Sequencing Data Report

Project: F24A430001331_MUSkbxhR

Date: 2025.4.23

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 27 samples in this project, the statistics of fastq data is shown below.

Sample	Length	Q20(%)	Q30(%)	GC Content(%)	Total Reads	Total Bases
1	150;150	96.76;94.46	90.12;85.74	45.99;46.07	30,035,250	9,010,575,000
11	150;150	96.71;93.71	89.94;84.04	47.05;47.27	14,810,656	4,443,196,800
13	150;150	96.89;93.42	90.47;83.48	46.89;47.09	24,301,943	7,290,582,900
14	150;150	96.86;93.43	90.33;83.34	47.36;47.51	17,596,207	5,278,862,100
15	150;150	96.96;93.65	90.61;83.66	45.64;45.83	19,985,496	5,995,648,800
2	150;150	96.67;93.6	89.81;83.54	46.23;46.38	13,556,020	4,066,806,000
6069_plus	150;150	97.56;96.04	92.51;89.26	42.67;42.67	16,055,725	4,816,717,500
Blast	150;150	98.67;93.29	95.28;80.79	37.12;36.68	75,958	22,787,400
CMV-PCR-Nla_Kl_1	150;150	97.47;95.6	92.2;87.79	48.71;49.24	941,729	282,518,700
CMV-PCR-Nla_Kl_2	150;150	97.64;95.11	92.73;86.83	50.03;50.65	1,209,480	362,844,000
Ezh1_K	150;150	96.96;94.47	91.03;85.92	43.5;43.6	3,899	1,169,700
Fed3cycle2_K	150;150	97.28;95.03	91.55;86.49	47.48;47.66	807,760	242,328,000
KM2	150;150	97.21;95.78	91.48;89.14	40.69;40.58	31,616,318	9,484,895,400
Lam1_K	150;150	97.48;95.23	92.25;87.62	46.25;46.69	11,353,701	3,406,110,300
MSQ_Chip-Seq_9	150;150	97.56;95.93	92.39;88.7	44.26;44.42	45,535,821	13,660,746,300
Mal3_K	150;150	96.91;95.0	90.5;86.39	43.9;43.94	4,456,179	1,336,853,700
PEF_6_B4	150;150	95.59;95.72	87.55;88.28	56.42;54.38	382,590	114,777,000
PEF_6_CM	150;150	96.46;96.9	89.47;90.86	41.58;42.83	165,157	49,547,100
PEF_6_GG	150;150	96.13;97.09	88.6;91.38	35.16;35.14	200,006	60,001,800
PEF_7_B4	150;150	95.35;95.43	86.87;87.62	56.39;54.45	275,215	82,564,500
PEF_7_CM	150;150	96.39;96.84	89.29;90.7	41.71;42.85	215,032	64,509,600
PEF_7_GG	150;150	95.95;96.73	88.11;90.42	35.38;35.32	328,581	98,574,300
Rosa	150;150	99.03;92.7	96.78;80.15	47.5;47.04	80,976	24,292,800
ZG	150;150	97.19;95.87	91.41;89.22	40.16;40.09	31,074,770	9,322,431,000
ZM1	150;150	97.37;96.05	91.93;89.79	39.99;39.91	33,850,526	10,155,157,800
ZM2	150;150	97.16;96.28	91.32;90.27	41.29;41.18	31,572,241	9,471,672,300

ench3-PCR-Dpn_Kl_3dtransf_4	150;150	97.17;93.26	91.7;83.7	51.59;52.89	999,006	299,701,800

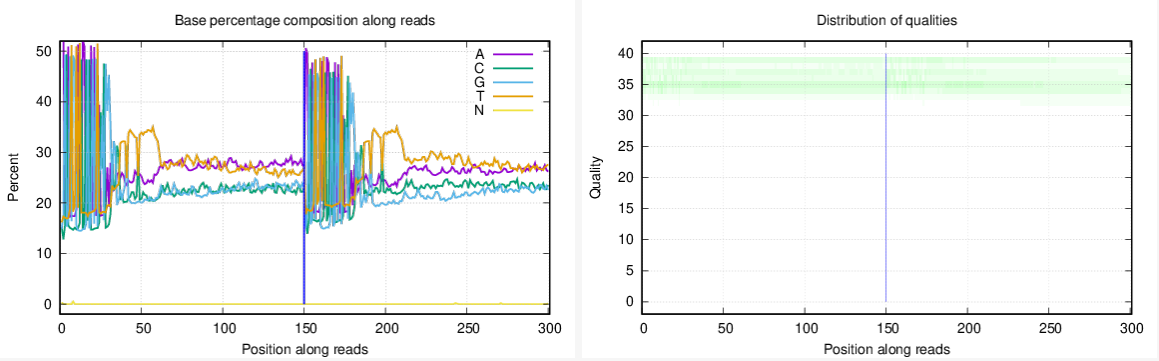
Table Format:

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

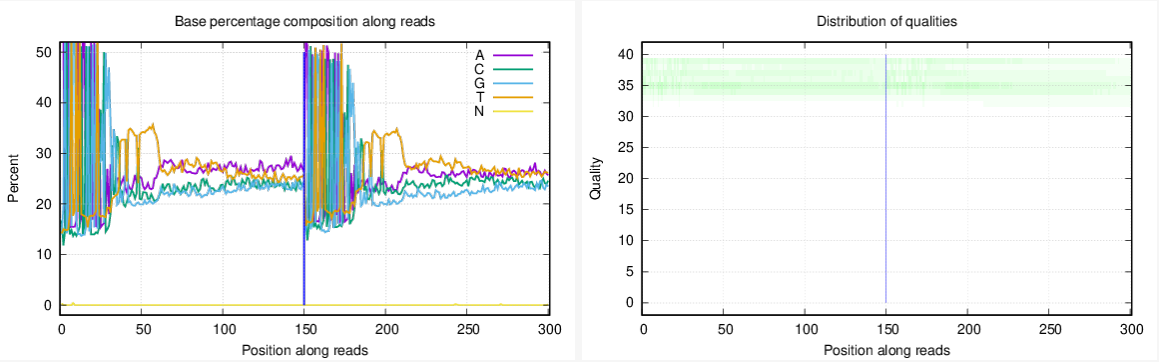
Data Quality Control

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.

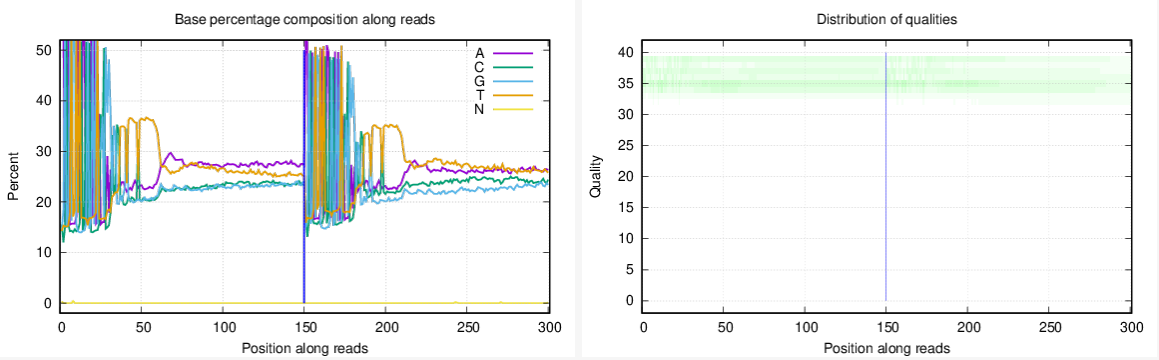
1



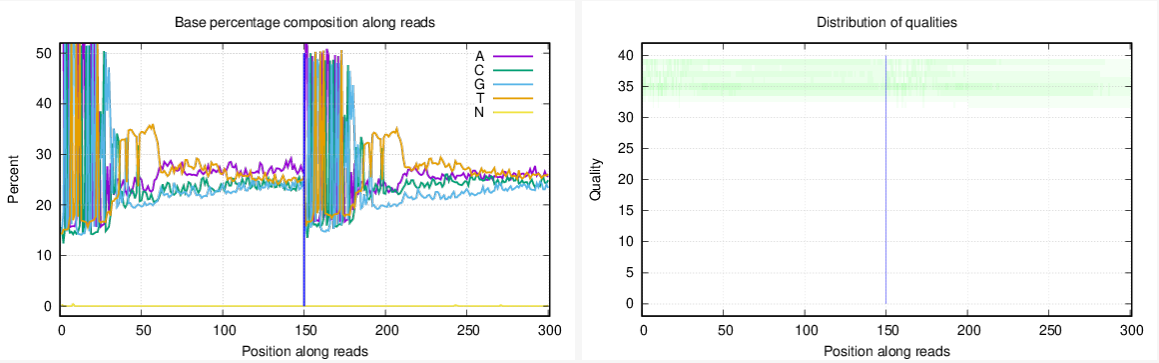
11



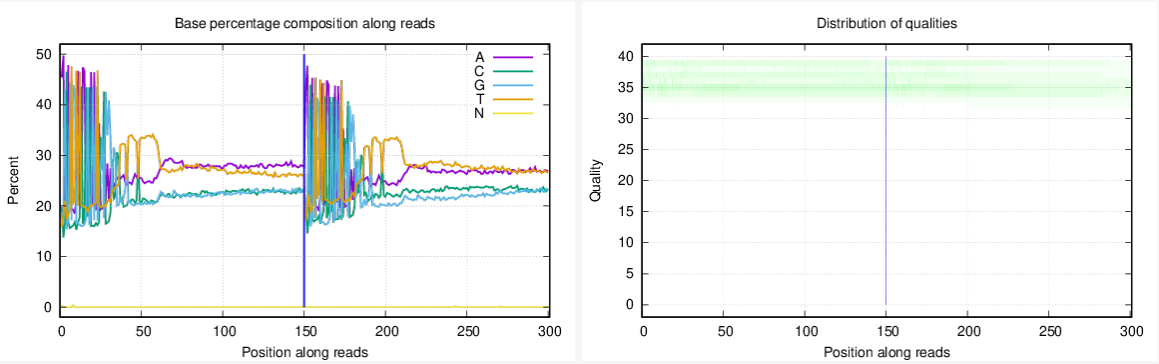
13



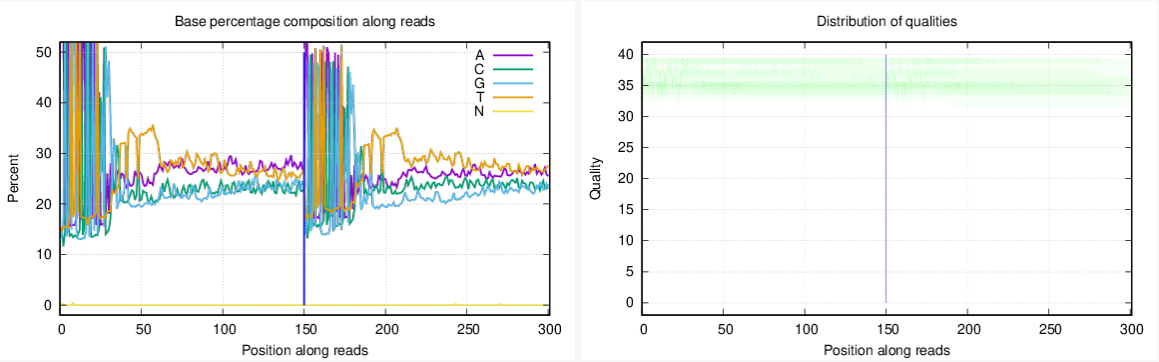
14



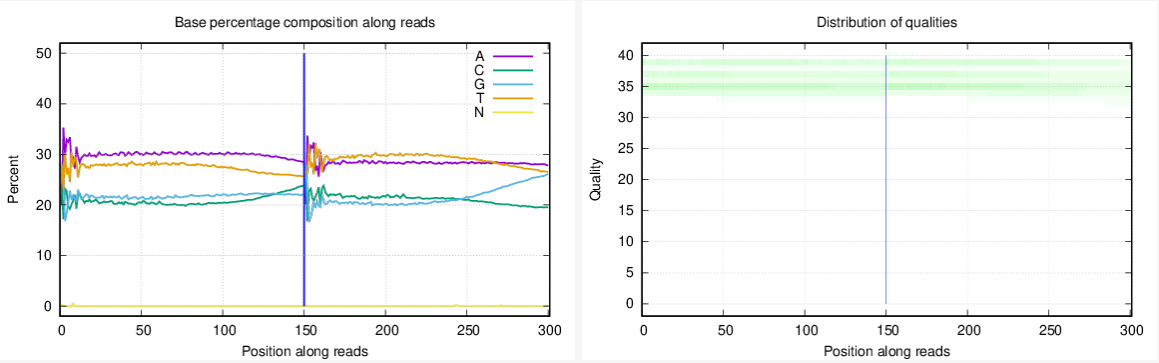
15



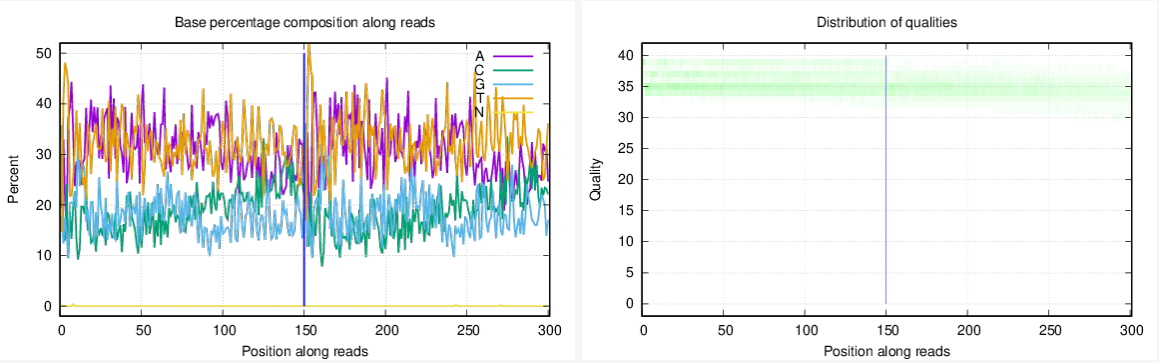
2



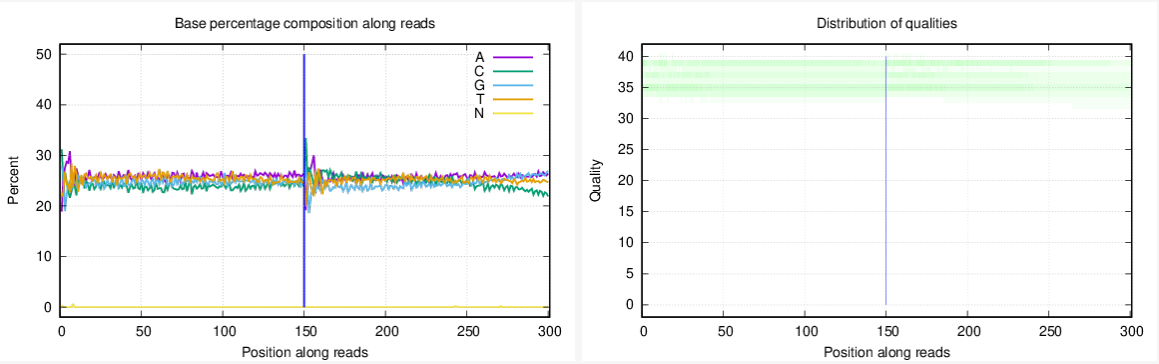
6069_plus



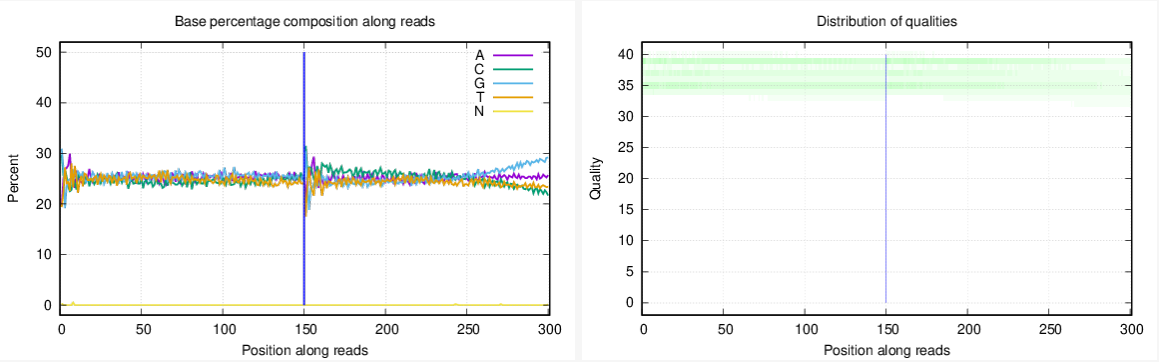
Blast



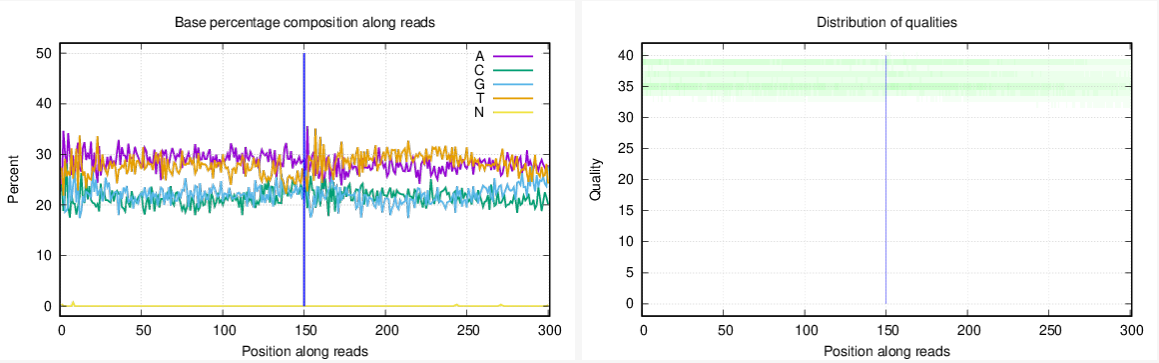
CMV-PCR-Nla_Kl_1



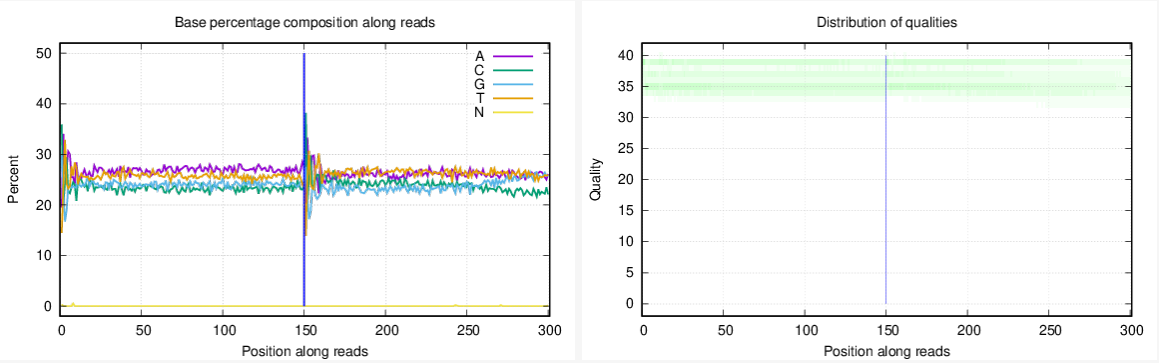
CMV-PCR-Nla_Kl_2



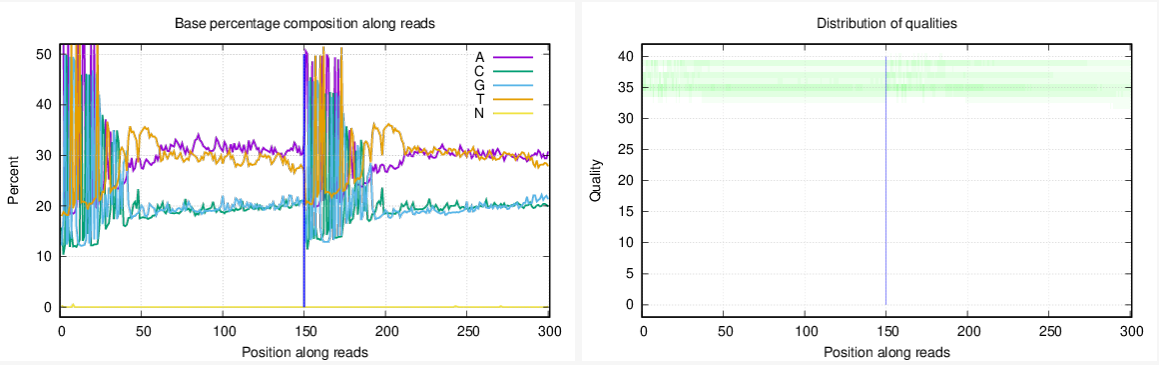
Ezh1_K



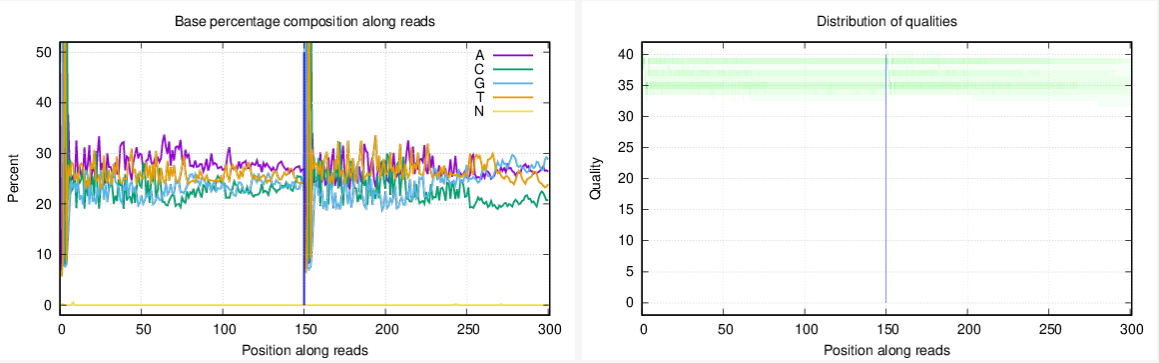
Fed3cycle2_K



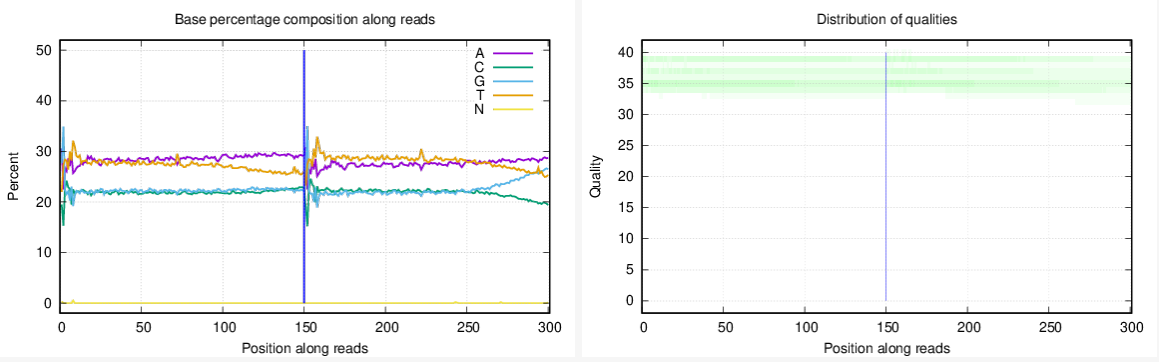
KM2



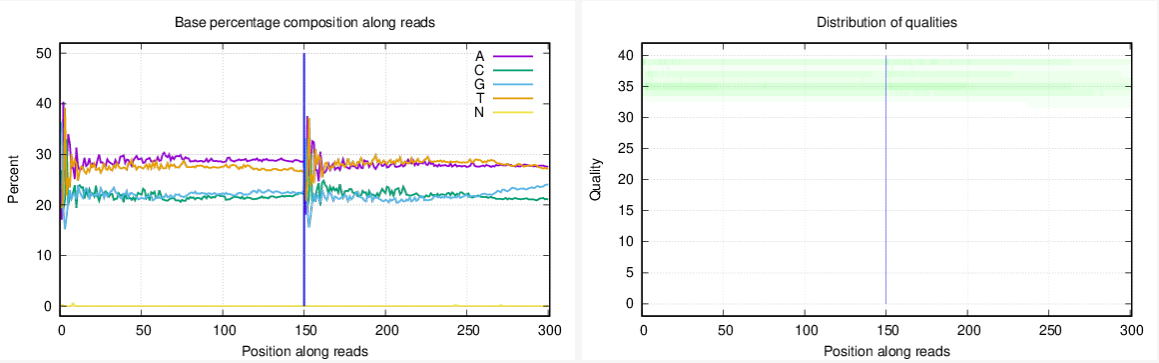
Lam1_K



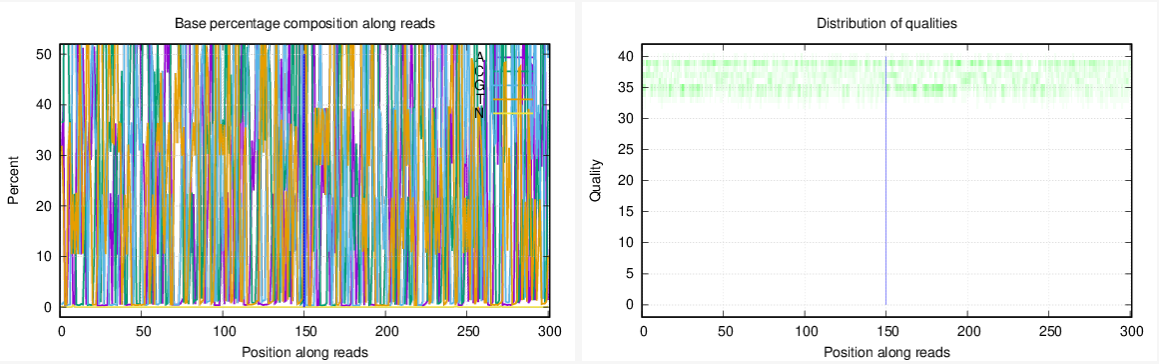
MSQ_Chip-Seq_9



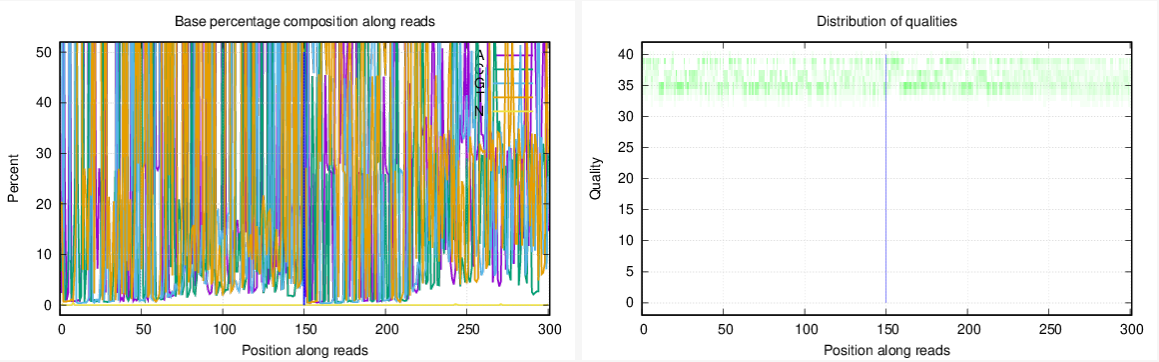
Mal3_K



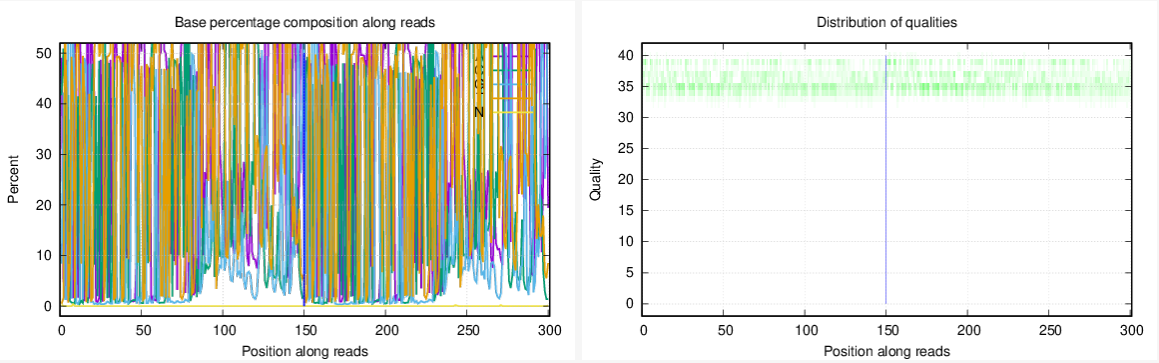
PEF_6_B4



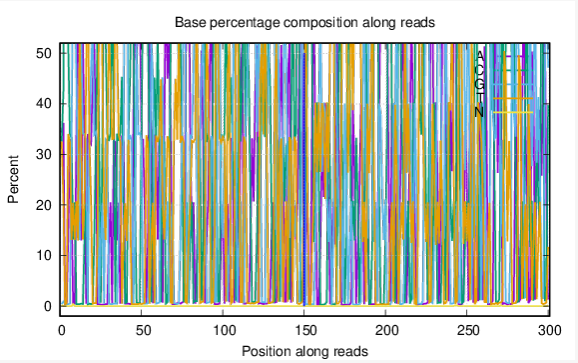
PEF_6_CM

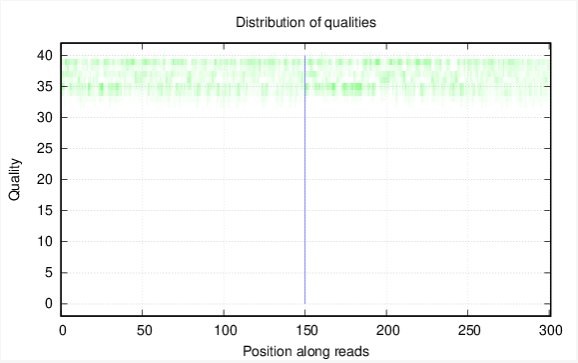


PEF_6_GG

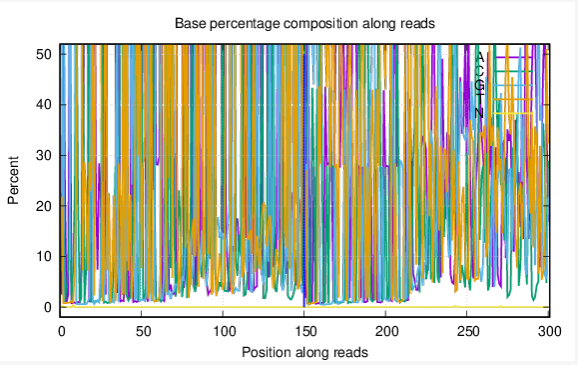


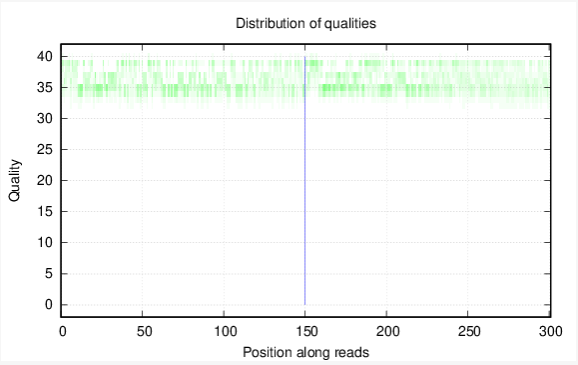
PEF_7_B4



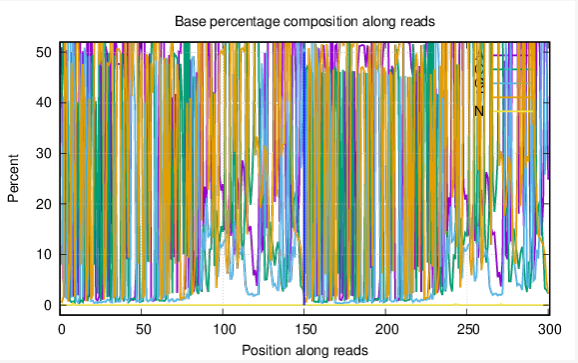


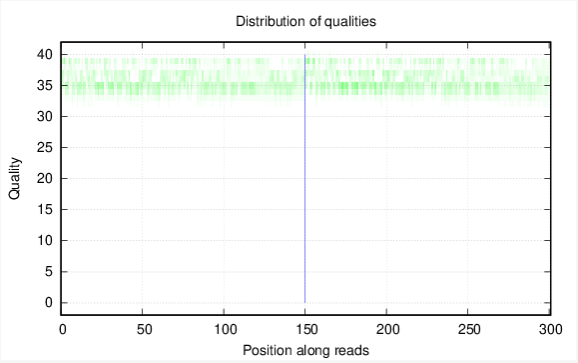
PEF_7_CM



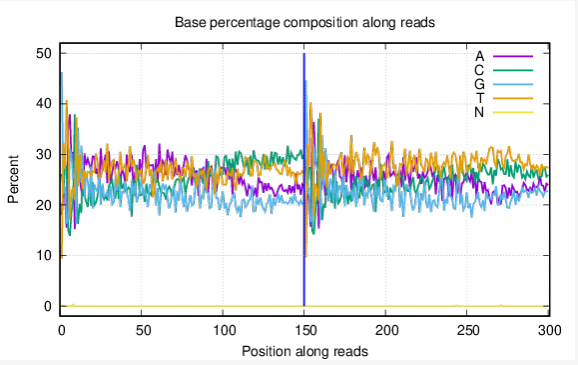


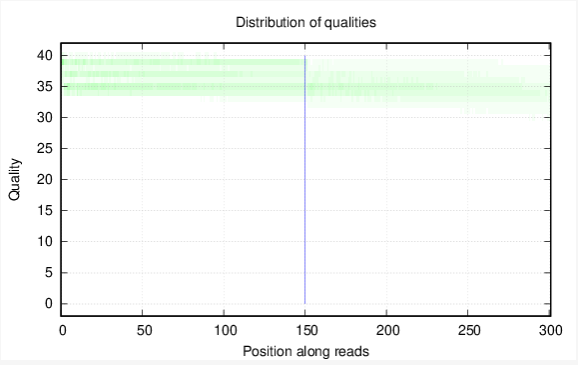
PEF_7_GG



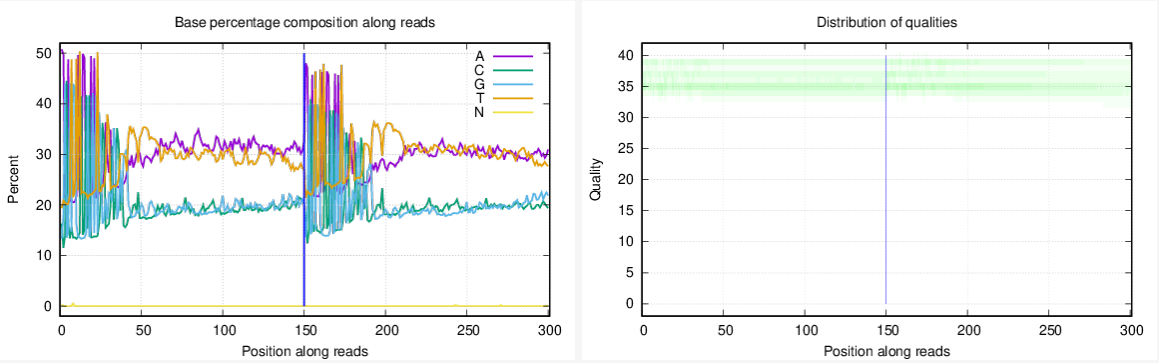


Rosa

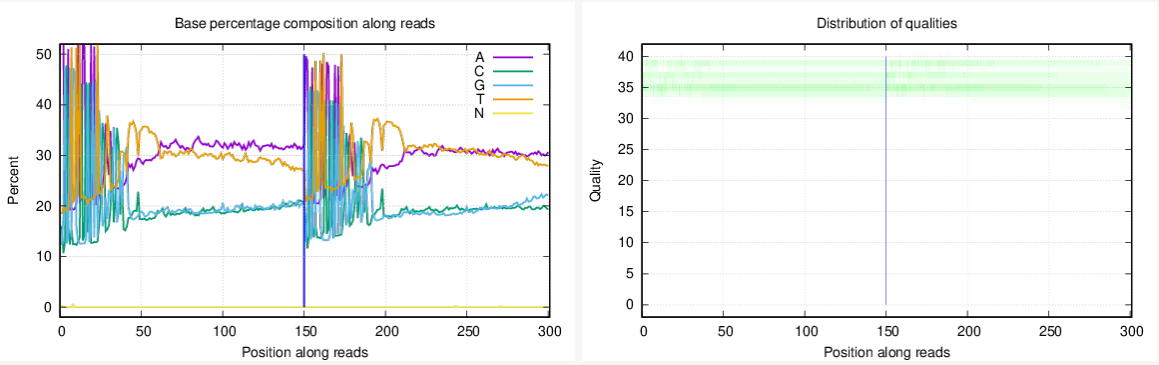




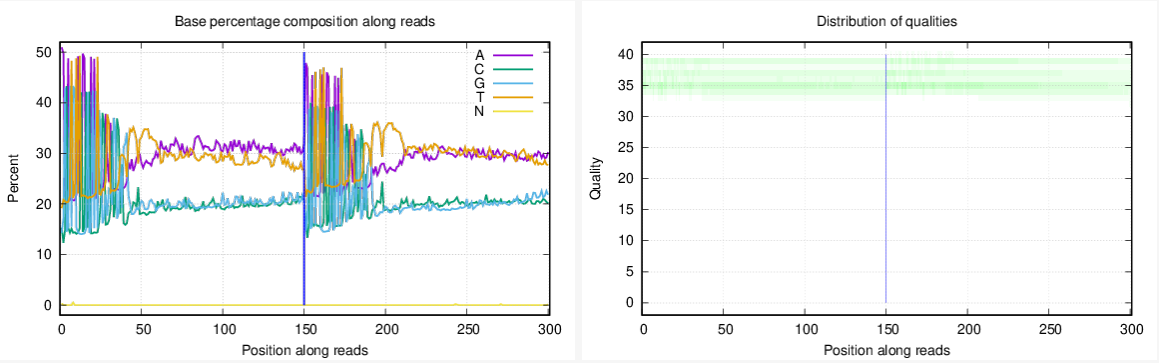
ZG



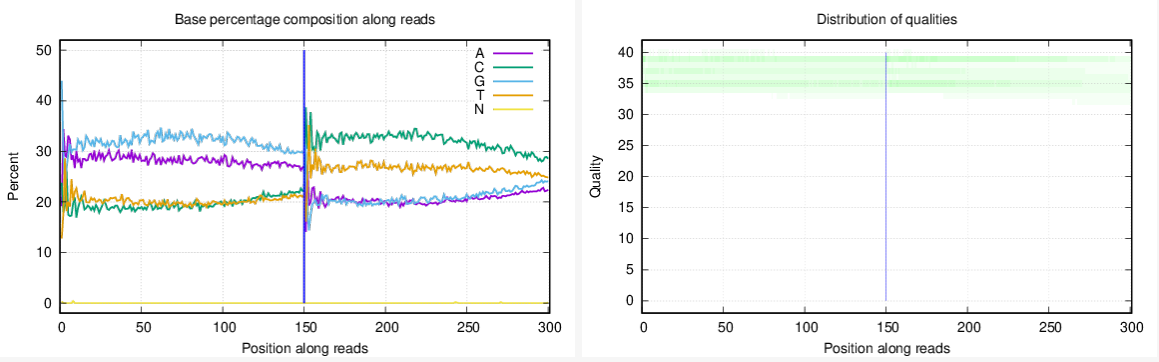
ZM1



ZM2



ench3-PCR-Dpn_Kl_3dtransf_4



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.