BGI Pre-made Library Sequencing Report

2022/11/7

Table of Contents

1 Project Information

Project code: F22FTSEUHT1331_MUSfhkyR
Sample number: 20

2 Data Statistics

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

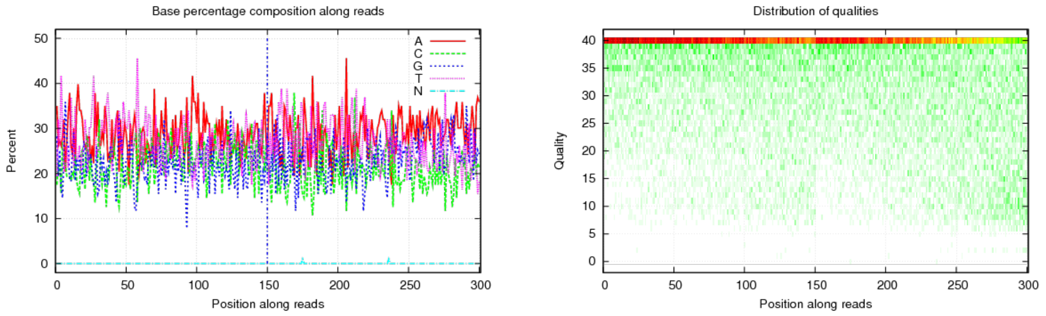
Sample	Length	Q20(%)	Q30(%)	GC Content(%)	Total Reads	Total Bases
CB_wt_GGCTACCA	150;150	91.14;84.39	77.46;66.79	44.49;44.52	103	30,900
CB_wt_TAGCTTCA	150;150	93.92;89.00	84.57;76.38	45.30;44.98	320	96,000
iN1-25_r1	150;150	97.51;96.08	91.98;89.46	42.95;42.98	19,321,227	5,796,368,100
iN_1-36_r1	150;150	97.65;96.54	92.45;90.86	43.08;43.14	20,682,542	6,204,762,600
iN_1-36_r2	150;150	96.95;93.72	90.79;84.37	49.28;49.40	6,039,239	1,811,771,700
iN_14-10_r1	150;150	96.52;93.85	89.33;83.91	46.94;47.10	36,233,465	10,870,039,500
iN_4-14_r1	150;150	97.70;96.27	92.57;90.05	43.62;43.66	20,560,992	6,168,297,600
iN_4-22_Cre2_r1	150;150	97.61;95.93	92.46;89.29	45.54;45.66	7,725,643	2,317,692,900
iN_4-22_Cre2_r2	150;150	97.68;95.48	92.49;87.94	45.52;45.57	10,455,490	3,136,647,000
iN_4-22_r1	150;150	97.67;95.86	92.40;88.63	44.22;44.33	13,496,878	4,049,063,400
iN_4-22_r2	150;150	97.75;95.88	92.74;88.84	44.97;45.04	10,487,016	3,146,104,800
input_iN1-25_r1	150;150	97.90;96.62	93.28;91.19	43.46;43.54	34,108,783	10,232,634,900
input_iN_1-25_r2	150;150	97.87;96.79	93.33;91.76	44.18;44.24	30,497,645	9,149,293,500
input_iN_1-36_r1	150;150	97.88;96.87	93.26;91.81	42.64;42.74	40,111,768	12,033,530,400
input_iN_1-36_r2	150;150	97.17;94.53	90.81;84.45	45.82;45.90	3,141,134	942,340,200
input_iN_14-10_r1	150;150	96.87;93.63	90.03;82.16	47.06;47.18	3,508,536	1,052,560,800
input_iN_14-10_r2	150;150	97.51;94.99	92.33;87.37	47.31;47.35	15,074,566	4,522,369,800
input_iN_14-8_r1	150;150	97.63;95.52	92.55;88.25	45.90;45.99	10,885,014	3,265,504,200
input_iN_4-22_Cre2_r1	150;150	97.72;96.40	92.66;90.49	42.83;42.93	14,956,920	4,487,076,000
input_iN_4-22_r1	150;150	98.01;96.54	93.68;91.00	42.60;42.61	31,549,349	9,464,804,700

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

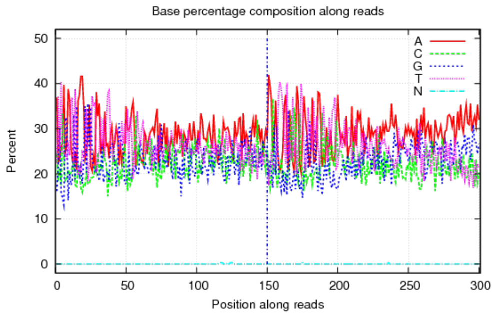
3 Data Quality Control

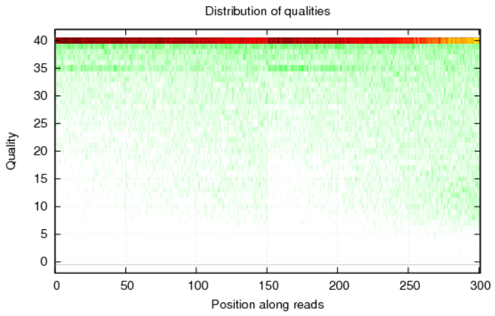
The distribution of base percentage and qualities along reads in data filtering are shown as following(If a sample has multiple lanes, only one of them will be displayed). The left picture is base percentage distribution along reads the sample, the right picture is distribution of qualities along reads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

Quality control of sample CB_wt_GGCTACCA

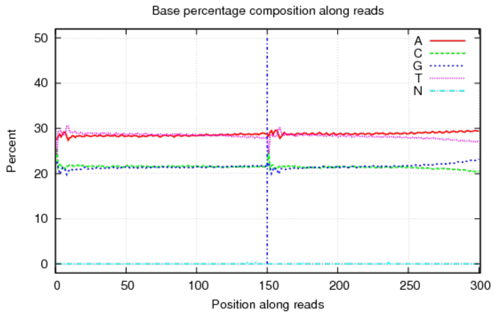


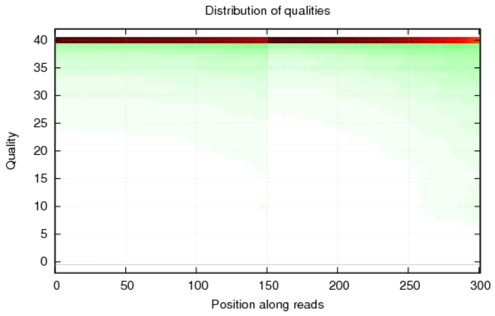
Quality control of sample CB_wt_TAGCTTCA


Base percentage composition along reads

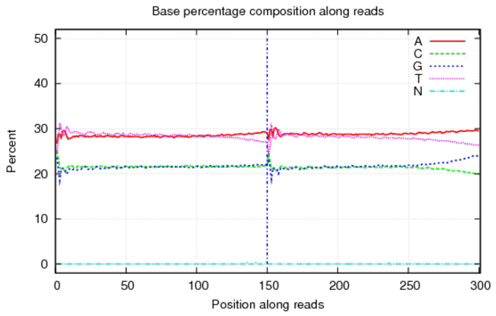

Distribution of qualities

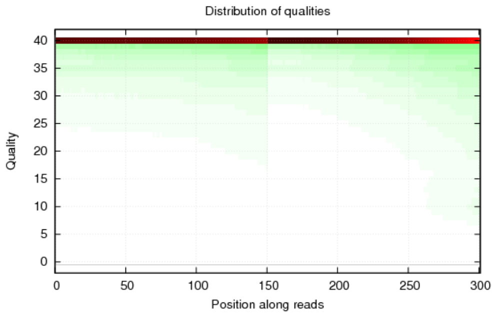
Quality control of sample iN1-25_r1


Base percentage composition along reads

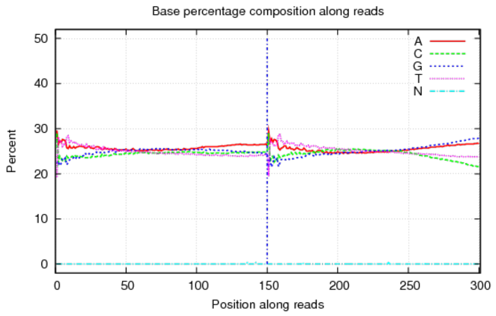

Distribution of qualities

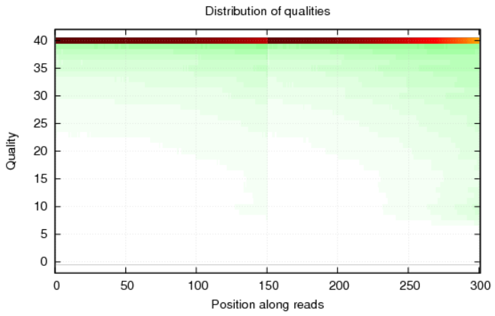
Quality control of sample iN_1-36_r1


Base percentage composition along reads

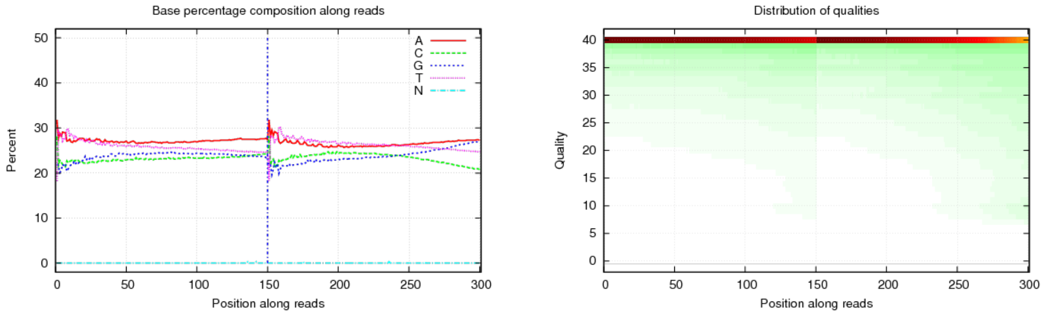

Distribution of qualities

Quality control of sample iN_1-36_r2

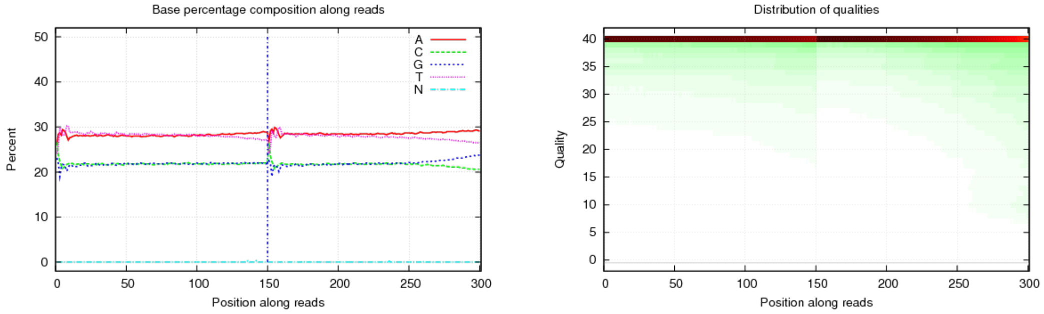

Base percentage composition along reads


Distribution of qualities

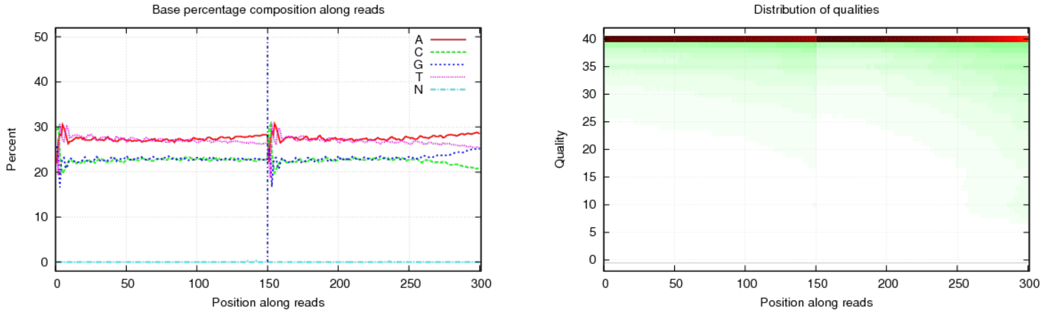
Quality control of sample iN_14-10_r1



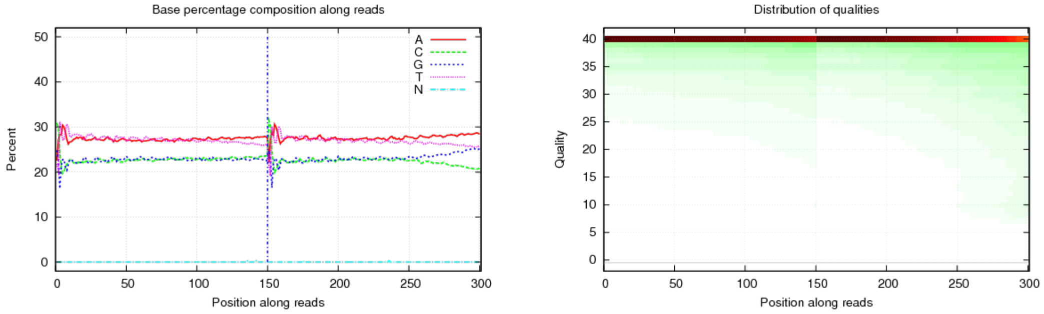
Quality control of sample iN_4-14_r1



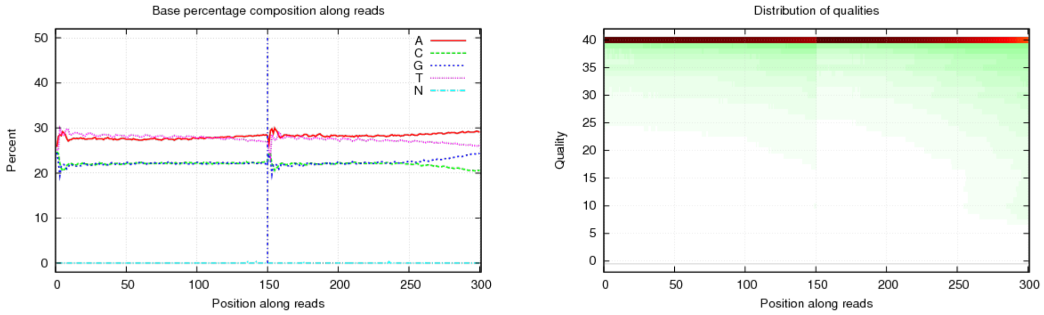
Quality control of sample iN_4-22_Cre2_r1



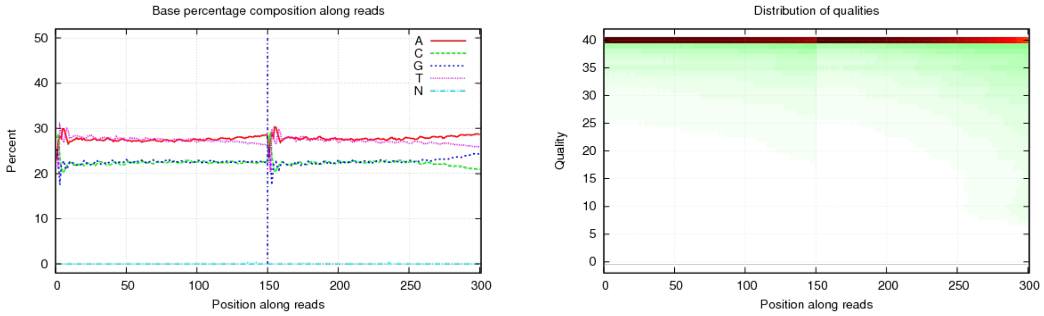
Quality control of sample iN_4-22_Cre2_r2



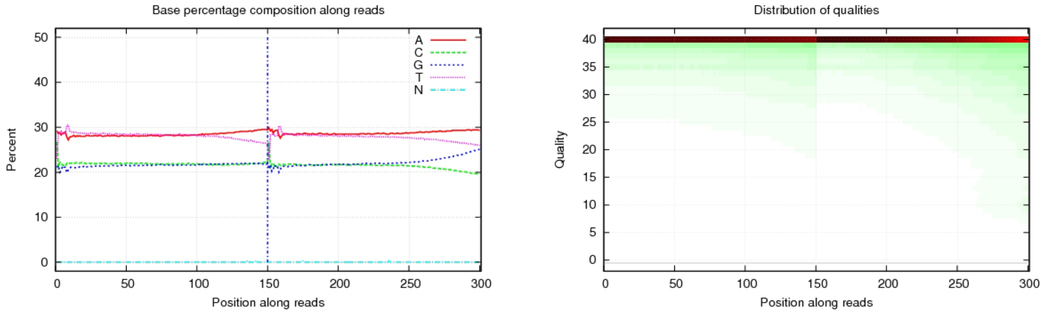
Quality control of sample iN_4-22_r1



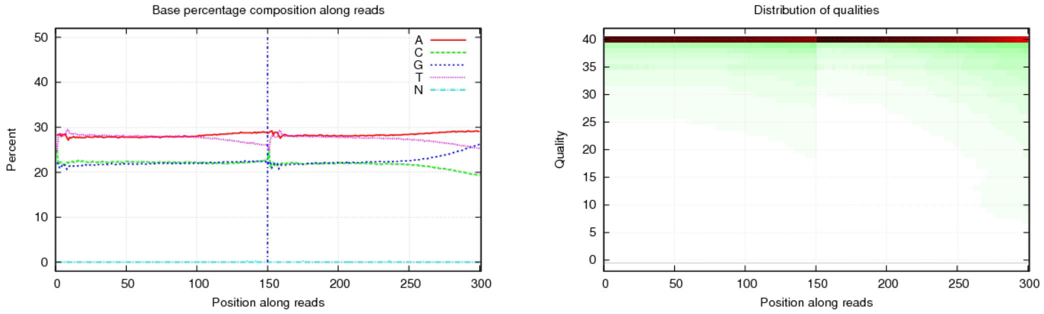
Quality control of sample iN_4-22_r2



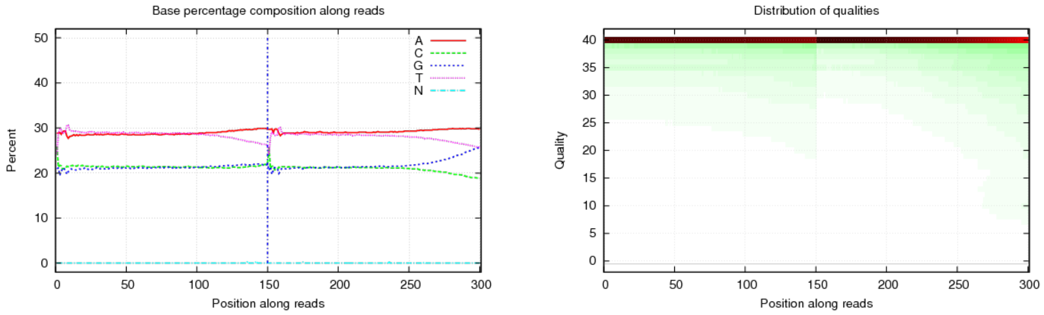
Quality control of sample input_iN1-25_r1



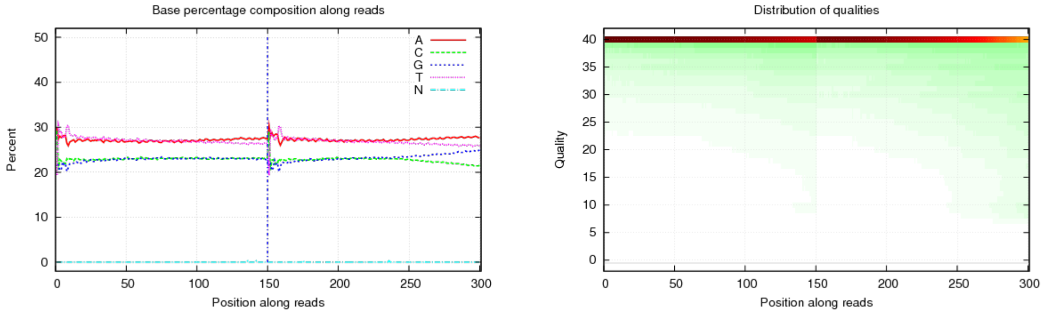
Quality control of sample input_iN_1-25_r2



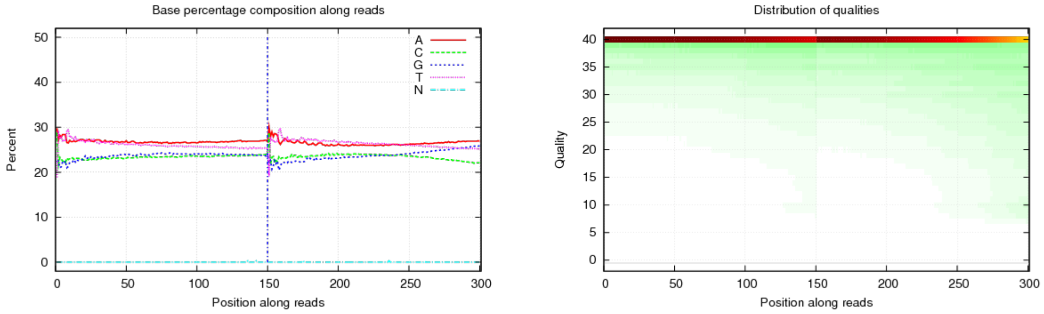
Quality control of sample input_iN_1-36_r1



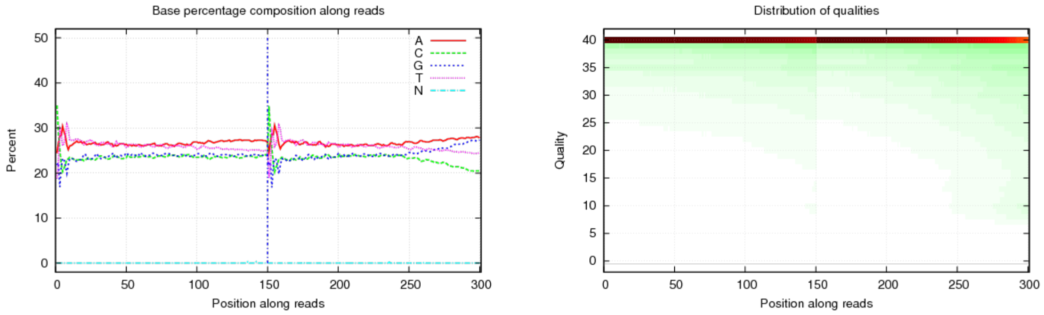
Quality control of sample input_iN_1-36_r2



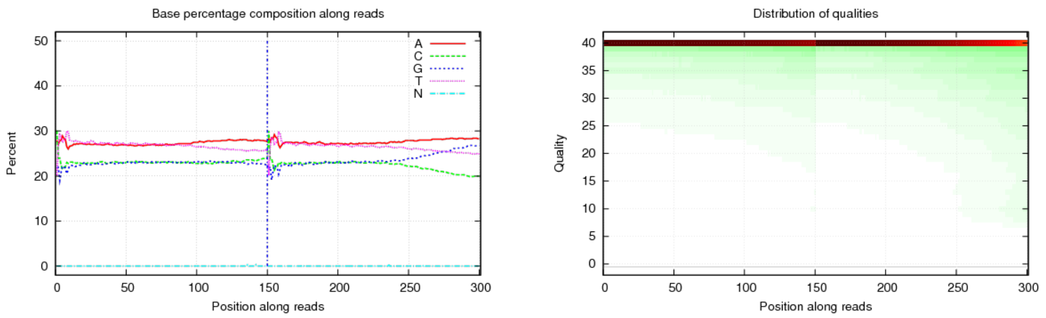
Quality control of sample input_iN_14-10_r1



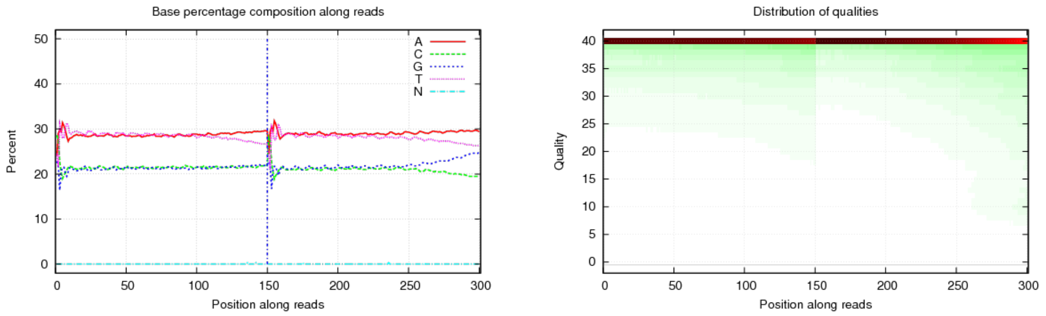
Quality control of sample input_iN_14-10_r2



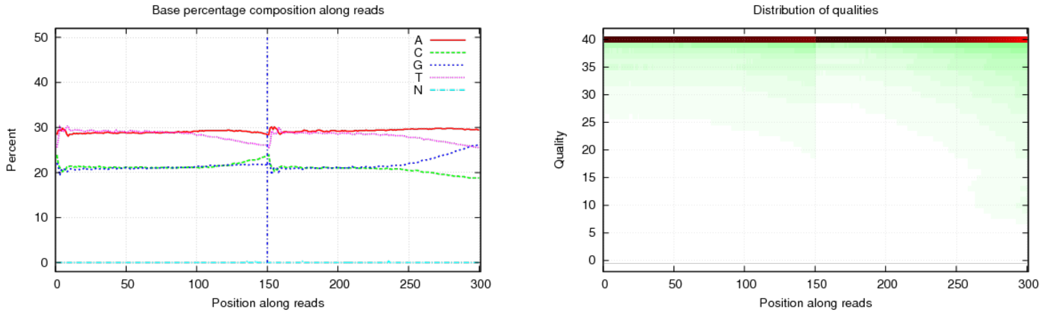
Quality control of sample input_iN_14-8_r1



Quality control of sample input_iN_4-22_Cre2_r1



Quality control of sample input_iN_4-22_r1



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	?