

Summary

- Basic Statistics
- Per base sequence quality
- Per tile sequence quality
- Per sequence quality scores
- Per base sequence content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Adapter Content
- Kmer Content

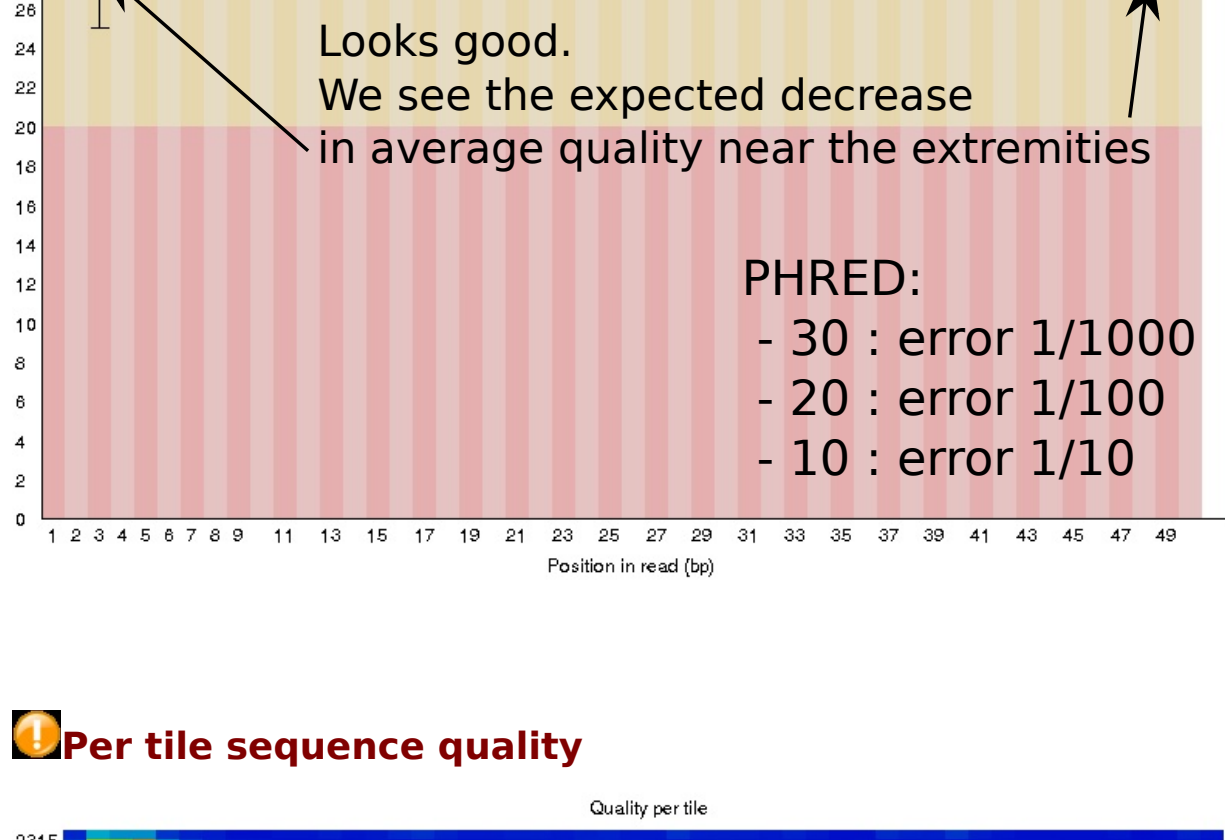
Some of the QC measures seems to be bad. We will check this.

Basic Statistics

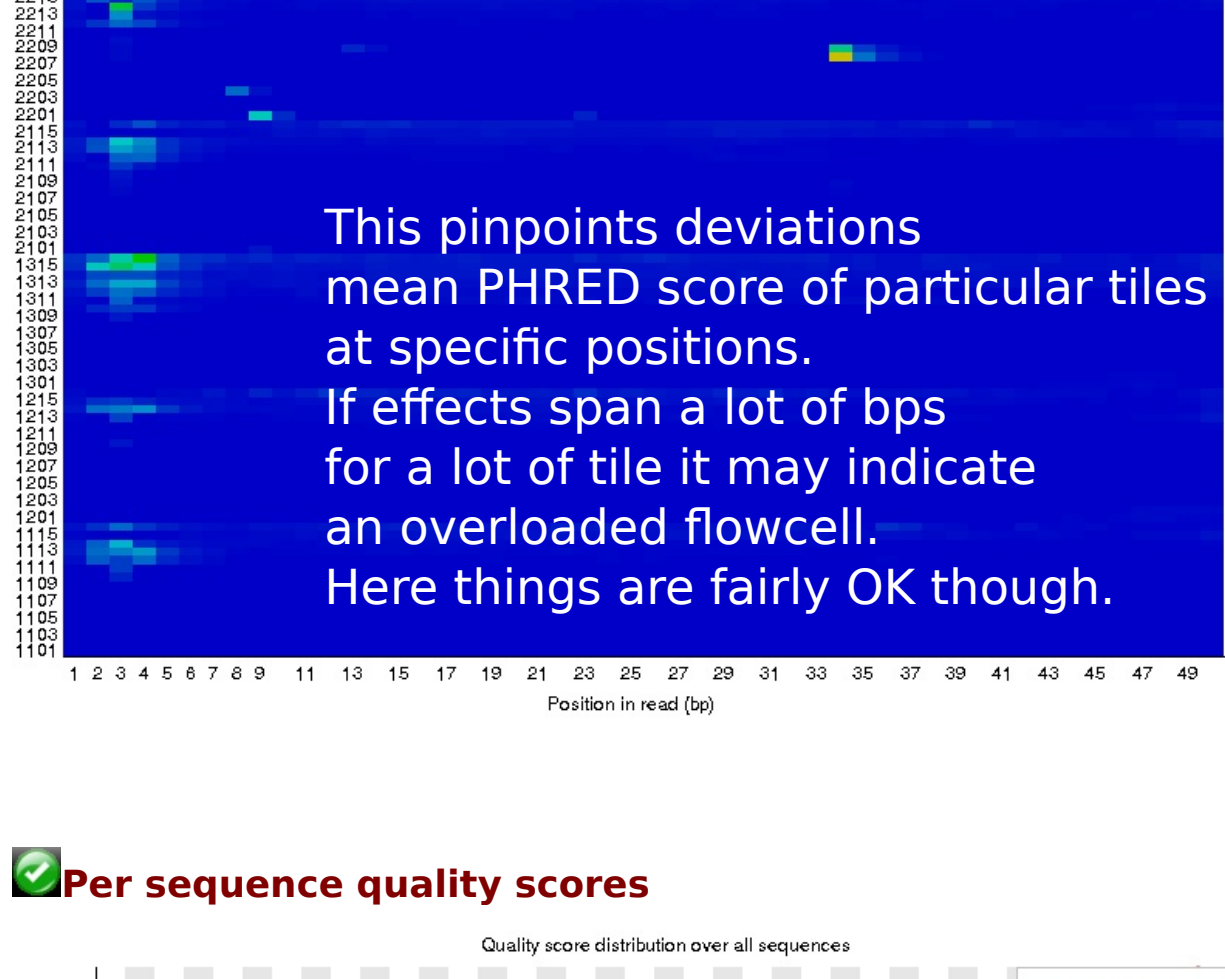
Measure	Value
Filename	SRR3180535_EtOH1_1.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	54952527
Sequences flagged as poor quality	0
Sequence length	50
%GC	51

Already we have some interesting information:
- ~55M reads
- 50bp/read

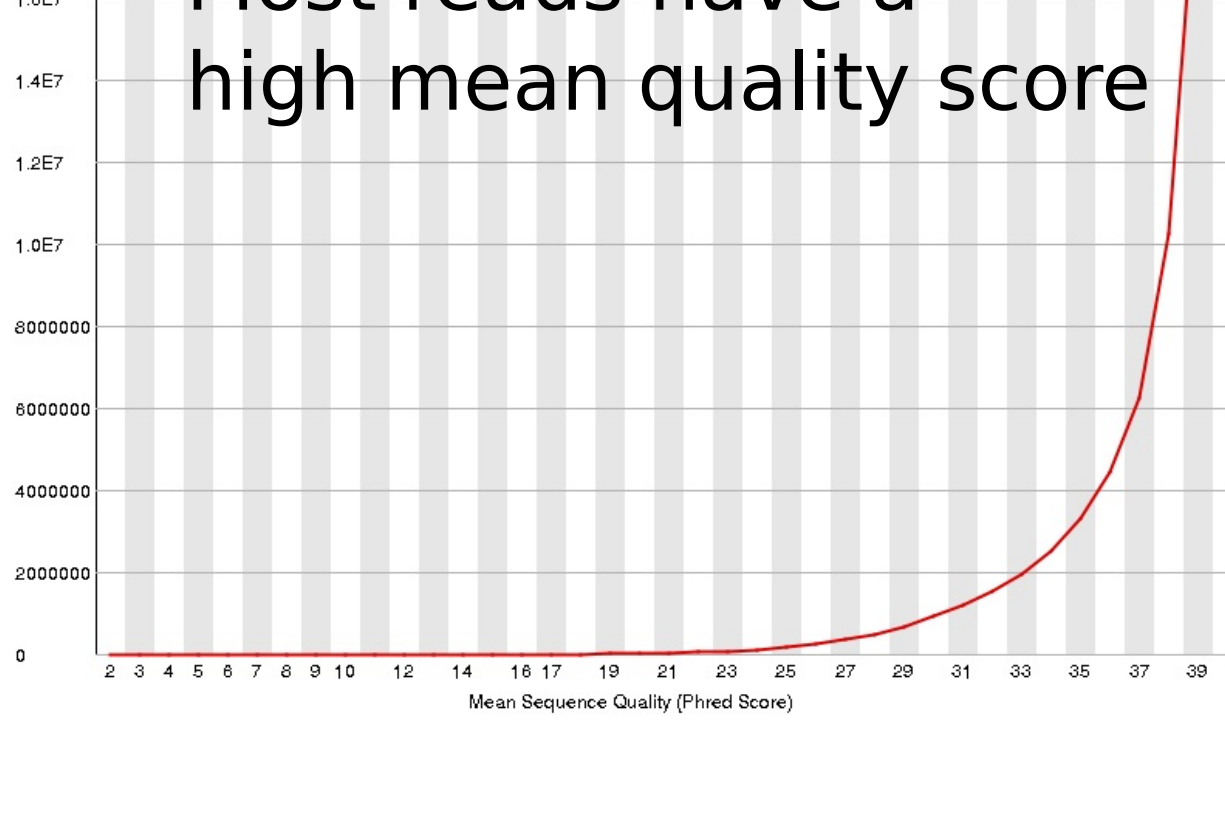
Per base sequence quality



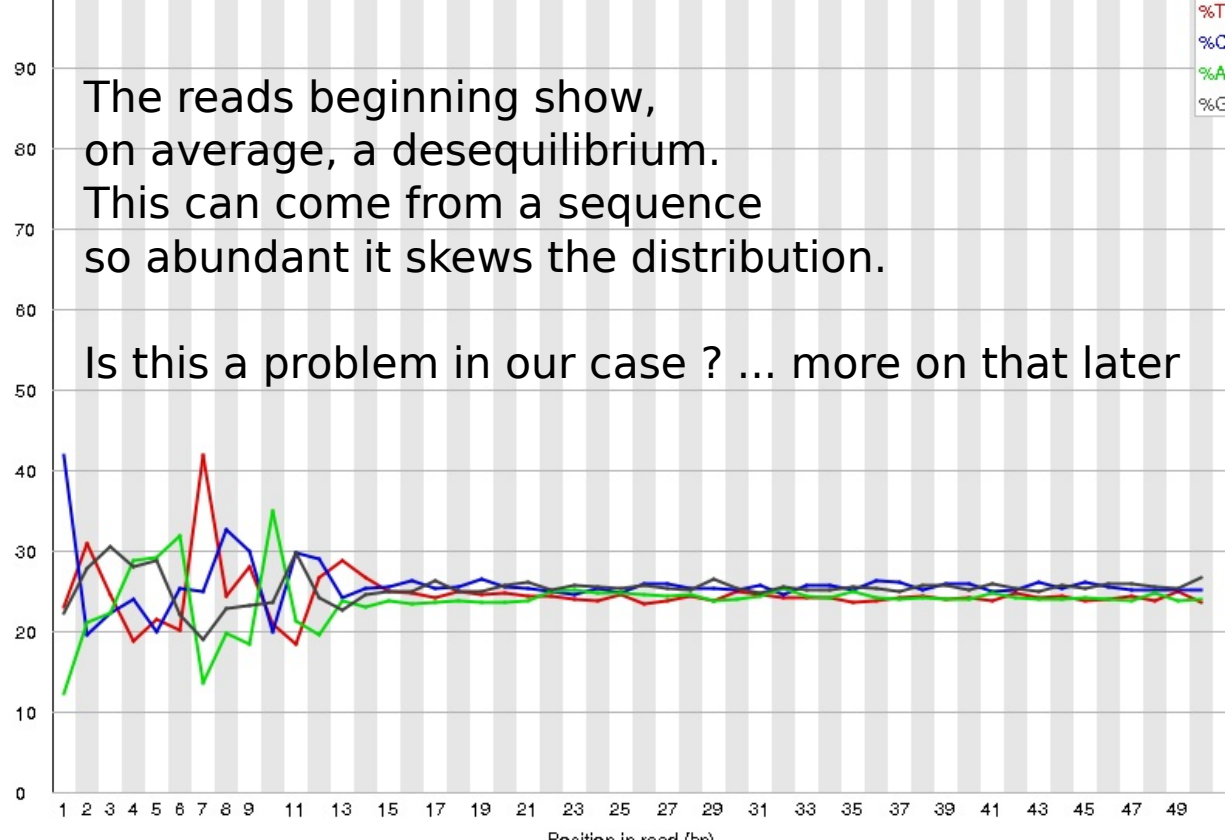
Per tile sequence quality



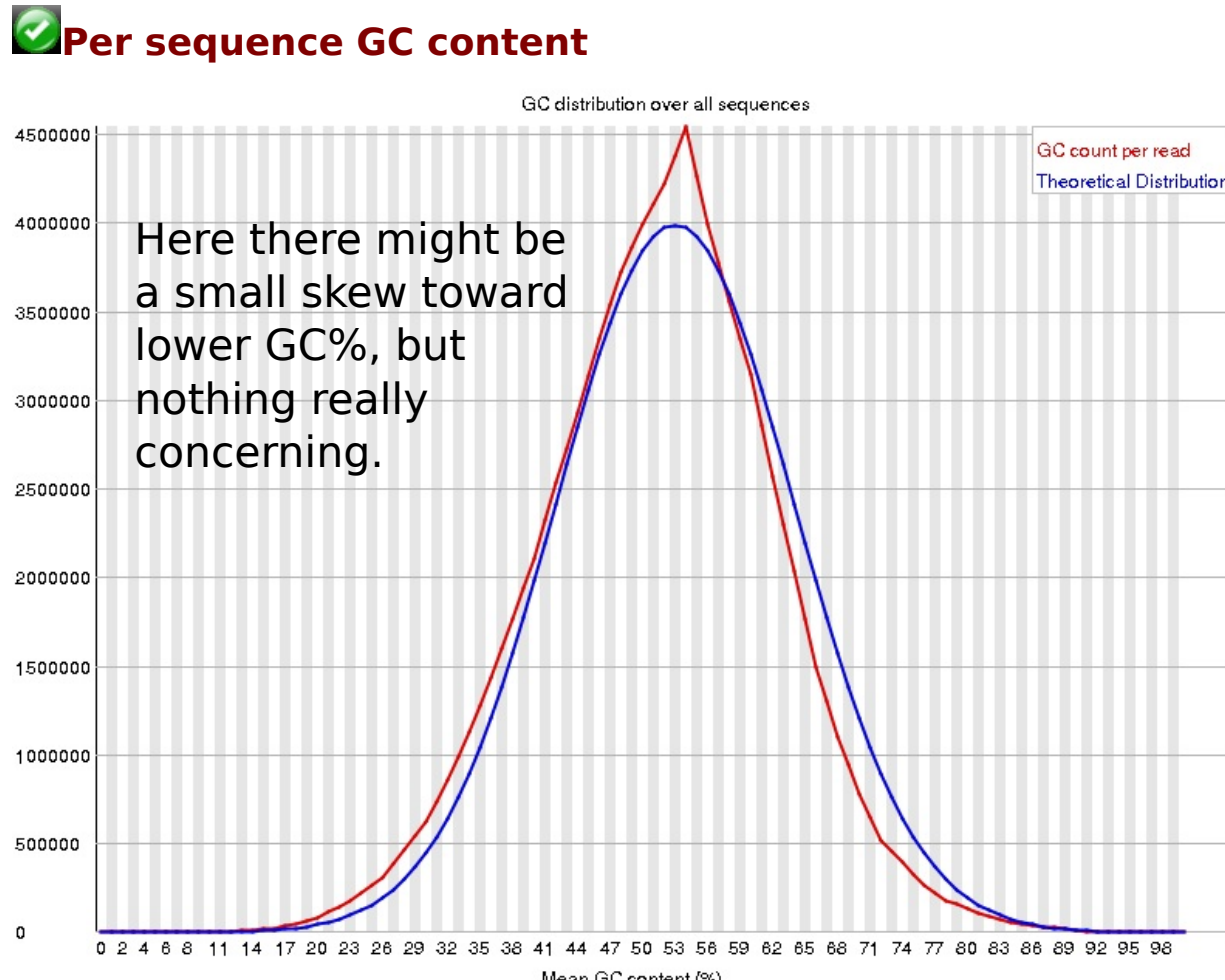
Per sequence quality scores



Per base sequence content



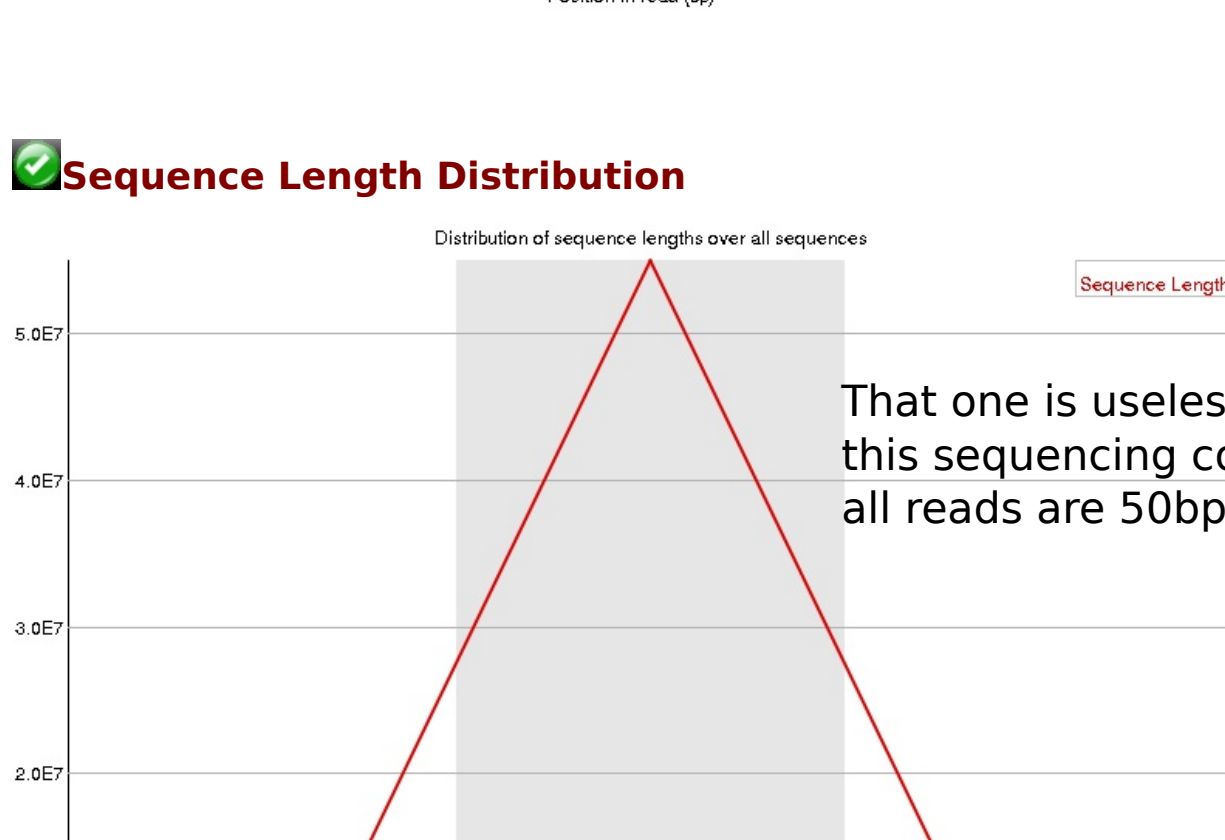
Per sequence GC content



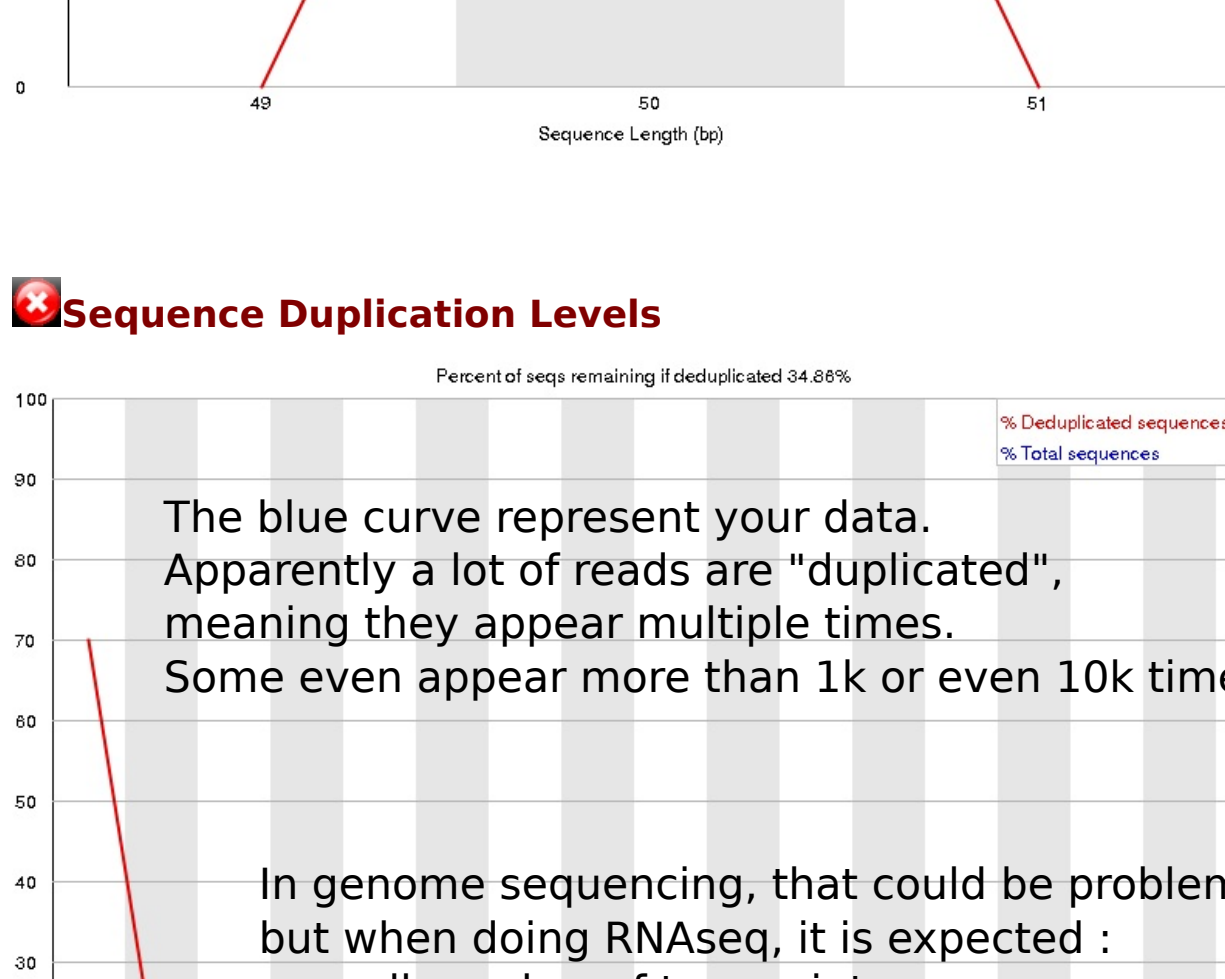
Per base N content



Sequence Length Distribution



Sequence Duplication Levels



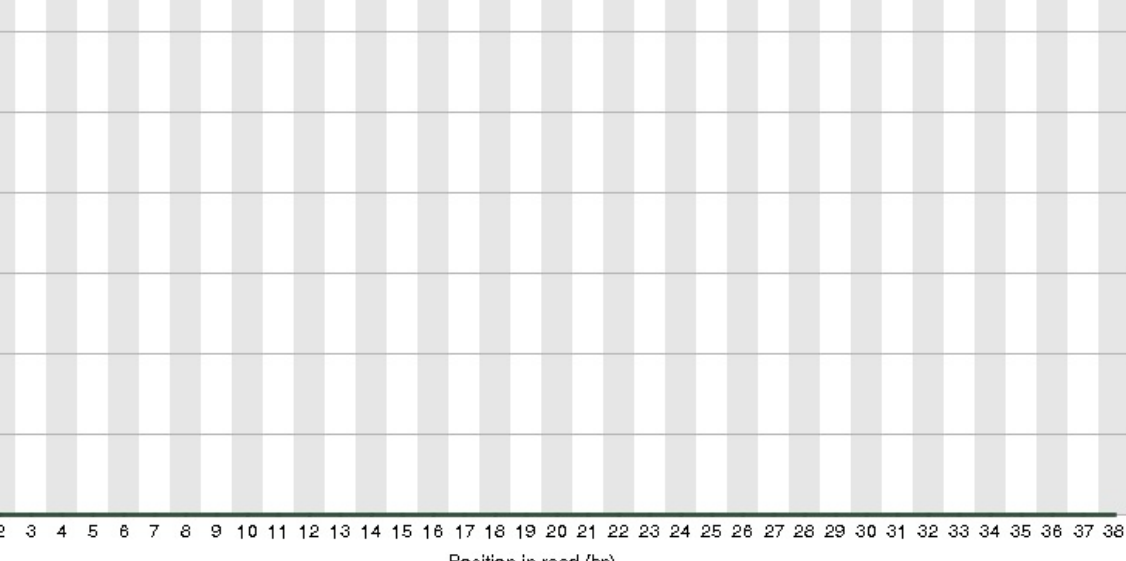
This is also what explains this bias in base % at the beginning of reads (as seen above)

Overrepresented sequences

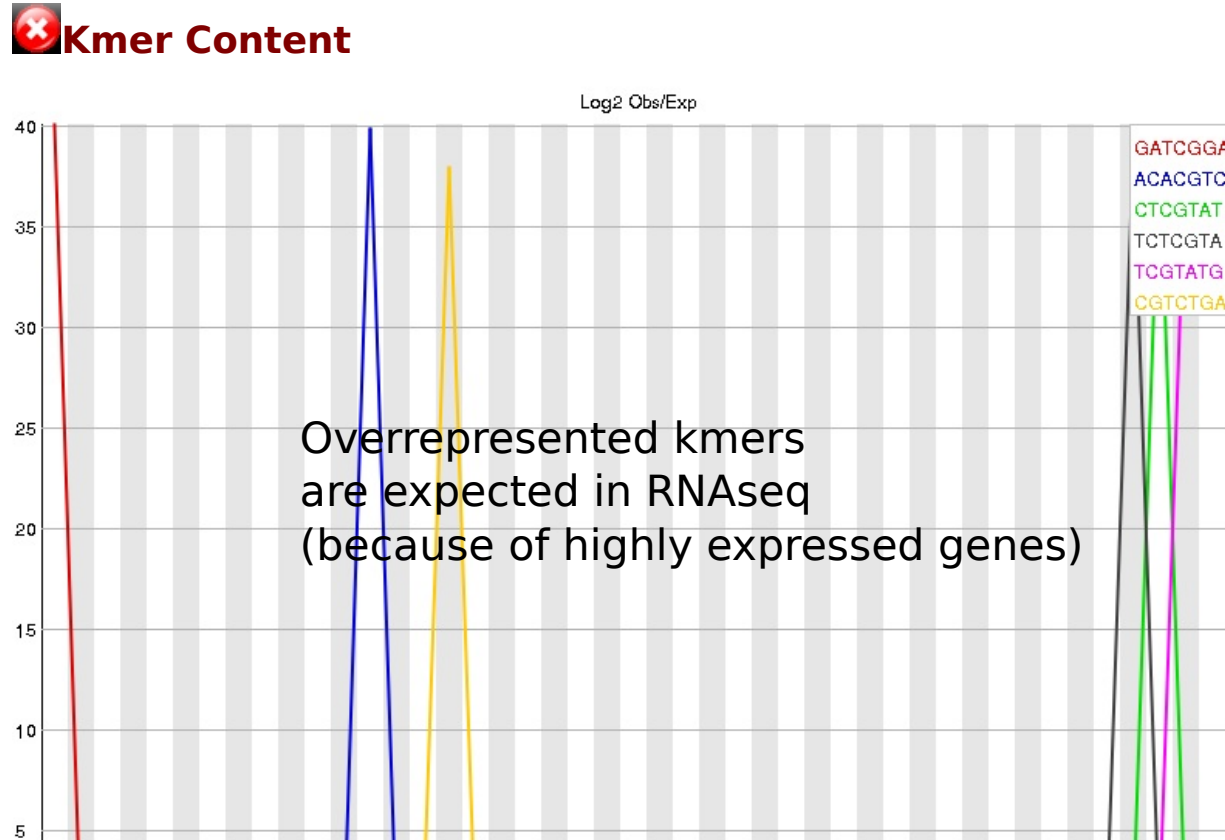
Sequence	Count	Percentage	Possible Source
GATCGGAAGACACACGCTCTGAACCTCCAGTCACCACTCCATCTCGTATG	367199	0.6682113898085922	TruSeq Adapter, Index 7 (97% over 35bp)

Well, also, there is some adapter sequence still left apparently. in 367k out of 55M sequences, that is <1% but still,

Adapter Content



Kmer Content



Sequence	Count	PValue	Obs/Exp Max	Max Obs/Exp Position
GATCGGA	44550	0.0	48.0678	1
ACACGTC	49620	0.0	39.8718	13
CTCGTAT	48985	0.0	39.219986	43
TCTCGTA	49430	0.0	38.83549	42
TCGTATG	49530	0.0	38.76692	44
CGTCTGA	52075	0.0	37.975723	16
CACACGT	52380	0.0	37.78761	12
ATCGGAA	47365	0.0	37.634815	2
CACGTCT	53955	0.0	36.69688	14
GCACACG	54520	0.0	36.481945	11
ACGTCTG	54380	0.0	36.36194	15
ATCTCGT	53780	0.0	35.84328	41
CATCTCT	57110	0.0	34.092354	40
CGGAAGA	59310	0.0	33.614536	4
TCGGAAG	53855	0.0	33.31565	3
GTCTGAA	62595	0.0	31.775808	17
AGACAC	63610	0.0	31.62487	8
AGCACAC	64310	0.0	31.168814	10
GAGCACA	65840	0.0	31.097986	9
TCTGAAC	63810	0.0	31.084723	18