FastQC
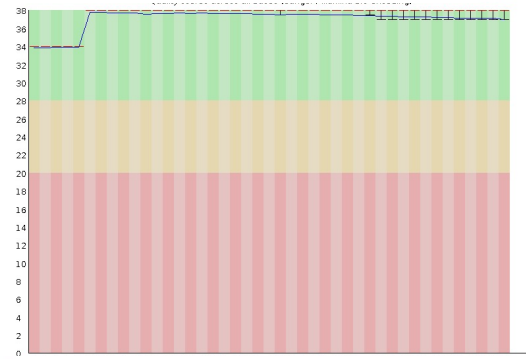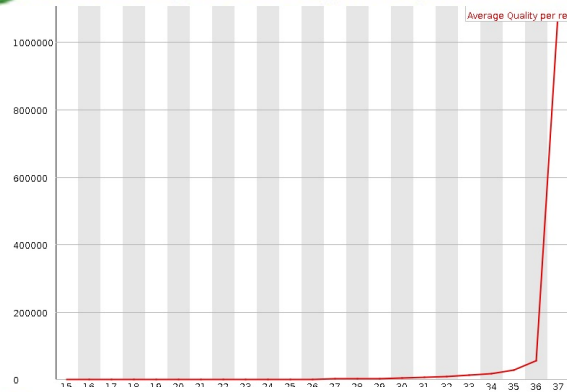
FastQC

Quality
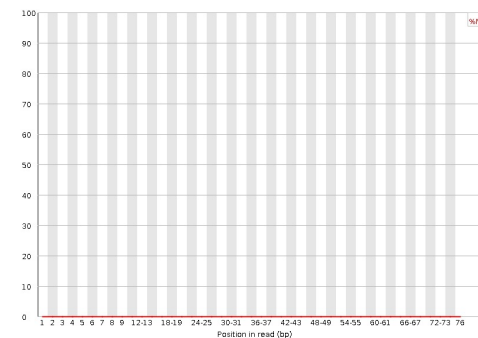
✅ **Per base sequence quality**
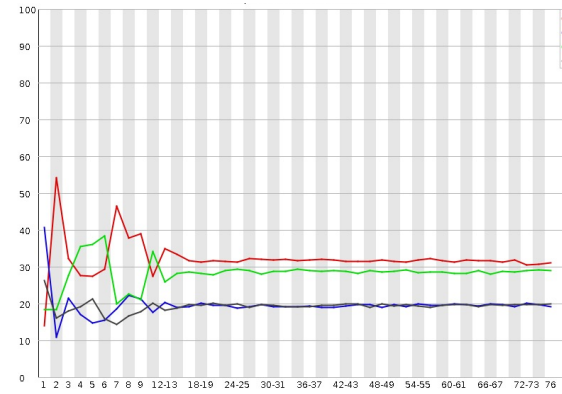

✅ **Per sequence quality scores**

✅ **Per base N content**

Characteristics

❌ **Per base sequence content**

❌ **Per sequence GC content**
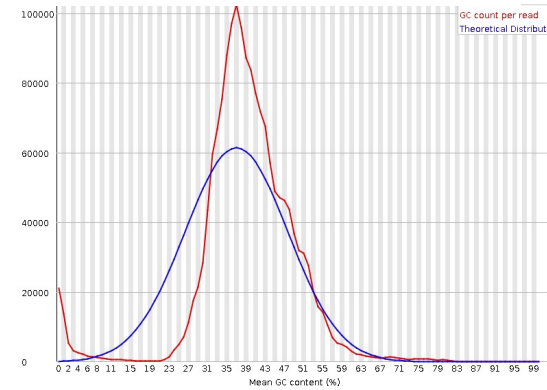
✅ **Basic Statistics**

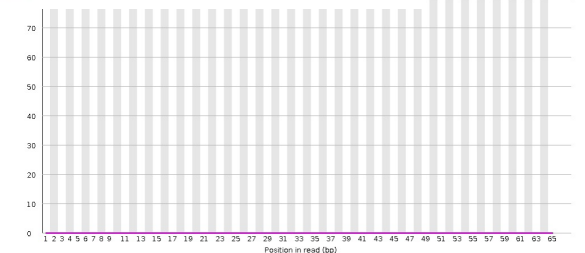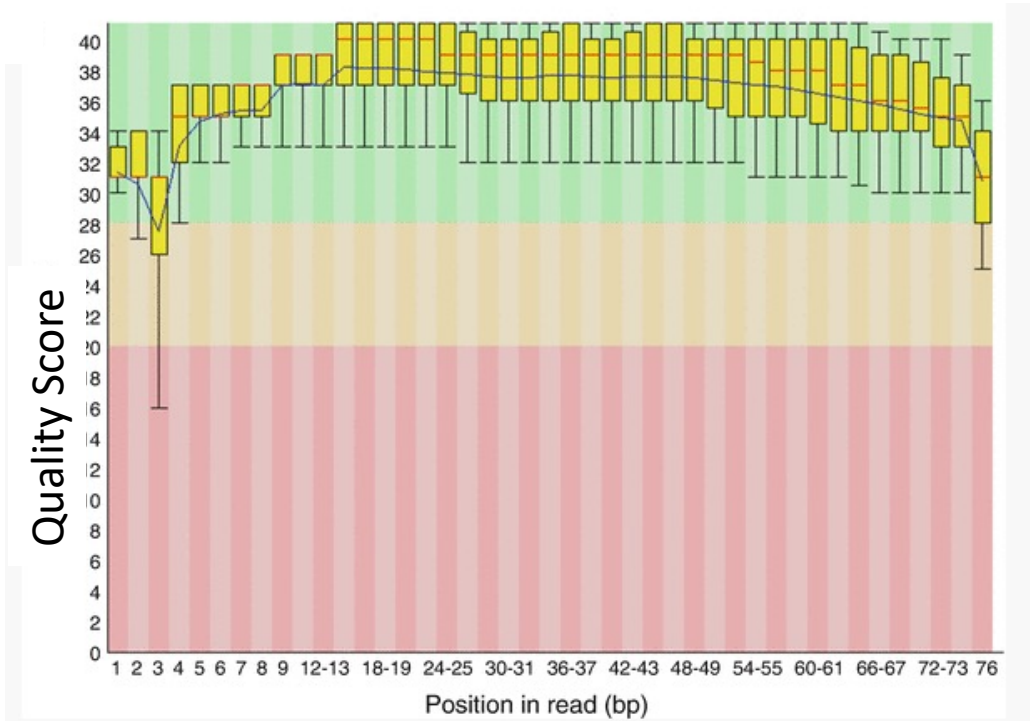| Measure	Value
Filename	forward.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	1264290
Sequences flagged as poor quality	0
Sequence length	76
%GC	39

✅ **Sequence Length Distribution**
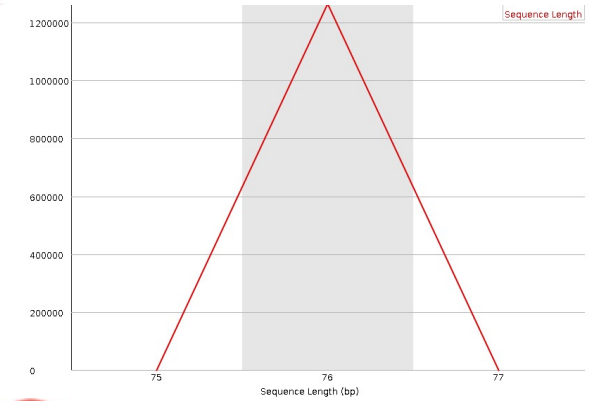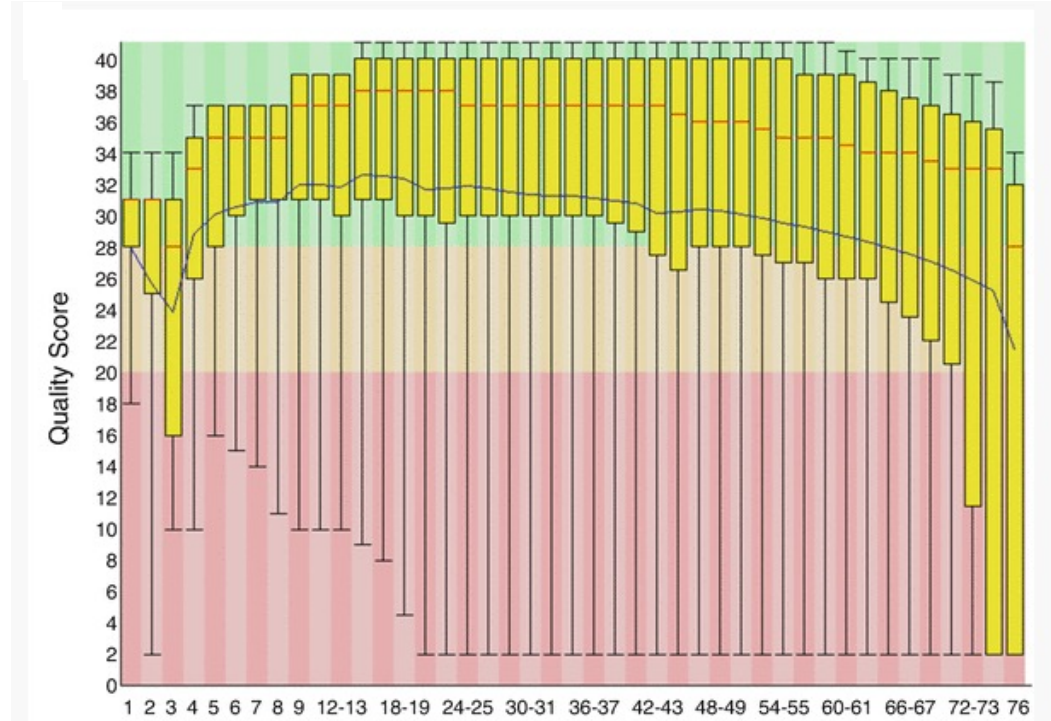

❌ **Sequence Duplication Levels**
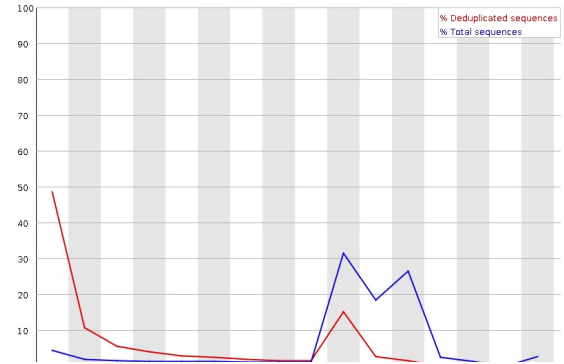
✅ **Adapter Content**

FastQC: Sequence Quality Histogram

Position in read (bp)

GOOD
High quality over the length of the read

BAD
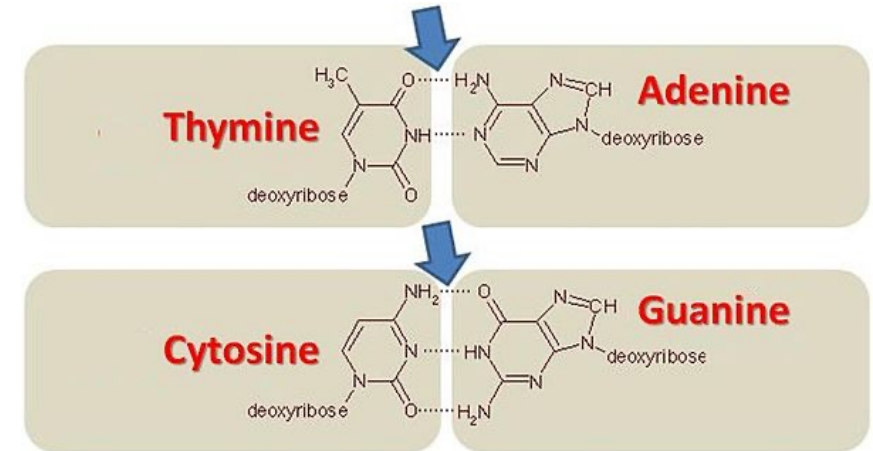Read quality drops at the beginning and end
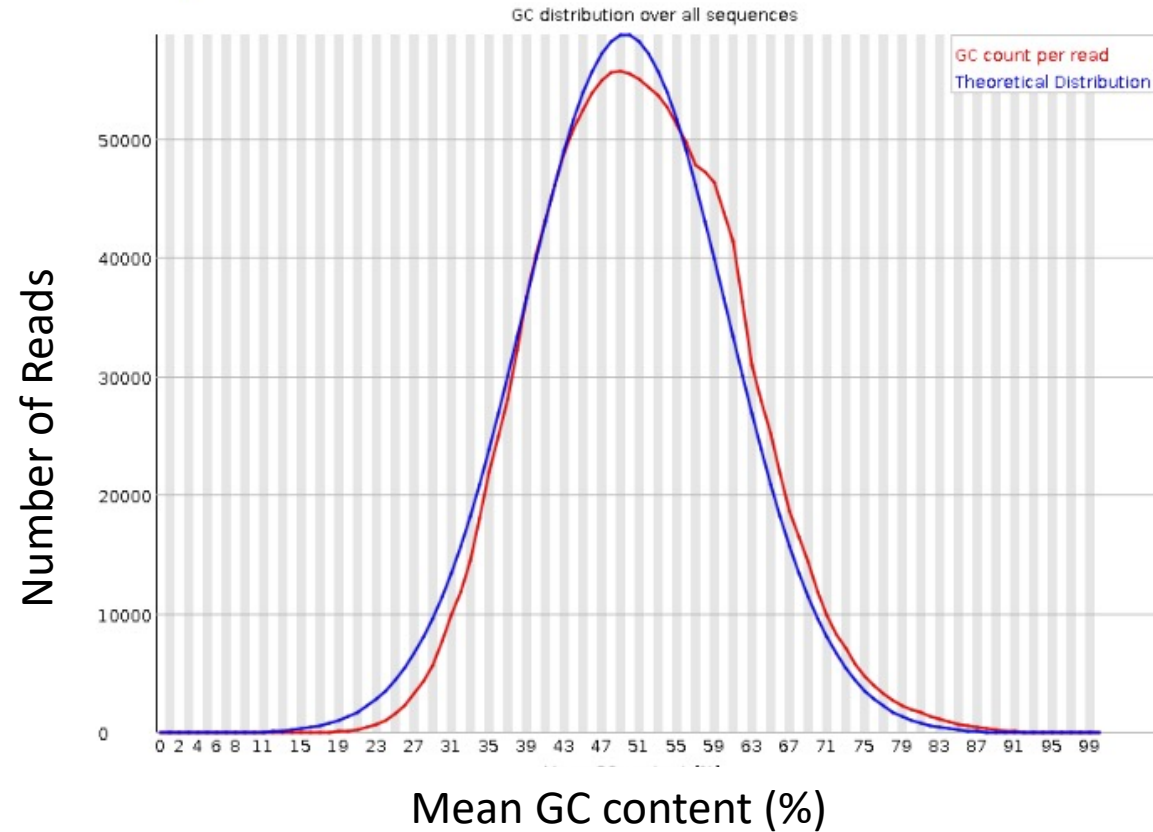
Sequencing

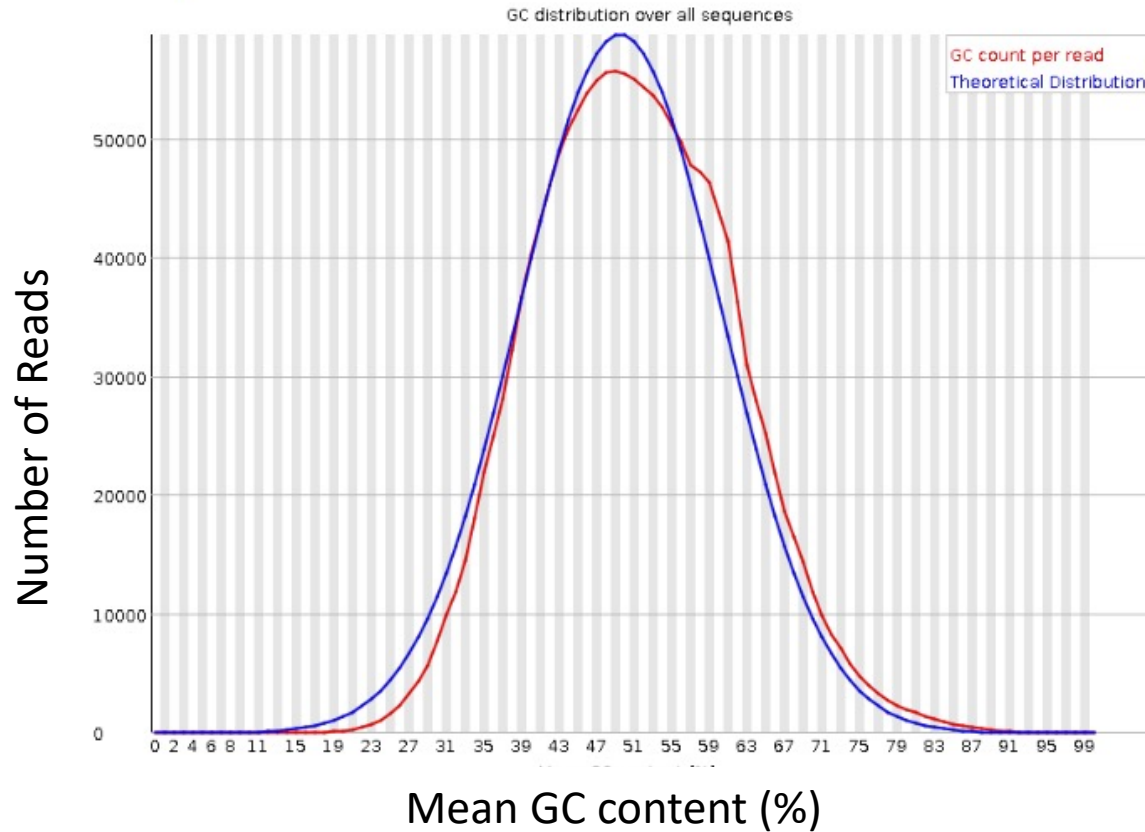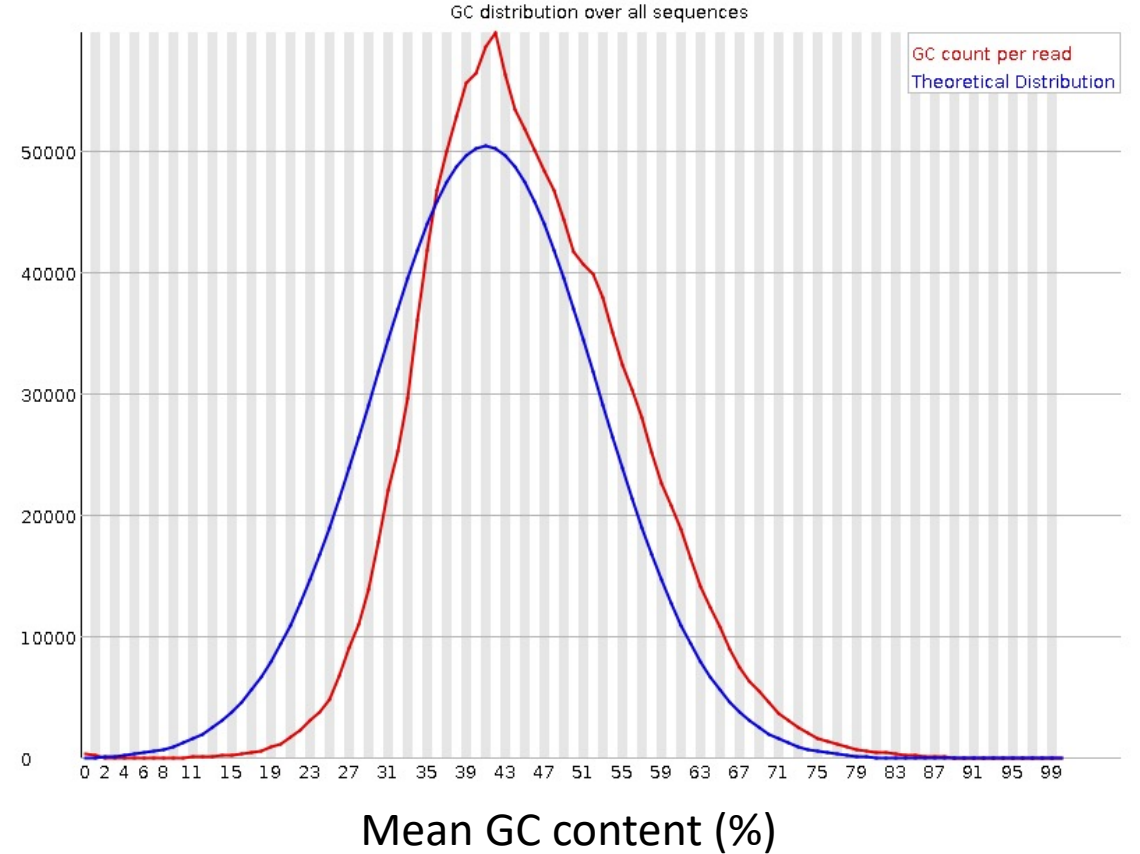# FastQC: Per sequence GC content
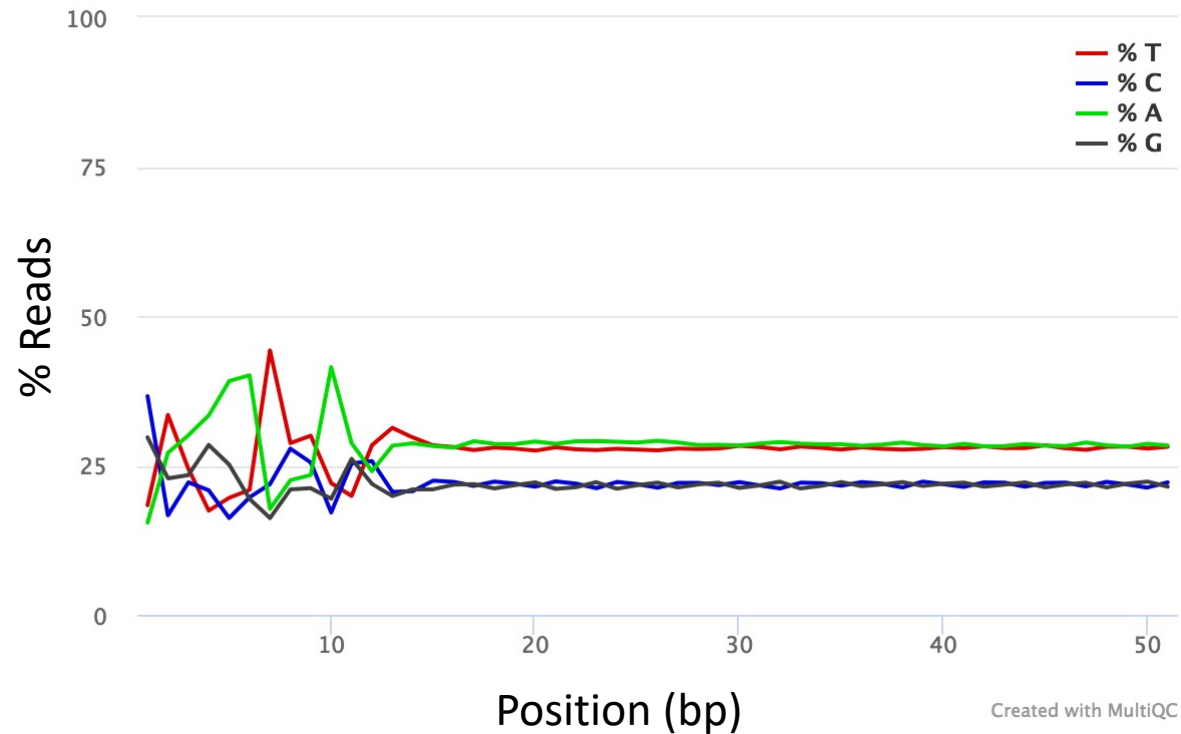
FastQC: Per sequence GC content

GOOD: follows normal distribution (sum of deviations is < 15% of reads)

BAD: can indicate contamination with adapter dimers, or another species

# FastQC: Per Base Sequence Content

- Proportion of each position for which each DNA base has been called
- RNAseq data tends to show a positional sequence bias in the first ~12 bases
- The "random" priming step during library construction is not truly random and certain hexamers are more prevalent than others

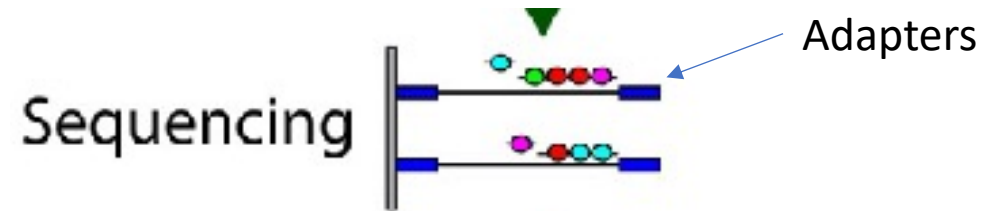sequencing.qcfail.com

FastQC: Per Base Sequence Content
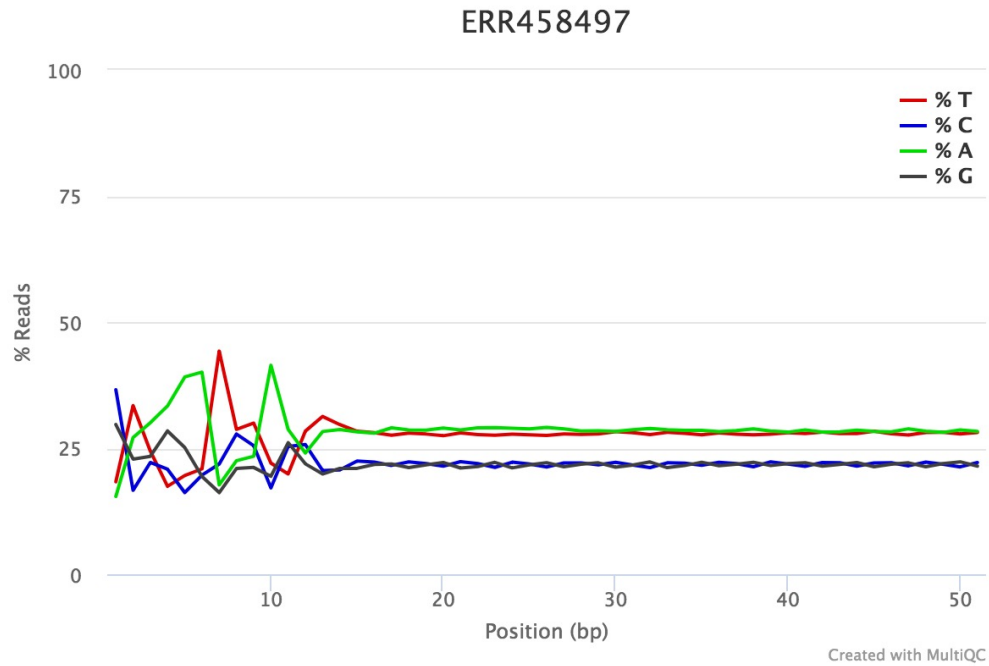
EXPECTED for RNAseq
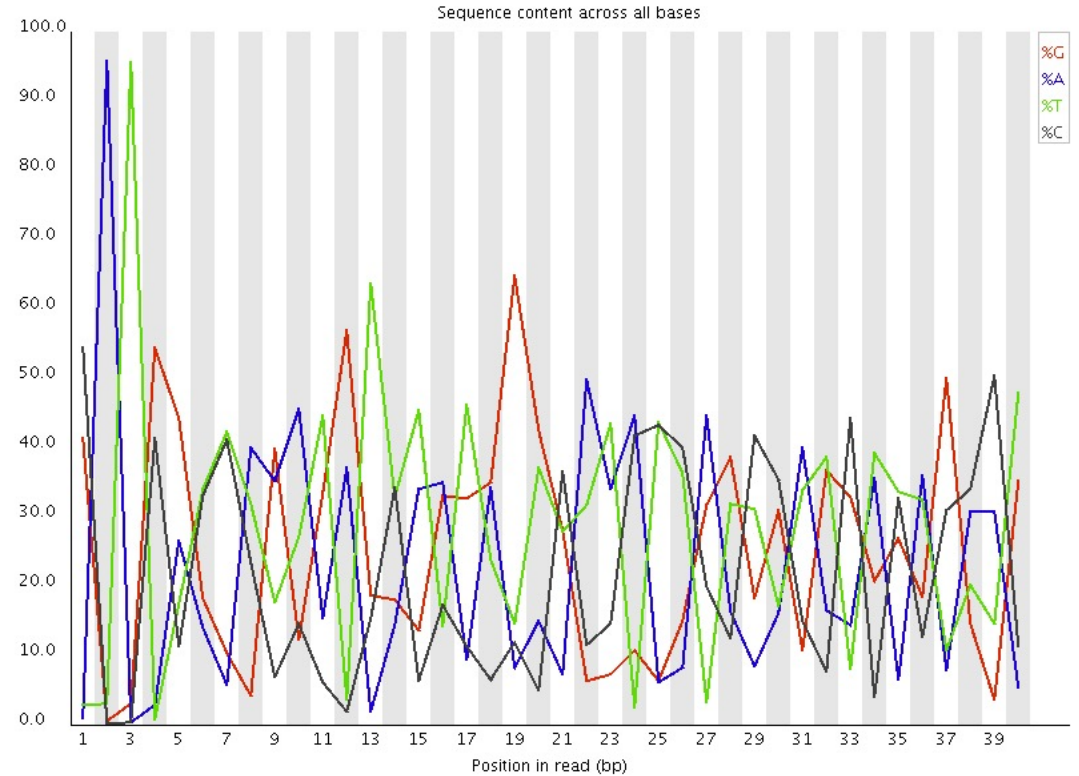
BAD:
Shows a strong positional bias throughout the reads, which in this case is due to the library having a certain sequence that is overrepresented

FastQC: Adapter content

Adapters

FastQC will scan each read for the presence of known adapter sequences

The plot shows that the adapter content rises over the course of the read

Solution – Adapter trimming!

% Adapter

Illumina Universal Adapter
Illumina Small RNA 3' Adapter
Illumina Small RNA 5' Adapter
Nextera Transposase Sequence
SOLID Small RNA Adapter

Position in read (bp)

sequencing.qcfail.com