Sequencing Data Report

Project: F24A430001351_HOMucbrN

Date: 2025.5.27

Note: For Research Use Only.

BGI Genomics Co.,Ltd.

Table of Contents

Data Statistics

Raw reads produced from sequencer contain adapters, unknown or low quality bases.

There are 4 samples in this project, the statistics of fastq data is shown below.

Sample	Length	Q20(%)	Q30(%)	GC Content(%)	Total Reads	Total Bases
NPC_mixA_cDNA	30;100	98.52;98.59	87.87;93.76	49.62;46.99	889,674,028	115,657,623,640
NPC_mixA_ologo	20;30	98.96;99.28	95.73;96.67	47.71;48.35	260,766,921	13,038,346,050
single_cell_1A_cDNA	30;100	98.55;98.33	88.29;93.33	49.81;46.88	770,731,614	100,195,109,820
single_cell_1A_ologo	20;30	98.96;99.28	95.75;96.68	47.6;48.0	218,724,924	10,936,246,200

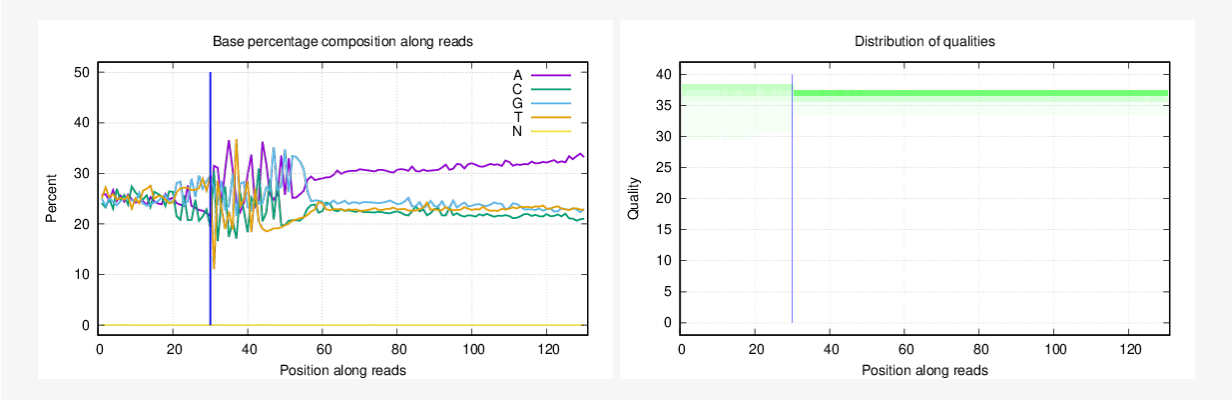
Table Format:

1. Sample: The name of sample
2. Length: The Length of reads
3. Q20 (%): The proportion of nucleotides with quality value larger than 20
4. Q30 (%): The proportion of nucleotides with quality value larger than 30
5. GC Content(%): The proportion of bases G and C
6. Total Reads: The total number of read pairs
7. Total Bases: The total nucleotides number of reads

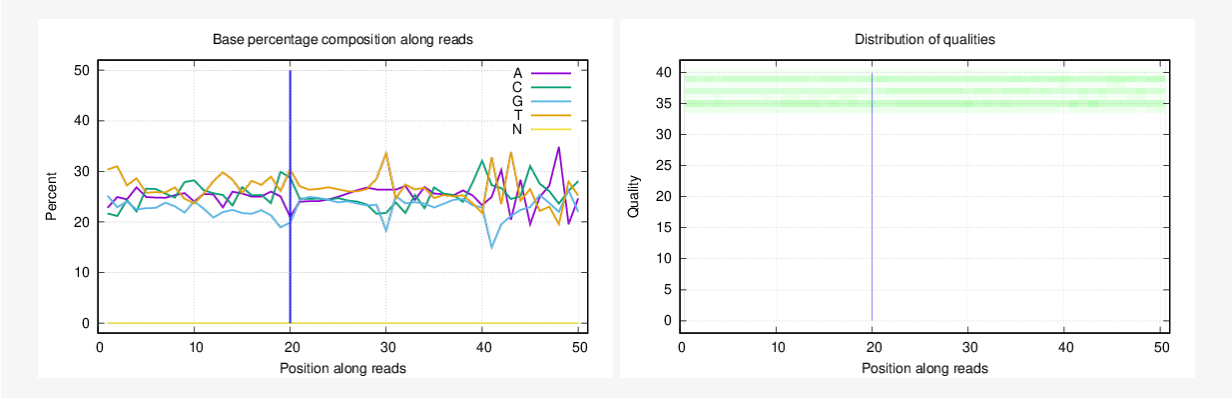
Data Quality Control

The distribution of base percentage and qualities along reads in data filtering are shown as following(If a sample has multiple lanes, only one of them will be displayed). The left picture is base percentage distribution along reads the sample, the right picture is distribution of qualities along reads of the sample.

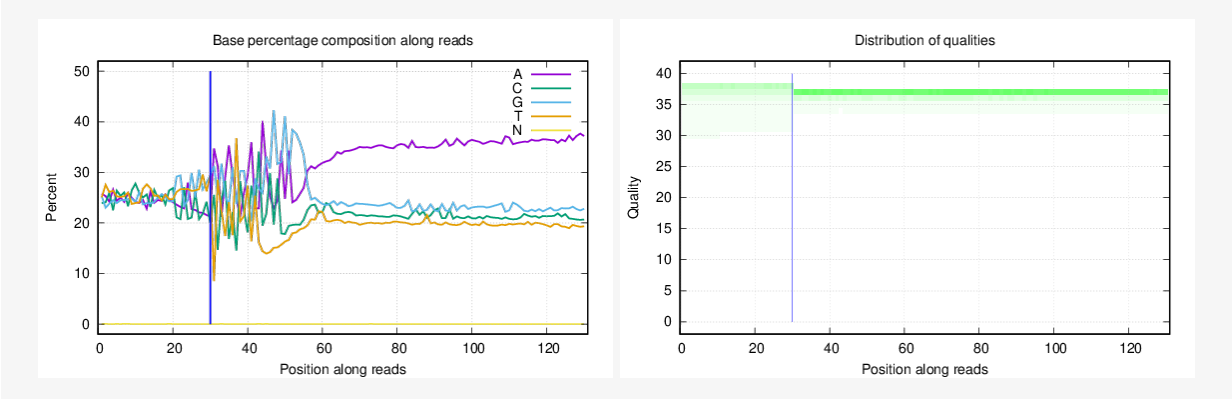
NPC_mixA_cDNA



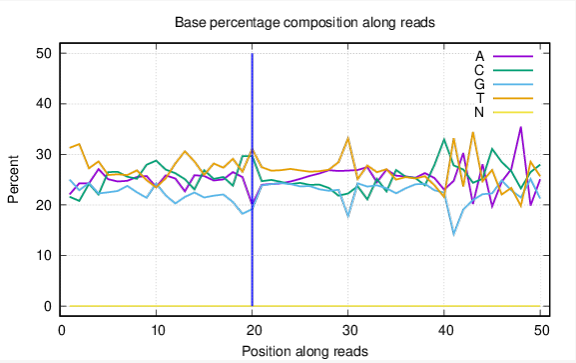
NPC_mixA_ologo

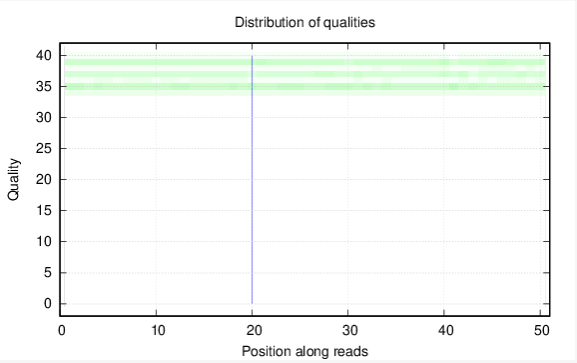


single_cell_1A_cDNA



single_cell_1A_ologo





Help Document

The original image data is transferred into sequence data via base calling, which is defined as raw data or raw reads and saved as FASTQ file. Each entry in a FASTQ files consists of 4 lines:

1. A sequence identifier with information about the sequencing run and the cluster. The exact contents of this line vary by based on the BCL to FASTQ conversion software used.
2. The sequence (the base calls; A, C, T, G and N).
3. A separator, which is simply a plus (+) sign.
4. The base call quality scores. These are Phred +33 encoded, using ASCII characters to represent the numerical quality scores.

Here is an example of a single entry in a FASTQ file:

```
@V300029029L1C001R0010000210/1
GCGACCCCAGGTCAGTCGGGACTACCCGCTGAAGTCGGAGGCCAAGCGGT
+
FFFCFFFFFFFFFDFEFFFFEFEF0FFFFEFFFFFFFEFFFFFECGFFFF
```

The relationship between DNBSEQ sequencer sequencing error rate and the sequencing quality value is shown in the following formula. Specifically, if the sequencing error rate is denoted as "E", DNBSEQ sequencer base quality value is denoted as "sQ", the relationship is as follows:

$$sQ = -10\log_{10} E$$

Sequencing error rate	Sequencing quality value	Character of Phred +33 quality system
5%	13	.
1%	20	5
0.1%	30	?

Omics For All

Contact us

Website: www.bgi.com
E-mail: info@bgi.com

For Research Use Only. Not for use in diagnostic procedures.