Sequencing Data Report

Project: F23A430000555_MUSducfR
Date: 2024.10.22
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 43 samples in this project, the statistics of fastq data is shown below.

Sample	Length	Q20(%)	Q30(%)	GC Content(%)	Total Reads	Total Bases
1_25_R1	150;150	97.51;97.09	92.16;90.53	45.78;45.72	1,118,866	335,659,800
1_25_R2	150;150	97.03;97.14	91.00;90.74	46.86;47.35	287,428	86,228,400
4C_Ph_MEF_1_fixed22	150;150	98.15;95.93	93.30;85.86	47.19;47.54	54,311	16,293,300
4C_Ph_MEF_2_fixed1	150;150	98.12;96.05	93.28;86.03	44.83;45.50	339,594	101,878,200
4C_Ph_MEF_3_fixed1	150;150	97.87;95.79	92.45;85.08	44.72;45.25	309,937	92,981,100
4C_Ph_MEF_4_fixed2	150;150	98.01;95.60	92.88;84.82	47.73;48.27	59,770	17,931,000
4C_Ph_MEF_5_fixed2	150;150	98.16;95.92	93.39;85.83	47.47;47.94	120,929	36,278,700
4_14_D_R1	150;150	97.54;96.98	92.36;90.20	45.75;45.69	677,593	203,277,900
4_14_D_R2	150;150	97.98;97.36	93.39;91.13	45.73;45.74	3,675,420	1,102,626,000
4_14_R1	150;150	97.82;97.44	93.03;91.56	44.80;44.70	3,547,634	1,064,290,200
4_14_R2	150;150	97.72;97.25	92.66;90.88	44.72;44.79	1,995,855	598,756,500
4_22_Ch_R1	150;150	97.86;96.95	93.12;90.10	46.03;45.80	1,473,099	441,929,700
4_22_Ch_R2	150;150	97.80;97.35	92.87;91.14	44.17;44.33	1,975,717	592,715,100
4_22_D_R1	150;150	97.83;97.47	93.20;91.66	47.24;47.47	454,001	136,200,300
4_22_D_R2	150;150	97.71;97.56	92.69;91.80	46.65;46.85	1,116,471	334,941,300
4_22_R1	150;150	97.55;97.23	92.31;90.92	46.02;46.09	831,703	249,510,900
4_22_R2	150;150	96.69;96.89	89.97;89.77	45.40;45.60	214,433	64,329,900
4_22_U_R1	150;150	97.57;97.01	92.35;90.18	45.21;45.36	1,678,501	503,550,300
4_22_U_R2	150;150	97.41;97.13	91.82;90.57	45.26;45.44	900,792	270,237,600
Bp	150;150	98.04;97.48	93.46;91.29	48.09;48.03	1,391,934	417,580,200
Cenp-B1	150;150	98.46;97.95	94.70;92.79	50.61;50.67	7,542,988	2,262,896,400
Cenp-B2	150;150	98.15;97.57	93.66;91.38	47.02;47.13	14,693,866	4,408,159,800
Cenp-P2	150;150	97.94;97.38	92.97;90.65	44.60;44.64	17,259,486	5,177,845,800
Ch_CTCF_hg_1	150;150	98.51;98.15	94.87;93.76	45.75;45.88	6,228,735	1,868,620,500
Ch_CTCF_hg_2	150;150	98.60;98.32	95.18;94.35	46.27;46.32	4,180,736	1,254,220,800
Ch_CTCF_hg_3	150;150	98.80;98.85	95.95;96.08	46.76;46.87	6,196,398	1,858,919,400

Ch_CTCF_hg_4	150;150	98.54;98.52	95.05;94.96	45.75;45.83	8,868,317	2,660,495,100
i_CTCF_hg_1	150;150	98.81;98.79	95.98;95.89	46.98;47.12	15,670,517	4,701,155,100
i_CTCF_hg_2	150;150	98.46;98.51	94.82;94.89	45.69;45.80	14,319,651	4,295,895,300
in1_25_R1	150;150	96.31;96.61	89.07;89.51	42.33;41.74	5,785,942	1,735,782,600
in1_25_R2	150;150	96.41;96.76	89.28;89.84	41.64;41.76	2,529,246	758,773,800
in4_14_D_R1	150;150	97.16;97.30	91.25;91.24	42.65;42.78	5,263,713	1,579,113,900
in4_14_D_R2	150;150	96.94;96.80	90.67;89.78	42.87;43.03	4,691,558	1,407,467,400
in4_14_R1	150;150	96.31;96.81	89.07;89.93	42.98;43.07	1,663,053	498,915,900
in4_14_R2	150;150	96.65;96.85	90.00;90.09	42.47;42.79	2,670,118	801,035,400
in4_22_Ch_R1	150;150	96.59;96.63	89.82;89.63	43.06;42.98	2,770,654	831,196,200
in4_22_Ch_R2	150;150	96.24;96.43	88.85;88.92	42.02;42.56	3,707,562	1,112,268,600
in4_22_D_R1	150;150	96.72;97.12	90.12;90.87	43.00;42.84	3,461,394	1,038,418,200
in4_22_D_R2	150;150	96.59;97.06	89.88;90.70	42.76;42.56	1,847,930	554,379,000
in4_22_R1	150;150	96.16;96.71	88.63;89.66	42.92;42.61	2,090,119	627,035,700
in4_22_R2	150;150	96.51;96.79	89.57;89.93	42.39;42.36	1,784,857	535,457,100
in4_22_U_R1	150;150	96.44;96.73	89.38;89.83	42.14;41.91	2,885,244	865,573,200
in4_22_U_R2	150;150	96.35;96.53	89.14;89.25	42.02;41.85	5,800,062	1,740,018,600

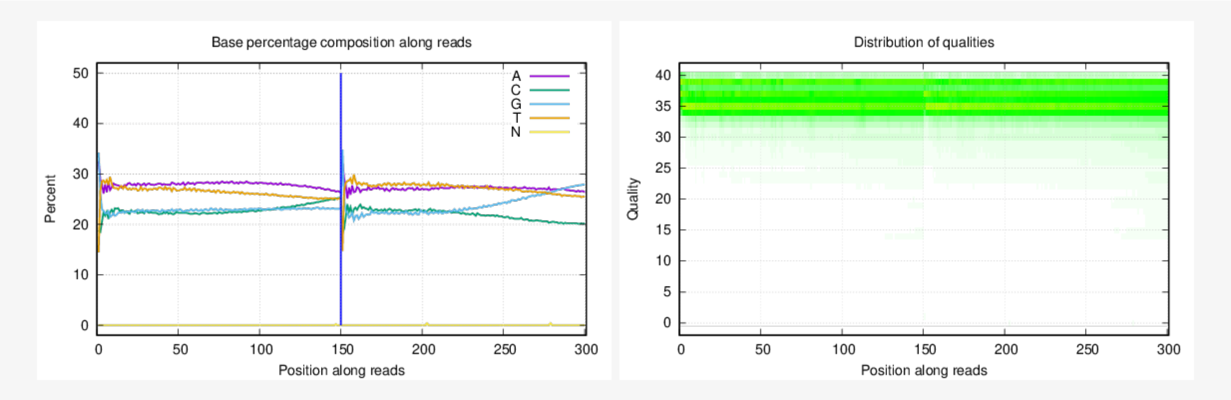
Table Format:

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

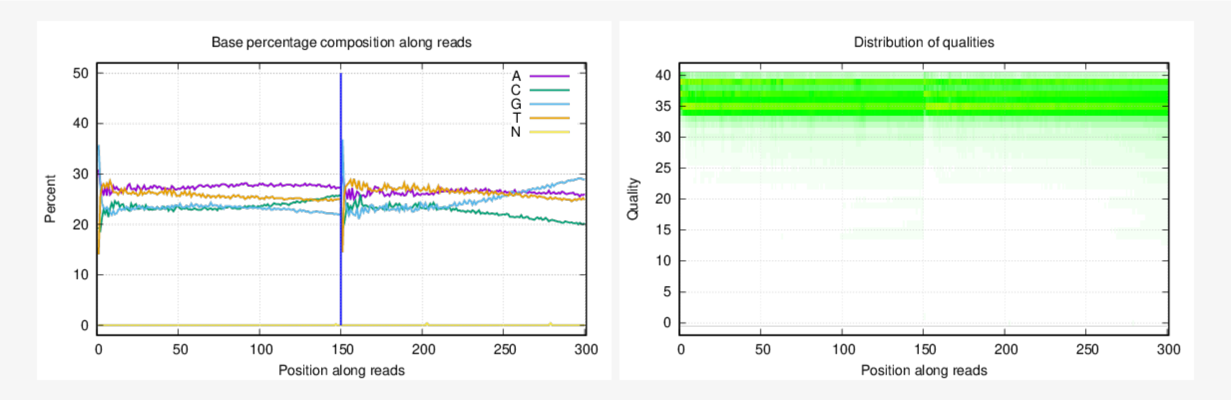
Data Quality Control

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.

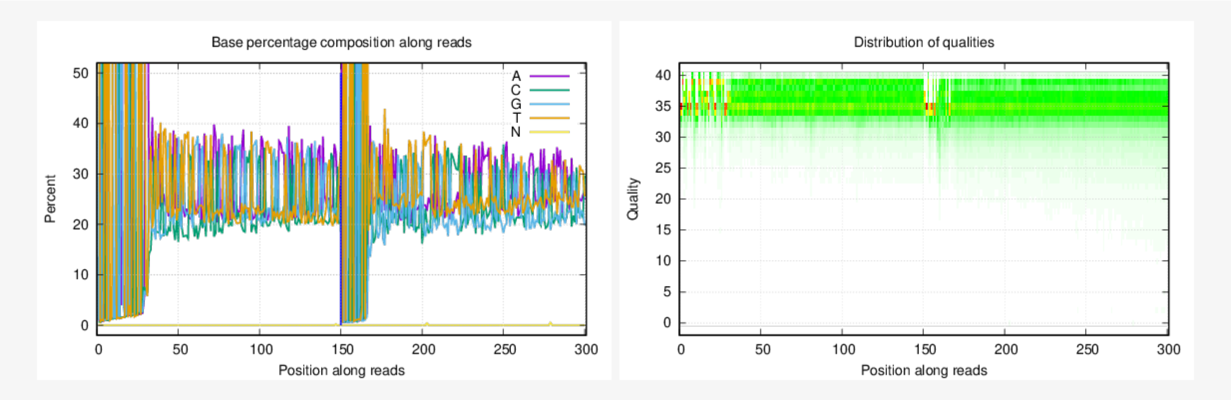
1_25_R1



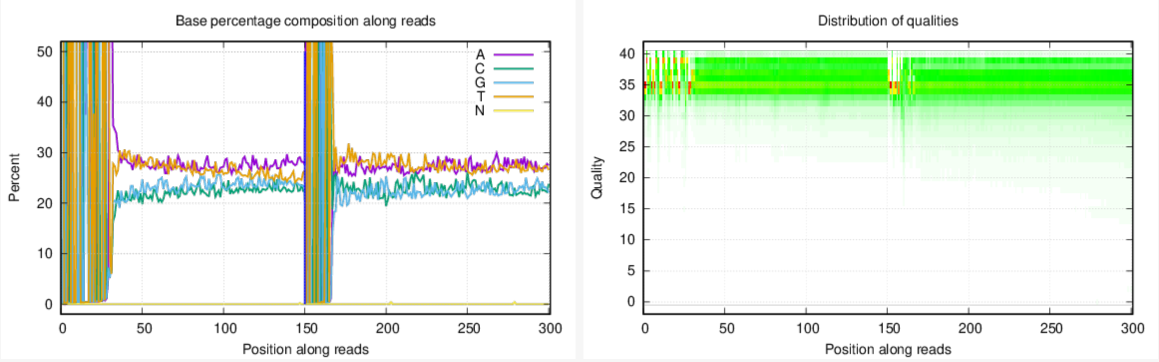
1_25_R2



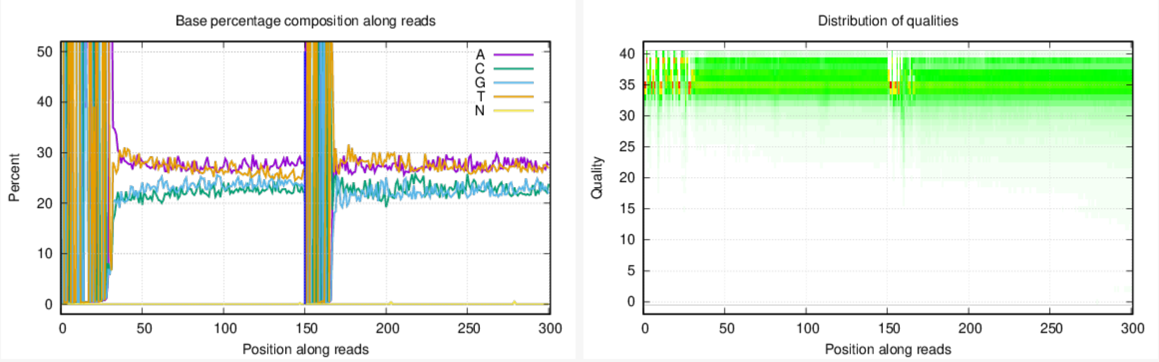
4C_Ph_MEF_1_fixed22



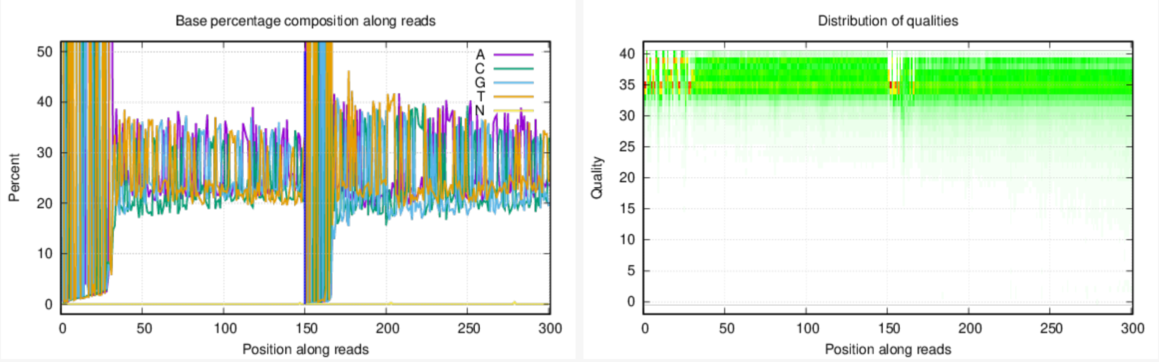
4C_Ph_MEF_2_fixed1



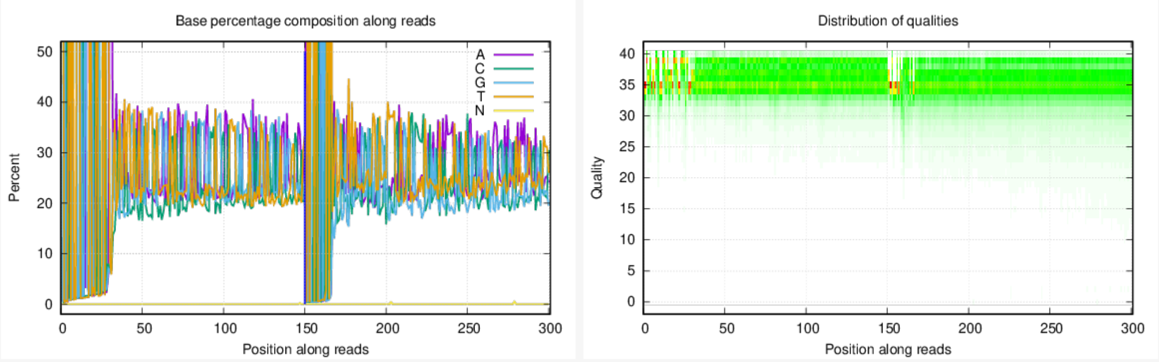
4C_Ph_MEF_3_fixed1



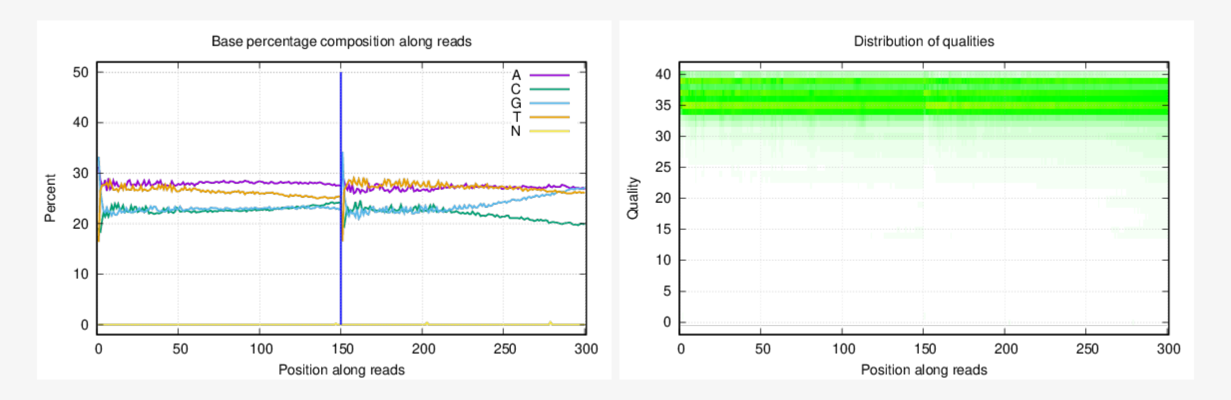
4C_Ph_MEF_4_fixed2



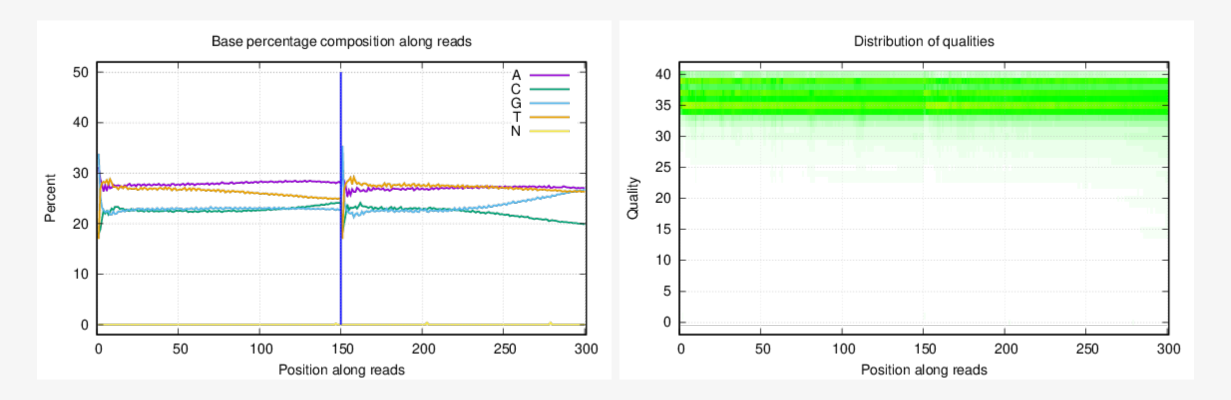
4C_Ph_MEF_5_fixed2



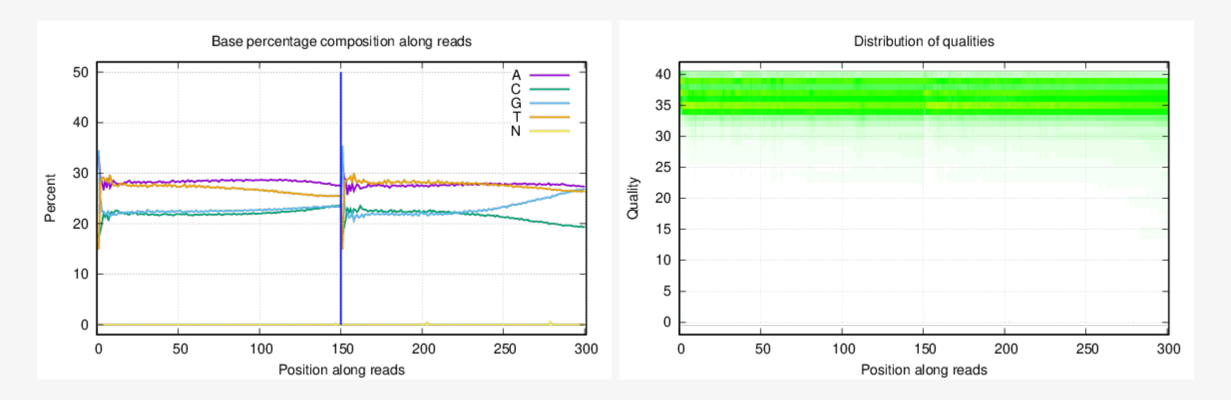
4_14_D_R1



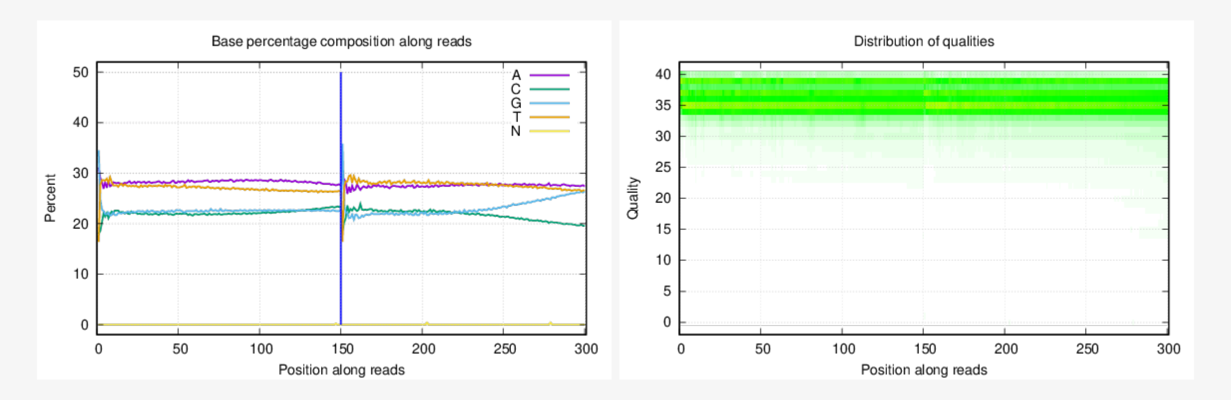
4_14_D_R2



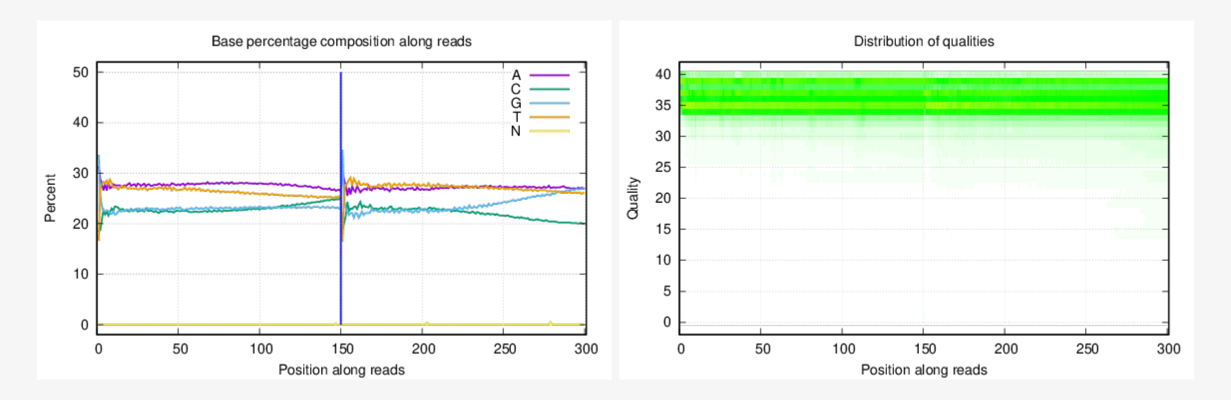
4_14_R1



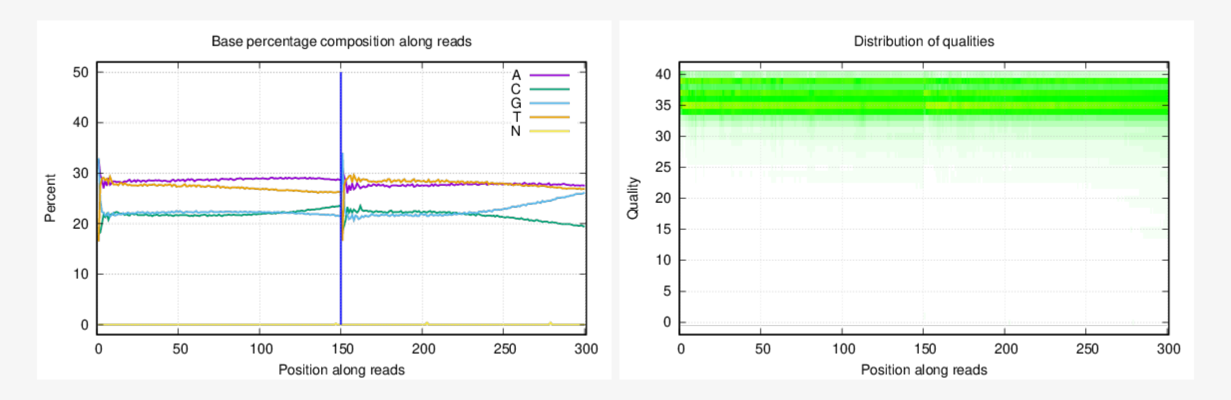
4_14_R2



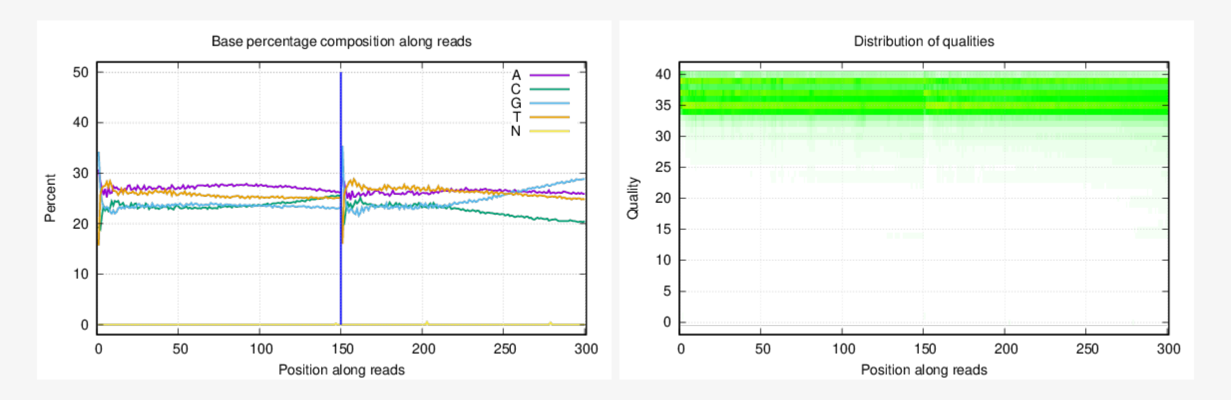
4_22_Ch_R1



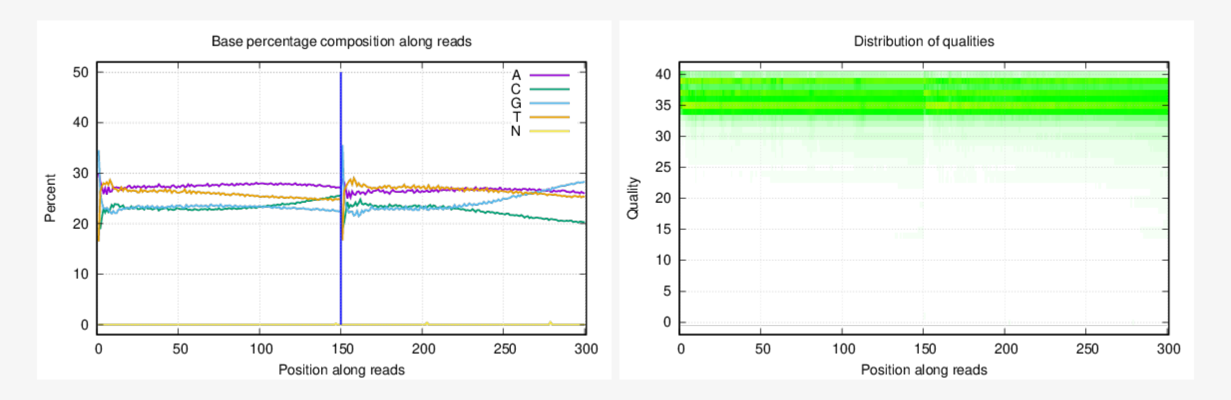
4_22_Ch_R2



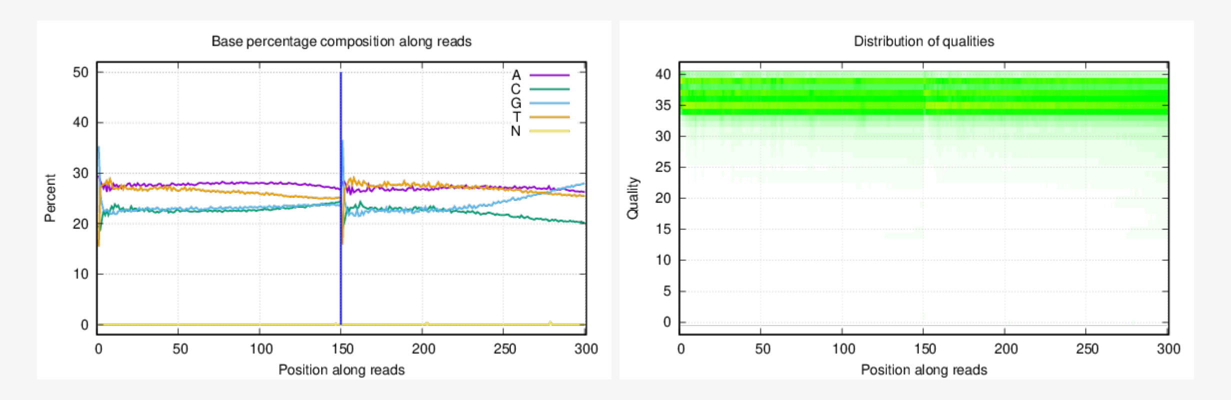
4_22_D_R1



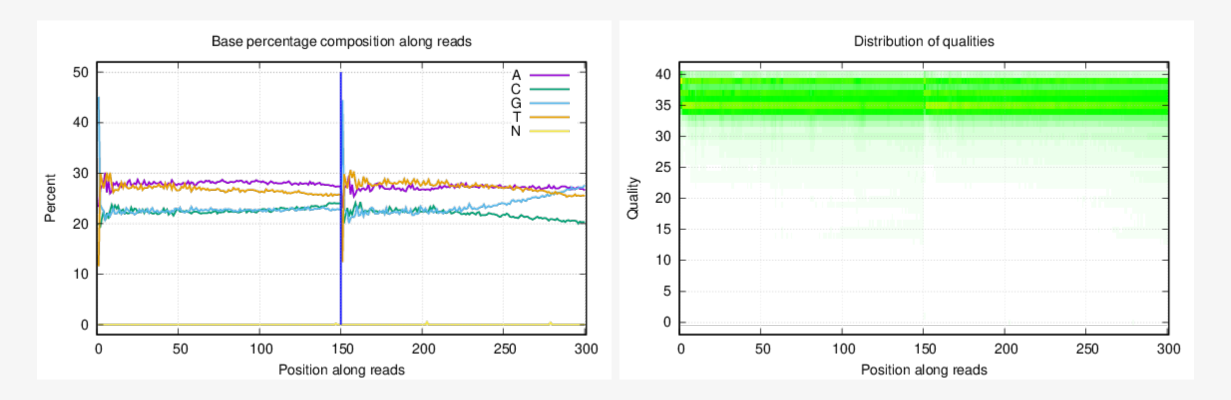
4_22_D_R2



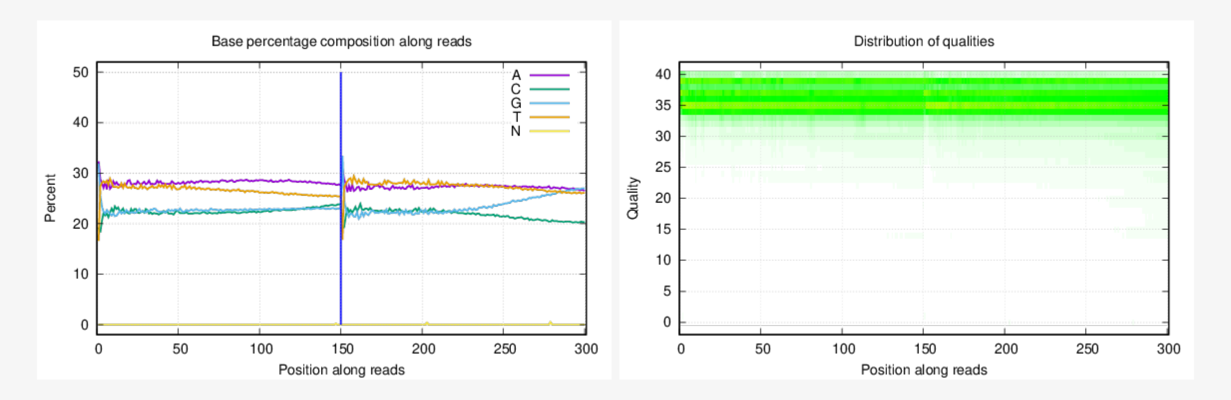
4_22_R1



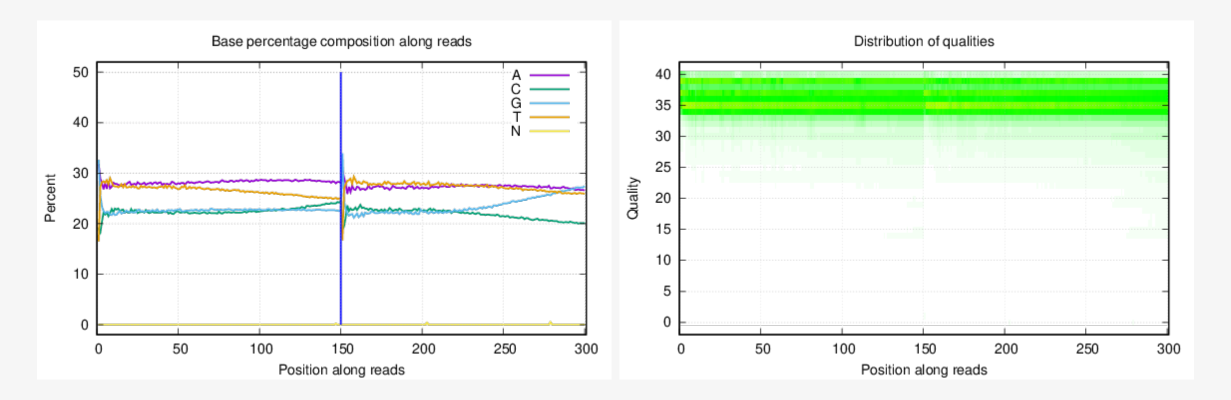
4_22_R2



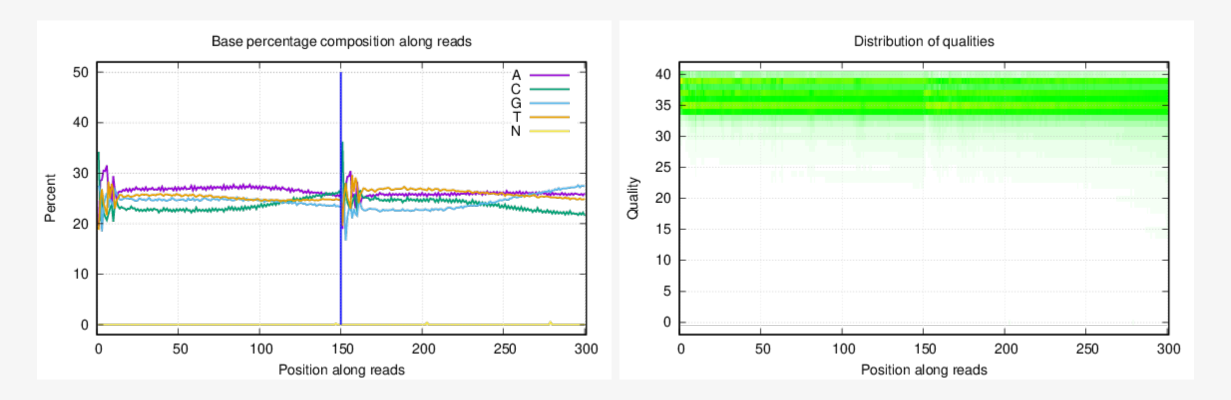
4_22_U_R1



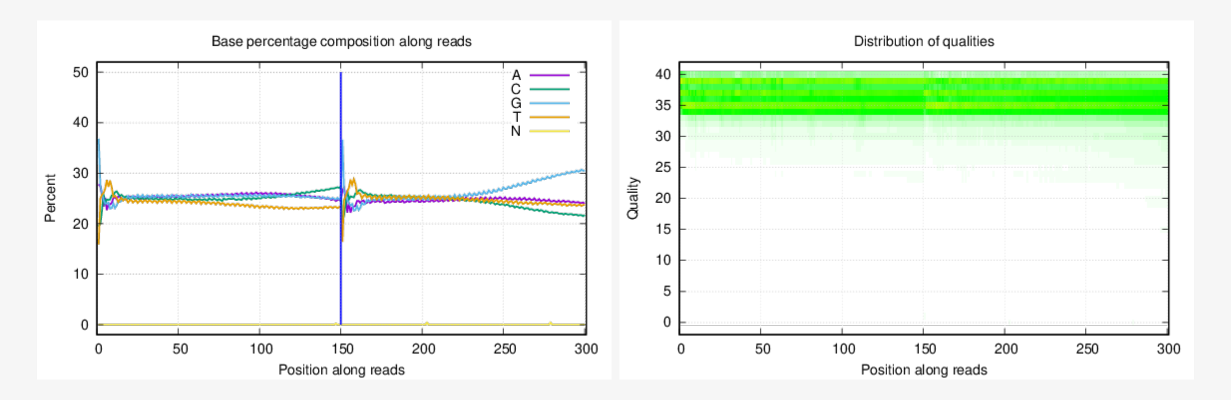
4_22_U_R2



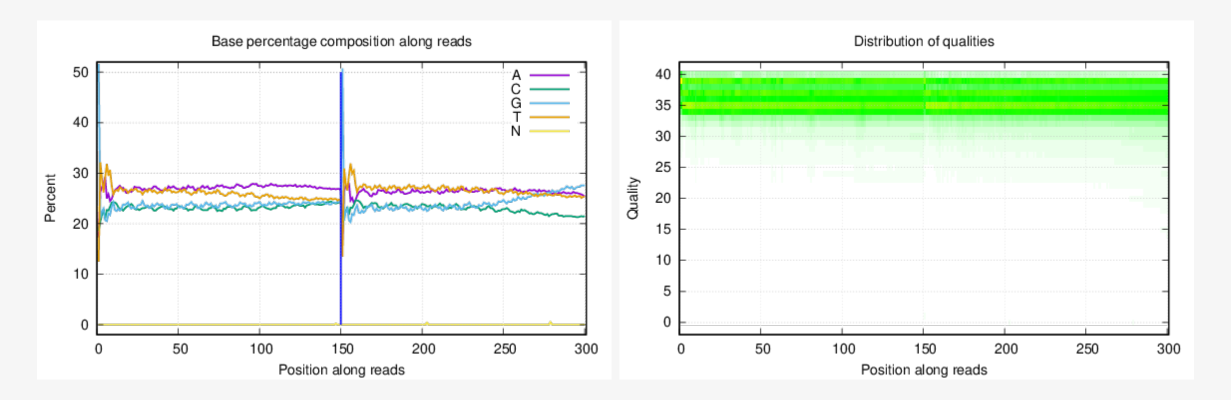
Bp



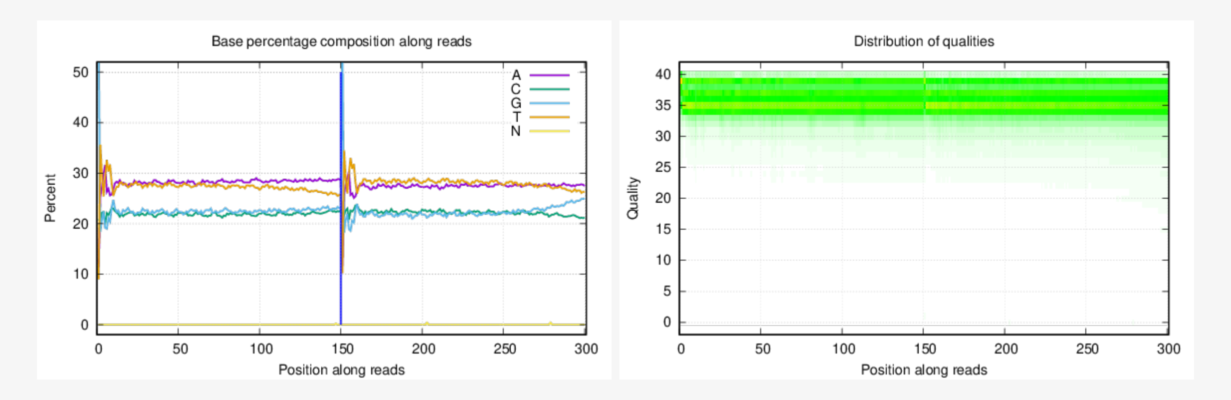
Cenp-B1



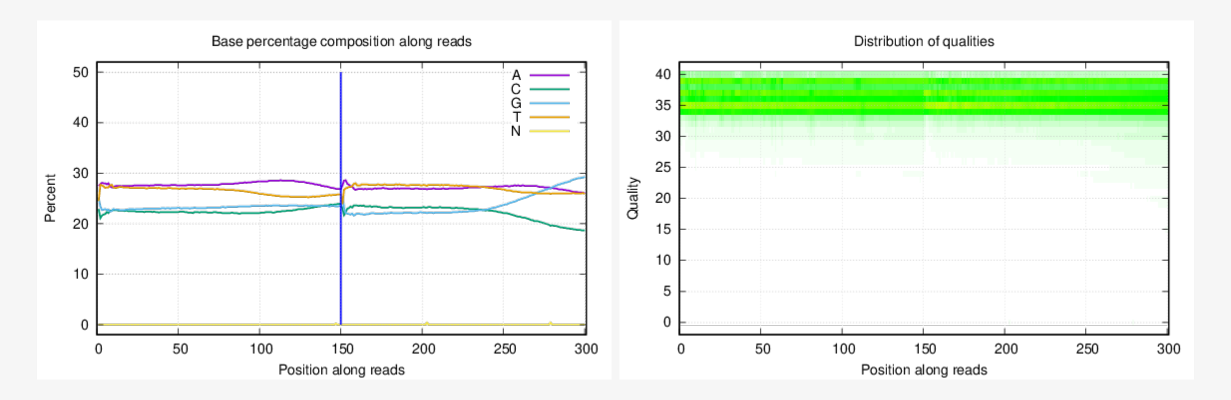
Cenp-B2



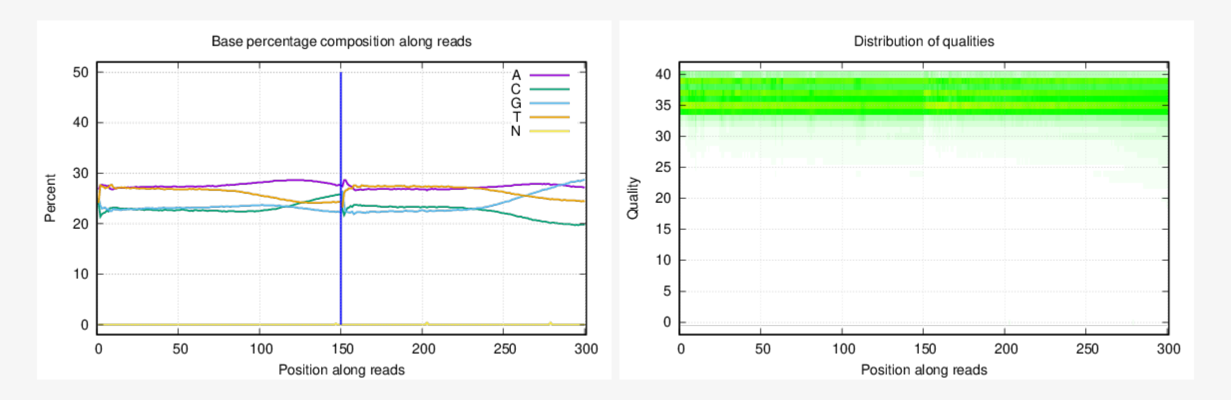
Cenp-P2



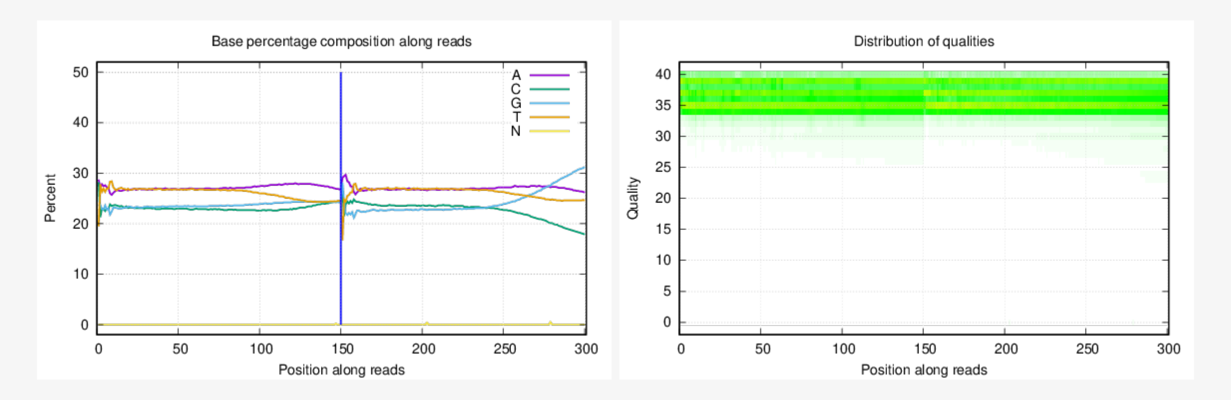
Ch_CTCF_hg_1



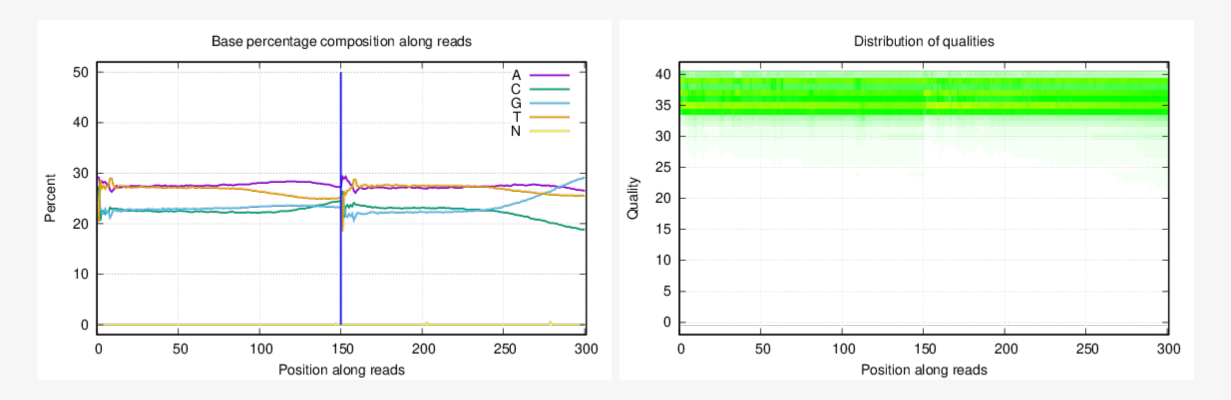
Ch_CTCF_hg_2



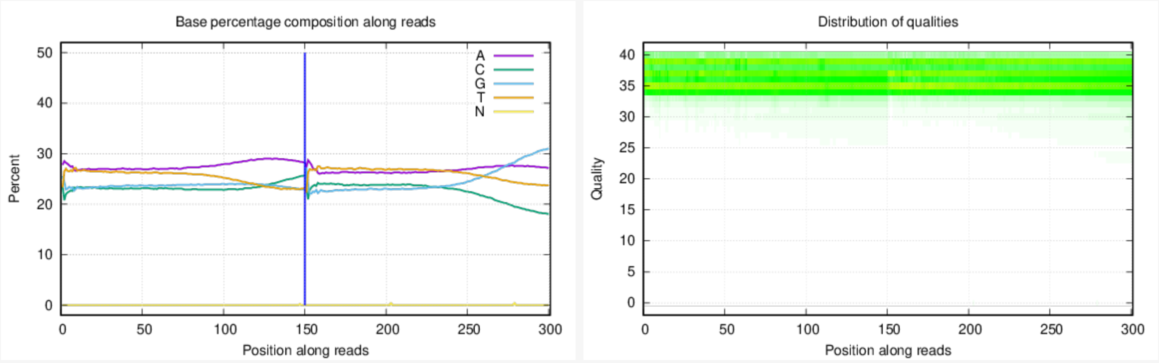
Ch_CTCF_hg_3



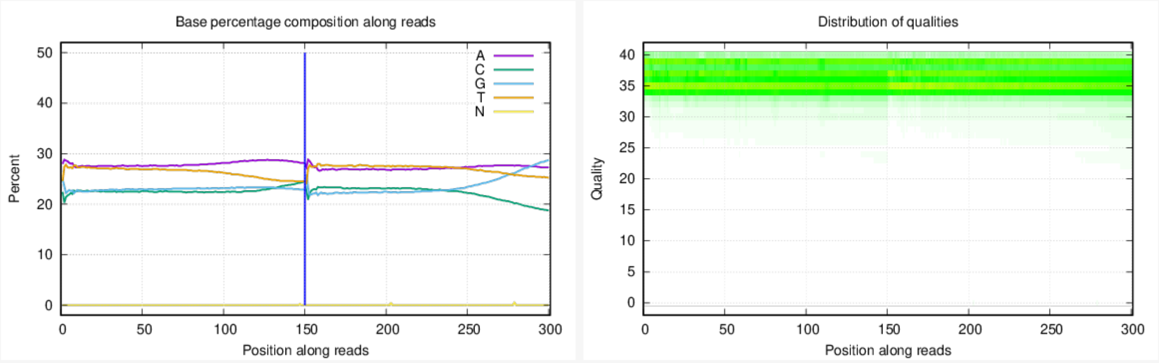
Ch_CTCF_hg_4



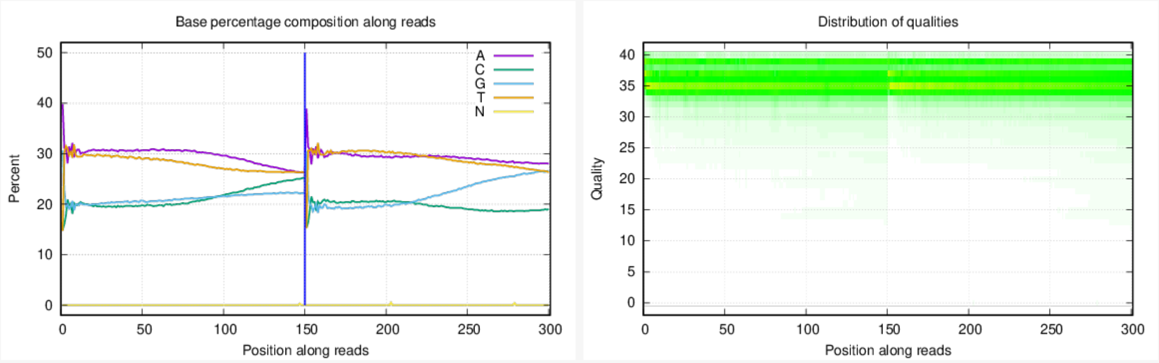
i_CTCF_hg_1



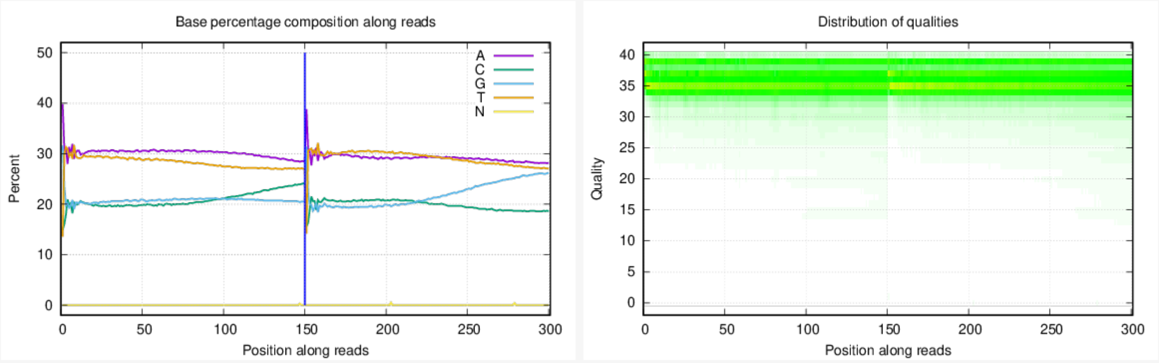
i_CTCF_hg_2



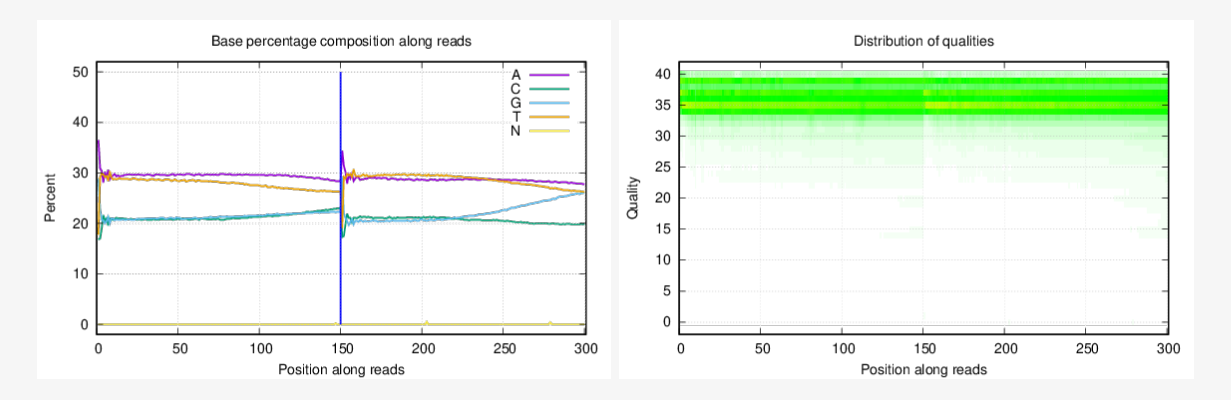
in1_25_R1



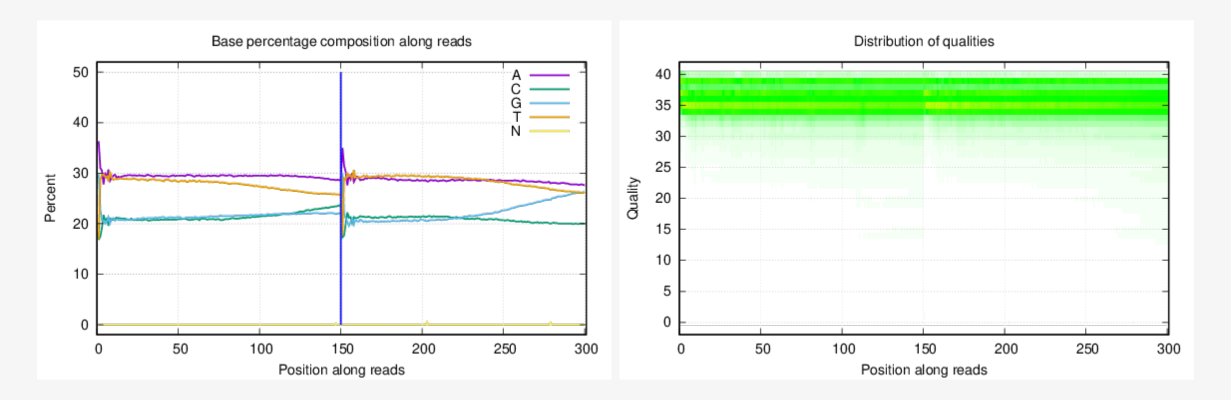
in1_25_R2



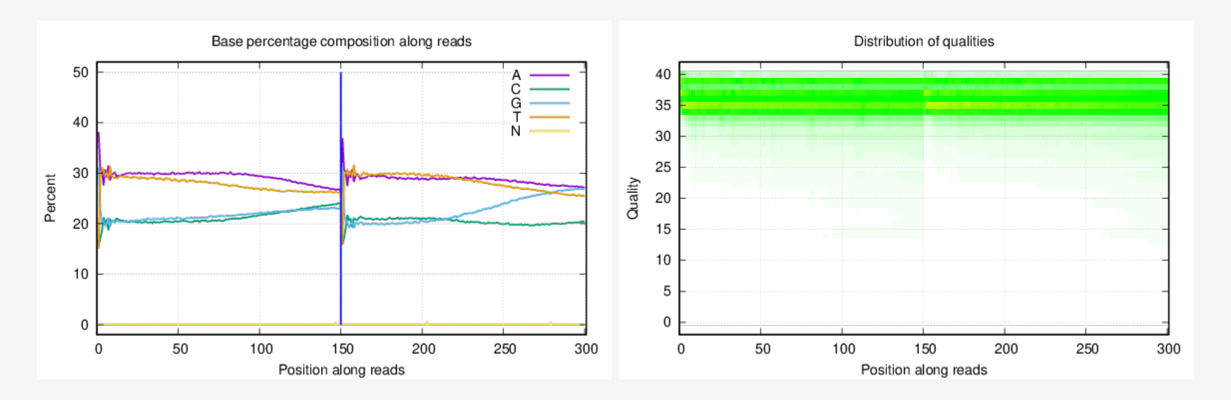
in4_14_D_R1



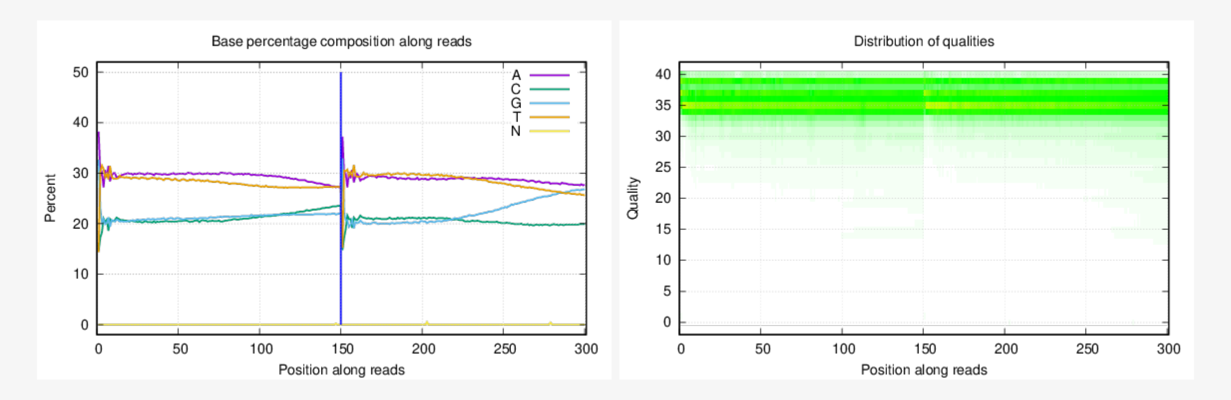
in4_14_D_R2



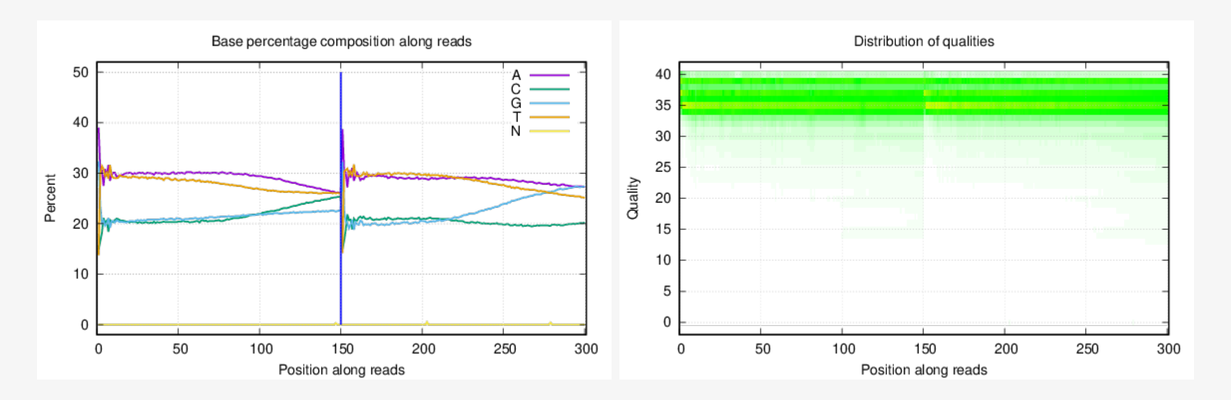
in4_14_R1



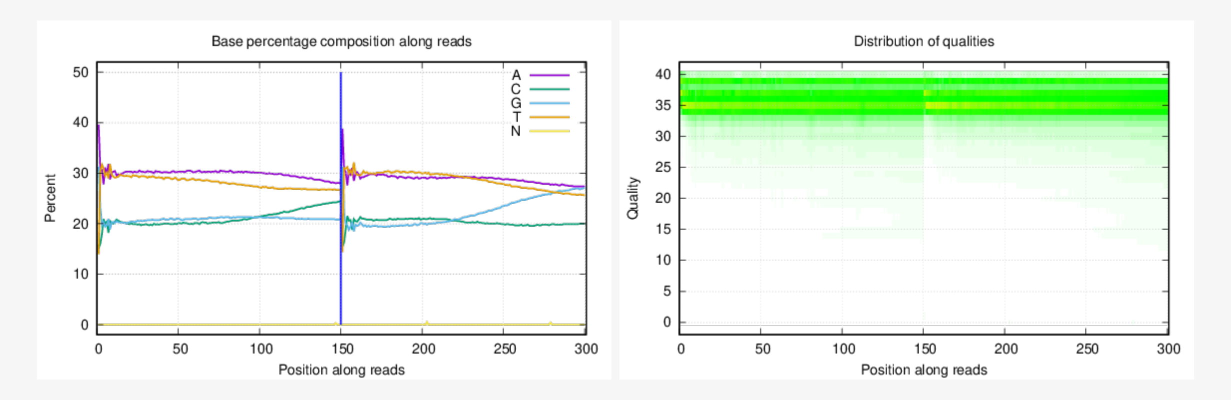
in4_14_R2



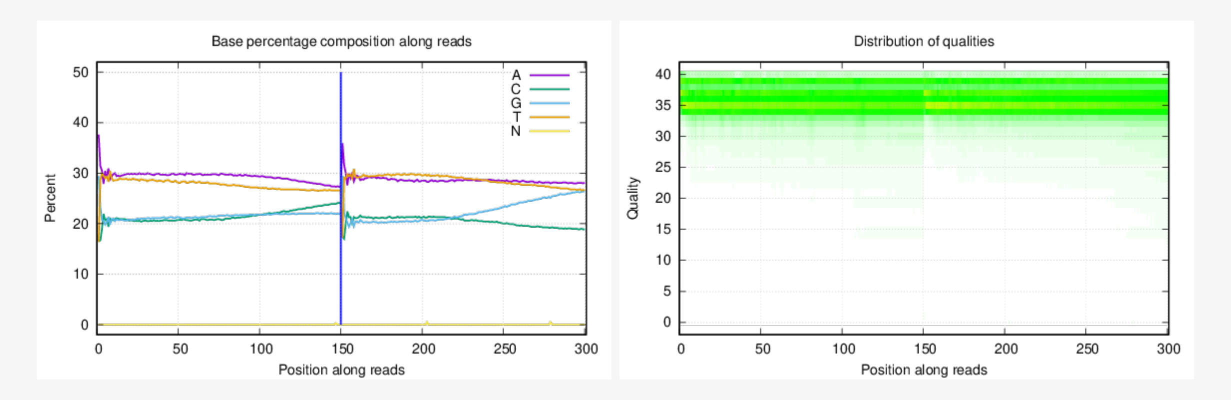
in4_22_Ch_R1



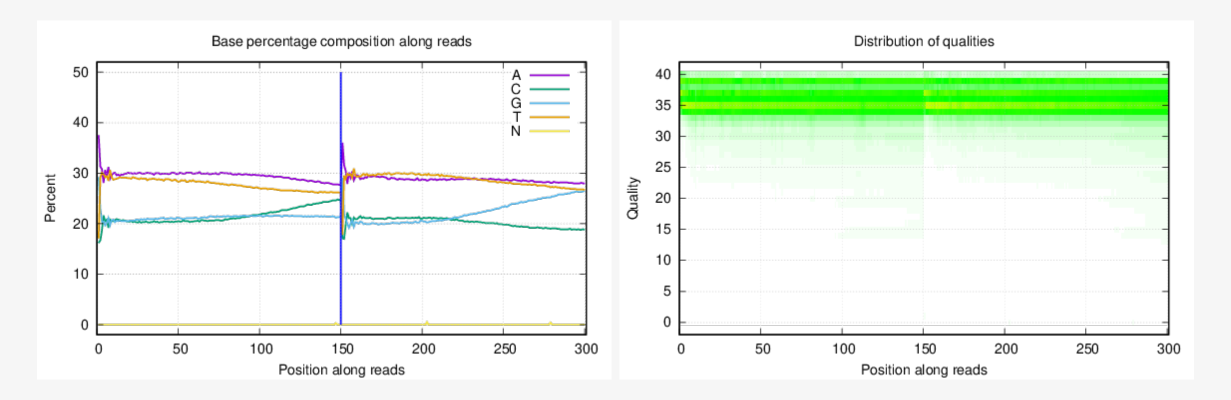
in4_22_Ch_R2



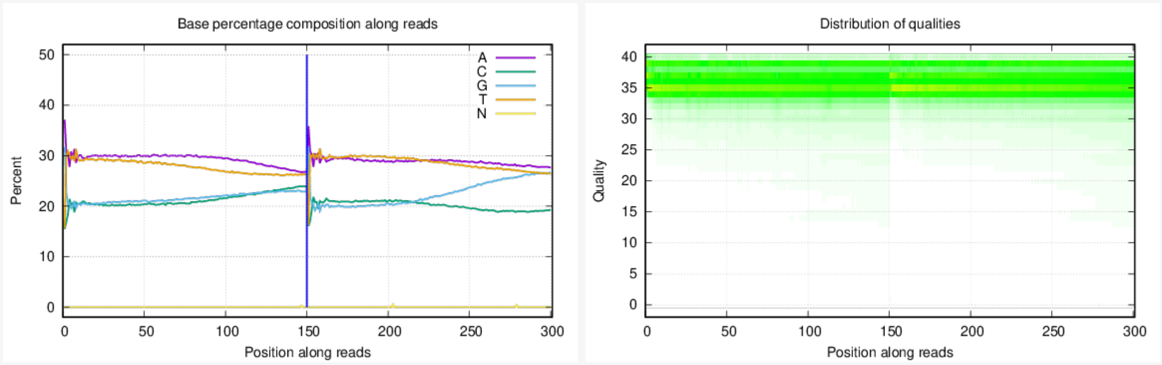
in4_22_D_R1



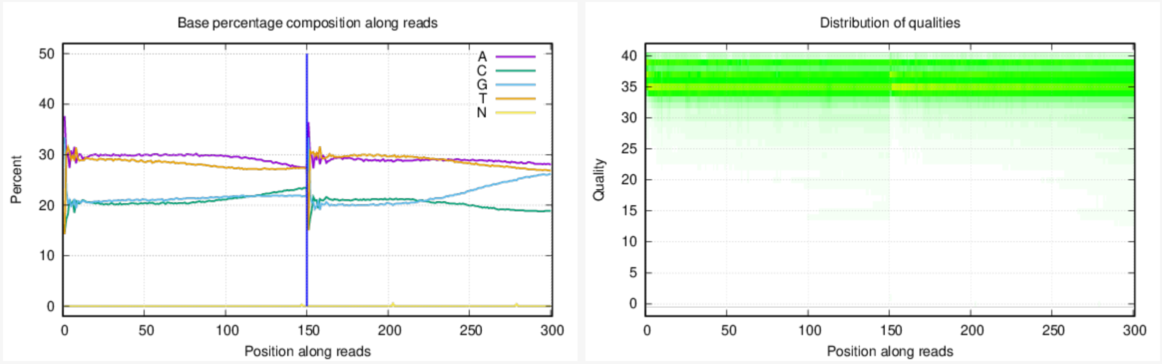
in4_22_D_R2



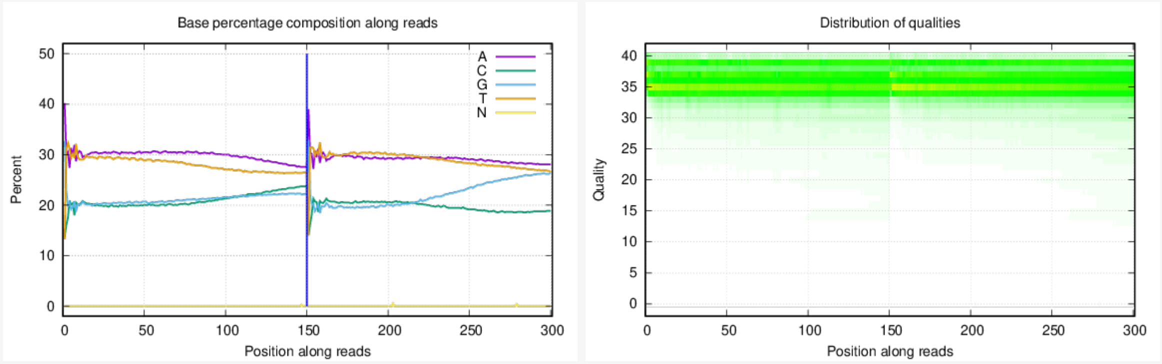
in4_22_R1



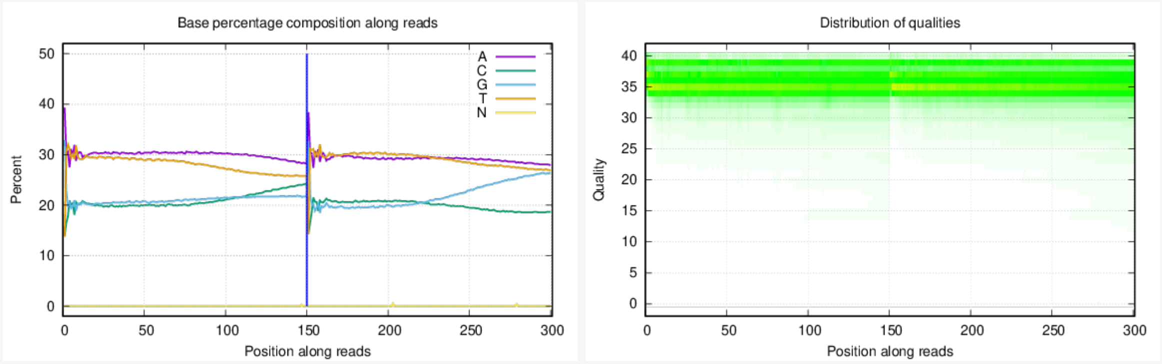
in4_22_R2



in4_22_U_R1



in4_22_U_R2



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