BitGenetic

Advancing life through genomic medicine

Bioinformatics QC Report

Report Serial Number: D2107190002

Contract ID: CYB21070023

Project Name: EU-Spain-BitGenetic-WES-WBI

Customer Name: Indiana Jones

Sample Receipt Date: 2021/07/18

Report Date: 2021/07/27

Index

1. RAW DATA 3

2. GENERAL SEQUENCING QUALITY INFORMATION... 4

3. STATISTICS OF COVERAGE.. 4

1. Raw Data

The original raw image data obtained from high throughput sequencing platforms (e.g. Illumina platform) is transformed to sequenced reads by base calling. The sequenced reads are regarded as raw data or raw reads, which is recorded in FASTQ file (fq) containing sequence information (reads) and corresponding sequencing quality information. Every read in FASTQ format is stored in four lines as follows:

```
@EAS139:136:FC706VJ:2:2104:15343:197393 1:Y:18: ATCACG
GCTCTTTGCCCTTCTCGTCGAAAATTGTCTCCTCATTCGAAAC
TTCTCTGT
+
@@CFFFDEHHHHFIJJJ@FHGIIIEHIIJBHHHIJJEGIIJJIGHIGHCCF
```

Line 1 beginning with a '@' character is followed by a sequence identifier and an optional description (like a FASTA title line). Line 2 is the raw sequence reads. Line 3 begins with a '+' character and is optionally followed by the same sequence identifier (and any description) again. Line 4 encodes the quality values for the sequence in Line 2, and must contain the same number of characters as bases in the sequence.

Table 1 Illumina sequence identifier details

EAS139	The unique instrument name
136	Run ID
FC706VJ	Flowcell ID
2	Flowcell lane
2104	Tile number within the flowcell lane
15343	'x'-coordinate of the cluster within the tile
197393	'y'-coordinate of the cluster within the tile
1	Member of a pair, 1 or 2 (paired-end or mate-pair reads only)
Y	Y if the read fails filter (read is bad), N otherwise
18	0 when none of the control bits are on, otherwise it is an even number
ATCACG	Index sequence

The ASCII value for every character at the fourth line minus 33 will be the corresponding sequencing base quality value at the second line. If the sequencing error rate is recorded by "e" and the base quality for Illumina platform is expressed as Q_{phred}, the equation 1 as below will be obtained:

Equation 1: $Q_{phred} = -10\log_{10}(e)$

The relationship between sequencing error rate (e) and sequencing base quality value (Q_{phred}) is listed as below (Table 2):

Table 2 Sequencing error rate and corresponding base quality value

Sequencing error rate	Sequencing quality value	Corresponding character
5%	13	.
1%	20	5
0.1%	30	?
0.01%	40	I

The higher the quality value, the lower the error rate and the higher the accuracy.

2. General Sequencing Quality Information

Table 3 The overview of sequencing quality

Sample Name	Sample ID	Library ID	Raw bases (bp)	Raw PE Reads	Q20(%)	Q30(%)	GC(%)	Meet Criteria?
BT-0000002_2F	KKHS210004318-1A	EDHE210009200-1A-7UDI1366-5UDI1366	13,994,022,300	46,646,741	98.25	94.75	49.77	YES
BT-0000003_3M	KKHS210004317-1A	EDHE210009200-1A-7UDI1364-5UDI1364	13,194,533,100	43,981,777	98.22	94.62	49.70	YES

Note:
(1) Sample ID: Sample ID.
(2) Library ID: Library ID.
(3) Raw bases: The original bases of sequence data.
(4) Raw PE reads: The number of sequencing reads pairs; four lines will be considered as one unit according to FASTQ format. (5) Q20: The percentage of bases with Phred score ≥20.
(6) Q30: The percentage of bases with Phred score ≥30.
(7) GC: The percentage of G and C in the total bases.

3. Statistics of Coverage

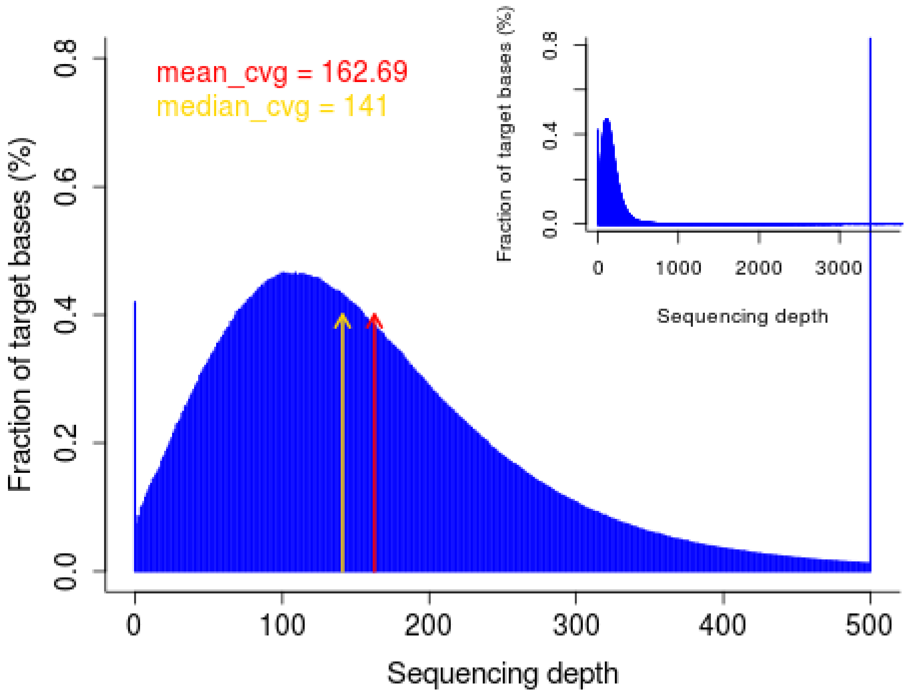
Table 4 The summary of mapping information

Sample Name	Sample ID	Library ID	Coverage of target	Average depth on	Mapping rate	% 4x Coverage	% 20x Coverage	Meet Criteria?

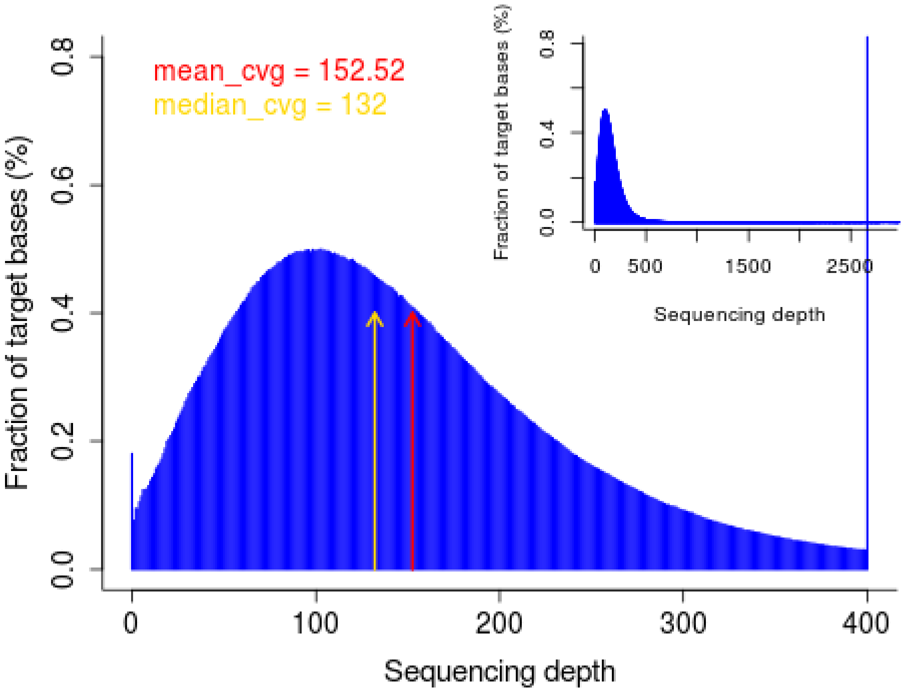
			region(%)	target:				
BT-000000 2_2F	KKHS2 100043 18-1A	EDHE2100 09200-1A-7UDI1366-5UDI1366	99.58	162.69	99.99	99.34	97.14	YES
BT-000000 3_3M	KKHS2 100043 17-1A	EDHE2100 09200-1A-7UDI1364-5UDI1364	99.82	152.52	99.99	99.55	97.19	YES

Note:
(1) Sample ID: Sample ID.
(2) Library ID: Library ID.
(3) Coverage of target region: The coverage percent of the target region.
(4) Average depth on target: The average sequencing depth on the target region. (5)
Mapping rate: The percent of total reads that mapped to the reference genome. (6)
% 4x Coverage: The fraction of target covered with at least 4x.
(7) % 20x Coverage: The fraction of target covered with at least 20x.

The follow figure is the ratio of bases with different sequencing depth. The x-axis is sequencing depth; the y-axis is the fraction of bases with the given sequencing depth. The curve follows a Poisson distribution around the average read depth.



The distribution of sequencing depth of KKHS210004318-1A.



The distribution of sequencing depth of KKHS210004317-1A.

METHODOLOGY

In this assay, human genomic DNA is sheared mechanically and prepared as libraries containing duel-indexed sequencing barcodes. The libraries are sequenced on a NovaSeq 6000 instrument (Illumina).

The sequence data are analyzed using a custom-developed bioinformatics pipeline which aligns sequence data to human genome (GRCh37/UCSC hg19) . The pipeline also performs QC analysis of the sequence data to ensure that the reported variant findings were obtained from quality sequencing data.

LIMITATIONS

This test analyzes whole genome regions. Some genes have inherent sequence properties (for example: repeats, homology, high GC content, rare polymorphisms) that may result in suboptimal data, and variants in those regions may not be reliably identified. At present, whole genome sequencing cannot consistently detect single and multi-exon deletions or duplications. In addition, whole genome sequencing does not provide complete coverage of all coding exons. It is possible that the genomic region where a disease-causing variant exists was not sufficiently sequenced in the current assay. Although rare, false positive or false negative results may occur. Results should be interpreted in the context of clinical findings, relevant history, and other laboratory data.

DISCLAIMER

The test results are for scientific research and clinical reference only. BitGenetic is only responsible for samples processed by BitGenetic. If you have any questions, please contact us within 7 workdays. If the analysis result is inaccurate due to the wrong information provided by the customer, the responsibility shall be borne by the customer. BitGenetic reserves all the right for the final explanation of this report.

Reported by

<Signature>

2021.07.27

<Date>