

Figure S1. Continued.

Fastp report

Summary

General

fastp version:	0.21.0 (https://github.com/OpenGene/fastp)
sequencing:	paired end (150 cycles + 150 cycles)
mean length before filtering:	150bp, 150bp
mean length after filtering:	149bp, 149bp
duplication rate:	17.419749%
Insert size peak:	248
Detected read2 adapter:	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

Before filtering

total reads:	5.355036 M
total bases:	803.255400 M
Q20 bases:	777.152110 M (96.750313%)
Q30 bases:	737.277038 M (91.786129%)
GC content:	43.008339%

After filtering

total reads:	5.296722 M
total bases:	792.941293 M
Q20 bases:	769.622446 M (97.059196%)
Q30 bases:	730.985294 M (92.186559%)
GC content:	42.957271%

Filtering result

reads passed filters:	5.296722 M (98.911044%)
reads with low quality:	57.868000 K (1.080628%)
reads with too many N:	172 (0.003212%)
reads too short:	274 (0.005117%)

Adapters

Adapter or bad ligation of read1

Input has little adapter percentage (~0.185076%), probably it's trimmed before.

Sequence	Occurrences
A	1892
AG	1802
AGA	1670
AGAT	1578
AGATC	1599
AGATCG	1437
AGATCGG	1349
AGATCGGA	1267
AGATCGGAA	1180
AGATCGGAAG	1130
AGATCGGAAGA	1170
AGATCGGAAGAG	1153
AGATCGGAAGAGC	1015
AGATCGGAAGAGCA	947
AGATCGGAAGAGCAC	916
AGATCGGAAGAGCACAC	823
AGATCGGAAGAGCACAC	856
AGATCGGAAGAGCACACG	738
AGATCGGAAGAGCACACGT	678
AGATCGGAAGAGCACACGTC	653
AGATCGGAAGAGCACACGTCT	649
AGATCGGAAGAGCACACGTCTG	611
AGATCGGAAGAGCACACGTCTGA	546
AGATCGGAAGAGCACACGTCTGAA	632
AGATCGGAAGAGCACACGTCTGAAC	546
AGATCGGAAGAGCACACGTCTGAACT	489
AGATCGGAAGAGCACACGTCTGAACTC	509
AGATCGGAAGAGCACACGTCTGAACTCC	428
AGATCGGAAGAGCACACGTCTGAACTCCA	417
AGATCGGAAGAGCACACGTCTGAACTCCAG	419
AGATCGGAAGAGCACACGTCTGAACTCCAGT	402
other adapter sequences	8778

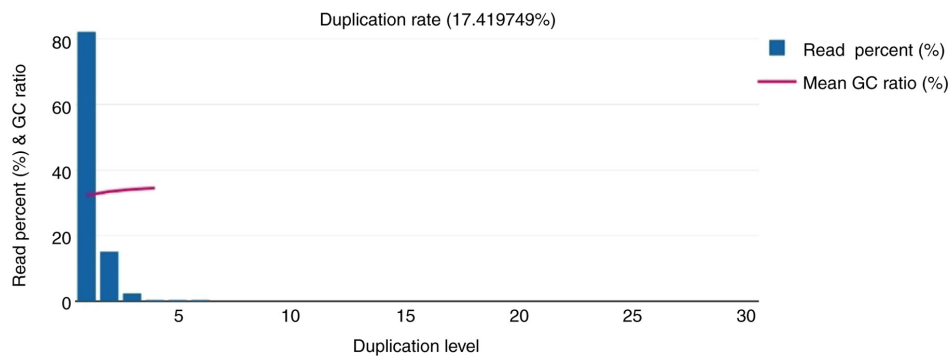
Figure S1. Continued.

Adapter or bad ligation of read2

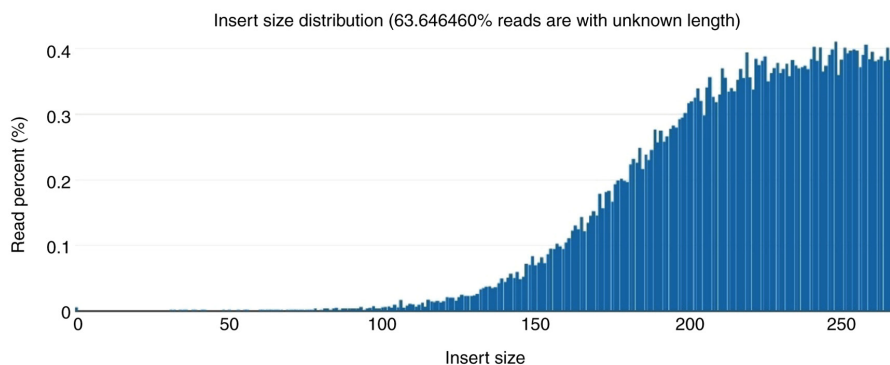
Input has little adapter percentage (~0.190723%), probably it's trimmed before.

Sequence	Occurrences
A	1893
AG	1822
AGA	1660
AGAT	1572
AGATC	3817
AGATCG	1618
AGATCGG	1400
AGATCGGA	1270
AGATCGGAA	1182
AGATCGGAAG	1142
AGATCGGAAGA	1157
AGATCGGAAGAG	1148
AGATCGGAAGAGC	1016
AGATCGGAAGAGCG	961
AGATCGGAAGAGCGT	918
AGATCGGAAGAGCGTC	830
AGATCGGAAGAGCGTCG	841
AGATCGGAAGAGCGTCGT	731
AGATCGGAAGAGCGTCGTG	680
AGATCGGAAGAGCGTCGTGT	650
AGATCGGAAGAGCGTCGTGTA	622
AGATCGGAAGAGCGTCGTGTAG	594
AGATCGGAAGAGCGTCGTGTAGG	542
AGATCGGAAGAGCGTCGTGTAGGG	605
AGATCGGAAGAGCGTCGTGTAGGGA	516
AGATCGGAAGAGCGTCGTGTAGGGA	455
AGATCGGAAGAGCGTCGTGTAGGAAA	480
other adapter sequences	11350

Duplication



Insert size estimation



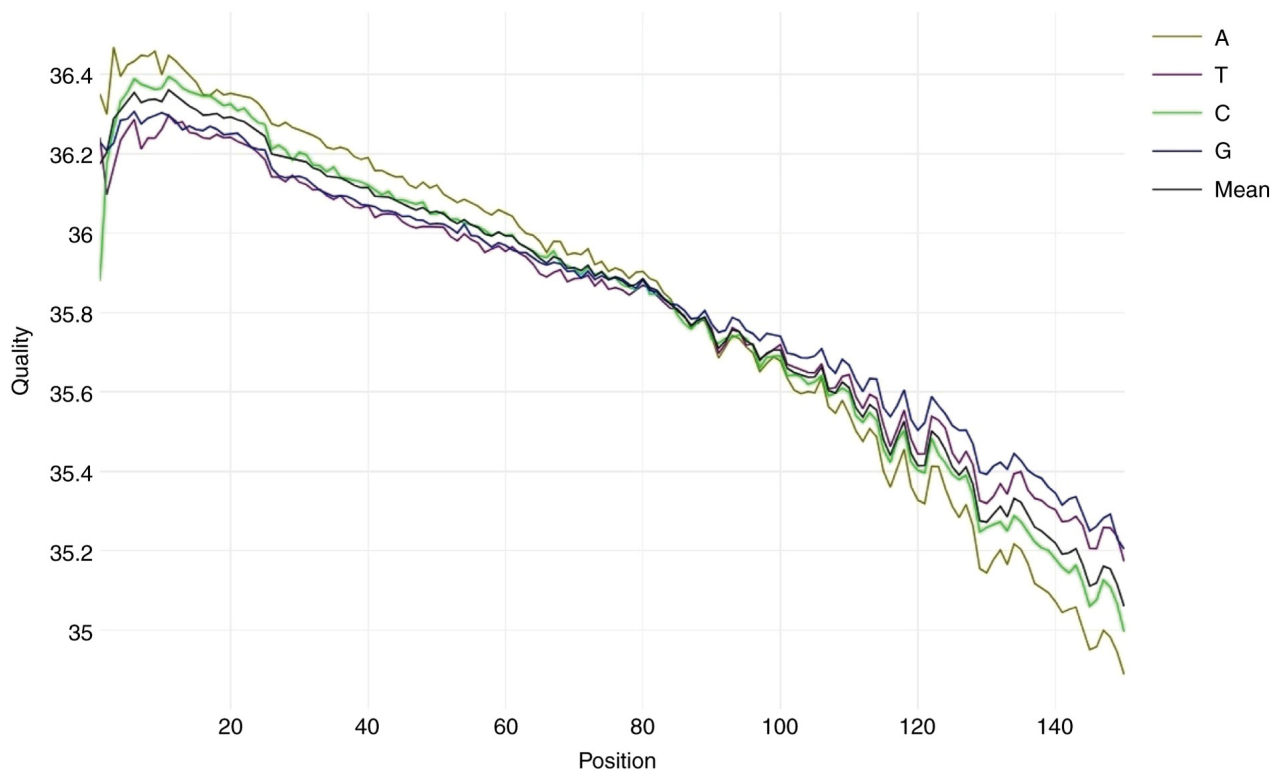
This estimation is based on paired-end overlap analysis, and there are 63.646460% reads found not overlapped. Nonoverlapped read pairs may have insert size <30 or >270, or contain too much sequencing errors to be detected as overlapped.

Figure S1. Continued.

Before filtering

Before filtering: read1: quality

Value of each position will be shown on mouse over.



Before filtering: read1: base contents

Value of each position will be shown on mouse over.

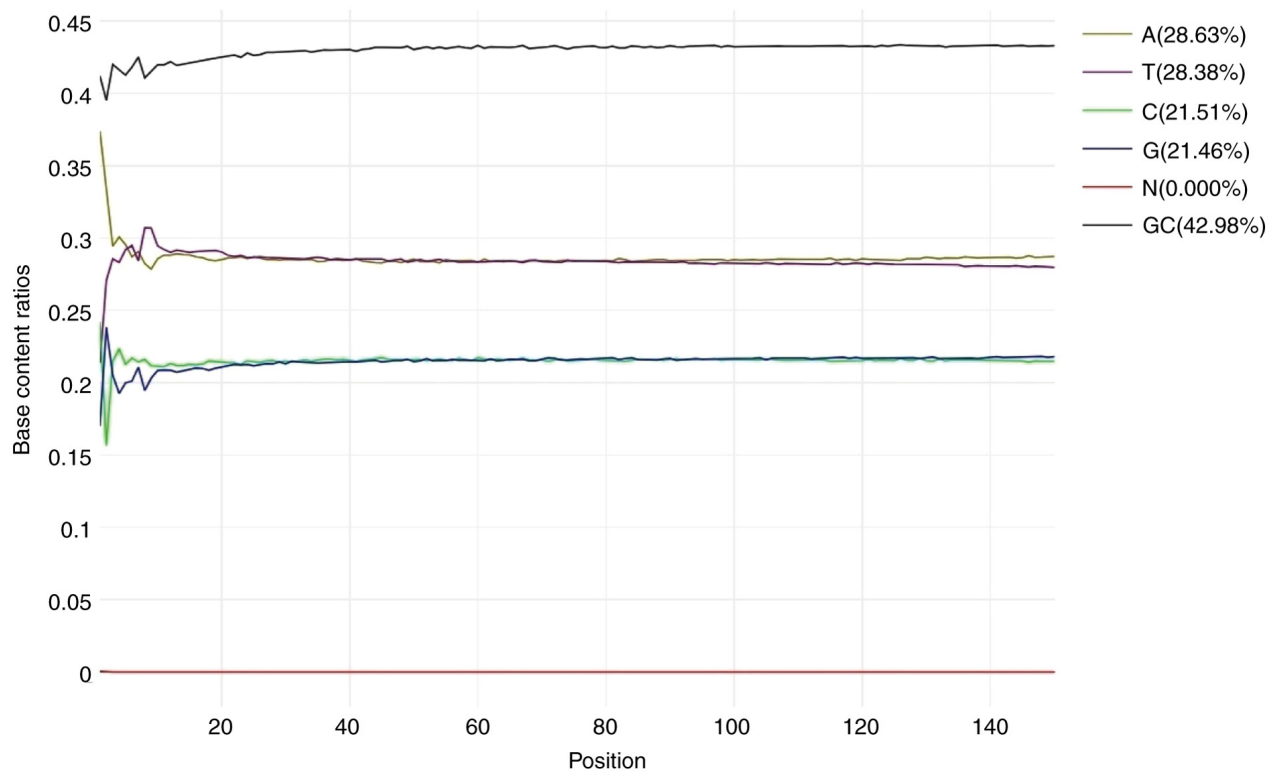
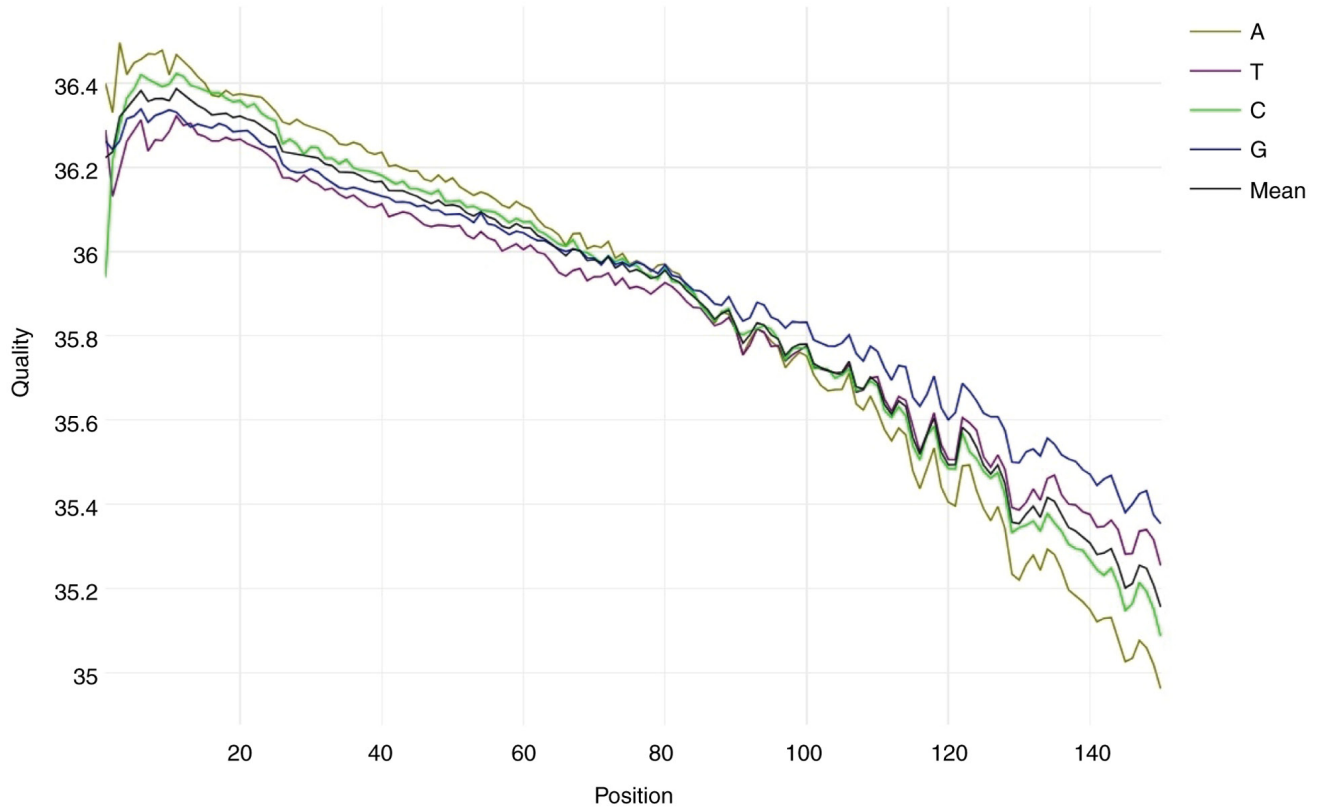


Figure S1. Continued.

After filtering

After filtering: read1: quality

Value of each position will be shown on mouse over.



After filtering: read1: base contents

Value of each position will be shown on mouse over.

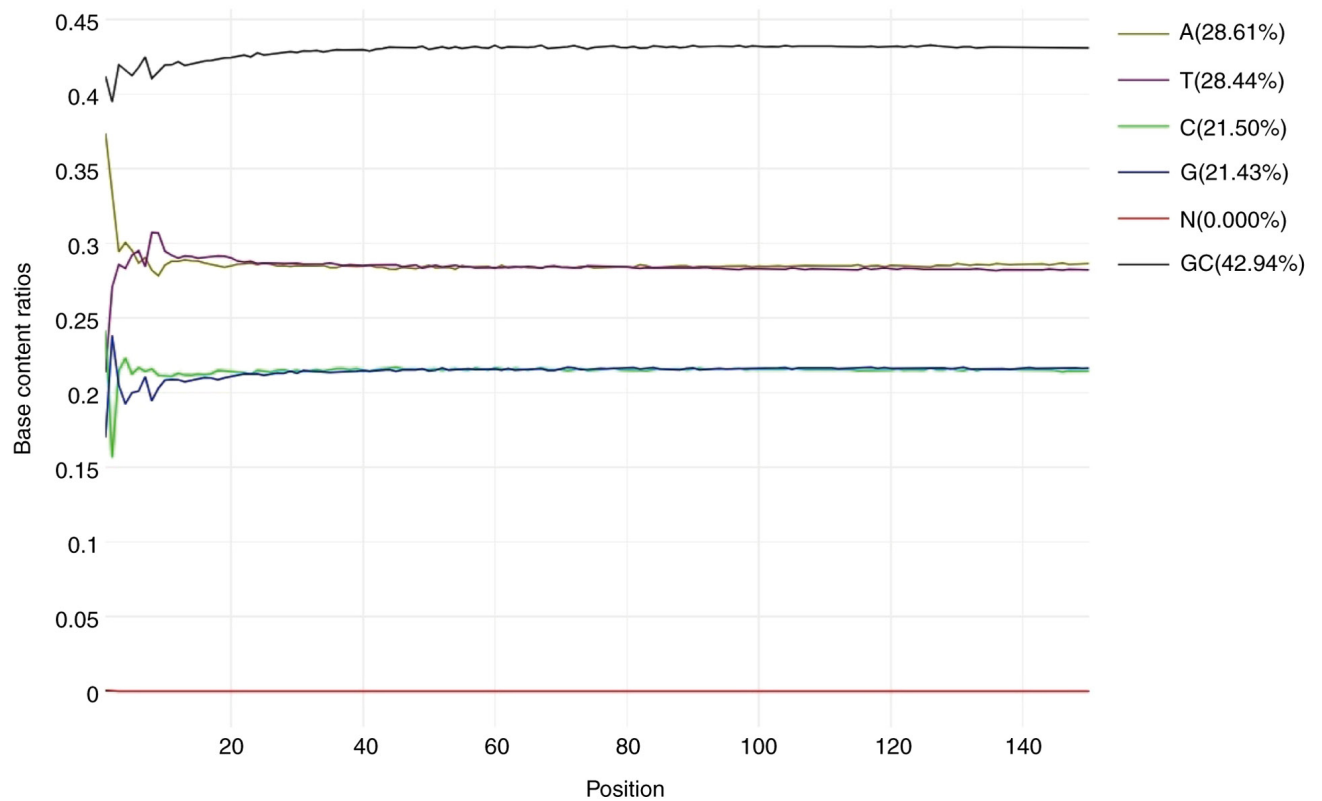
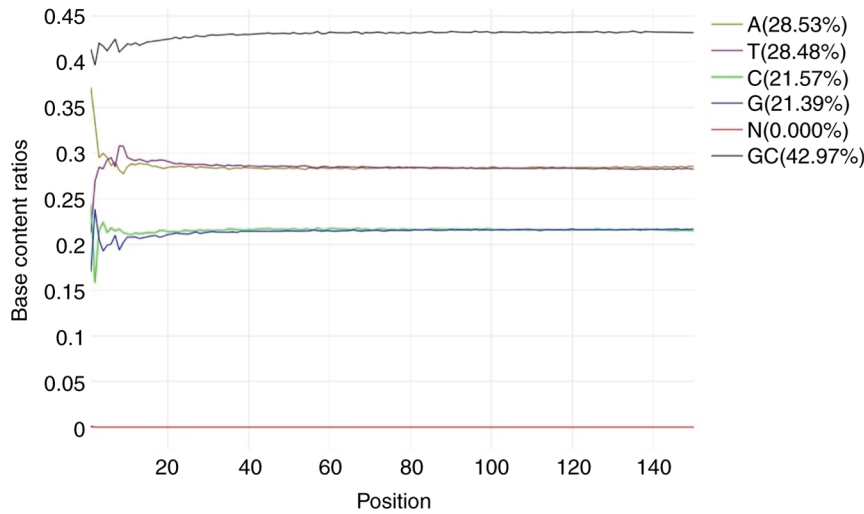


Figure S1. Fastp report for the biopsy specimen (the first technical replicate).

After filtering: read2: base contents

Value of each position will be shown on mouse over.



After filtering: read2: KMER counting

Darker background means larger counts. The count will be shown on mouse over.

AA	AT	AC	AG	TA	TT	TC	TG	CA	CT	CC	CG	GA	GT	GC	GG
AAA	AAAA	AAAT	AAAC	AAAG	AAATA	AAATT	AAATC	AAATG	AAACA	AAACT	AAACC	AAACG	AAAGA	AAAGT	AAAGC
AAT	AATA	AATAT	AATAC	AATAG	AATTA	AATTT	AATTC	AATTG	AATCA	AATCT	AATCC	AATCG	AATGA	AATGT	AATGC
AAC	AACA	AACAT	AACAC	AACAG	AACATA	AACATT	AACATC	AACATG	AACACA	AACACT	AACACC	AACACG	AACAGA	AACAGT	AACAGC
AAG	AAGA	AAGAT	AAGAC	AAGAG	AAGATA	AAGATT	AAGATC	AAGATG	AAGACA	AAGACT	AAGACC	AAGACG	AAGAGA	AAGAGT	AAGAGC
ATA	ATAA	ATAAT	ATAAC	ATAAG	ATAATA	ATAATT	ATAATC	ATAATG	ATAACA	ATAACT	ATAACC	ATAACG	ATAAGA	ATAAGT	ATAAGC
ATT	ATTA	ATTAT	ATTAC	ATTAG	ATTATA	ATTATT	ATTATC	ATTATG	ATTACA	ATTACT	ATTACC	ATTACG	ATTAGA	ATTAGT	ATTAGC
ATC	ATCA	ATCAT	ATCAC	ATCAG	ATCTA	ATCTT	ATCTC	ATCTG	ATCCA	ATCCT	ATCCC	ATCCG	ATCGA	ATCGT	ATCCG
ATG	ATGA	ATGAT	ATGAC	ATGAG	ATGATA	ATGATT	ATGATC	ATGATG	ATGACA	ATGACT	ATGCC	ATGCG	ATGGA	ATGGT	ATGGC
ACA	ACAA	ACAAT	ACAAC	ACAG	ACATA	ACATT	ACATC	ACATG	ACACA	ACACT	ACACC	ACACG	ACAGA	ACAGT	ACAGC
ACT	ACTA	ACTAT	ACTAC	ACTAG	ACTATA	ACTATT	ACTATC	ACTATG	ACTACA	ACTACT	ACTACC	ACTACG	ACTAGA	ACTAGT	ACTAGC
ACC	ACCA	ACCAT	ACCAC	ACCAG	ACCTA	ACCTT	ACCTC	ACCTG	ACCCA	ACCTT	ACCCC	ACCCG	ACCGA	ACCGT	ACCCG
ACG	ACGA	ACGAT	ACGAC	ACGAG	ACGTA	ACGTT	ACGTC	ACGTG	ACGCA	ACGCT	ACGCC	ACGCG	ACGGA	ACGGT	ACGGC
AGT	AGTA	AGTAT	AGTAC	AGTAG	AGTATA	AGTATT	AGTATC	AGTATG	AGTACA	AGTACT	AGTACC	AGTACG	AGTAGA	AGTAGT	AGTAGC
AGC	AGCA	AGCAT	AGCAC	AGCAG	AGCTA	AGCTT	AGCTC	AGCTG	AGCCA	AGCCT	AGCCC	AGCCG	AGCGA	AGCGT	AGCCG
AGS	AGSA	AGSAT	AGSAC	AGSAG	AGSATA	AGSATT	AGSATC	AGSATG	AGSACA	AGSACT	AGSACC	AGSACG	AGSAGA	AGSAGT	AGSAGC
TAA	TAAA	TAAAT	TAAAC	TAAAG	TAAATA	TAAATT	TAAATC	TAAATG	TAAACA	TAAACT	TAAACC	TAAACG	TAAAGA	TAAAGT	TAAAGC
TAT	TATA	TATAT	TATAC	TATAG	TATATA	TATATT	TATATC	TATATG	TATACA	TATACT	TATACC	TATACG	TATAGA	TATAGT	TATAGC
TAC	TACA	TACAT	TACAC	TACAG	TACTA	TACTT	TACTC	TACTG	TACCA	TACCT	TACCC	TACCG	TACGA	TACGT	TACGC
TAG	TAGA	TAGAT	TAGAC	TAGAG	TAGTA	TAGTT	TAGTC	TAGTG	TAGCA	TAGCT	TAGCC	TAGCG	TAGGA	TAGGT	TAGGC
TAA	TAAA	TAAAT	TAAAC	TAAAG	TAAATA	TAAATT	TAAATC	TAAATG	TAAACA	TAAACT	TAAACC	TAAACG	TAAAGA	TAAAGT	TAAAGC
TTA	TTAA	TTAAT	TTAAC	TTAG	TTATA	TTATT	TTATC	TTATG	TTACA	TTACT	TTACC	TTACG	TTAGA	TTAGT	TTAGC
TTT	TTTA	TTTAT	TTTAC	TTTAG	TTTTA	TTTTT	TTTTC	TTTTG	TTTCA	TTTCT	TTTCC	TTTCG	TTTGA	TTTGT	TTTGC
TTG	TTGA	TTGAT	TTGAC	TTGAG	TTGTA	TTGTT	TTGTC	TTGTG	TTGCA	TTGCT	TTGCC	TTGCG	TTGGA	TTGGT	TTGGC
TCA	TCAA	TCAAT	TCAAC	TCAAG	TCATA	TCATT	TCATC	TCATG	TCACA	TCACT	TCACC	TCACG	TCAGA	TCAGT	TCAGC
TCT	TCTA	TCTAT	TCTAC	TCTAG	TCTTA	TCTTT	TCTTC	TCTTG	TCTCA	TCTCT	TCTCC	TCTCG	TCTGA	TCTGT	TCTGC
TCC	TCCA	TCCAT	TCCAC	TCCAG	TCCTA	TCCTT	TCCTC	TCCTG	TCCCA	TCCCT	TCCCC	TCCCG	TCCGA	TCCGT	TCCGC
TCS	TCSA	TCSAT	TCSAC	TCSAG	TCSATA	TCSATT	TCSATC	TCSATG	TCSACA	TCSACT	TCSACC	TCSACG	TCSAGA	TCSAGT	TCSAGC
TGA	TGAA	TGAAT	TGAAC	TGAG	TGATA	TGATT	TGATC	TGATG	TGACA	TGACT	TGACC	TGACG	TGAGA	TGAGT	TGAGC
TGT	TGTA	TGTAT	TGTAC	TGTAG	TGTATA	TGTATT	TGTATC	TGTATG	TGTACA	TGTACT	TGTACC	TGTACG	TGTAGA	TGTAGT	TGTAGC
TGC	TGCA	TGCAT	TGCAC	TGCAG	TGCTA	TGCTT	TGCTC	TGCTG	TGCCA	TGCTT	TGCCC	TGCCG	TGCCA	TGCCG	TGCCG
TGG	TGGA	TGGAT	TGGAC	TGGAG	TGGTA	TGGTT	TGGTC	TGGTG	TGGCA	TGGCT	TGGCC	TGGCG	TGGGA	TGGGT	TGGGC
CAA	CAAA	CAAAAT	CAAAAC	CAAAAG	CAATA	CAATT	CAATC	CAATG	CACAA	CACAT	CACAC	CACAG	CACAGA	CACAGT	CACAGC
CAT	CATA	CATAT	CATAC	CATAG	CATATA	CATATT	CATATC	CATATG	CATACA	CATACT	CATACC	CATACG	CATAGA	CATAGT	CATAGC
CAC	CACA	CACAT	CACAC	CACAG	CACATA	CACATT	CACATC	CACATG	CACACA	CACACT	CACACC	CACACG	CACAGA	CACAGT	CACAGC
CAG	CAGA	CAGAT	CAGAC	CAGAG	CAGATA	CAGATT	CAGATC	CAGATG	CAGACA	CAGACT	CAGACC	CAGACG	CAGAGA	CAGAGT	CAGAGC
CTA	CTAA	CTAAT	CTAAC	CTAG	CTATA	CTATT	CTATC	CTATG	CTACA	CTACT	CTACC	CTACG	CTAGA	CTAGT	CTAGC
CTT	CTTA	CTTAT	CTTAC	CTTAG	CTTTA	CTTTT	CTTTC	CTTTG	CTTCA	CTTCT	CTTCC	CTTCG	CTTGA	CTTGT	CTTGC
CTC	CTCA	CTCAT	CTCAC	CTCAG	CTCTA	CTCTT	CTCTC	CTCTG	CTCCA	CTCCT	CTCCC	CTCCG	CTCGA	CTCGT	CTCCG
CTG	CTGA	CTGAT	CTGAC	CTGAG	CTGTA	CTGTT	CTGTC	CTGTG	CTGCA	CTGCT	CTGCC	CTGCG	CTGGA	CTGGT	CTGGC
CGA	CGAA	CGAAT	CGAAC	CGAG	CGATA	CGATT	CGATC	CGATG	CGACA	CGACT	CGACC	CGACG	CGAGA	CGAGT	CGAGC
CGT	CGTA	CGTAT	CGTAC	CGTAG	CGTTA	CGTTT	CGTTC	CGTTG	CGTCA	CGTCT	CGTCC	CGTCG	CGTGA	CGTGT	CGTGC
CGC	CGCA	CGCAT	CGCAC	CGCAG	CGCTA	CGCTT	CGCTC	CGCTG	CGCCA	CGCCT	CGCCC	CGCCG	CGCGA	CGCGT	CGCCG
CGG	CGGA	CGGAT	CGGAC	CGGAG	CGGTA	CGGTT	CGGTC	CGGTG	CGGCA	CGGCT	CGGCC	CGGCG	CGGGA	CGGGT	CGGGC
GAA	GAAA	GAAAT	GAAAC	GAAAG	GAAATA	GAAATT	GAAATC	GAAATG	GAAACA	GAAACT	GAAACC	GAAACG	GAAAGA	GAAAGT	GAAAGC
GAT	GATA	GATAT	GATAC	GATAG	GATATA	GATATT	GATATC	GATATG	GATACA	GATACT	GATACC	GATACG	GATAGA	GATAGT	GATAGC
GAC	GACA	GACAT	GACAC	GACAG	GACATA	GACATT	GACATC	GACATG	GACACA	GACACT	GACACC	GACACG	GACAGA	GACAGT	GACAGC
GAG	GAGA	GAGAT	GAGAC	GAGAG	GAGATA	GAGATT	GAGATC	GAGATG	GAGACA	GAGACT	GAGACC	GAGACG	GAGAGA	GAGAGT	GAGAGC
GTA	GTAA	GTAAT	GTAAC	GTAG	GTATA	GTAATT	GTAATC	GTAATG	GTACA	GTACT	GTACC	GTAACG	GTAGA	GTAGT	GTAGC
GTT	GTTA	GTTAT	GTTAC	GTTAG	GTTTA	GTTTT	GTTTC	GTTTG	GTTCA	GTTCT	GTTCC	GTTTCG	GTTGA	GTTGT	GTTGC
GTC	GTCA	GTCAT	GTCAC	GTCAG	GCTA	GCTT	GCTC	GCTG	GTTCA	GTTCT	GTTCC	GTTTCG	GTTGA	GTTGT	GTTGC
GTG	GTGA	GTGAT	GTGAC	GTGAG	GTGTA	GTGTT	GTGTC	GTGTG	GTGCA	GTGCT	GTGCC	GTGCG	GTGGA	GTGGT	GTGGC
GCA	GCAA	GCAAT	GCAAC	GCAAG	GCAATA	GCAATT	GCAATC	GCAATG	GCAACA	GCAACT	GCAACC	GCAACG	GCAAGA	GCAAGT	GCAAGC
GCT	GCTA	GCTAT	GCTAC	GCTAG	GCTTA	GCTTT	GCTTC	GCTTG	GCTCA	GCTCT	GCTCC	GCTTCG	GCTGA	GCTGT	GCTGC
GCC	GCCA	GCCAT	GCCAC	GCCAG	GCCTA	GCCTT	GCCTC	GCCTG	GCCCA	GCCCT	GCCCC	GCCCG	GCCGA	GCCGT	GCCGC
GCS	GCSA	GCSAT	GCSAC	GCSAG	GCSATA	GCSATT	GCSATC	GCSATG	GCSACA	GCSACT	GCSACC	GCSACG	GCSAGA	GCSAGT	GCSAGC
GGA	GAAA	GAAAT	GAAAC	GAAAG	GAAATA	GAAATT	GAAATC	GAAATG	GAAACA	GAAACT	GAAACC	GAAACG	GAAAGA	GAAAGT	GAAAGC
GGT	GATA	GATAT	GATAC	GATAG	GATATA	GATATT	GATATC	GATATG	GATACA	GATACT	GATACC	GATACG	GATAGA	GATAGT	GATAGC
GGC	GACA	GACAT	GACAC	GACAG	GACATA	GACATT	GACATC	GACATG	GACACA	GACACT	GACACC	GACACG	GACAGA	GACAGT	GACAGC
GGS	GAGA	GAGAT	GAGAC	GAGAG	GAGATA	GAGATT	GAGATC	GAGATG	GAGACA	GAGACT	GAGACC	GAGACG	GAGAGA	GAGAGT	GAGAGC
GGA	GAAA	GAAAT	GAAAC	GAAAG	GAAATA	GAAATT	GAAATC	GAAATG	GAAACA	GAAACT	GAAACC	GAAACG	GAAAGA	GAAAGT	GAAAGC
GGA	GAAA	GAAAT	GAAAC	GAAAG	GAAATA	GAAATT	GAAATC	GAAATG	GAAACA	GAAACT	GAAACC	GAAACG	GAAAGA	GAAAGT	GAAAGC

```
fastp -i /TJPROJ4/X3/departament_data-nova/shared/231022_A01426_0733_AHNTWDBX5-new/F1M230026084-1A-7UXI0522-SUXI0522/F1M230026084-1A-7UXI0522-SUXI0522_I3_1.fq.gz -I /TJPROJ4/X3/departament_data-nova/shared/231022_A01426_0733_AHNTWDBX5-new/F1M230026084-1A-7UXI0522-SUXI0522/F1M230026084-1A-7UXI0522-SUXI0522_I3_2.fq.gz -o /TJPROJ13/CR/Production/produce/20231025000600/novopm2_tia_50/TXH230050813-1A/TXH230050813-1AG/01.GC/F1M230026084-1A-7UXI0522-SUXI0522_I3_1_clean.fq -o /TJPROJ13/CR/Production/produce/20231025000600/novopm2_tia_50/TXH230050813-1A/TXH230050813-1AG/01.GC/F1M230026084-1A-7UXI0522-SUXI0522_I3_2_clean.fq -h /TJPROJ13/CR/Production/produce/20231025000600/novopm2_tia_50/TXH230050813-1A/TXH230050813-1AG/01.GC/F1M230026084-1A-7UXI0522-SUXI0522_I3_fastp.html -j /TJPROJ13/CR/Production/produce/20231025000600/novopm2_tia_50/TXH230050813-1A/TXH230050813-1AG/01.GC/F1M230026084-1A-7UXI0522-SUXI0522_I3_fastp.json --detect_adapter_for_pe -w 12
```

fastp 0.21.0, at 2023-10-25 18:08:25

Figure S2. Continued.

Fastp report

Summary

General	
fastp version:	0.21.0 (https://github.com/OpenGene/fastp)
sequencing:	paired end (150 cycles + 150 cycles)
mean length before filtering:	150bp, 150bp
mean length after filtering:	148bp, 148bp
duplication rate:	21.392373%
Insert size peak:	203
Detected read1 adapter:	AGATCGGAAGAGCACACGTCTGAACTCCAGTCA
Detected read2 adapter:	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

Before filtering	
total reads:	22.537354 M
total bases:	3.380603 G
Q20 bases:	3.288208 G (97.266902%)
Q30 bases:	3.141848 G (92.937504%)
GC content:	43.437434%

After filtering	
total reads:	22.329500 M
total bases:	3.312397 G
Q20 bases:	3.231493 G (97.557544%)
Q30 bases:	3.090959 G (93.314853%)
GC content:	43.322502%

Filtering result	
reads passed filters:	22.329500 M (99.077736%)
reads with low quality:	205.514000 K (0.911882%)
reads with too many N:	782 (0.003470%)
reads too short:	1.558000 K (0.006913%)

Adapters

Adapter or bad ligation of read1

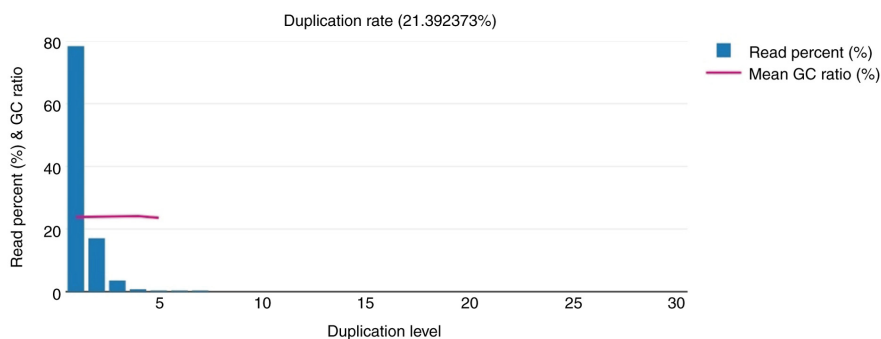
Sequence	Occurrences
A	29625
AG	28272
AGA	27542
AGAT	26912
AGATC	34975
AGATCG	25737
AGATCGG	23840
AGATCGGA	23315
AGATCGGAA	22118
AGATCGGAAG	21764
AGATCGGAAGA	20958
AGATCGGAAGAG	19817
AGATCGGAAGAGC	19116
AGATCGGAAGAGCA	18935
AGATCGGAAGAGCAC	18068
AGATCGGAAGAGCACAC	17356
AGATCGGAAGAGCACACAC	16622
AGATCGGAAGAGCACACAG	16029
AGATCGGAAGAGCACACAGT	15798
AGATCGGAAGAGCACACAGTC	14813
AGATCGGAAGAGCACACAGTCT	14326
AGATCGGAAGAGCACACAGTCTG	13939
AGATCGGAAGAGCACACAGTCTGA	13415
AGATCGGAAGAGCACACAGTCTGAA	12667
AGATCGGAAGAGCACACAGTCTGAAC	12425
AGATCGGAAGAGCACACAGTCTGAACT	11843
AGATCGGAAGAGCACACAGTCTGAACTC	11437
AGATCGGAAGAGCACACAGTCTGAACTCC	10871
AGATCGGAAGAGCACACAGTCTGAACTCCA	10445
AGATCGGAAGAGCACACAGTCTGAACTCCAG	10151
AGATCGGAAGAGCACACAGTCTGAACTCCAGT	9799
AGATCGGAAGAGCACACAGTCTGAACTCCAGTC	9914
AGATCGGAAGAGCACACAGTCTGAACTCCAGTCA	8820
AGATCGGAAGAGCACACAGTCTGAACTCCAGTCAC	8736
other adapter sequences	223213

Figure S2. Continued.

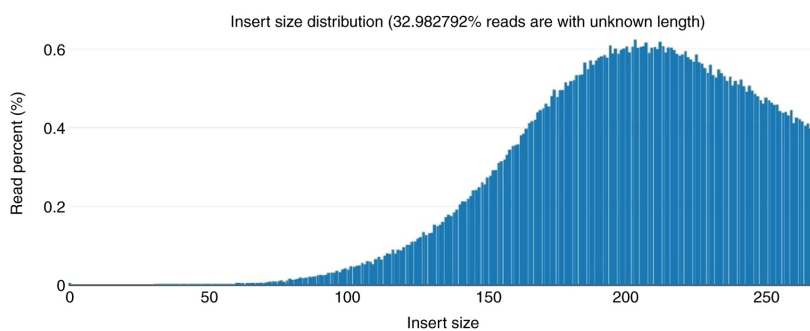
Adapter or bad ligation of read2

Sequence	Occurrences
A	29699
AG	28471
AGA	27777
AGAT	27214
AGATC	35399
AGATCG	26002
AGATCGG	24184
AGATCGGA	23629
AGATCGGAA	22020
AGATCGGAAG	21665
AGATCGGAAGA	20949
AGATCGGAAGAG	19943
AGATCGGAAGAGC	19089
AGATCGGAAGAGCG	18939
AGATCGGAAGAGCGCT	17968
AGATCGGAAGAGCGCTC	17341
AGATCGGAAGAGCGCTCG	16723
AGATCGGAAGAGCGCTCGT	15795
AGATCGGAAGAGCGCTCGTG	15631
AGATCGGAAGAGCGCTCGTGT	14556
AGATCGGAAGAGCGCTCGTGTG	13885
AGATCGGAAGAGCGCTCGTGTAG	13359
AGATCGGAAGAGCGCTCGTGTAGG	13004
AGATCGGAAGAGCGCTCGTGTAGGG	12224
AGATCGGAAGAGCGCTCGTGTAGGGA	11942
AGATCGGAAGAGCGCTCGTGTAGGGA	10840
AGATCGGAAGAGCGCTCGTGTAGGGA	10534
AGATCGGAAGAGCGCTCGTGTAGGGAAG	9981
AGATCGGAAGAGCGCTCGTGTAGGGAAGA	9773
AGATCGGAAGAGCGCTCGTGTAGGGAAGAG	9364
AGATCGGAAGAGCGCTCGTGTAGGGAAGAGT	8946
AGATCGGAAGAGCGCTCGTGTAGGGAAGAGTG	8533
other adapter sequences	245327

Duplication



Insert size estimation



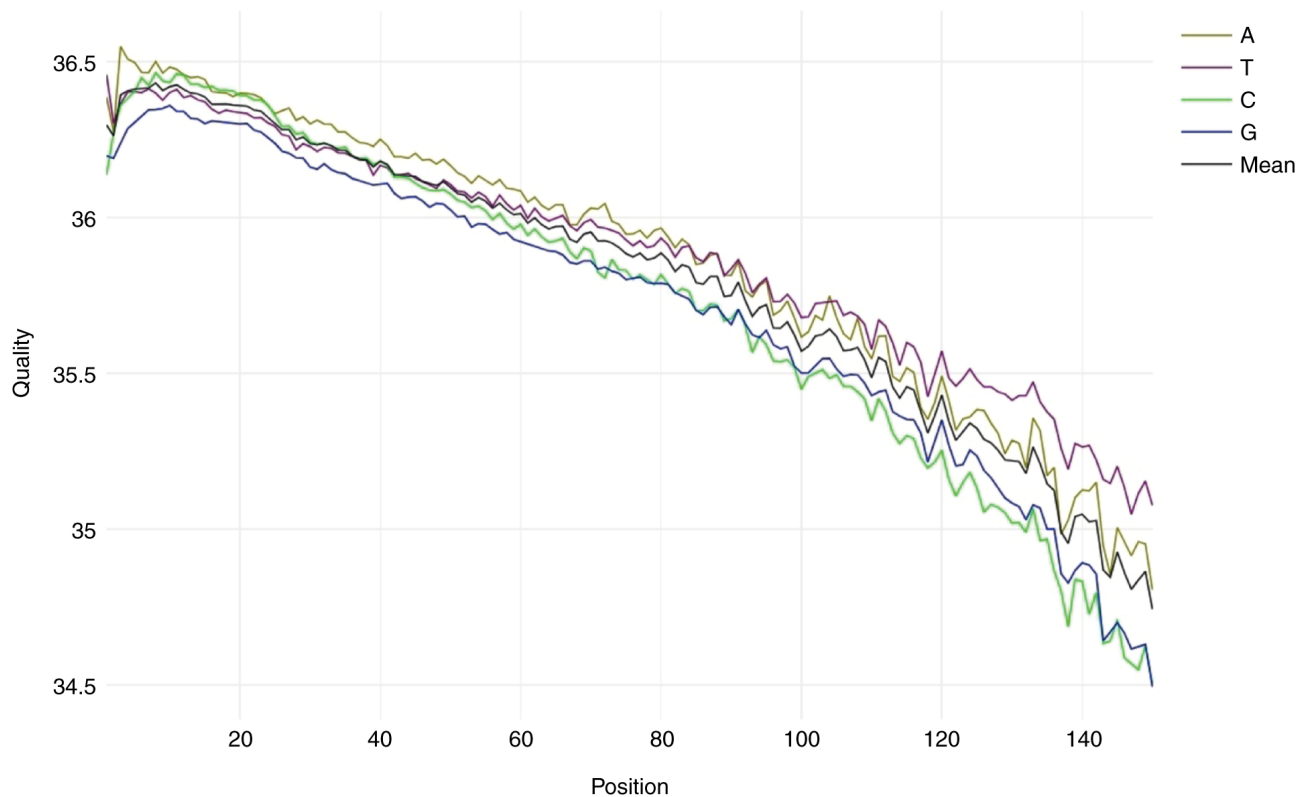
This estimation is based on paired-end overlap analysis, and there are 32.982792% reads found not overlapped. Nonoverlapped read pairs may have insert size <30 or >270, or contain too much sequencing errors to be detected as overlapped.

Figure S2. Continued.

Before filtering

Before filtering: read1: quality

Value of each position will be shown on mouse over.



Before filtering: read1: base contents

Value of each position will be shown on mouse over.

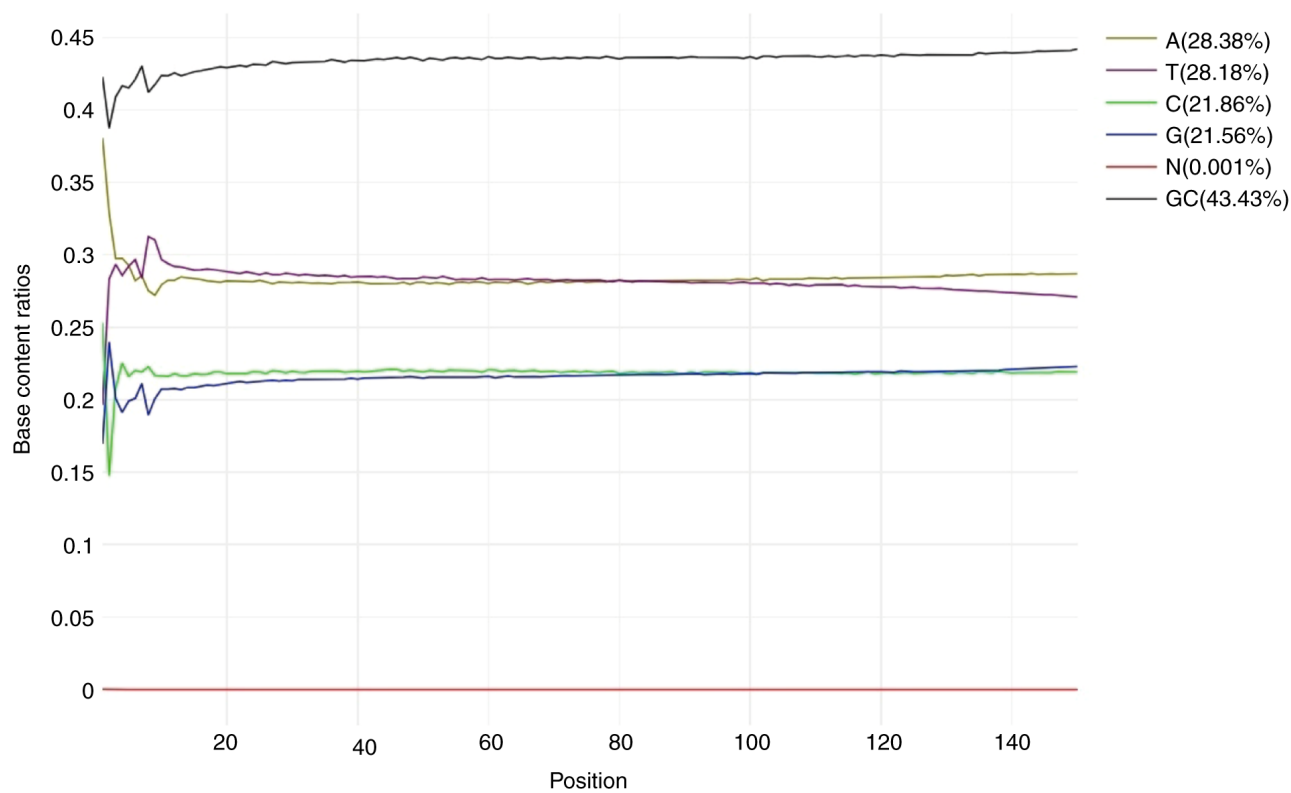


Figure S2. Continued.

Before filtering: read1: KMER counting

Darker background means larger counts. The count will be shown on mouse over.

	AA	AT	AC	AG	TA	TT	TC	TG	CA	CT	CC	CG	GA	GT	GC	GG
AAA	AAAA	AAAT	AAAC	AAAG	AAATA	AAATT	AAATC	AAATG	AAACA	AAACT	AAACC	AAACG	AAAGA	AAAGT	AAAGC	AAAGG
AAT	AAATA	AAATAT	AAATAC	AAATAG	AAATTA	AAATTT	AAATTC	AAATTG	AAATCA	AAATCT	AAATCC	AAATCG	AAATGA	AAATGT	AAATGC	AAATGG
AAC	AACAA	AACAT	AACAC	AACAG	AACATA	AACATT	AACATC	AACATG	AACACA	AACACT	AACACC	AACACG	AACAGA	AACAGT	AACAGC	AACAGG
AAG	AAGAA	AAGAT	AAGAC	AAGAG	AAGATA	AAGATT	AAGATC	AAGATG	AAGACA	AAGACT	AAGACC	AAGACG	AAGAGA	AAGAGT	AAGAGC	AAGAGG
ATA	ATAAA	ATAAT	ATAAC	ATAAG	ATATA	ATATT	ATATC	ATATG	ATACA	ATACT	ATACC	ATACG	ATAGA	ATAGT	ATAGC	ATAGG
ATT	ATTAA	ATTAT	ATTAC	ATTAG	ATTTA	ATTTT	ATTTC	ATTTG	ATTCA	ATTCT	ATTCC	ATTTCG	ATTGA	ATTGT	ATTGC	ATTGG
ATC	ATCAA	ATCAT	ATCAC	ATCAG	ATCATA	ATCATT	ATCATC	ATCATG	ATCACA	ATCACT	ATCACC	ATCACG	ATCAGA	ATCAGT	ATCAGC	ATCAGG
ATG	ATGAA	ATGAT	ATGAC	ATGAG	ATGATA	ATGATT	ATGATC	ATGATG	ATGACA	ATGACT	ATGACC	ATGACG	ATGAGA	ATGAGT	ATGAGC	ATGAGG
ACA	ACAAA	ACAAT	ACAAC	ACAAG	ACATA	ACATT	ACATC	ACATG	ACACA	ACACT	ACACC	ACACG	ACAGA	ACAGT	ACAGC	ACAGG
ACT	ACTAA	ACTAT	ACTAC	ACTAG	ACTTA	ACTTT	ACTTC	ACTTG	ACTCA	ACTCT	ACTCC	ACTTCG	ACTGA	ACTGT	ACTGC	ACTGG
ACC	ACCAA	ACCAT	ACCAC	ACCAG	ACCTA	ACCTT	ACCTC	ACCTG	ACCCA	ACCTCT	ACCTCC	ACCTCG	ACCTGA	ACCTGT	ACCTGC	ACCTGG
ACG	ACGAA	ACGAT	ACGAC	ACGAG	ACGATA	ACGATT	ACGATC	ACGATG	ACGACA	ACGACT	ACGACC	ACGACG	ACGAGA	ACGAGT	ACGAGC	ACGAGG
AGA	AGAAA	AGAAT	AGAAC	AGAAG	AGATA	AGATT	AGATC	AGATG	AGACA	AGACT	AGACC	AGACG	AGAGA	AGAGT	AGAGC	AGAGG
AGT	AGTAA	AGTAT	AGTAC	AGTAG	AGTTA	AGTTT	AGTTC	AGTTG	AGTCA	AGTCT	AGTCC	AGTTTCG	AGTGA	AGTGT	AGTGC	AGTGG
AGC	AGCAA	AGCAT	AGCAC	AGCAG	AGCTA	AGCTT	AGCTC	AGCTG	AGCCA	AGCTCT	AGCTCC	AGCTCG	AGCTGA	AGCTGT	AGCTGC	AGCTGG
AGG	AGGAA	AGGAT	AGGAC	AGGAG	AGGATA	AGGATT	AGGATC	AGGATG	AGGACA	AGGACT	AGGACC	AGGACG	AGGAGA	AGGAGT	AGGAGC	AGGAGG
TAA	TAAAA	TAAAT	TAAAC	TAAAG	TAATA	TAATT	TAATC	TAATG	TAAACA	TAAACT	TAAACC	TAAACG	TAAAGA	TAAAGT	TAAAGC	TAAAGG
TAT	TATAA	TATAT	TATAC	TATAG	TATTA	TATTT	TATTC	TATTG	TATCA	TATCT	TATCC	TATTCG	TATGA	TATGT	TATGC	TATGG
TAC	TACAA	TACAT	TACAC	TACAG	TACTA	TACTT	TACTC	TACTG	TACCA	TACTCT	TACTCC	TACTCG	TACTGA	TACTGT	TACTGC	TACTGG
TAG	TAGAA	TAGAT	TAGAC	TAGAG	TAGTA	TAGTT	TAGTC	TAGTG	TAGCA	TAGCT	TAGCC	TAGTCG	TAGGA	TAGGT	TAGGC	TAGGG
TTA	TTAAA	TTAAAT	TTAAAC	TTAAAG	TTATA	TTATT	TTATC	TTATG	TTACA	TTACT	TTACC	TTATTCG	TTTGA	TTTGT	TTTGC	TTTGG
TTT	TTTAA	TTTAT	TTTAC	TTTAG	TTTTA	TTTTT	TTTTC	TTTTG	TTTCA	TTTCT	TTTCC	TTTTCG	TTTGA	TTTGT	TTTGC	TTTGG
TTG	TTGAA	TTGAT	TTGAC	TTGAG	TTGTA	TTGTT	TTGTC	TTGTG	TTGCA	TTGCT	TTGCC	TTGTCG	TTTGA	TTTGT	TTTGC	TTTGG
TTG	TTGAA	TTGAT	TTGAC	TTGAG	TTGTA	TTGTT	TTGTC	TTGTG	TTGCA	TTGCT	TTGCC	TTGTCG	TTTGA	TTTGT	TTTGC	TTTGG
TGA	TGAAA	TGAAT	TGAAC	TGAAG	TGATA	TGATT	TGATC	TGATG	TGACA	TGACT	TGACC	TGACG	TGAGA	TGAGT	TGAGC	TGAGG
TGT	TGTAA	TGTAT	TGTAC	TGTAG	TGTTA	TGTTT	TGTTC	TGTTG	TGTCA	TGTCT	TGTCC	TGTTCG	TGTGA	TGTGT	TGTGC	TGTGG
TGC	TGCAA	TGCAT	TGCAC	TGCAG	TGCTA	TGCTT	TGCTC	TGCTG	TGCCA	TGCTCT	TGCTCC	TGCTTCG	TGCTGA	TGCTGT	TGCTGC	TGCTGG
TGC	TGCAA	TGCAT	TGCAC	TGCAG	TGCTA	TGCTT	TGCTC	TGCTG	TGCCA	TGCTCT	TGCTCC	TGCTTCG	TGCTGA	TGCTGT	TGCTGC	TGCTGG
TGG	TGAAA	TGAAT	TGAAC	TGAAG	TGATA	TGATT	TGATC	TGATG	TGACA	TGACT	TGACC	TGACG	TGAGA	TGAGT	TGAGC	TGAGG
TGA	TGAAA	TGAAT	TGAAC	TGAAG	TGATA	TGATT	TGATC	TGATG	TGACA	TGACT	TGACC	TGACG	TGAGA	TGAGT	TGAGC	TGAGG
TGT	TGTAA	TGTAT	TGTAC	TGTAG	TGTTA	TGTTT	TGTTC	TGTTG	TGTCA	TGTCT	TGTCC	TGTTCG	TGTGA	TGTGT	TGTGC	TGTGG
TGC	TGCAA	TGCAT	TGCAC	TGCAG	TGCTA	TGCTT	TGCTC	TGCTG	TGCCA	TGCTCT	TGCTCC	TGCTTCG	TGCTGA	TGCTGT	TGCTGC	TGCTGG
TGG	TGAAA	TGAAT	TGAAC	TGAAG	TGATA	TGATT	TGATC	TGATG	TGACA	TGACT	TGACC	TGACG	TGAGA	TGAGT	TGAGC	TGAGG
CAA	CAAAA	CAAAAT	CAAAAC	CAAAAG	CAATA	CAATT	CAATC	CAATG	CAACA	CAACT	CAACC	CAACG	CAAGA	CAAGT	CAAGC	CAAGG
CAT	CATAA	CATAT	CATAC	CATAG	CATTA	CATTT	CATTC	CATTG	CATCA	CATCT	CATCC	CATTCG	CATGA	CATGT	CATGC	CATGG
CAC	CACAA	CACAT	CACAC	CACAG	CACATA	CACATT	CACATC	CACATG	CACACA	CACACT	CACACC	CACACG	CACAGA	CACAGT	CACAGC	CACAGG
CAG	CAGAA	CAGAT	CAGAC	CAGAG	CAGTA	CAGTT	CAGTC	CAGTG	CAGCA	CAGCT	CAGCC	CAGTCG	CAGGA	CAGGT	CAGGC	CAGGG
CTA	CTAAA	CTAAT	CTAAC	CTAAG	CTATA	CTATT	CTATC	CTATG	CTACA	CTACT	CTACC	CTATTCG	CTTGA	CTTGT	CTTGC	CTTGG
CTT	CTTAA	CTTAT	CTTAC	CTTAG	CTTTA	CTTTT	CTTTC	CTTTG	CTTCA	CTTCT	CTTCC	CTTTCG	CTTGA	CTTGT	CTTGC	CTTGG
CTC	CTCAA	CTCAT	CTCAC	CTCAG	CTCTA	CTCTT	CTCTC	CTCTG	CTCCA	CTCTCT	CTCTCC	CTCTTCG	CTCTGA	CTCTGT	CTCTGC	CTCTGG
CTG	CTGAA	CTGAT	CTGAC	CTGAG	CTGTA	CTGTT	CTGTC	CTGTG	CTGCA	CTGCT	CTGCC	CTGTCG	CTTGA	CTTGT	CTTGC	CTTGG
CCA	CCAAA	CCAAAT	CCAAAC	CCAAAG	CCATA	CCATT	CCATC	CCATG	CCACA	CCACT	CCACC	CCACG	CCAGA	CCAGT	CCAGC	CCAGG
CCT	CCTAA	CCTAT	CCTAC	CCTAG	CCTTA	CCTTT	CCTTC	CCTTG	CCTCA	CCTCT	CCTCC	CCTTCG	CCTGA	CCTGT	CCTGC	CCTGG
CCC	CCCAA	CCCAAT	CCCAAC	CCCAAG	CCATA	CCATT	CCATC	CCATG	CCACA	CCACT	CCACC	CCACG	CCAGA	CCAGT	CCAGC	CCAGG
CCG	CCGAA	CCGAT	CCGAC	CCGAG	CCGTA	CCGTT	CCGTC	CCGTG	CCGCA	CCGCT	CCGCC	CCGTCG	CCGGA	CCGGT	CCGGC	CCGGG
CCG	CCGAA	CCGAT	CCGAC	CCGAG	CCGTA	CCGTT	CCGTC	CCGTG	CCGCA	CCGCT	CCGCC	CCGTCG	CCGGA	CCGGT	CCGGC	CCGGG
CGA	CGAAA	CGAAT	CGAAC	CGAAG	CGATA	CGATT	CGATC	CGATG	CGACA	CGACT	CGACC	CGACG	CGAGA	CGAGT	CGAGC	CGAGG
CGT	CGTAA	CGTAT	CGTAC	CGTAG	CGTTA	CGTTT	CGTTC	CGTTG	CGTCA	CGTCT	CGTCC	CGTTCG	CGTGA	CGTGT	CGTGC	CGTGG
CGC	CGCAA	CGCAT	CGCAC	CGCAG	CGCTA	CGCTT	CGCTC	CGCTG	CGCCA	CGCTCT	CGCTCC	CGCTTCG	CGCTGA	CGCTGT	CGCTGC	CGCTGG
CGG	CGGAA	CGGAT	CGGAC	CGGAG	CGGTA	CGGTT	CGGTC	CGGTG	CGGCA	CGGCT	CGGCC	CGGTCG	CGGGA	CGGGT	CGGGC	CGGGG
GAA	GAAAA	GAAAT	GAAAC	GAAAG	GAATA	GAATT	GAATC	GAATG	GAACA	GAACT	GAACC	GAACG	GAAGA	GAAGT	GAAGC	GAAGG
GAT	GATAA	GATAT	GATAC	GATAG	GATTA	GATTT	GATTC	GATTG	GATCA	GATCT	GATCC	GATTCG	GATGA	GATGT	GATGC	GATGG
GAC	GACAA	GACAT	GACAC	GACAG	GACTA	GACTT	GACTC	GACTG	GACCA	GACTCT	GACTCC	GACTTCG	GACTGA	GACTGT	GACTGC	GACTGG
GAG	GAGAA	GAGAT	GAGAC	GAGAG	GAGTA	GAGTT	GAGTC	GAGTG	GAGCA	GAGCT	GAGCC	GAGTCG	GAGGA	GAGGT	GAGGC	GAGGG
GTA	GTAOA	GTAAT	GTAAC	GTAAG	GTATA	GTATT	GTATC	GTATG	GTACA	GTACT	GTACC	GTATTCG	GTAGA	GTAGT	GTAGC	GTAGG
GTT	GTTAA	GTTAT	GTTAC	GTTAG	GTTTA	GTTTT	GTTTC	GTTTG	GTTCA	GTTCT	GTTCC	GTTTCG	GTTGA	GTTGT	GTTGC	GTTGG
GTC	GTCOA	GTCAT	GTCAC	GTCAG	GCTTA	GCTTT	GCTTC	GCTTG	GTTCA	GTTCT	GTTCC	GTTTCG	GTTGA	GTTGT	GTTGC	GTTGG
GTC	GTCOA	GTCAT	GTCAC	GTCAG	GCTTA	GCTTT	GCTTC	GCTTG	GTTCA	GTTCT	GTTCC	GTTTCG	GTTGA	GTTGT	GTTGC	GTTGG
GTG	GTAAA	GTAAAT	GTAAAC	GTAAAG	GTATA	GTATT	GTATC	GTATG	GTACA	GTACT	GTACC	GTATTCG	GTAGA	GTAGT	GTAGC	GTAGG
GCA	GCAAA	GCAAT	GCAAC	GCAAG	GCATA	GCATT	GCATC	GCATG	GCACA	GCACT	GCACC	GCACG	GCAGA	GCAGT	GCAGC	GCAGG
GCT	GCTAA	GCTAT	GCTAC	GCTAG	GCTTA	GCTTT	GCTTC	GCTTG	GCTCA	GCTCT	GCTCC	GCTTCG	GCTGA	GCTGT	GCTGC	GCTGG
GCC	GCCAA	GCCAT	GCCAC	GCCAG	GCCTA	GCCTT	GCCTC	GCCTG	GCCCA	GCCCT	GCCCC	GCCTCG	GCCGA	GCCGT	GCCGC	GCCGG
GCG	GCGAA	GCGAT	GCGAC	GCGAG	GCGTA	GCGTT	GCGTC	GCGTG	GCGCA	GCGCT	GCGCC	GCGTCG	GCGGA	GCGGT	GCGGC	GCGGG
GGA	GGAOA	GGAAT	GGAAC	GGAAG	GGATA	GGATT	GGATC	GGATG	GGACA	GGACT	GGACC	GGATTCG	GGAGA	GGAGT	GGAGC	GGAGG
GGT	GGTAA	GGTAT	GGTAC	GGTAG	GGTTA	GGTTT	GGTTC	GGTTG	GGTCA	GGTCT	GGTCC	GGTTCG	GGTGA	GGTGT	GGTGC	GGTGG
GGC	GGCAA	GGCAT	GGCAC	GGCAG	GGCTA	GGCTT	GGCTC	GGCTG	GGCCA	GGCTCT	GGCTCC	GGCTTCG	GGCTGA	GGCTGT	GGCTGC	GGCTGG
GGG	GGGAA	GGGAT	GGGAC	GGGAG	GGGTA	GGGTT	GGGTC	GGGTG	GGGCA	GGGCT	GGGCC	GGGTCG	GGGGA	GGGGT	GGGGC	GGGGG

Before filtering: read2: quality

Value of each position will be shown on mouse over.

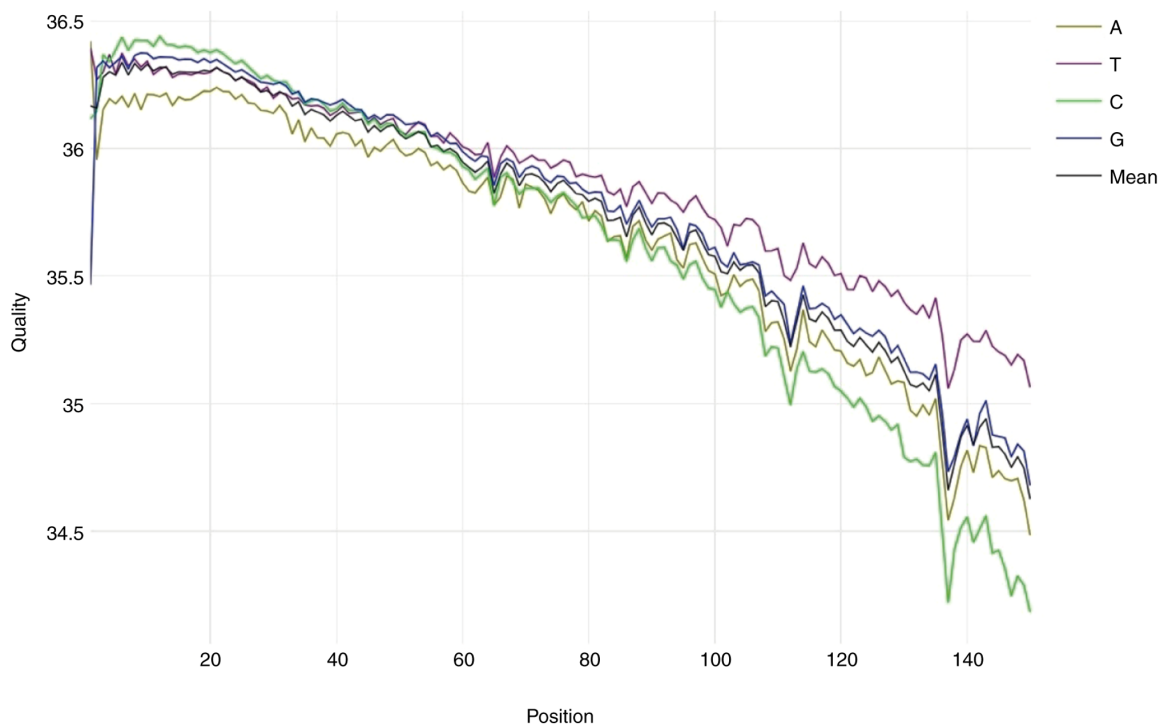
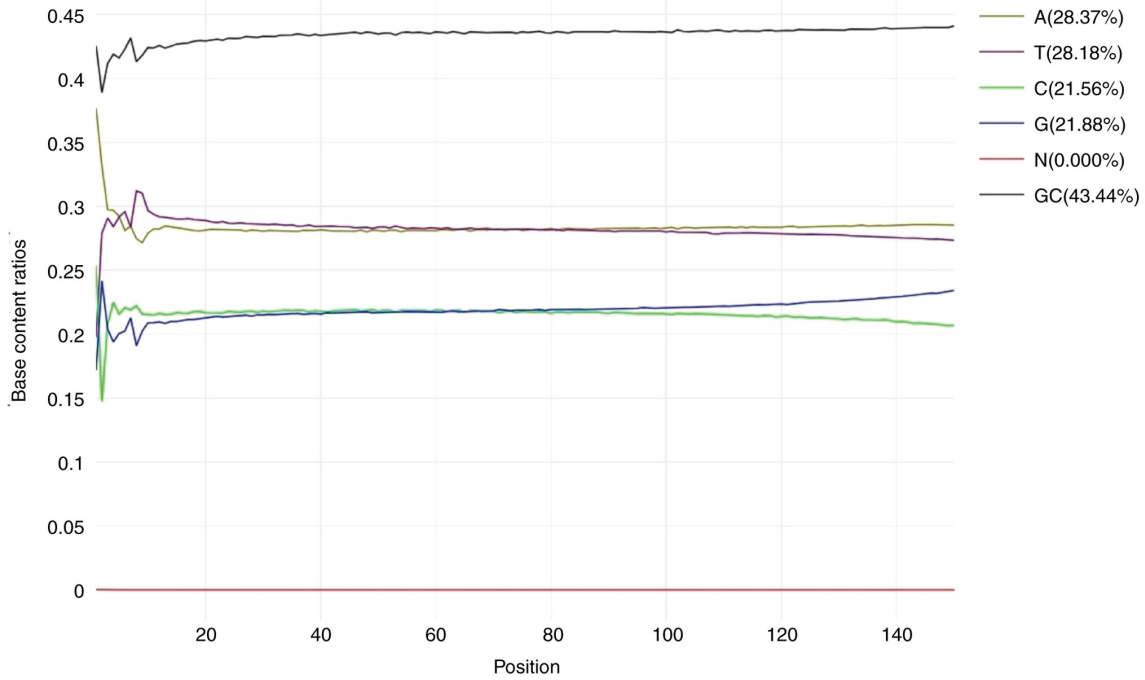


Figure S2. Continued.

Before filtering: read2: base contents
Value of each position will be shown on mouse over.



Before filtering: read2: KMER counting
Darker background means larger counts. The count will be shown on mouse over.

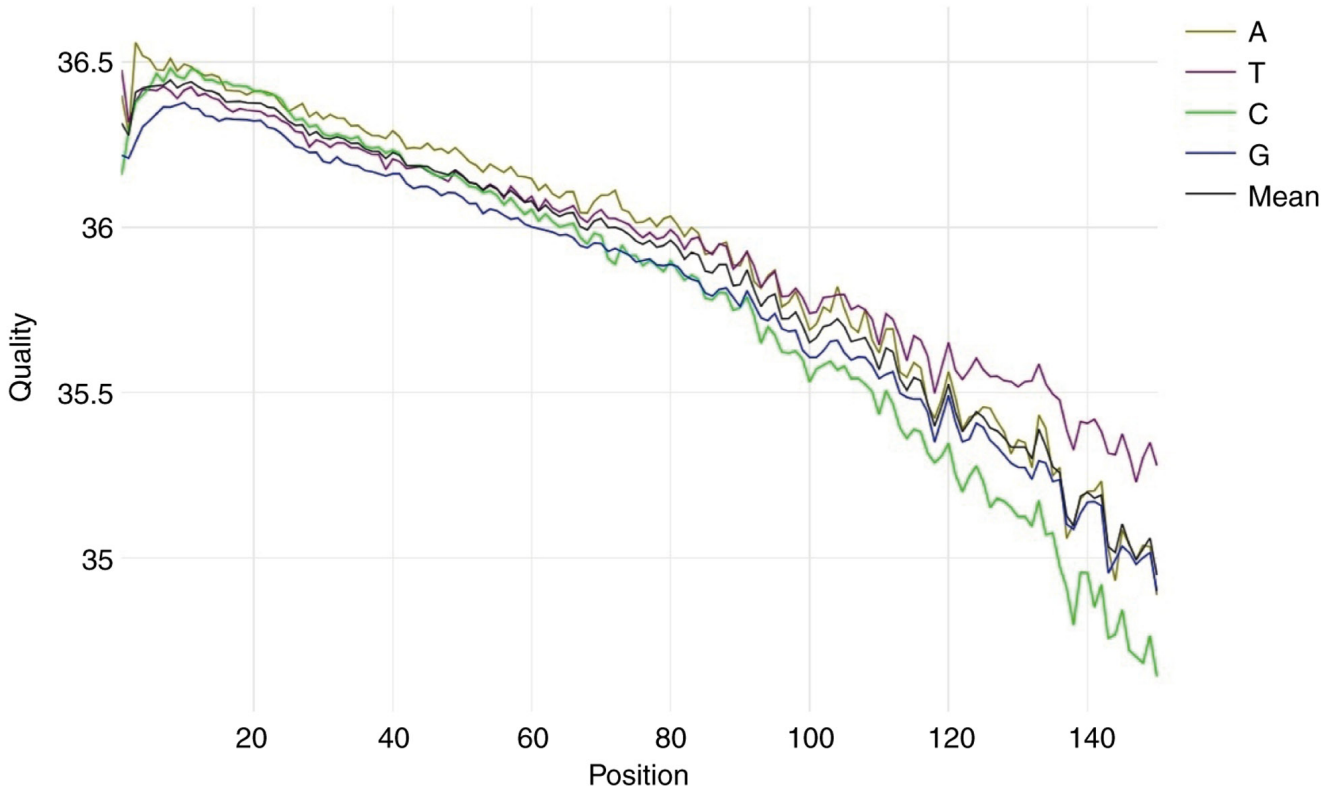
	AA	AT	AC	AG	TA	TT	TC	TG	CA	CT	CC	CG	GA	GT	GC	GG
AAA	AAAA	AAAAT	AAAAC	AAANG	AAATA	AAATT	AAATC	AAATG	AAACA	AAACT	AAACC	AAACG	AAAGA	AAAGT	AAAGC	AAAGG
AAI	AAATA	AAATAT	AAATAC	AAATAG	AAATTA	AAATTT	AAATTC	AAATTG	AAATCA	AAATCT	AAATCC	AAATCG	AAATGA	AAATGT	AAATGC	AAATGG
AAC	AACAA	AACAAT	AACAC	AACAG	AACATA	AACATT	AACATC	AACATG	AACACA	AACACT	AACACC	AACACG	AACAGA	AACAGT	AACAGC	AACAGG
AAG	AAGAA	AAGAAT	AAGAC	AAGAG	AAGATA	AAGATT	AAGATC	AAGATG	AAGACA	AAGACT	AAAGCC	AAAGCG	AAAGGA	AAAGGT	AAAGGC	AAAGGG
ATA	ATAAA	ATAAAT	ATAAC	ATAAG	ATAATA	ATAATT	ATAATC	ATAATG	ATAACA	ATAACT	ATAACC	ATAACG	ATAAGA	ATAAGT	ATAAGC	ATAAGG
ATT	ATTAA	ATTAAAT	ATTAC	ATTAG	ATTATA	ATTATT	ATTATC	ATTATG	ATTACA	ATTACT	ATTACC	ATTACG	ATTAGA	ATTAGT	ATTAGC	ATTAGG
ATC	ATCAA	ATCAAT	ATCAC	ATCAG	ATCATA	ATCATT	ATCATC	ATCATG	ATCACA	ATCACT	ATCACC	ATCACG	ATCAGA	ATCAGT	ATCAGC	ATCAGG
ATG	ATGAA	ATGAAT	ATGAC	ATGAG	ATGATA	ATGATT	ATGATC	ATGATG	ATGACA	ATGACT	ATGACC	ATGACG	ATGAGA	ATGAGT	ATGAGC	ATGAGG
ACA	ACAAA	ACAAAT	ACAAC	ACAAG	ACATA	ACATT	ACATC	ACATG	ACACA	ACACT	ACACC	ACACG	ACAGA	ACAGT	ACAGC	ACAGG
ACT	ACTAA	ACTAAT	ACTAC	ACTAG	ACTATA	ACTATT	ACTATC	ACTATG	ACTACA	ACTACT	ACTACC	ACTACG	ACTAGA	ACTAGT	ACTAGC	ACTAGG
ACC	ACCAA	ACCAAT	ACCAC	ACCAG	ACCTA	ACCTT	ACCTC	ACCTG	ACCCA	ACCCCT	ACCCCC	ACCCCG	ACCGA	ACCGT	ACCGC	ACCGG
ACG	ACGAA	ACGAAT	ACGAC	ACGAG	ACGATA	ACGATT	ACGATC	ACGATG	ACGACA	ACGACT	ACGACC	ACGACG	ACGAGA	ACGAGT	ACGAGC	ACGAGG
AGA	AGAAA	AGAAAT	AGAAC	AGAG	AGATA	AGATT	AGATC	AGATG	AGACA	AGACT	AGACC	AGACG	AGAGA	AGAGT	AGAGC	AGAGG
AGT	AGTAA	AGTAAAT	AGTAC	AGTAG	AGTATA	AGTATT	AGTATC	AGTATG	AGTACA	AGTACT	AGTACC	AGTACG	AGTAGA	AGTAGT	AGTAGC	AGTAGG
AGC	AGCAA	AGCAAT	AGCAC	AGCAG	AGCTA	AGCTT	AGCTC	AGCTG	AGCCA	AGCCCT	AGCCCC	AGCCCG	AGCGA	AGCGT	AGCGC	AGCGG
AGG	AGGAA	AGGAAT	AGGAC	AGGAG	AGGATA	AGGATT	AGGATC	AGGATG	AGGACA	AGGACT	AGGACC	AGGACG	AGGAGA	AGGAGT	AGGAGC	AGGAGG
TAA	TAAAA	TAAAAT	TAAAC	TAAAG	TAAATA	TAAATT	TAAATC	TAAATG	TAAACA	TAAACT	TAAACC	TAAACG	TAAAGA	TAAAGT	TAAAGC	TAAAGG
TAT	TATAA	TATAAT	TATAC	TATAG	TATATA	TATATT	TATATC	TATATG	TATACA	TATACT	TATACC	TATACG	TATAGA	TATAGT	TATAGC	TATAGG
TAC	TACAA	TACAAT	TACAC	TACAG	TACTA	TACTT	TACTC	TACTG	TACCA	TACCT	TACCC	TACCCG	TACCGA	TACCGT	TACCGC	TACCGG
TAG	TAGAA	TAGAAAT	TAGAC	TAGAG	TAGATA	TAGATT	TAGATC	TAGATG	TAGACA	TAGACT	TAGACC	TAGACG	TAGAGA	TAGAGT	TAGAGC	TAGAGG
TAA	TAAAA	TAAAAT	TAAAC	TAAAG	TAAATA	TAAATT	TAAATC	TAAATG	TAAACA	TAAACT	TAAACC	TAAACG	TAAAGA	TAAAGT	TAAAGC	TAAAGG
TAT	TATAA	TATAAT	TATAC	TATAG	TATATA	TATATT	TATATC	TATATG	TATACA	TATACT	TATACC	TATACG	TATAGA	TATAGT	TATAGC	TATAGG
TAC	TACAA	TACAAT	TACAC	TACAG	TACTA	TACTT	TACTC	TACTG	TACCA	TACCT	TACCC	TACCCG	TACCGA	TACCGT	TACCGC	TACCGG
TAG	TAGAA	TAGAAAT	TAGAC	TAGAG	TAGATA	TAGATT	TAGATC	TAGATG	TAGACA	TAGACT	TAGACC	TAGACG	TAGAGA	TAGAGT	TAGAGC	TAGAGG
TAA	TAAAA	TAAAAT	TAAAC	TAAAG	TAAATA	TAAATT	TAAATC	TAAATG	TAAACA	TAAACT	TAAACC	TAAACG	TAAAGA	TAAAGT	TAAAGC	TAAAGG
TAT	TATAA	TATAAT	TATAC	TATAG	TATATA	TATATT	TATATC	TATATG	TATACA	TATACT	TATACC	TATACG	TATAGA	TATAGT	TATAGC	TATAGG
TTC	TTCAA	TTCAAT	TTCAC	TTCAG	TTCATA	TTCATT	TTCATC	TTCATG	TTCACA	TTCACT	TTCACC	TTCACG	TTCAGA	TTCAGT	TTCAGC	TTCAGG
TTG	TTGAA	TTGAAAT	TTGAC	TTGAG	TTGATA	TTGATT	TTGATC	TTGATG	TTGACA	TTGACT	TTGACC	TTGACG	TTGAGA	TTGAGT	TTGAGC	TTGAGG
TCA	TCAAA	TCAAAT	TCAAC	TCAAG	TCATA	TCAATT	TCAATC	TCAATG	TCACA	TCACT	TCACC	TCACCG	TCCGA	TCCGT	TCCGC	TCCGG
TCT	TCTAA	TCTAAT	TCTAC	TCTAG	TCTATA	TCTATT	TCTATC	TCTATG	TCTACA	TCTACT	TCTACC	TCTACG	TCTAGA	TCTAGT	TCTAGC	TCTAGG
TCC	TCCAA	TCCAAT	TCCAC	TCCAG	TCCATA	TCCATT	TCCATC	TCCATG	TCCACA	TCCACT	TCCACC	TCCACG	TCCAGA	TCCAGT	TCCAGC	TCCAGG
TCC	TCCAA	TCCAAT	TCCAC	TCCAG	TCCATA	TCCATT	TCCATC	TCCATG	TCCACA	TCCACT	TCCACC	TCCACG	TCCAGA	TCCAGT	TCCAGC	TCCAGG
TGA	TGAAA	TGAAAT	TGAAC	TGAG	TGATA	TGATT	TGATC	TGATG	TGACA	TGACT	TGACC	TGACG	TGAGA	TGAGT	TGAGC	TGAGG
TGT	TGTAA	TGTAAAT	TGTAC	TGTAG	TGTATA	TGTATT	TGTATC	TGTATG	TGTACA	TGTACT	TGTACC	TGTACG	TGTAGA	TGTAGT	TGTAGC	TGTAGG
TGC	TGCAA	TGCAAT	TGCAC	TGCAG	TGCTA	TGCTT	TGCTC	TGCTG	TGCCA	TGCCCT	TGCCCC	TGCCCG	TGCCGA	TGCCGT	TGCCGC	TGCCGG
TGG	TGAAA	TGAAAT	TGAAC	TGAG	TGATA	TGATT	TGATC	TGATG	TGACA	TGACT	TGACC	TGACG	TGAGA	TGAGT	TGAGC	TGAGG
CAA	CAAAA	CAAAAT	CAAAC	CAAG	CAATA	CAATT	CAATC	CAATG	CAACA	CAACT	CAACC	CAACG	CAAGA	CAAGT	CAAGC	CAAGG
CAI	CATAA	CATAAT	CATAC	CATAG	CATATA	CATATT	CATATC	CATATG	CATACA	CATACT	CATACC	CATACG	CATAGA	CATAGT	CATAGC	CATAGG
CAC	CACAA	CACAAT	CACAC	CACAG	CACTA	CACTT	CACTC	CACTG	CACCA	CACCT	CACCC	CACCCG	CACCGA	CACCGT	CACCGC	CACCGG
CAG	CAGAA	CAGAAAT	CAGAC	CAGAG	CAGATA	CAGATT	CAGATC	CAGATG	CAGACA	CAGACT	CAGACC	CAGACG	CAGAGA	CAGAGT	CAGAGC	CAGAGG
CTA	CTAAA	CTAAAT	CTAAC	CTAG	CTATA	CTATT	CTATC	CTATG	CTACA	CTACT	CTACC	CTACG	CTAGA	CTAGT	CTAGC	CTAGG
CTI	CTTAA	CTTAAAT	CTTAC	CTTAG	CTTATA	CTTATT	CTTATC	CTTATG	CTTACA	CTTACT	CTTACC	CTTACG	CTTAGA	CTTAGT	CTTAGC	CTTAGG
CTC	CTCAA	CTCAAT	CTCAC	CTCAG	CITTA	CITTT	CITTC	CITTG	CTCCA	CTCCT	CTCCC	CTCCCG	CTCCGA	CTCCGT	CTCCGC	CTCCGG
CTG	CTGAA	CTGAAAT	CTGAC	CTGAG	CITGTA	CITGTT	CITGTC	CITGTG	CTGCA	CTGCT	CTGCC	CTGCCG	CTGCCGA	CTGCCGT	CTGCCGC	CTGCCGG
CCA	CCAAA	CCAAAT	CCAAC	CCAG	CCATA	CCATT	CCATC	CCATG	CCACA	CCACT	CCACC	CCACG	CCAGA	CCAGT	CCAGC	CCAGG
CCI	CCTAA	CCTAAT	CCTAC	CCTAG	CCTATA	CCTATT	CCTATC	CCTATG	CCTACA	CCTACT	CCTACC	CCTACG	CCTAGA	CCTAGT	CCTAGC	CCTAGG
CCC	CCCAA	CCCAAT	CCCAC	CCCG	CCCTA	CCCTT	CCCTC	CCCTG	CCCCA	CCCCCT	CCCCC	CCCCCG	CCCCGA	CCCCGT	CCCCGC	CCCCGG
CCG	CCGAA	CCGAAT	CCGAC	CCGAG	CCGTA	CCGTT	CCGTC	CCGTG	CCGCA	CCGCT	CCGCC	CCGCCG	CCGCCGA	CCGCCGT	CCGCCGC	CCGCCGG
CGA	CGAAA	CGAAAT	CGAAC	CGAG	CGATA	CGATT	CGATC	CGATG	CGACA	CGACT	CGACC	CGACG	CGAGA	CGAGT	CGAGC	CGAGG
CGI	CGTAA	CGTAAAT	CGTAC	CGTAG	CGTATA	CGTATT	CGTATC	CGTATG	CGTACA	CGTACT	CGTACC	CGTACG	CGTAGA	CGTAGT	CGTAGC	CGTAGG
CGC	CGCAA	CGCAAT	CGCAC	CGCAG	CGCTA	CGCTT	CGCTC	CGCTG	CGCCA	CGCCT	CGCCC	CGCCCG	CGCCGA	CGCCGT	CGCCGC	CGCCGG
CGG	CGGAA	CGGAAAT	CGGAC	CGGAG	CGGTA	CGGTT	CGGTC	CGGTG	CGGCA	CGGCT	CGGCC	CGGCCG	CGGCCGA	CGGCCGT	CGGCCGC	CGGCCGG
GAA	GAAAA	GAAAAT	GAAAC	GAAAG	GAAATA	GAAATT	GAAATC	GAAATG	GAAACA	GAAACT	GAAACC	GAAACG	GAAAGA	GAAAGT	GAAAGC	GAAAGG
GAI	GATAA	GATAAT	GATAC	GATAG	GATATA	GATATT	GATATC	GATATG	GATACA	GATACT	GATACC	GATACG	GATAGA	GATAGT	GATAGC	GATAGG
GAC	GACAA	GACAAT	GACAC	GACAG	GACTA	GACTT	GACTC	GACTG	GACCA	GACCT	GACCC	GACCCG	GACCGA	GACCGT	GACCGC	GACCGG
GAG	GAGAA	GAGAAAT	GAGAC	GAGAG	GAGATA	GAGATT	GAGATC	GAGATG	GAGACA	GAGACT	GAGACC	GAGACG	GAGAGA	GAGAGT	GAGAGC	GAGAGG
GTA	GTA AA	GTA AAT	GTA AC	GTA AG	GTA ATA	GTA ATT	GTA ATC	GTA ATG	GTA ACA	GTA ACT	GTA ACC	GTA ACG	GTA AGA	GTA AGT	GTA AGC	GTA AGG
GTT	GTTAA	GTTAAAT	GTTAC	GTTAG	GTTATA	GTTATT	GTTATC	GTTATG	GTTACA	GTTACT	GTTACC	GTTACG	GTTAGA	GTTAGT	GTTAGC	GTTAGG
GTC	GTCAA	GTCAAT	GTCAC	GTCAG	GICTA	GICTT	GICTC	GICTG	GITCA	GICTT	GITCC	GITCCG	GITCGA	GITCGT	GITCGC	GITCGG
GTG	GTGAA	GTGAAAT	GTGAC	GTGAG	GITGTA	GITGTT	GITGTC	GITGTG	GITGA	GITGT	GITGC	GITGCC	GITGCCA	GITGCCGT	GITGCCGC	GITGCCGG
GCA	GCAAA	GCAAAT	GCAAC	GCAAG	GCAATA	GCAATT	GCAATC	GCAATG	GCAACA	GCAACT	GCAACC	GCAACG	GCAAGA	GCAAGT	GCAAGC	GCAAGG
GCT	GCTAA	GCTAAT	GCTAC	GCTAG	GCTATA	GCTATT	GCTATC	GCTATG	GCTACA	GCTACT	GCTACC	GCTACG	GCTAGA	GCTAGT	GCTAGC	GCTAGG
GCC	GCCAA	GCCAAT	GCCAC	GCCAG	GCCTA	GCCTT	GCCTC	GCCTG	GCCCA	GCCCT	GCCCC	GCCCCG	GCCCGA	GCCCGT	GCCCGC	GCCCGG
GCG	GCGAA	GCGAAT	GCGAC	GCGAG	GCGTA	GCGTT	GCGTC	GCGTG	GCGCA	GCGCT	GCGCC	GCGCCG	GCGCCGA	GCGCCGT	GCGCCGC	GCGCCGG
GGA	GGA AA	GGA AAT	GGA AC	GGA AG	GGA ATA	GGA ATT	GGA ATC	GGA ATG	GGA ACA	GGA ACT	GGA ACC	GGA ACG	GGA AGA	GGA AGT	GGA AGC	GGA AGG
GGI	GGTAA	GGTAAAT	GGTAC	GGTAG	GGTATA	GGTATT	GGTATC	GGTATG	GGTACA	GGTACT	GGTACC	GGTACG	GGTAGA	GGTAGT	GGTAGC	GGTAGG
GGC	GCCAA	GCCAAT	GCCAC	GCCAG	GGCTA	GGCTT	GGCTC	GGCTG	GGCCA	GGCCT	GGCCC	GGCCCG	GGCCGA	GGCCGT	GGCCGC	GGCCGG
GGG	GGGAA	GGGAAAT	GGGAC	GGGAG	GGGTA	GGGTT	GGGTC	GGGTG	GGGCA	GGGCT	GGGCC	GGGCCG	GGGCCGA	GGGCCGT	GGGCCGC	GGGCCGG

Figure S2. Continued.

After filtering

After filtering: read1: quality

Value of each position will be shown on mouse over.



After filtering: read1: base contents

Value of each position will be shown on mouse over.

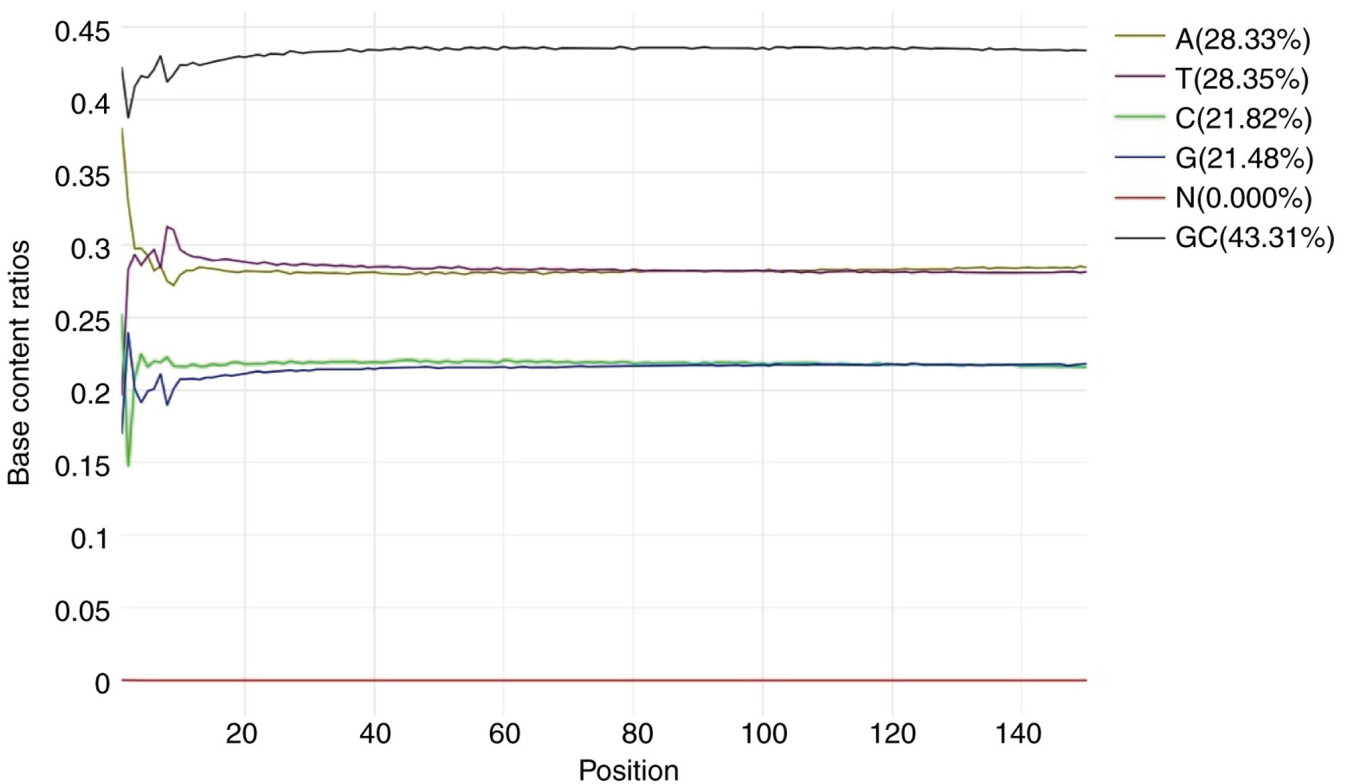


Figure S2. Continued.

After filtering: read1: KMER counting

Darker background means larger counts. The count will be shown on mouse over.

	AA	AT	AC	AG	TA	TT	TC	TS	CA	CT	CC	CG	GA	GT	GC	GG
AAA	AAAA	AAAT	AAAC	AAAG	AAATA	AAATT	AAATC	AAATG	AAACA	AAACT	AAACC	AAACG	AAAGA	AAAGT	AAAGC	AAAGG
AAT	AAATA	AAAT	AAATC	AAATG	AAATA	AAATT	AAATC	AAATG	AAATA	AAACT	AAACC	AAACG	AAAGA	AAAGT	AAAGC	AAAGG
AAC	AAACA	AACAT	AACAC	AACAG	AACATA	AACATT	AACTC	AACTG	AACCA	AACCT	AACCC	AACCG	AACGA	AACGT	AACGC	AACGG
AAG	AAAGA	AAGAT	AAGAC	AAGAG	AAGATA	AAGATT	AAGTC	AAGTG	AAGCA	AAGCT	AAGCC	AAGCG	AAGGA	AAGGT	AAGGC	AAGGG
ATA	ATAAA	ATAAT	ATAAC	ATAAG	ATAATA	ATAATT	ATAATC	ATAATG	ATAACA	ATAACT	ATAACC	ATAACG	ATAAGA	ATAAGT	ATAAGC	ATAAGG
ATT	ATATA	ATAT	ATATC	ATATG	ATATA	ATATT	ATATC	ATATG	ATATA	ATACT	ATACC	ATACG	ATAGA	ATAGT	ATAGC	ATAGG
ATC	ATCAA	ATCAT	ATCAC	ATCAG	ATCTA	ATCTT	ATCTC	ATCTG	ATCCA	ATCCT	ATCCC	ATCCG	ATCGA	ATCGT	ATCGC	ATCCG
ATG	ATGAA	ATGAT	ATGAC	ATGAG	ATGATA	ATGATT	ATGTC	ATGTG	ATGCA	ATGCT	ATGCC	ATGGG	ATGGA	ATGGT	ATGGC	ATGGG
ACA	ACAAA	ACAAT	ACAAC	ACAAG	ACAATA	ACAATT	ACAATC	ACAATG	ACAACA	ACAAC	ACAAC	ACAACG	ACAAGA	ACAAGT	ACAAGC	ACAAGG
ACT	ACTAA	ACTAT	ACTAC	ACTAG	ACTATA	ACTATT	ACTTC	ACTTG	ACTCA	ACTCT	ACTCC	ACTCG	ACTGA	ACTGT	ACTGC	ACTGG
ACC	ACCAA	ACCAT	ACCAC	ACCAG	ACCTA	ACCTT	ACCTC	ACCTG	ACCCA	ACCTT	ACCCC	ACCCG	ACCGA	ACCGT	ACCGC	ACCCG
ACG	ACGAA	ACGAT	ACGAC	ACGAG	ACGATA	ACGATT	ACGTC	ACGTG	ACGCA	ACGCT	ACGCC	ACGGG	ACGGA	ACGGT	ACGGC	ACGGG
AGA	AGAAA	AGAA	AGAAC	AGAAG	AGATA	AGATT	AGATC	AGATG	AGACA	AGACT	AGACC	AGACG	AGAGA	AGAGT	AGAGC	AGAGG
AGT	AGTAA	AGTAT	AGTAC	AGTAG	AGTATA	AGTATT	AGTTC	AGTTG	AGTCA	AGTCT	AGTCC	AGTCG	AGTGA	AGTGT	AGTGC	AGTGG
AGC	AGCAA	AGCAT	AGCAC	AGCAG	AGCTA	AGCTT	AGCTC	AGCTG	AGCCA	AGCCT	AGCCC	AGCCG	AGCGA	AGCGT	AGCGC	AGCCG
AGG	AGGAA	AGGAT	AGGAC	AGGAG	AGGATA	AGGATT	AGGTC	AGGTG	AGGCA	AGGCT	AGGCC	AGGGG	AGGGA	AGGGT	AGGGC	AGGGG
TAA	TAAAA	TAAAT	TAAAC	TAAAG	TAAATA	TAAATT	TAAATC	TAAATG	TAAACA	TAAACT	TAAACC	TAAACG	TAAAGA	TAAAGT	TAAAGC	TAAAGG
TAT	TATAA	TATAT	TATAC	TATAG	TATATA	TATATT	TATATC	TATATG	TATACA	TATCT	TATCC	TATCG	TATGA	TATGT	TATGC	TATGG
TAC	TACAA	TACAT	TACAC	TACAG	TACATA	TACATT	TACTC	TACTG	TACCA	TACCT	TACCC	TACCG	TACGA	TACGT	TACGC	TACGG
TAG	TAGAA	TAGAT	TAGAC	TAGAG	TAGATA	TAGATT	TAGTC	TAGTG	TAGCA	TAGCT	TAGCC	TAGCG	TAGGA	TAGGT	TAGGC	TAGGG
TTA	TTAAA	TTAAT	TTAAC	TTAAG	TTATA	TTATT	TTATC	TTATG	TTACA	TTACT	TTACC	TTACG	TTAGA	TTAGT	TTAGC	TTAGG
TTT	TTTAA	TTTAT	TTTAC	TTTAG	TTTATA	TTTATT	TTTTC	TTTTG	TTTCA	TTTCT	TTTCC	TTTCG	TTTGA	TTTGT	TTTGC	TTTGG
TTG	TTGAA	TTGAT	TTGAC	TTGAG	TTGATA	TTGATT	TTGTC	TTGTG	TTGCA	TTGCT	TTGCC	TTGGG	TTGGA	TTGGT	TTGGC	TTGGG
TCA	TCAAA	TCAAT	TCAAC	TCAAG	TCAATA	TCAATT	TCAATC	TCAATG	TCAACA	TCAACT	TCAACC	TCAACG	TCAAGA	TCAAGT	TCAAGC	TCAAGG
TCT	TCTAA	TCTAT	TCTAC	TCTAG	TCTATA	TCTATT	TCTTC	TCTTG	TCTCA	TCTCT	TCTCC	TCTCG	TCTGA	TCTGT	TCTGC	TCTGG
TCC	TCCAA	TCCAT	TCCAC	TCCAG	TCCATA	TCCATT	TCCTC	TCCTG	TCCCA	TCCCT	TCCCC	TCCCG	TCCGA	TCCGT	TCCGC	TCCGG
TCCG	TCCGA	TCCAT	TCCAC	TCCAG	TCCATA	TCCATT	TCCTC	TCCTG	TCCCA	TCCCT	TCCCC	TCCCG	TCCGA	TCCGT	TCCGC	TCCGG
TGA	TGAAA	TGAAT	TGAAC	TGAAG	TGATA	TGATT	TGATC	TGATG	TGACA	TGACT	TGACC	TGACG	TGAGA	TGAGT	TGAGC	TGAGG
TGT	TGTAA	TGTAT	TGTAC	TGTAG	TGTATA	TGTATT	TGTTC	TGTTG	TGTCA	TGTCT	TGTCC	TGTCTG	TGTGA	TGTGT	TGTGC	TGTGG
TGC	TGCAA	TGCAT	TGCAC	TGCAG	TGCTA	TGCTT	TGCTC	TGCTG	TGCCA	TGCTT	TGCCC	TGCCG	TGCGA	TGCGT	TGCGC	TGCCG
TGG	TGGA	TGGAT	TGGAC	TGGAG	TGGATA	TGGATT	TGGTC	TGGTG	TGGCA	TGGCT	TGGCC	TGGCG	TGGGA	TGGGT	TGGGC	TGGGG
CAA	CAAAA	CAAA	CAAAC	CAAG	CAATA	CAATT	CAATC	CAATG	CAACA	CAACT	CAACC	CAACG	CAAGA	CAAGT	CAAGC	CAAGG
CAT	CATAA	CATAT	CATAC	CATAG	CATATA	CATATT	CATTC	CATTG	CATCA	CATCT	CATCC	CATCG	CATGA	CATGT	CATGC	CATGG
CAC	CACAA	CACAT	CACAC	CACAG	CACATA	CACATT	CACTC	CACTG	CACCA	CACCT	CACCC	CACCG	CACGA	CACGT	CACGC	CACGG
CAG	CAGAA	CAGAT	CAGAC	CAGAG	CAGATA	CAGATT	CAGTC	CAGTG	CAGCA	CAGCT	CAGCC	CAGCG	CAGGA	CAGGT	CAGGC	CAGGG
CCT	CTAAA	CTAAT	CTAAC	CTAAG	CTATA	CTATT	CTATC	CTATG	CTACA	CTACT	CTACC	CTACG	CTAGA	CTAGT	CTAGC	CTAGG
CTA	CTTAA	CTTAT	CTTAC	CTTAG	CTTATA	CTTATT	CTTTC	CTTTG	CTTCA	CTTCT	CTTCC	CTTCG	CTTGA	CTTGT	CTTGC	CTTGG
CTC	CTGAA	CTGAT	CTGAC	CTGAG	CTGATA	CTGATT	CTGTC	CTGTG	CTGCA	CTGCT	CTGCC	CTGGG	CTGGA	CTGGT	CTGGC	CTGGG
CTG	CTGAA	CTGAT	CTGAC	CTGAG	CTGATA	CTGATT	CTGTC	CTGTG	CTGCA	CTGCT	CTGCC	CTGGG	CTGGA	CTGGT	CTGGC	CTGGG
CCA	CCAAA	CCAA	CCAAC	CCAAG	CCATA	CCATT	CCATC	CCATG	CCACA	CCAAT	CCACC	CCACG	CCAGA	CCAGT	CCAGC	CCAGG
CCT	CCTAA	CCTAT	CCTAC	CCTAG	CCTATA	CCTATT	CCTTC	CCTTG	CCTCA	CCTCT	CCTCC	CCTCG	CCTGA	CCTGT	CCTGC	CCTGG
CCC	CCCAA	CCCAT	CCCAC	CCCAG	CCCTA	CCCTT	CCCTC	CCCTG	CCCCA	CCCTT	CCCCC	CCCCG	CCCGA	CCCGT	CCCGC	CCCCG
CCG	CCGAA	CCGAT	CCGAC	CCGAG	CCGATA	CCGATT	CCGTC	CCGTG	CCGCA	CCGCT	CCGCC	CCGGG	CCGGA	CCGGT	CCGGC	CCGGG
CGA	CGAAA	CGAAT	CGAAC	CGAAG	CGATA	CGATT	CGATC	CGATG	CGACA	CGACT	CGACC	CGACG	CGAGA	CGAGT	CGAGC	CGAGG
CGT	CGTAA	CGTAT	CGTAC	CGTAG	CGTATA	CGTATT	CGTTC	CGTTG	CGTCA	CGTCT	CGTCC	CGTCG	CGTGA	CGTGT	CGTGC	CGTGG
CGC	CGCAA	CGCAT	CGCAC	CGCAG	CGCTA	CGCTT	CGCTC	CGCTG	CGCCA	CGCCT	CGCCC	CGCCG	CGCGA	CGCGT	CGCGC	CGCCG
CGG	CGGAA	CGGAT	CGGAC	CGGAG	CGGATA	CGGATT	CGGTC	CGGTG	CGGCA	CGGCT	CGGCC	CGGGG	CGGGA	CGGGT	CGGGC	CGGGG
GAA	GAAAA	GAAAT	GAAAC	GAAAG	GAAATA	GAAATT	GAAATC	GAAATG	GAAACA	GAAACT	GAAACC	GAAACG	GAAAGA	GAAAGT	GAAAGC	GAAAGG
GA	GATAA	GATAT	GATAC	GATAG	GATATA	GATATT	GATTC	GATTG	GATCA	GATCT	GATCC	GATCG	GATGA	GATGT	GATGC	GATGG
GAC	GACAA	GACAT	GACAC	GACAG	GACATA	GACATT	GACTC	GACTG	GACCA	GACCT	GACCC	GACCG	GACGA	GACGT	GACGC	GACGG
GAG	GAGAA	GAGAT	GAGAC	GAGAG	GAGATA	GAGATT	GAGTC	GAGTG	GAGCA	GAGCT	GAGCC	GAGCG	GAGGA	GAGGT	GAGGC	GAGGG
GTA	GTA	GTAAT	GTAAC	GTAAG	GTAATA	GTAATT	GTAATC	GTAATG	GTAACA	GTAACT	GTAACC	GTAACG	GTAAGA	GTAAGT	GTAAGC	GTAAGG
GTT	GTTAA	GTTAT	GTTAC	GTTAG	GTTATA	GTTATT	GTTTC	GTTTG	GTTCA	GTTCT	GTTCC	GTTCTG	GTTGA	GTTGT	GTTGC	GTTGG
GTC	GTC	GTCAT	GTCAC	GTCAG	GTCATA	GTCATT	GTCTC	GTC	GTC	GTCCT	GTC	GTC	GTC	GTC	GTC	GTC
GTG	GTGAA	GTGAT	GTGAC	GTGAG	GTGATA	GTGATT	GTGTC	GTGTG	GTGCA	GTGCT	GTGCC	GTGGG	GTGGA	GTGGT	GTGGC	GTGGG
GTA	GTAAA	GTAAT	GTAAC	GTAAG	GTAATA	GTAATT	GTAATC	GTAATG	GTAACA	GTAACT	GTAACC	GTAACG	GTAAGA	GTAAGT	GTAAGC	GTAAGG
GCT	GCTAA	GCTAT	GCTAC	GCTAG	GCTATA	GCTATT	GCTTC	GCTTG	GCTCA	GCTCT	GCTCC	GCTCG	GCTGA	GCTGT	GCTGC	GCTGG
GCC	GCCAA	GCCAT	GCCAC	GCCAG	GCCATA	GCCATT	GCCTC	GCCTG	GCCCA	GCCCT	GCCCC	GCCCG	GCCGA	GCCGT	GCCGC	GCCGG
GCG	GCGAA	GCGAT	GCGAC	GCGAG	GCGATA	GCGATT	GCGTC	GCGTG	GCGCA	GCGCT	GCGCC	GCGCG	GCGGA	GCGGT	GCGGC	GCGGG
GGA	GGAAA	GGAA	GGAAC	GGAAG	GGATA	GGATT	GGATC	GGATG	GGACA	GGACT	GGACC	GGACG	GGAGA	GGAGT	GGAGC	GGAGG
GGT	GGTAA	GGTAT	GGTAC	GGTAG	GGTATA	GGTATT	GGTTC	GGTTG	GGTCA	GGTCT	GGTCC	GGTCG	GGTGA	GGTGT	GGTGC	GGTGG
GGC	GSCAA	GSCAT	GSCAC	GSCAG	GSCATA	GSCATT	GSCTC	GSCTG	GSCCA	GSCCT	GSCCC	GSCCG	GSCGA	GSCGT	GSCGC	GSCGG
GGG	GSGAA	GSGAT	GSGAC	GSGAG	GSGATA	GSGATT	GSGTC	GSGTG	GSGCA	GSGCT	GSGCC	GSGCG	GSGGA	GSGGT	GSGGC	GSGGG

After filtering: read2: quality

Value of each position will be shown on mouse over.

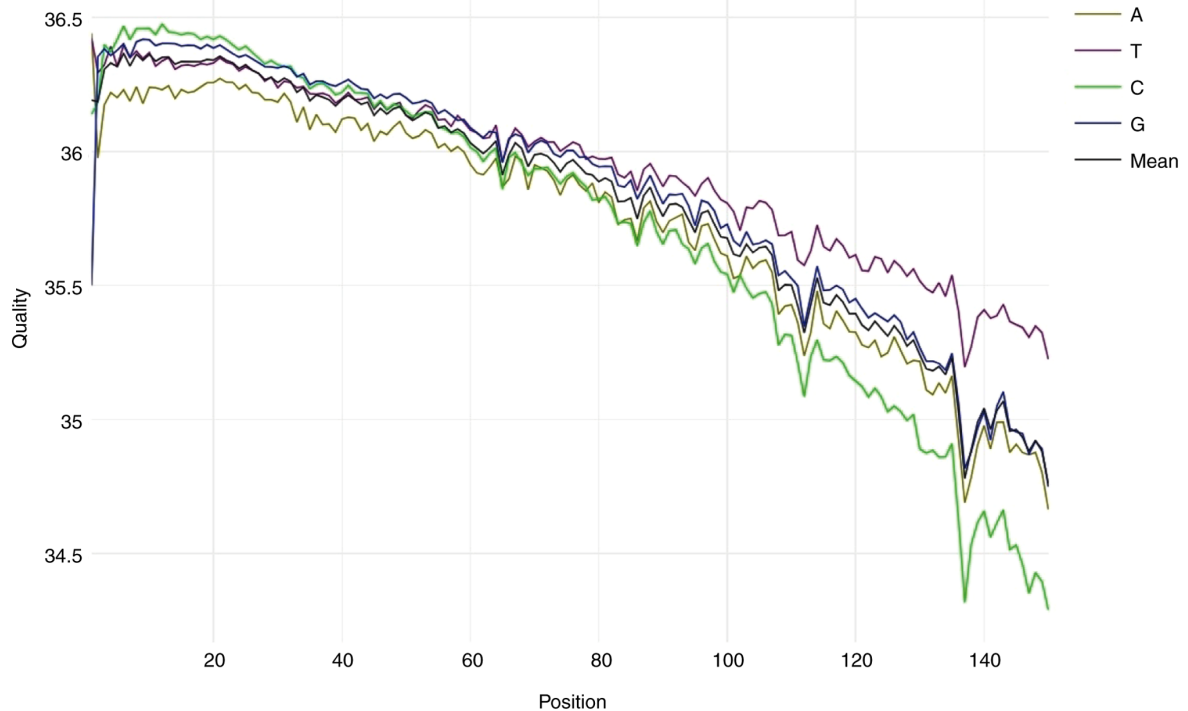


Figure S3. Positive immunohistochemistry results of the rectal specimen. Positive expression for (A) DNA mismatch repair protein Mlh1, (B) DNA mismatch repair protein Msh2, (C) DNA mismatch repair protein Msh6 and (D) mismatch repair endonuclease PMS2.

