BGI Pre-made Library Sequencing Report

2022/7/19

Table of Contents

1 Project Information 3
2 Data Statistics 3
3 Data Quality Control 3
4 Help Document 8

1 Project Information

Project code: Default
Sample number: 21

2 Data Statistics

Raw reads produced from sequencer contain adapters, unknown or low quality bases. The statistics of raw data is shown below.

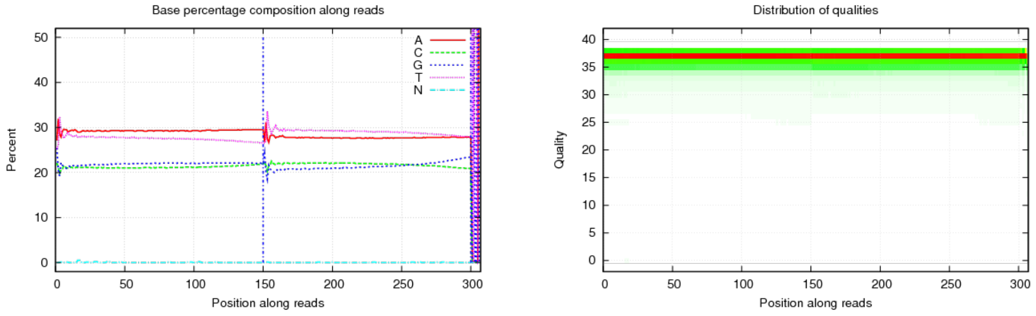
Sample	Length	Q20(%)	Q30(%)	GC Content(%)	Total Reads	Total Bases
GFP-Pei	150;150;6	96.50;96.68;65.01	90.78;90.56;56.30	47.24;47.14;67.52	46,975	14,374,350
MS22	150;150;6	96.78;98.43;92.43	91.48;95.49;85.60	36.73;37.34;49.72	8,230,122	2,518,417,332
MS23	150;150;6	97.99;97.91;85.46	94.24;94.20;72.91	49.60;50.14;50.37	2,481,656	759,386,736
MS24	150;150;6	98.08;98.14;89.27	94.54;94.82;79.60	48.92;49.51;50.28	2,237,195	684,581,670
MS25	150;150;6	98.06;98.05;87.52	94.28;94.55;77.30	50.18;50.65;48.67	3,029,000	926,874,000
MS26	150;150;6	96.93;98.50;93.59	91.80;95.50;87.39	32.70;33.03;49.96	8,807,374	2,695,056,444
MS27	150;150;6	96.66;98.37;92.38	91.13;95.07;86.27	30.63;31.14;33.91	7,249,856	2,218,455,936
Tn5-EC	150;150;6	97.36;96.53;28.88	92.72;90.86;15.74	44.93;44.73;55.96	52,986	16,213,716
Tn5-EP	150;150;6	92.82;96.19;82.48	81.49;88.27;77.88	40.51;40.49;49.46	196,593	60,157,458
Tn5-KP	150;150;6	95.50;97.59;86.76	88.03;92.45;81.71	42.35;42.37;50.44	220,746	67,548,276
Tn5-L	150;150;6	97.46;97.32;97.17	93.03;92.60;94.67	43.16;44.42;50.04	2,148,302	657,380,412
a	150;150;6	97.98;98.18;97.05	94.20;94.48;93.16	43.13;43.23;33.53	14,540,106	4,449,272,436
b	150;150;6	97.91;98.42;95.63	94.01;95.20;90.93	42.87;42.94;49.85	10,147,031	3,104,991,486
c	150;150;6	98.06;98.07;97.38	94.35;94.15;93.84	45.89;45.97;49.99	10,328,265	3,160,449,090
d	150;150;6	98.06;98.47;90.78	94.59;95.48;82.74	44.56;44.24;83.21	17,743,683	5,429,566,998
dMcphC1	150;150;6	98.04;98.07;96.00	94.27;94.08;90.84	43.77;43.82;66.22	77,186,065	23,618,935,890
dMcphC2	150;150;6	98.06;98.24;96.16	94.31;94.55;90.95	43.50;43.55;50.18	76,998,905	23,561,664,930
dMcphRA1	150;150;6	97.94;98.22;97.66	93.97;94.49;93.96	42.76;42.82;50.05	71,590,959	21,906,833,454
dMcphRA2	150;150;6	98.07;98.24;98.54	94.37;94.62;96.41	42.92;42.97;49.97	44,720,493	13,684,470,858
e	150;150;6	98.28;98.57;91.53	95.24;95.82;83.55	44.01;43.78;82.65	22,826,546	6,984,923,076
f	150;150;6	97.98;98.64;97.21	94.27;95.88;93.97	42.47;42.58;50.02	23,779,271	7,276,456,926

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
4. GC Content(%): The proportion of bases G and C
5. Total Reads: The total number of raw reads
6. Total Bases: The total nucleotides number of raw reads

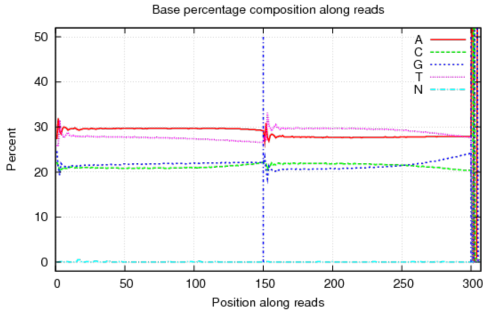
3 Data Quality Control

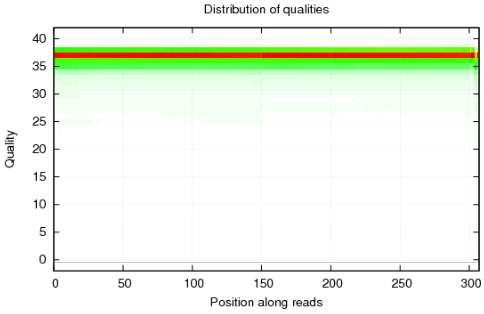
The distribution of base percentage and qualities along reads in data filtering are shown as following(If a sample has multiple lanes, only one of them will be displayed). The left picture is base percentage distribution along reads the sample, the right picture is distribution of qualities along reads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

Quality control of sample a

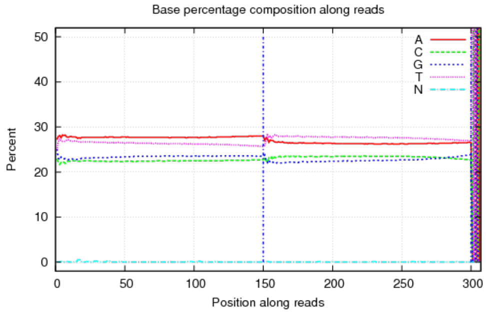


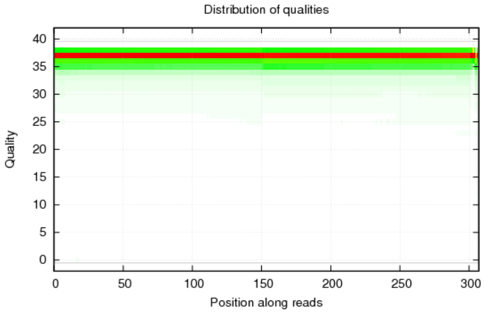
Quality control of sample b



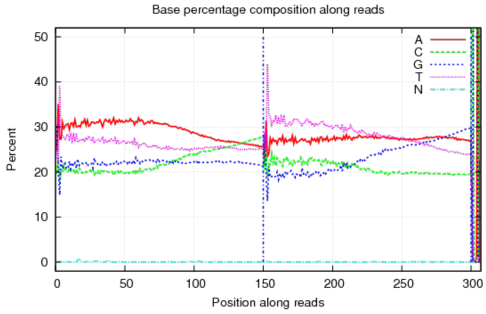


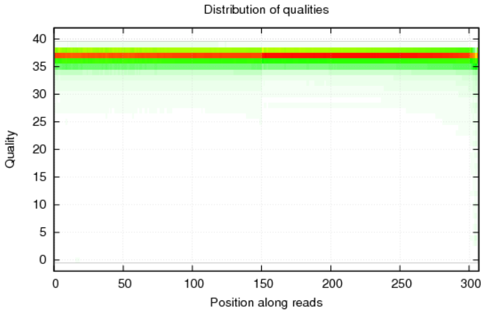
Quality control of sample c



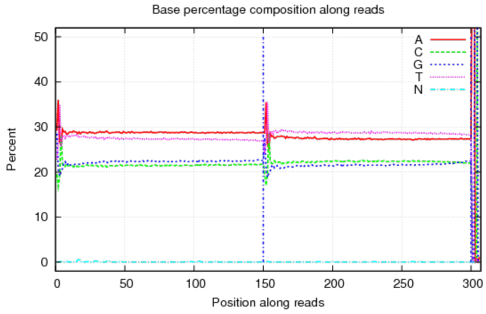


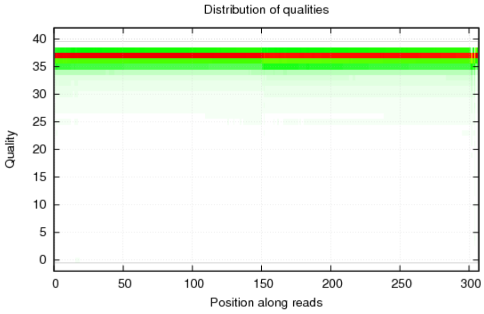
Quality control of sample d



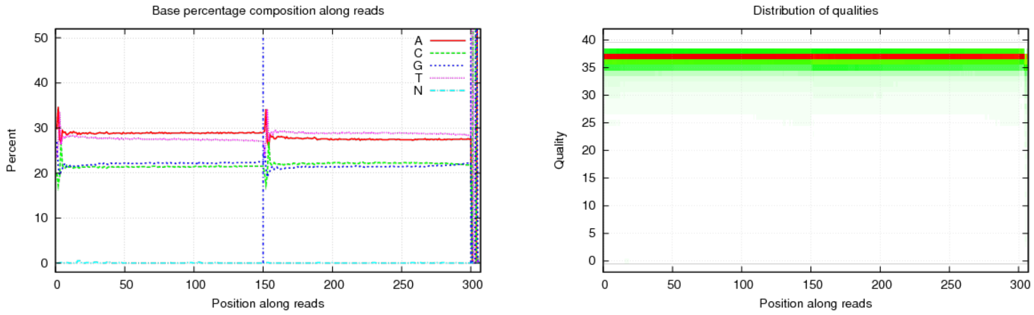


Quality control of sample dMcphC1

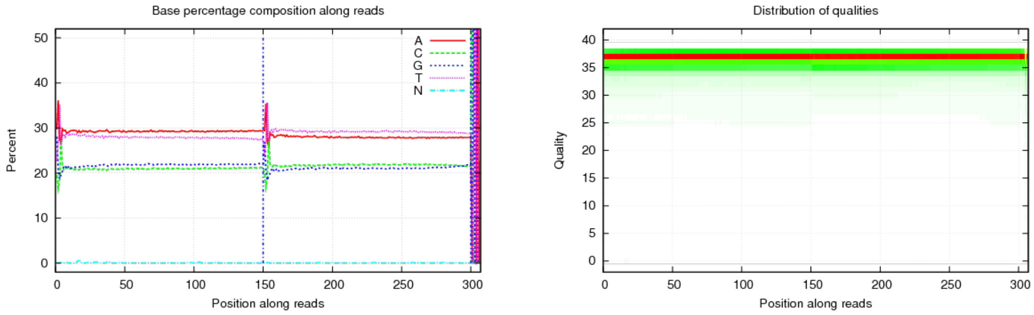




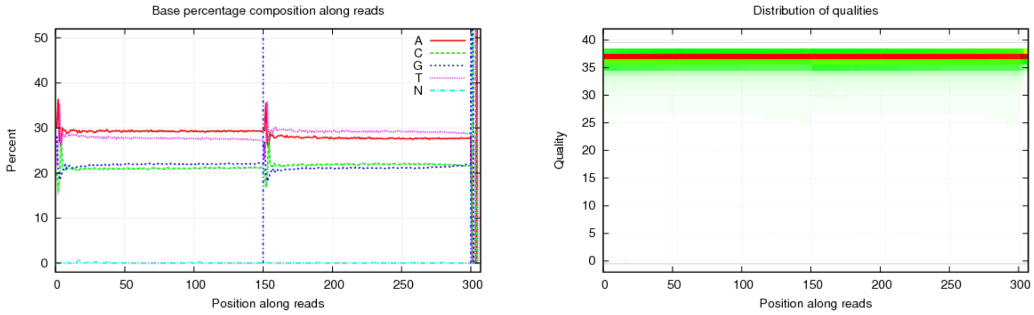
Quality control of sample dMcphC2



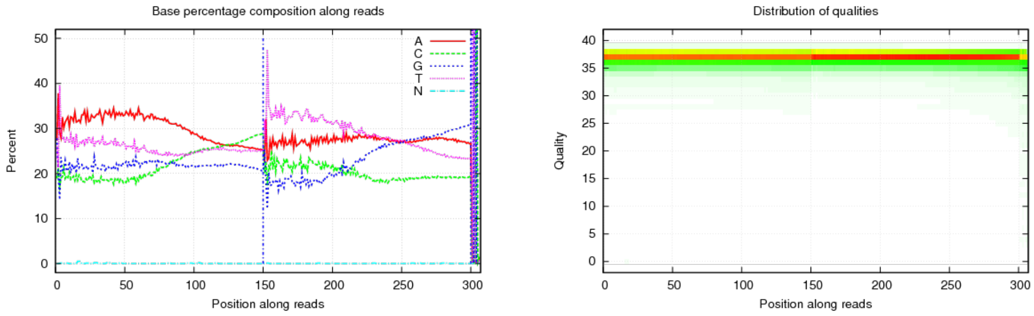
Quality control of sample dMcphRA1



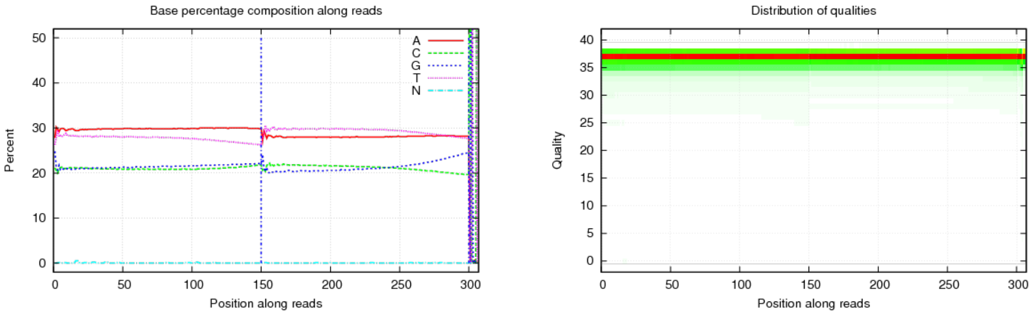
Quality control of sample dMcphRA2



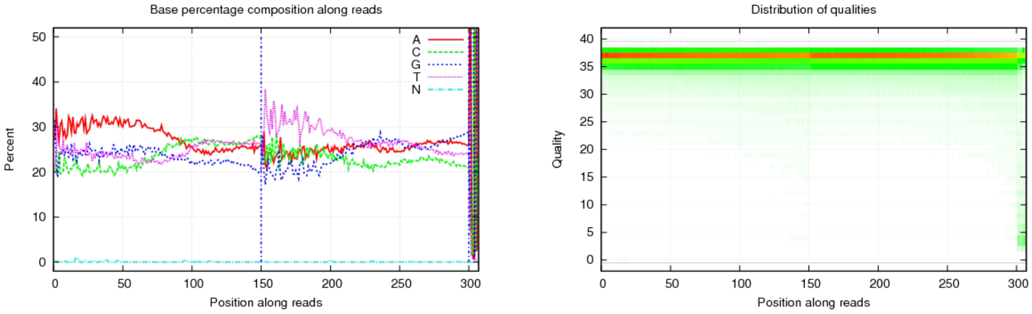
Quality control of sample e



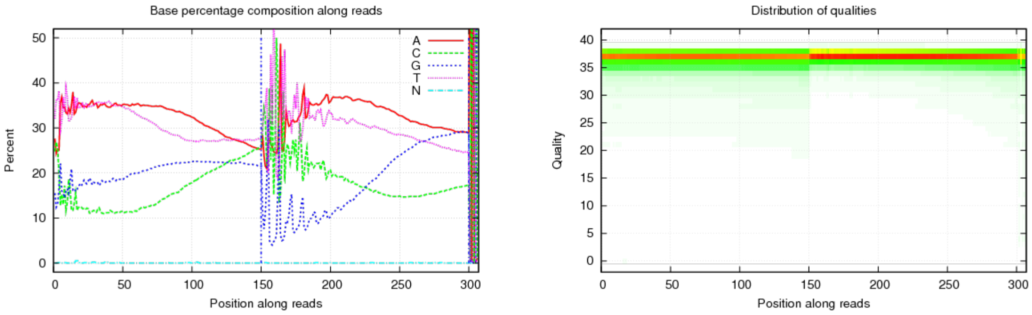
Quality control of sample f



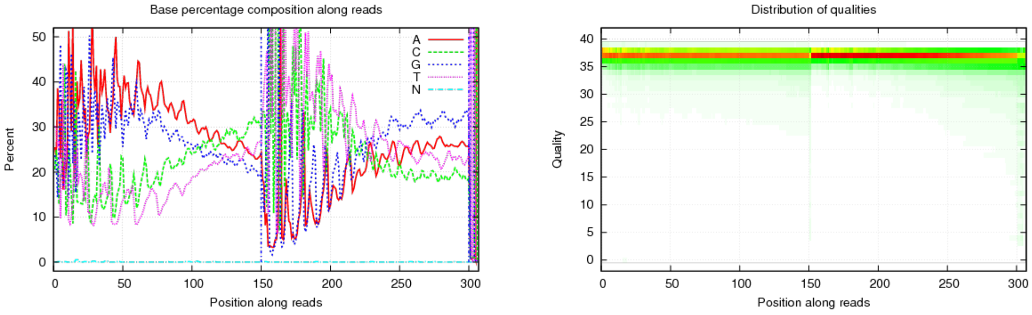
Quality control of sample GFP-Pei



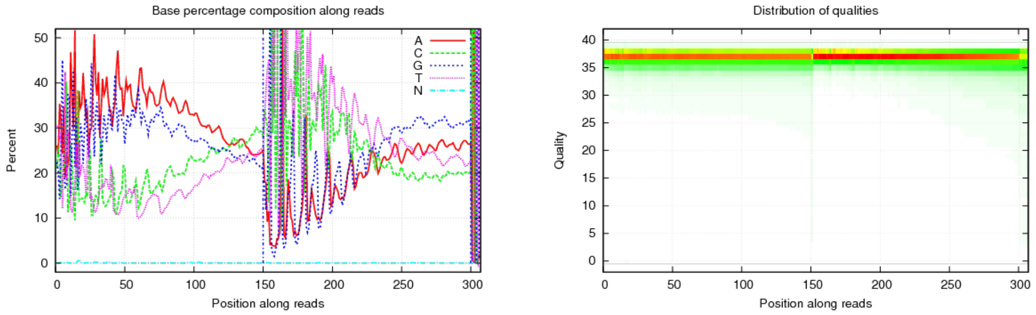
Quality control of sample MS22



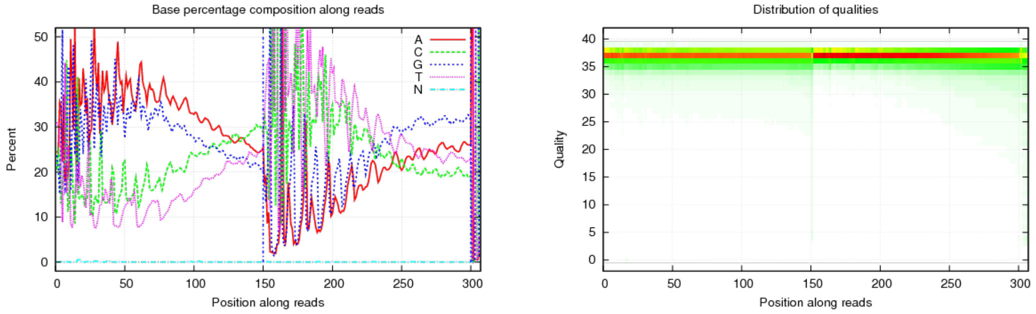
Quality control of sample MS23



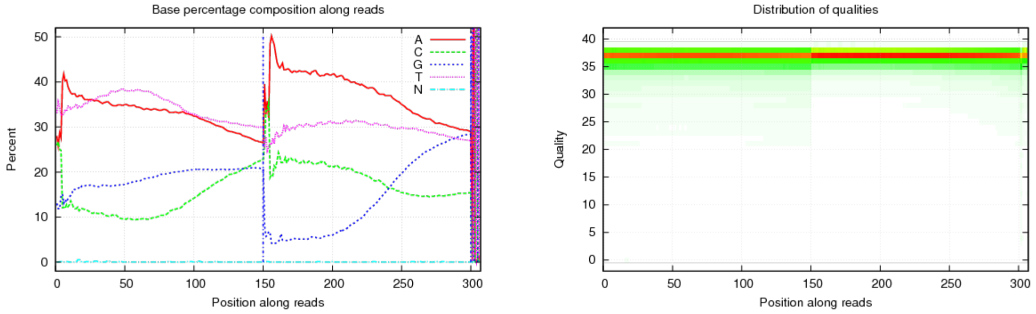
Quality control of sample MS24



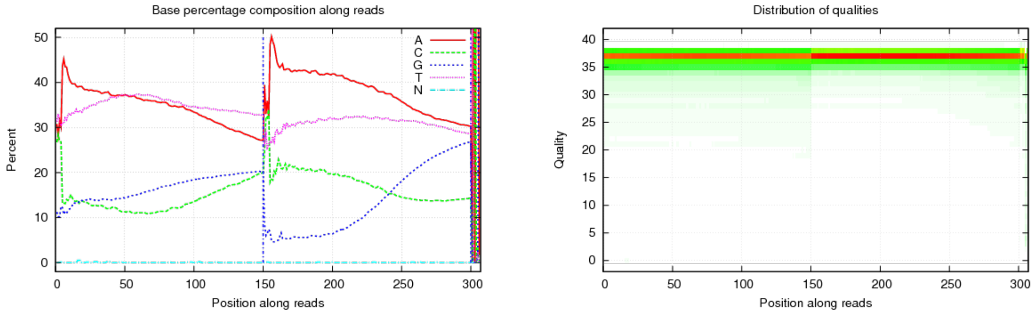
Quality control of sample MS25



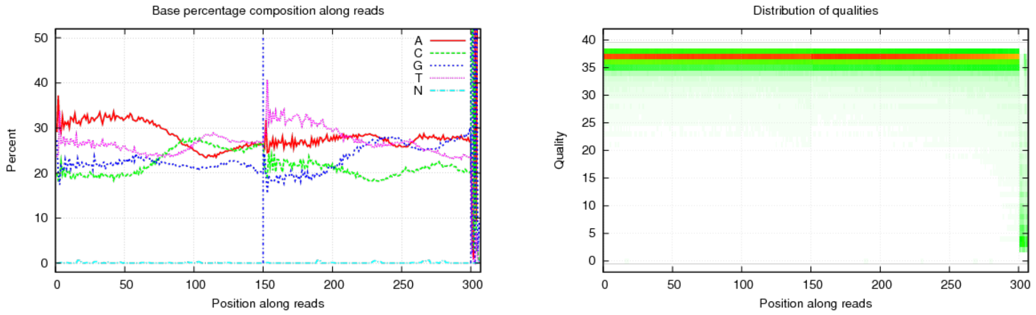
Quality control of sample MS26



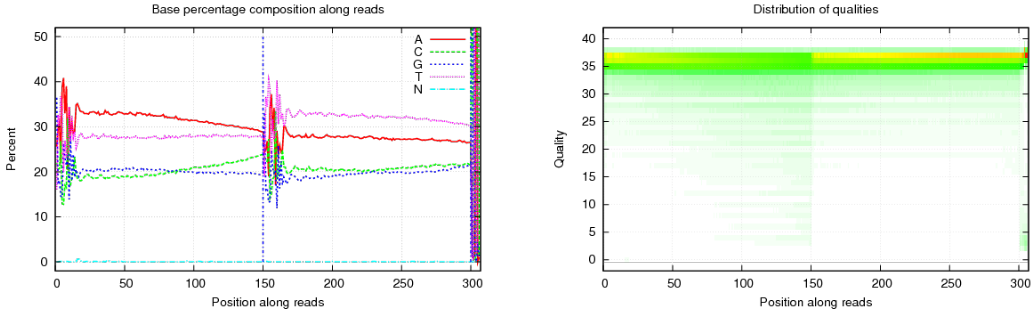
Quality control of sample MS27



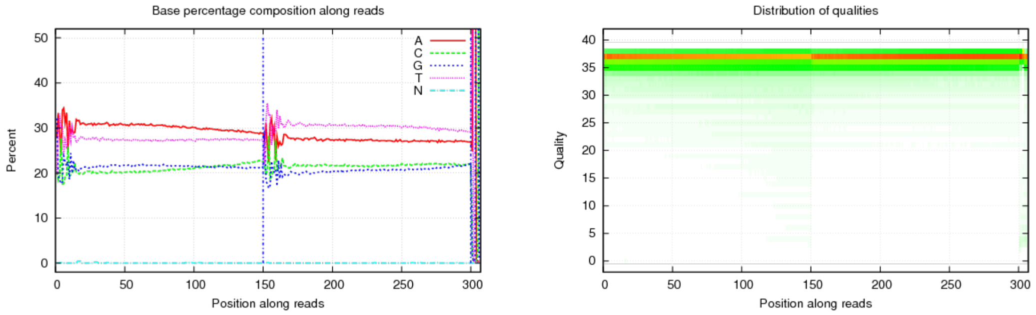
Quality control of sample Tn5-EC



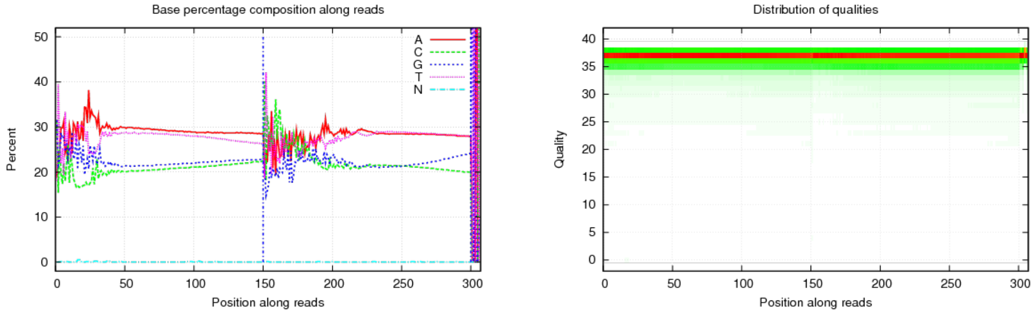
Quality control of sample Tn5-EP



Quality control of sample Tn5-KP



Quality control of sample Tn5-L



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	?