BGI Sequencing Data Report

2023/9/21

Table of Contents

1 Project Information

Project code: F22FTSEUHT2056_MUSuyisR
Sample number: 30

2 Data Statistics

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

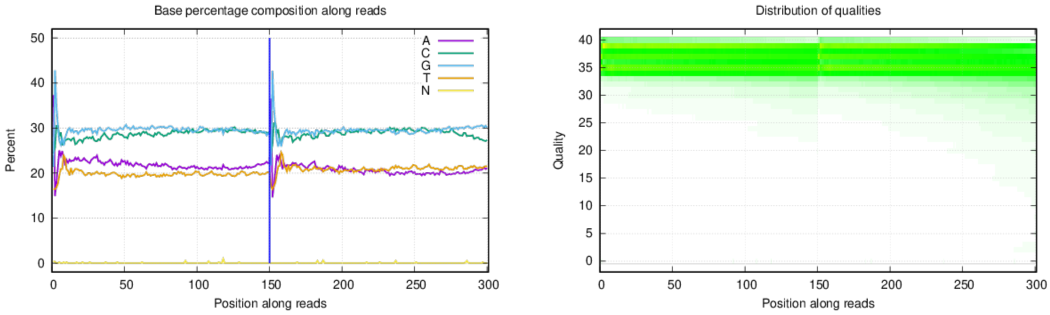
Sample	Length	Q20(%)	Q30(%)	GC Content(%)	Total Reads	Total Bases
Chal_K1	150;150	98.26;97.79	94.43;92.85	45.39;45.42	128,719	38,615,700
Chal_K2	150;150	98.09;97.20	93.79;90.91	42.89;42.86	77,494	23,248,200
Chal_e	150;150	97.91;97.89	93.33;93.05	41.20;41.26	168,191	50,457,300
Dik_K2	150;150	98.14;97.14	93.98;90.87	44.71;44.72	70,538	21,161,400
Dik_e	150;150	98.16;97.33	94.05;91.37	43.64;43.65	105,661	31,698,300
Droz_2_K	150;150	98.19;97.99	94.17;93.37	41.07;41.14	226,791	68,037,300
Kaz3_K	150;150	98.23;97.20	94.22;90.97	43.85;43.79	5,167,769	1,550,330,700
Mat2_K	150;150	98.26;97.22	94.32;91.23	48.46;48.39	6,683,703	2,005,110,900
Mel6_K	150;150	98.31;97.51	94.49;91.92	44.77;44.76	5,827,546	1,748,263,800
Mog1_e	150;150	98.31;97.54	94.56;92.15	47.57;47.61	3,673,548	1,102,064,400
Pct_1	150;150	98.12;97.85	93.91;92.96	42.95;43.51	44,340	13,302,000
Pct_2	150;150	97.96;97.59	93.40;92.17	42.86;43.48	40,293	12,087,900
Sach2_K	150;150	98.40;98.02	94.85;93.61	44.22;44.35	6,166,425	1,849,927,500
Vla1_e	150;150	97.93;96.93	93.24;90.07	42.05;42.10	5,661,042	1,698,312,600
Vla2_e	150;150	98.35;97.48	94.64;91.92	47.25;47.36	5,232,144	1,569,643,200
cen1	150;150	98.12;96.60	94.13;89.74	58.30;58.36	21,624,951	6,487,485,300
cen2	150;150	98.48;98.35	95.04;94.52	46.17;46.17	28,865,458	8,659,637,400
cen3	150;150	98.43;98.58	94.95;95.27	46.54;46.54	32,842,779	9,852,833,700
cen4	150;150	98.63;98.70	95.54;95.66	46.56;46.58	35,350,309	10,605,092,700
cen5	150;150	98.40;98.26	94.80;94.26	46.45;46.45	23,304,626	6,991,387,800
cenR1	150;150	98.54;98.50	95.27;95.08	45.87;45.88	34,321,442	10,296,432,600
cenR2	150;150	98.54;98.18	95.21;94.08	45.47;45.49	34,235,587	10,270,676,100
cenit1	150;150	98.60;98.67	95.43;95.68	45.95;46.01	38,550,933	11,565,279,900
chip_U87-dax_1	150;150	98.25;98.11	94.45;93.93	46.27;46.31	31,009,853	9,302,955,900
chip_U87-dax_2	150;150	97.60;97.97	92.59;93.54	45.98;45.93	19,745,724	5,923,717,200
input_U87-dax_1	150;150	98.45;98.57	95.14;95.39	44.63;44.68	14,579,467	4,373,840,100
input_U87-dax_2	150;150	98.60;98.75	95.57;95.91	44.04;44.08	2,846,445	853,933,500
input_h3k27ac_1	150;150	92.67;74.00	73.33;39.33	63.33;46.00	1	300
input_h3k27ac_2	150;150	98.00;98.52	93.72;95.14	42.67;42.68	46,009,157	13,802,747,100
kit30+_h3k27ac_2	150;150	98.28;96.64	94.61;90.00	53.77;53.73	20,889,304	6,266,791,200

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

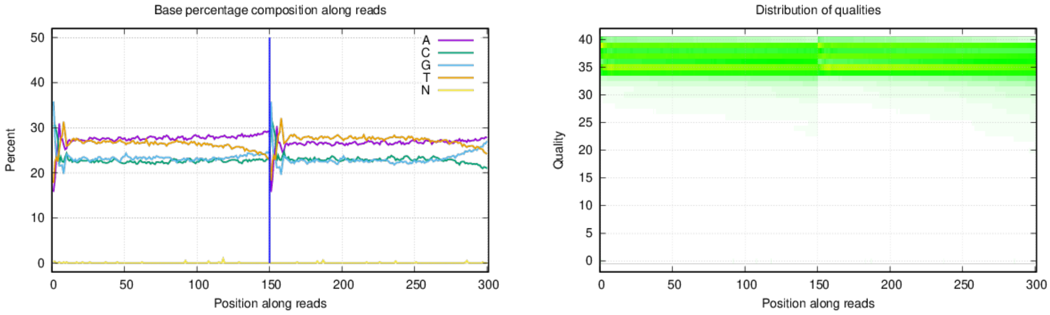
3 Data Quality Control

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.

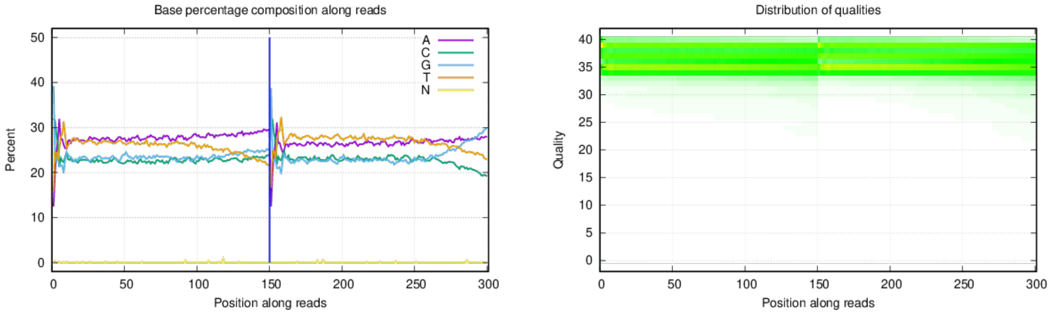
Quality control of sample cen1



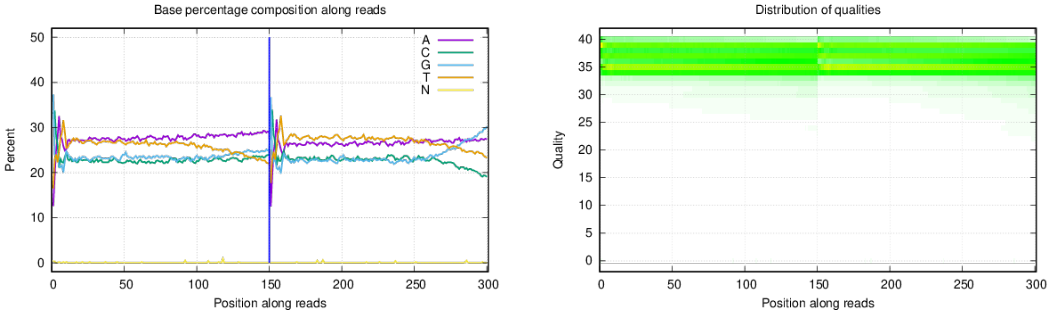
Quality control of sample cen2



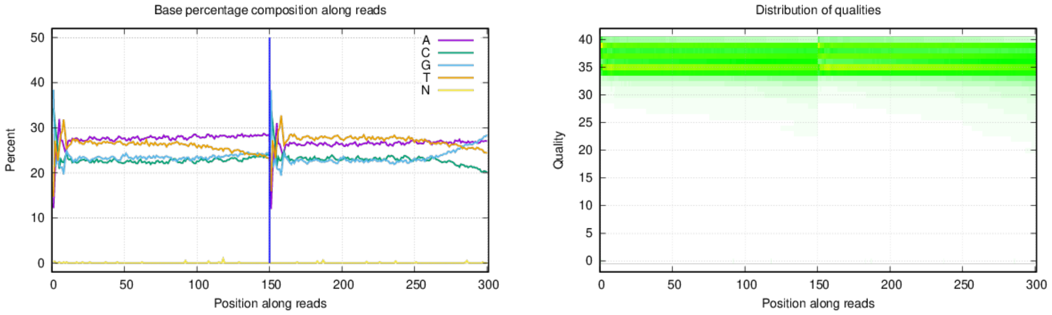
Quality control of sample cen3



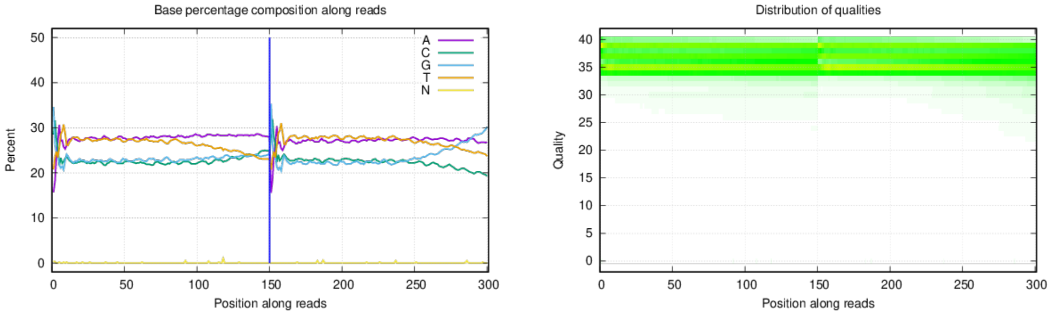
Quality control of sample cen4



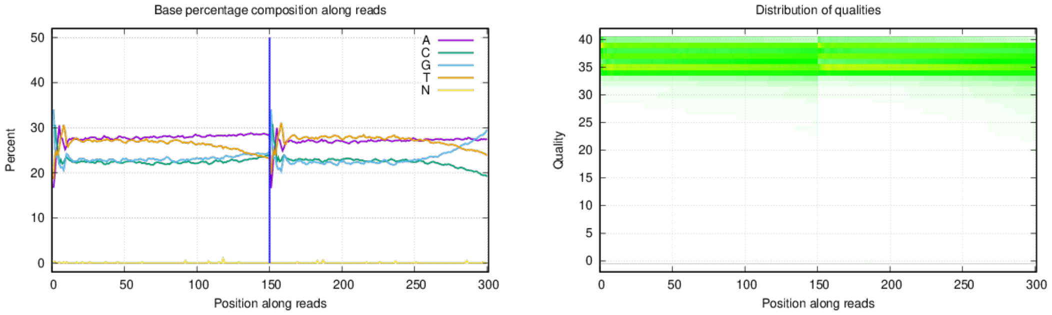
Quality control of sample cen5



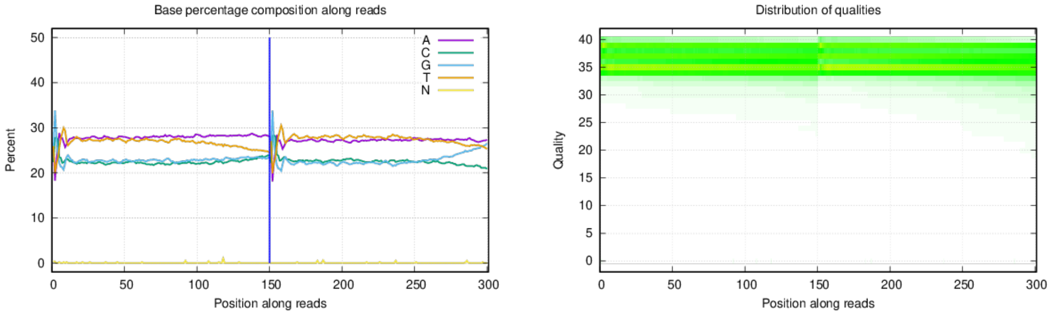
Quality control of sample cenit1



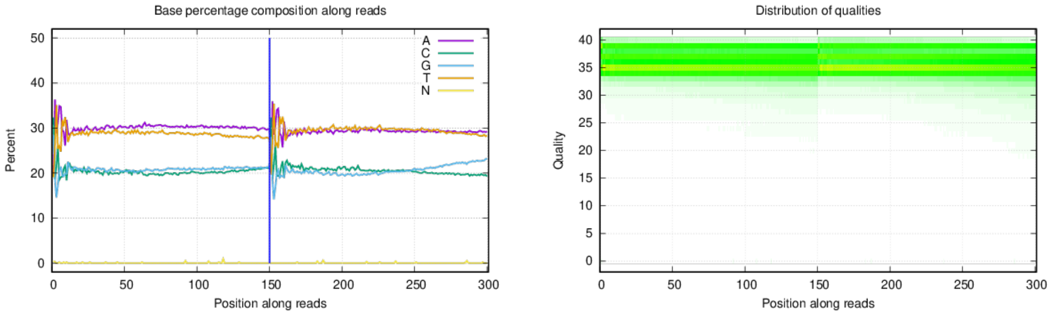
Quality control of sample cenR1



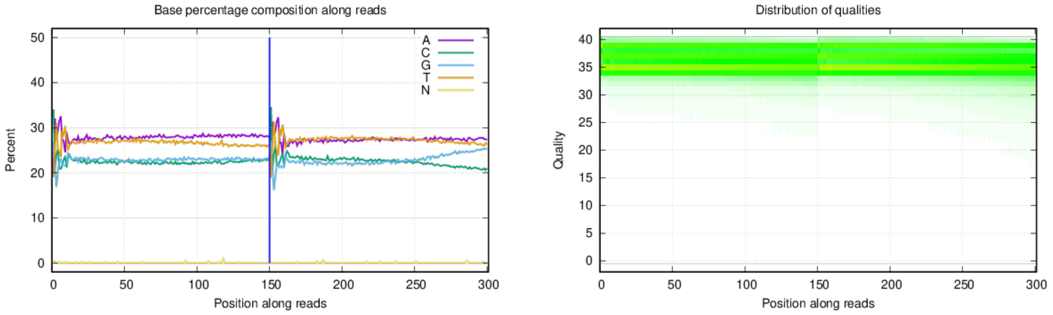
Quality control of sample cenR2



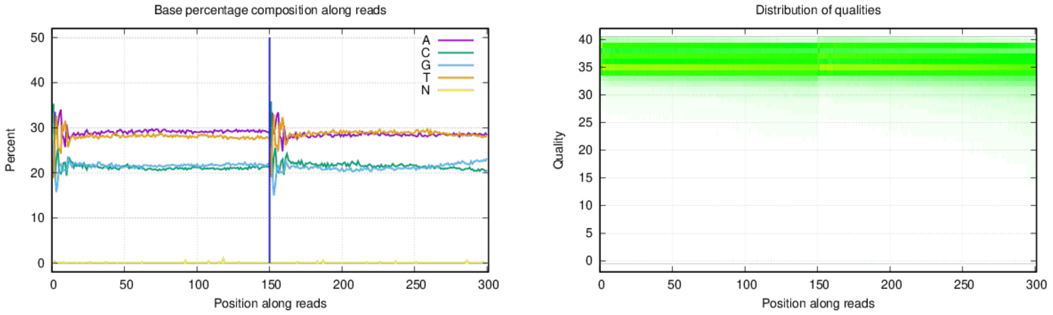
Quality control of sample Chal_e



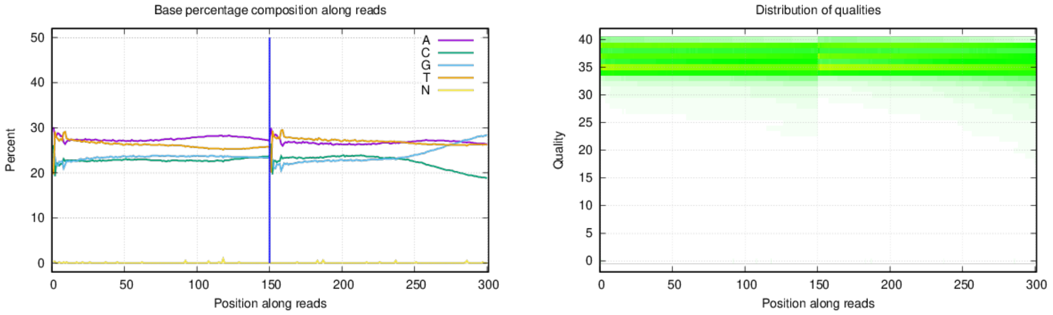
Quality control of sample Chal_K1



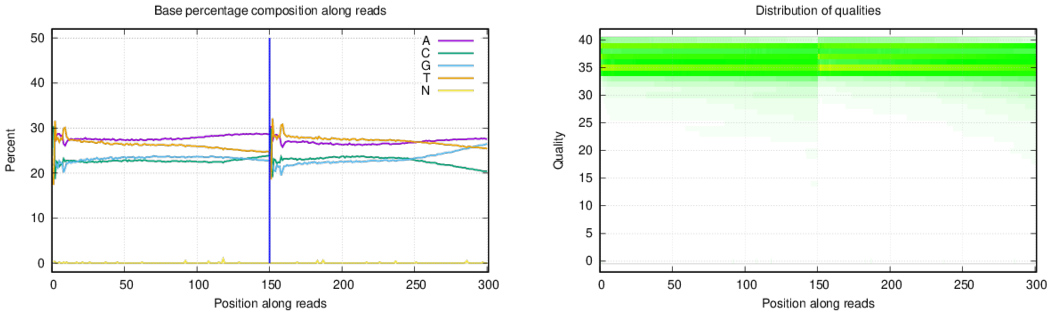
Quality control of sample Chal_K2



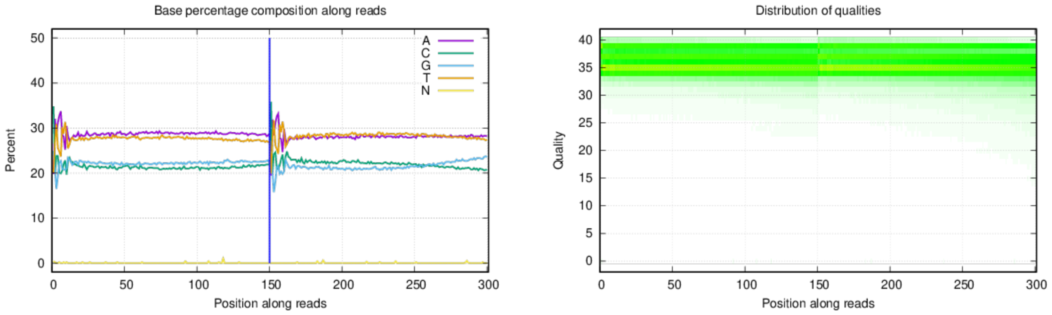
Quality control of sample chip_U87-dax_1



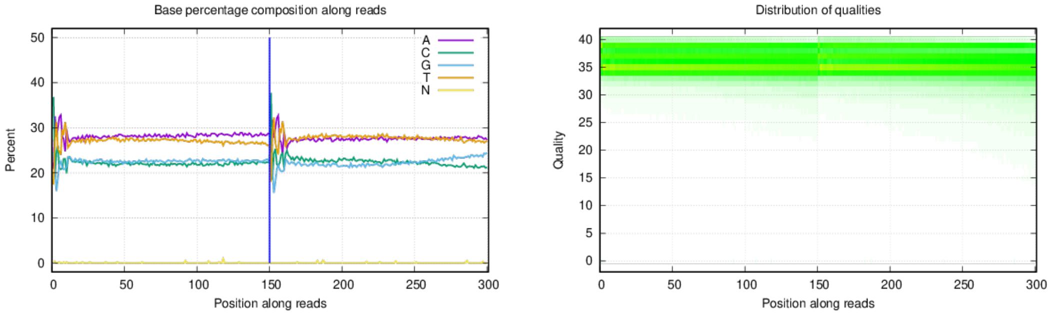
Quality control of sample chip_U87-dax_2



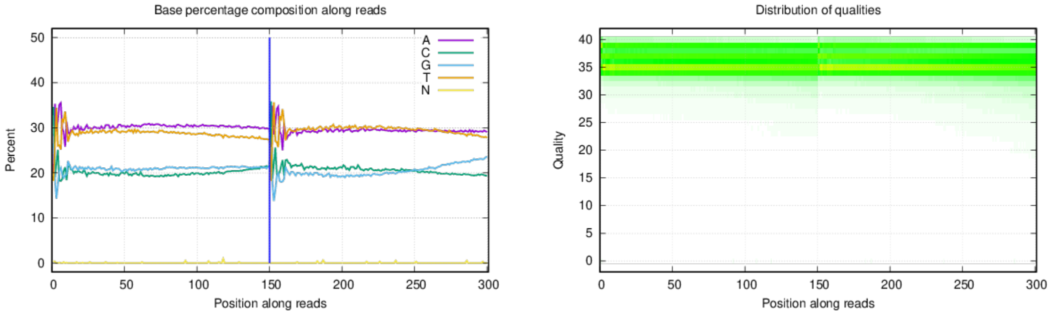
Quality control of sample Dik_e



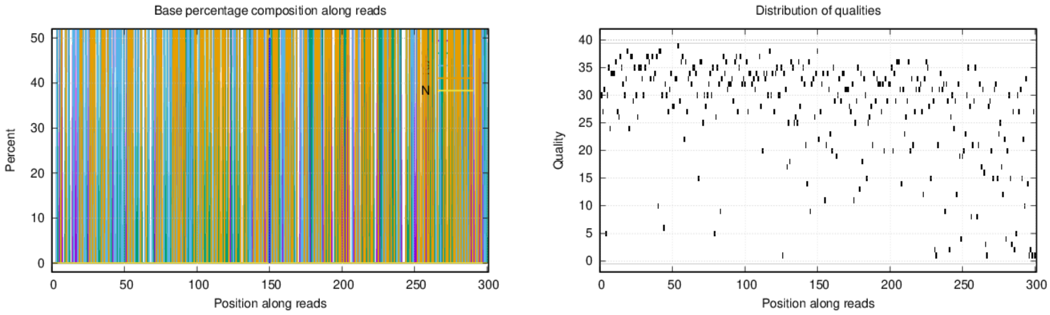
Quality control of sample Dik_K2



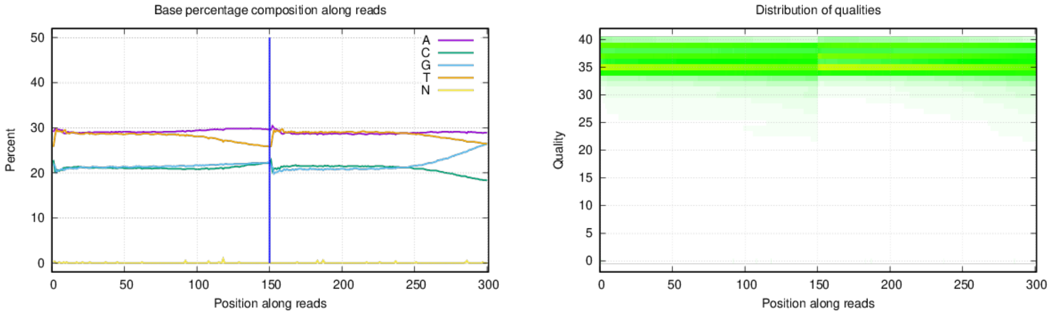
Quality control of sample Droz_2_K



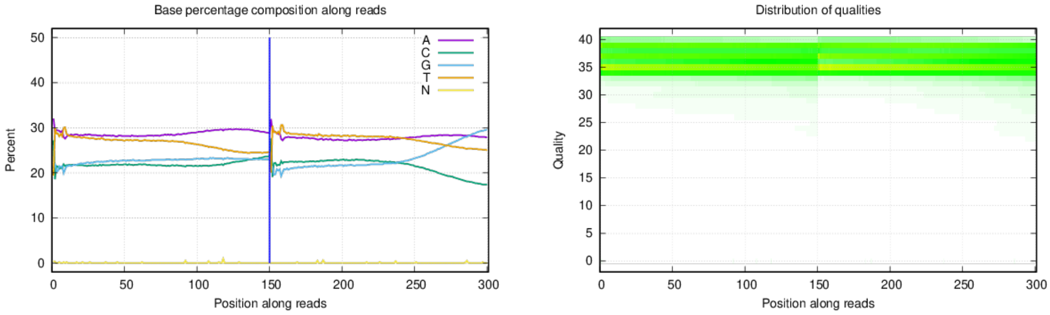
Quality control of sample input_h3k27ac_1



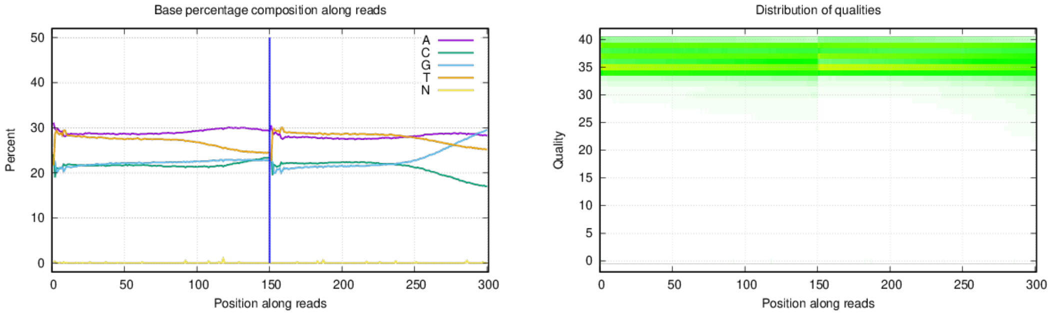
Quality control of sample input_h3k27ac_2



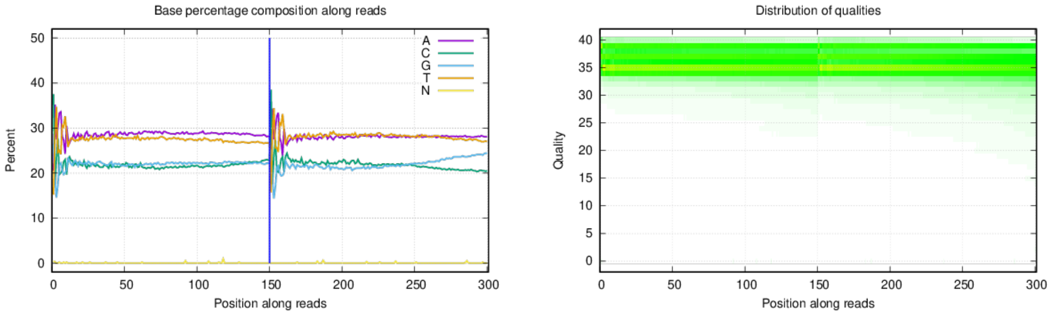
Quality control of sample input_U87-dax_1



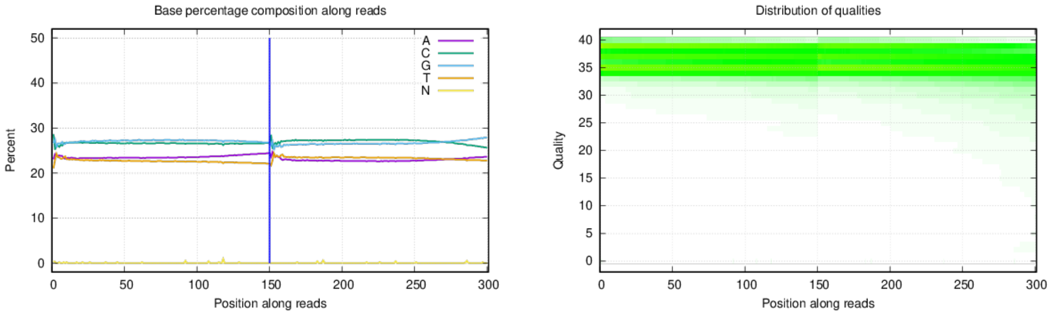
Quality control of sample input_U87-dax_2



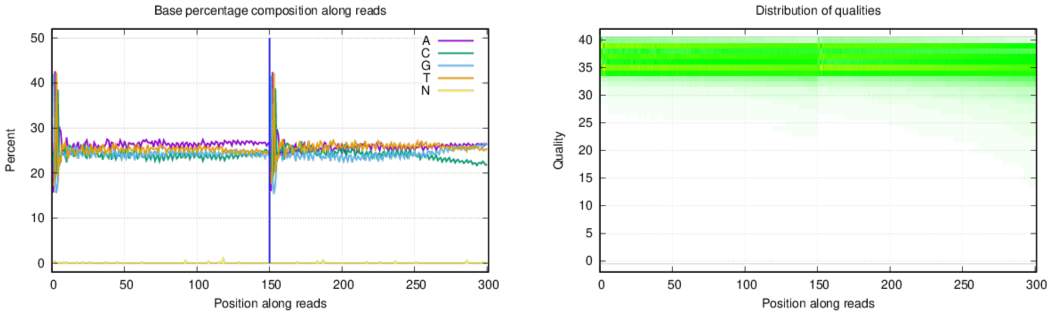
Quality control of sample Kaz3_K



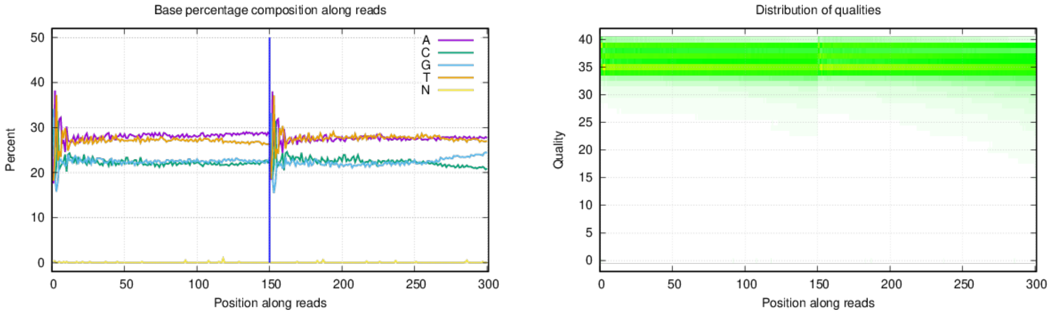
Quality control of sample kit30+_h3k27ac_2



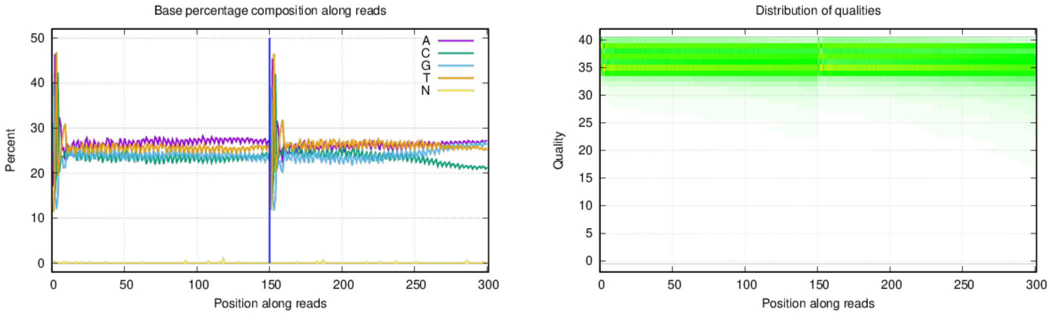
Quality control of sample Mat2_K



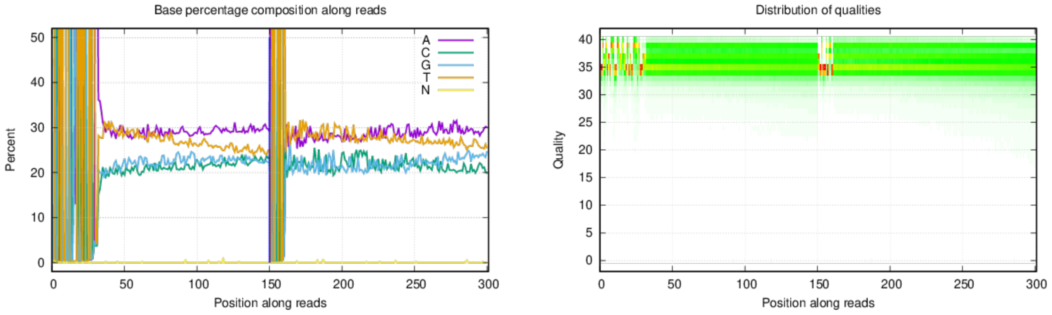
Quality control of sample Mel6_K



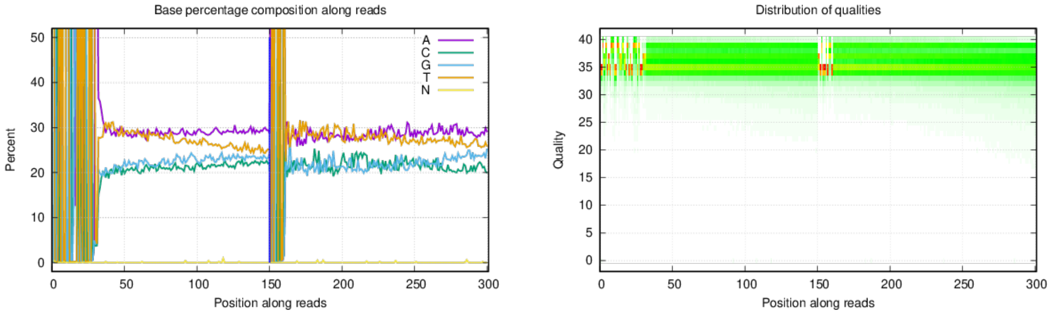
Quality control of sample Mog1_e



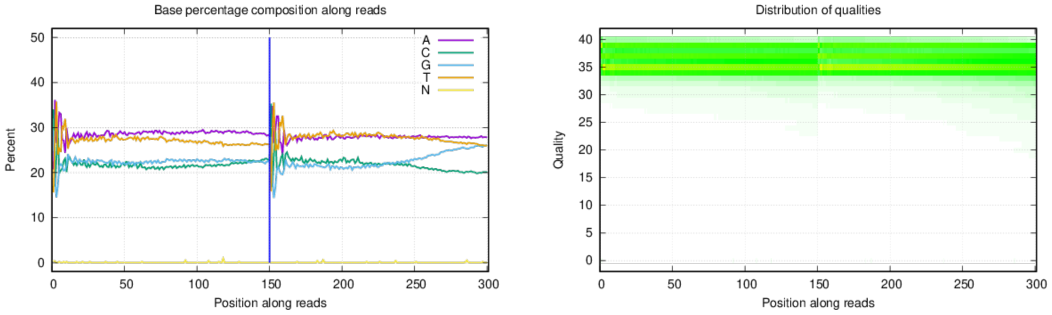
Quality control of sample Pct_1



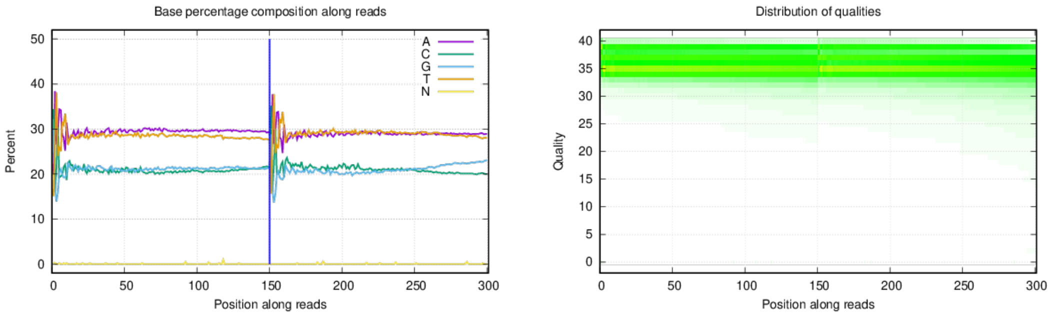
Quality control of sample Pct_2



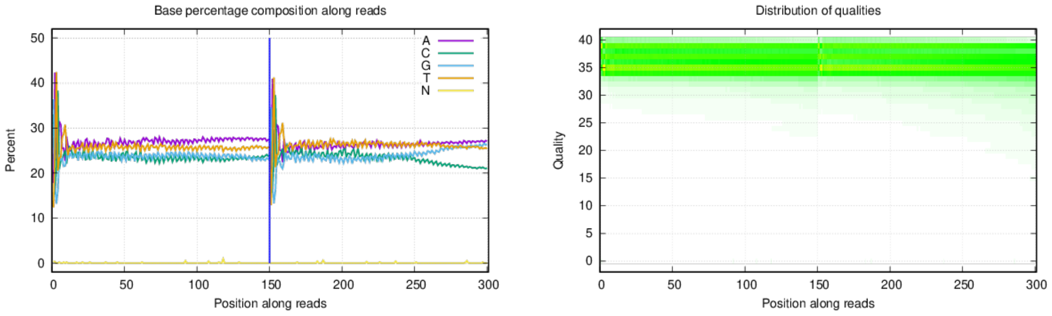
Quality control of sample Sach2_K



Quality control of sample Vla1_e



Quality control of sample Vla2_e



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
+
FFFCFFFFFFFFFFDFEFFFFFEFEF0FFFFEFFFFFFFFEFFFFFECGFFFF
```

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	?