence. There are five variant types including SNP, INS, DEL, INDEL, and MNP(Multiple Nucleotide Polymorphism).



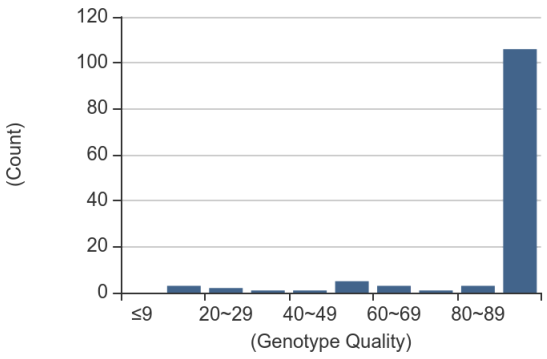
Variant Quality by Depth Distribution

Variant Quality by Depth puts the variant confidence into perspective by normalizing for the amount of coverage available. The higher value reflects the variant calling result is more reliable.



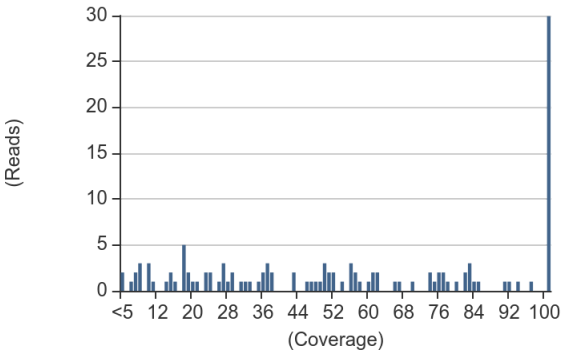
Genotype Quality Score

The genotype quality score represents the Phred-scaled confidence that the called genotype is the true. Higher values reflect more accurate genotype calls.



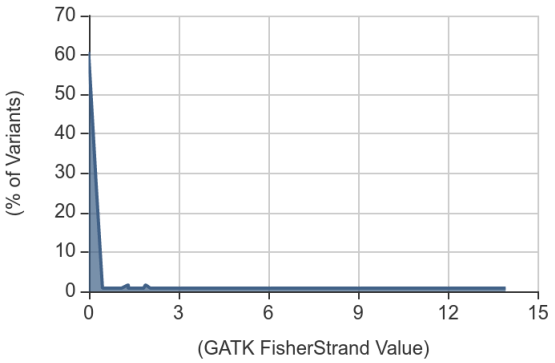
Depth of Variants

The number of aligned reads after filtering on a given position of the target region. For detecting human genome SNVs or structure variation.



Strand Bias Distribution

Strand bias is a type of sequencing bias in which one DNA strand is favored over the other, which can result in incorrect evaluation of the amount of evidence observed for one allele vs. the other. The higher the output value, the more likely there is to be bias. Each laboratory must define the tolerance level for strand bias and outline specific criteria for when alternate testing should be instituted.



References

1. Use of Standards in FDA Regulatory Oversight of Next Generation Sequencing (NGS)-Based In Vitro Diagnostics (IVDs) Used for Diagnosing Germline Diseases. FDA, Jul. 2016
2. Use of Public Human Genetic Variant Databases to Support Clinical Validity for Next Generation Sequencing (NGS)-Based In Vitro Diagnostics. FDA, Jul. 2016
3. Clinical laboratory standards for next-generation sequencing. ACMG, Sep. 2013.
4. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. ACMG, Mar. 2015.
5. Good laboratory practice for clinical next-generation sequencing informatics pipelines. CDC, Jul. 2015.
6. "Next Generation" Sequencing (NGS) guidelines for somatic genetic variant detection. NYSDOH, Mar. 2015.
7. Guidelines for Validation of Next-Generation Sequencing Based Oncology Panels. AMP/CAP, May. 2017.
8. Guidelines for development and validation of software, with particular focus on bioinformatics pipelines for processing NGS data. ACGS, Sep. 2016.
9. Guidelines for diagnostic next-generation sequencing. EuroGentest, Oct. 2015.
10. Massively Parallel Sequencing Implementation Guidelines. RCPA, May. 2014.
11. Assuring the Quality of Next-Generation Sequencing in Clinical Laboratory Practice. CDC, Nov. 2012.