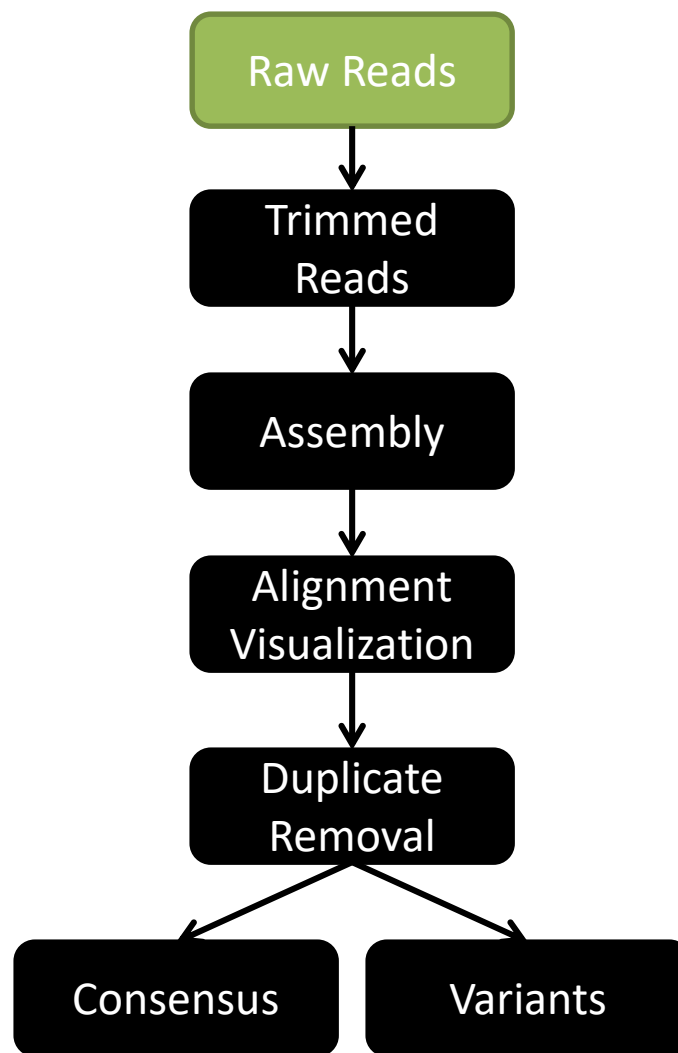


Quality Control and Visualization

Quality Check

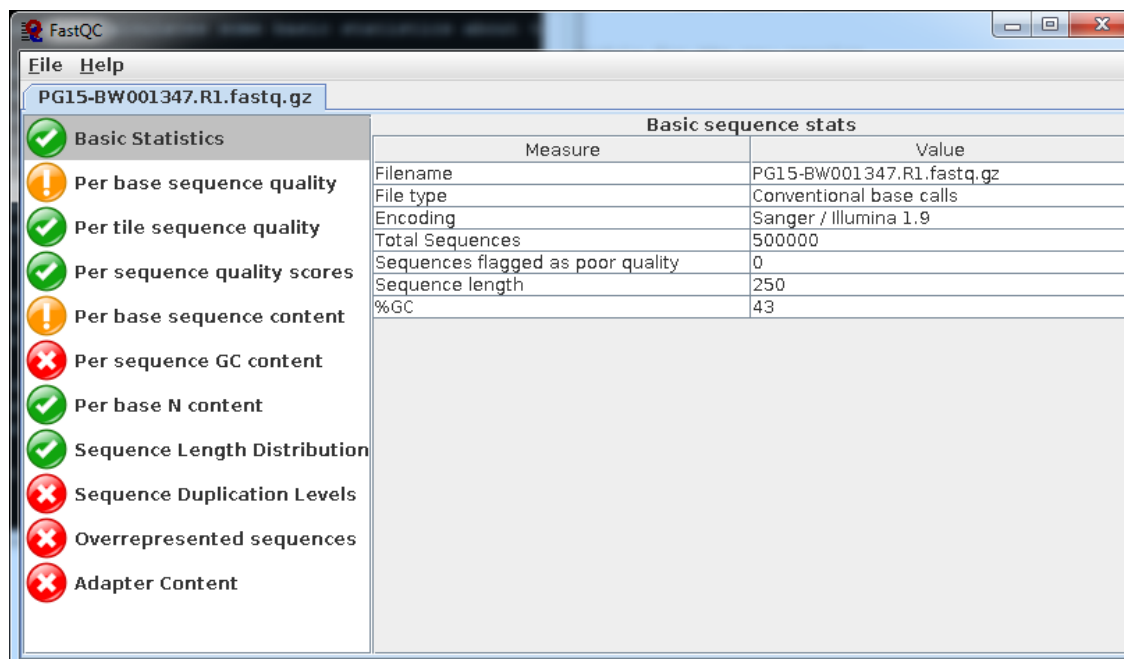
- Garbage in Garbage out
- Low quality data generated towards the end of reads.
- High level of duplication and low quality reads can effect variant calling.
- To identify read through into adapters.
- Basic statistics of a run.

Where do we start?



Quality Check - FASTQC

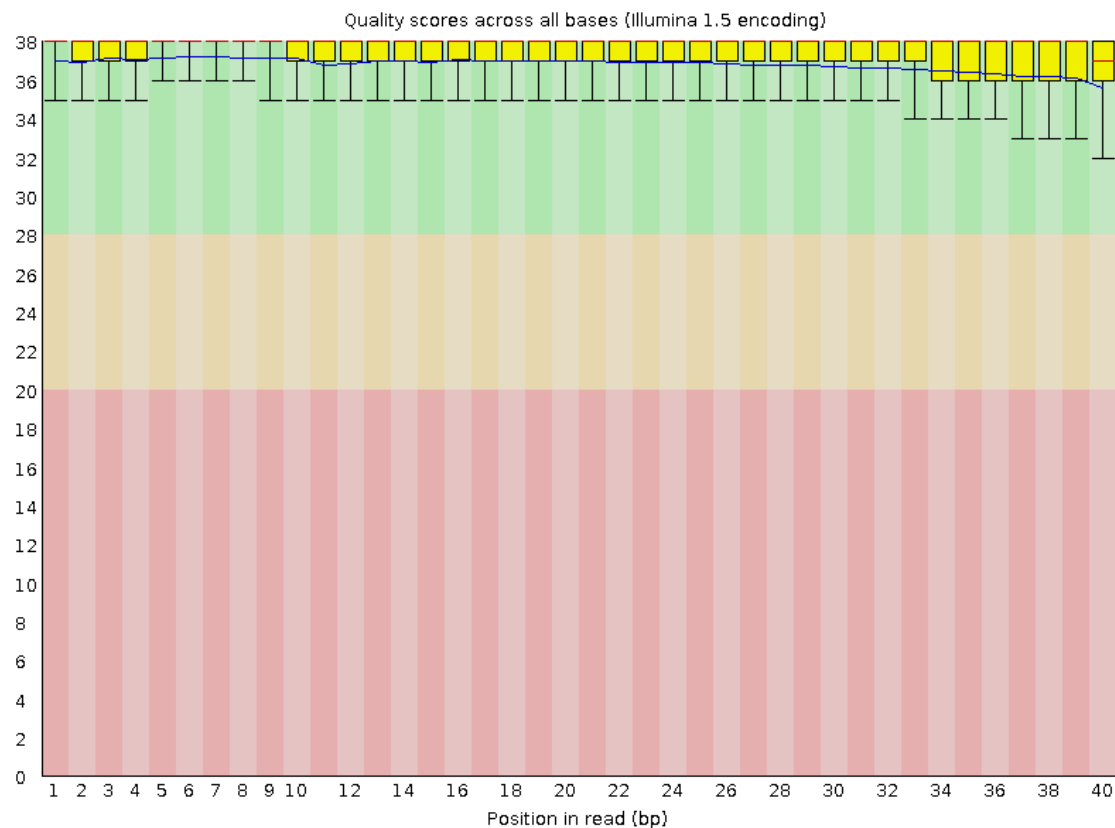
- Developed by Babraham Institute.
- Data can be input in FASTQ, SAM/BAM format.
- Visual overview of issues with data
- Data can be exported in HTML format



The screenshot shows the FastQC application window. The title bar reads 'FastQC'. The menu bar has 'File' and 'Help'. The file name 'PG15-BW001347.R1.fastq.gz' is displayed in the top bar. On the left, a list of modules is shown with status icons: Basic Statistics (green check), Per base sequence quality (yellow exclamation mark), Per tile sequence quality (green check), Per sequence quality scores (green check), Per base sequence content (yellow exclamation mark), Per sequence GC content (red X), Per base N content (green check), Sequence Length Distribution (green check), Sequence Duplication Levels (red X), Overrepresented sequences (red X), and Adapter Content (red X). The main panel is titled 'Basic sequence stats' and contains a table with the following data:

Measure	Value
Filename	PG15-BW001347.R1.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	500000
Sequences flagged as poor quality	0
Sequence length	250
%GC	43

Quality Scores



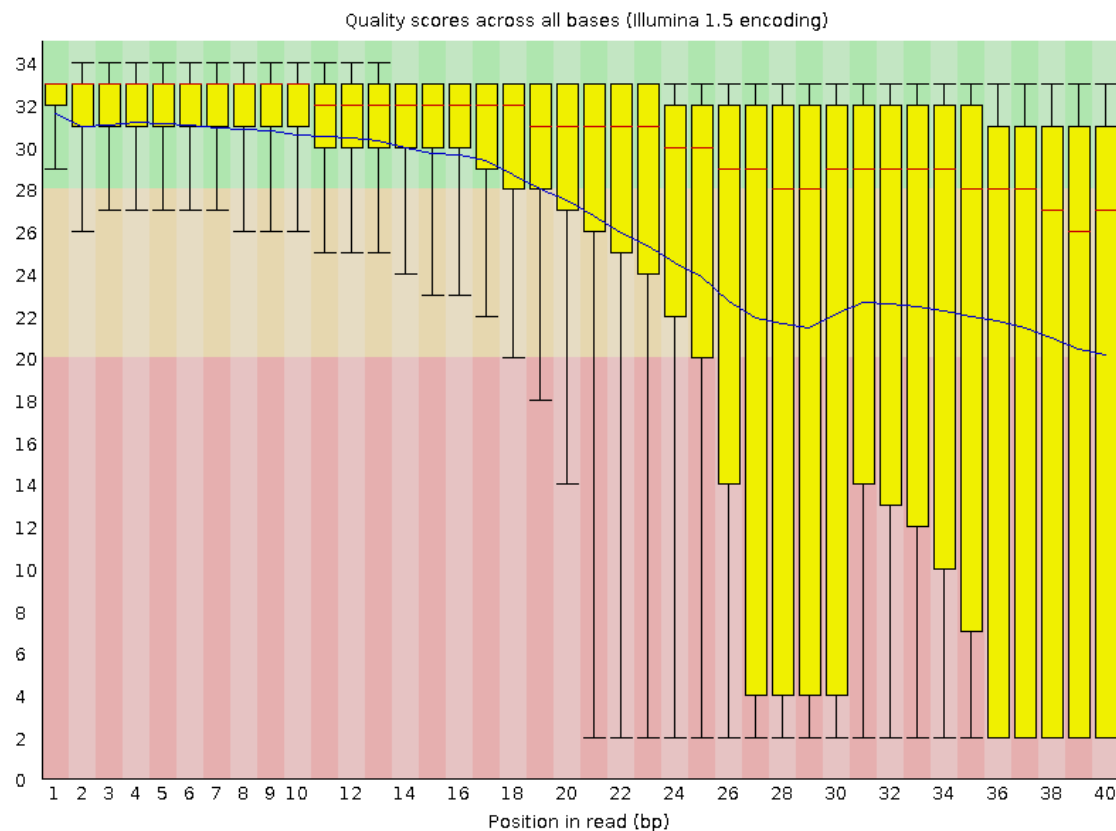
Q40 = 1 in 10,000

Q30 = 1 in 1,000

Q20 = 1 in 100

Q10 = 1 in 10

Quality Scores



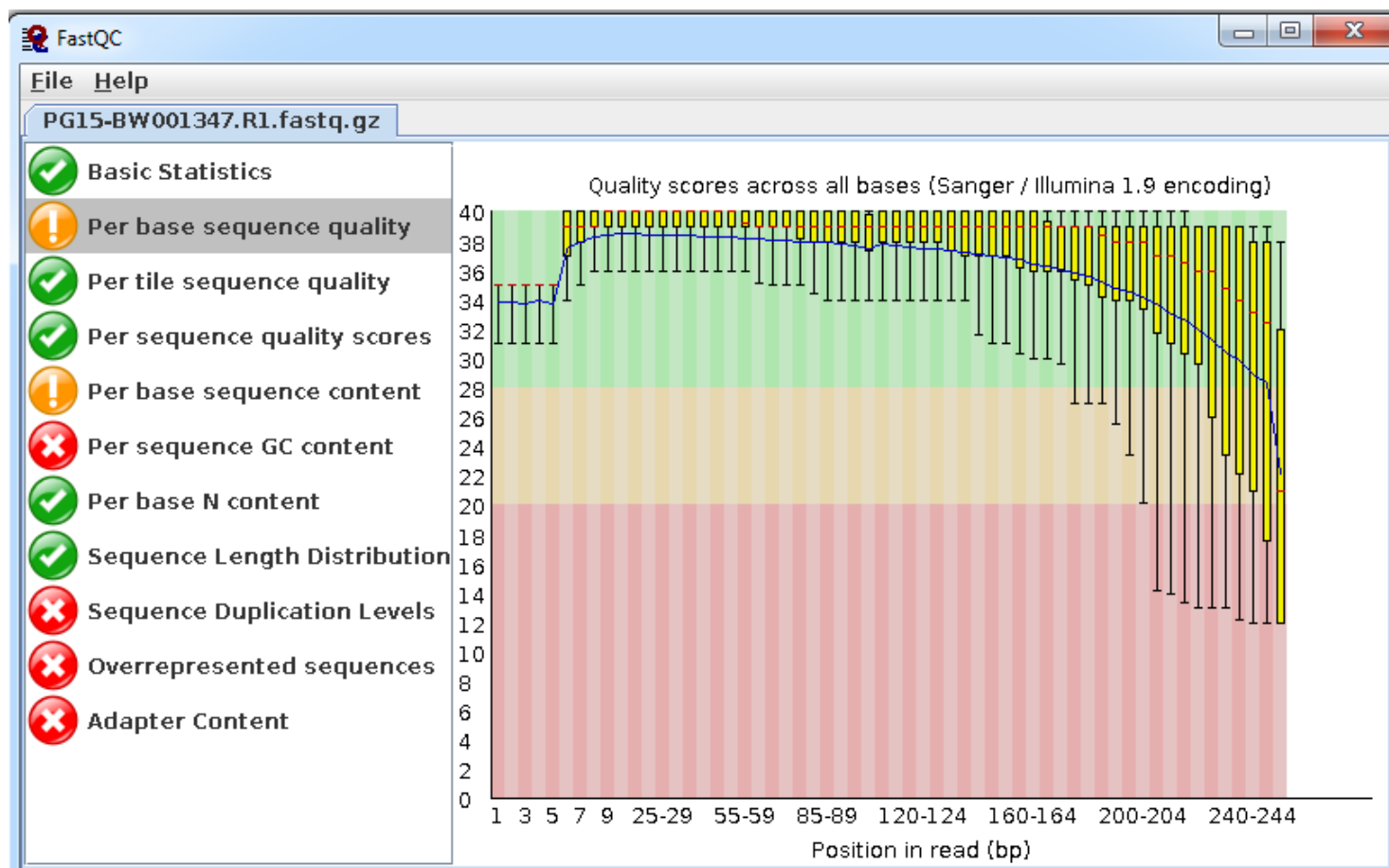
Q40 = 1 in 10,000

Q30 = 1 in 1,000

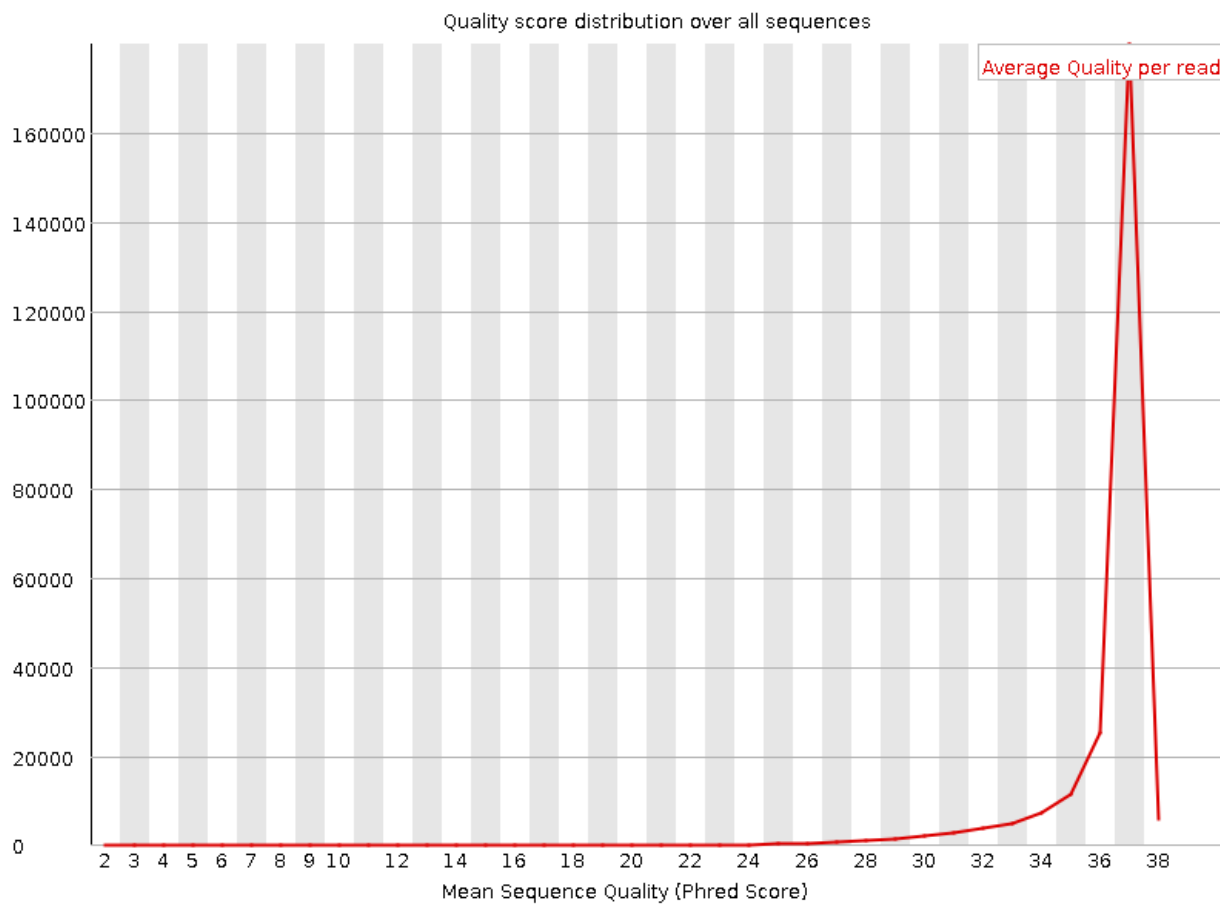
Q20 = 1 in 100

Q10 = 1 in 10

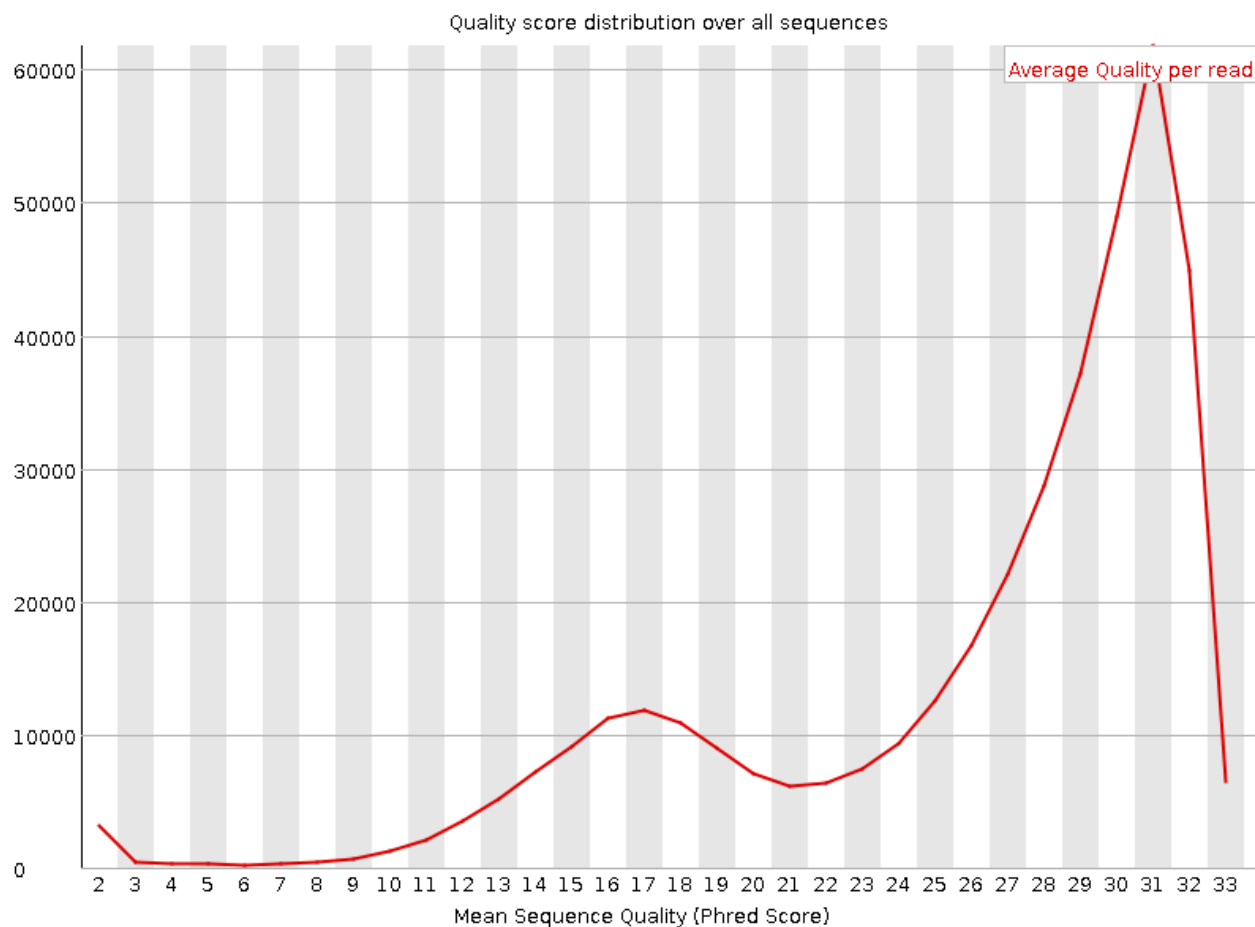
Quality Scores



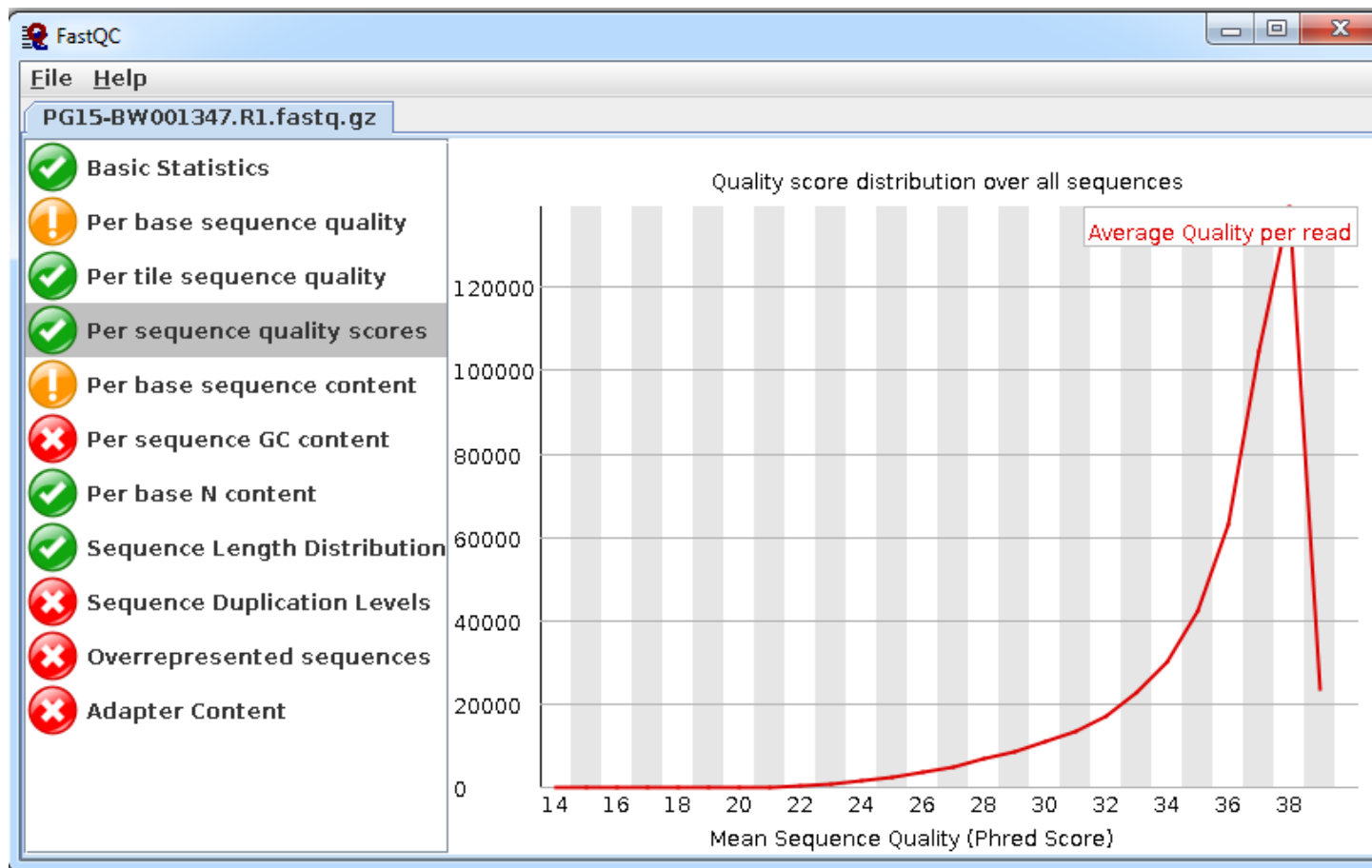
Average Quality



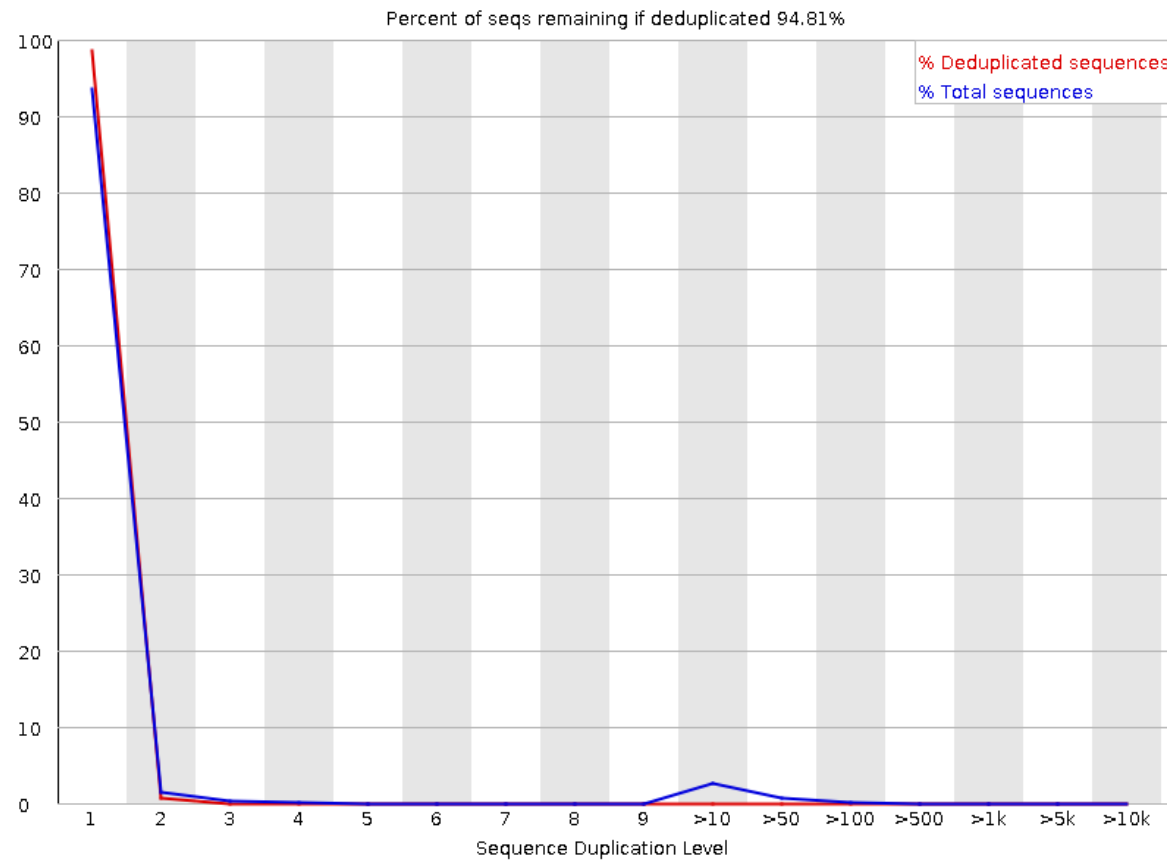
Average Quality



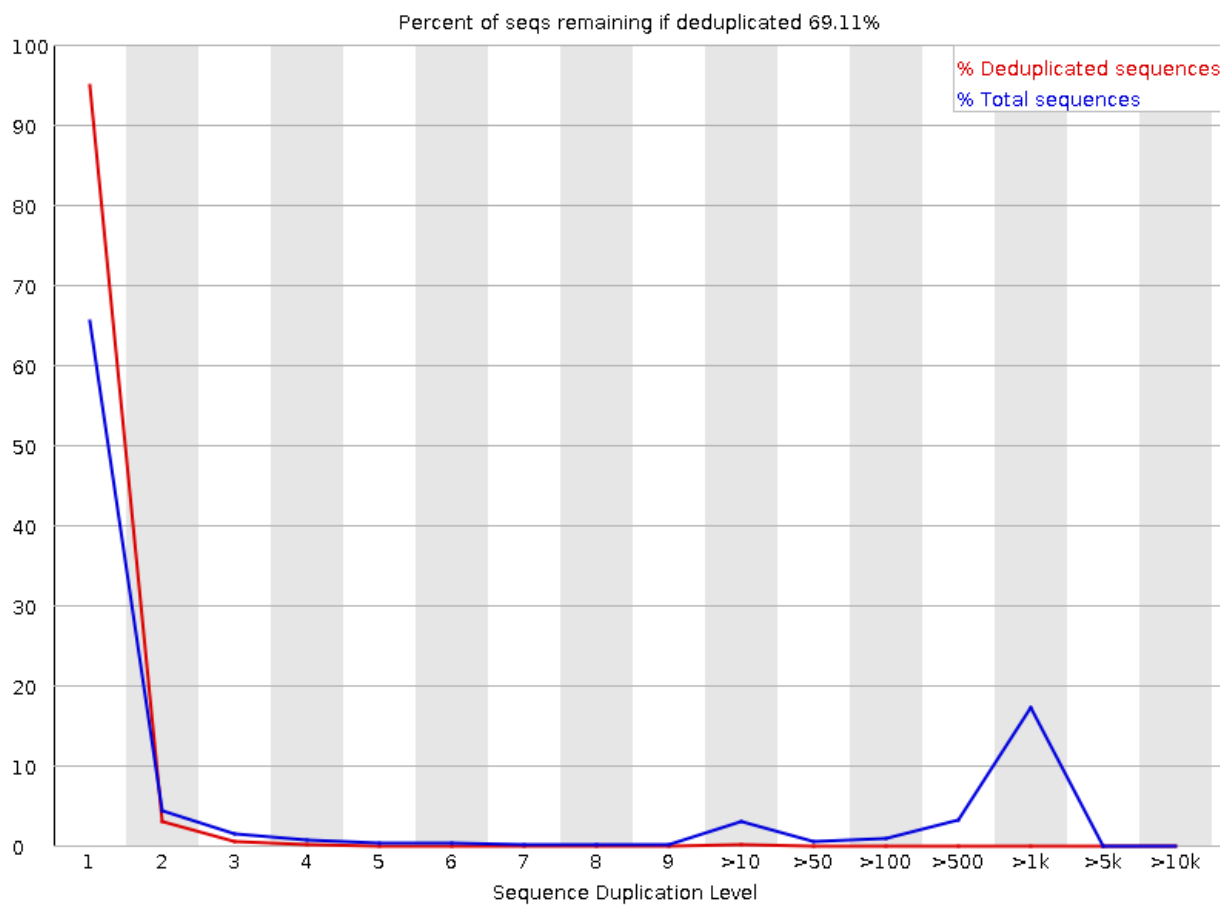
Average Quality



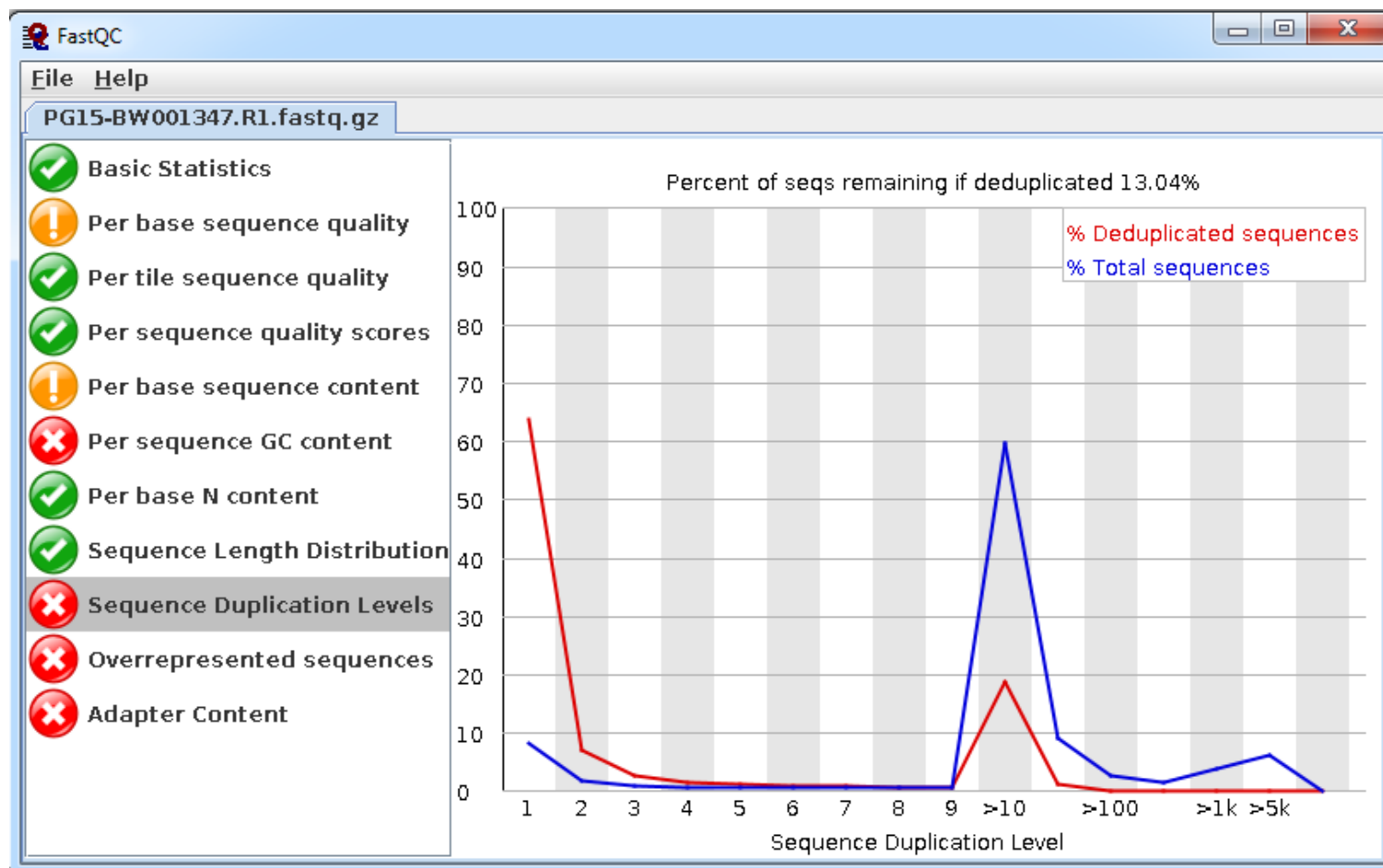
Duplicate reads



Duplicate reads



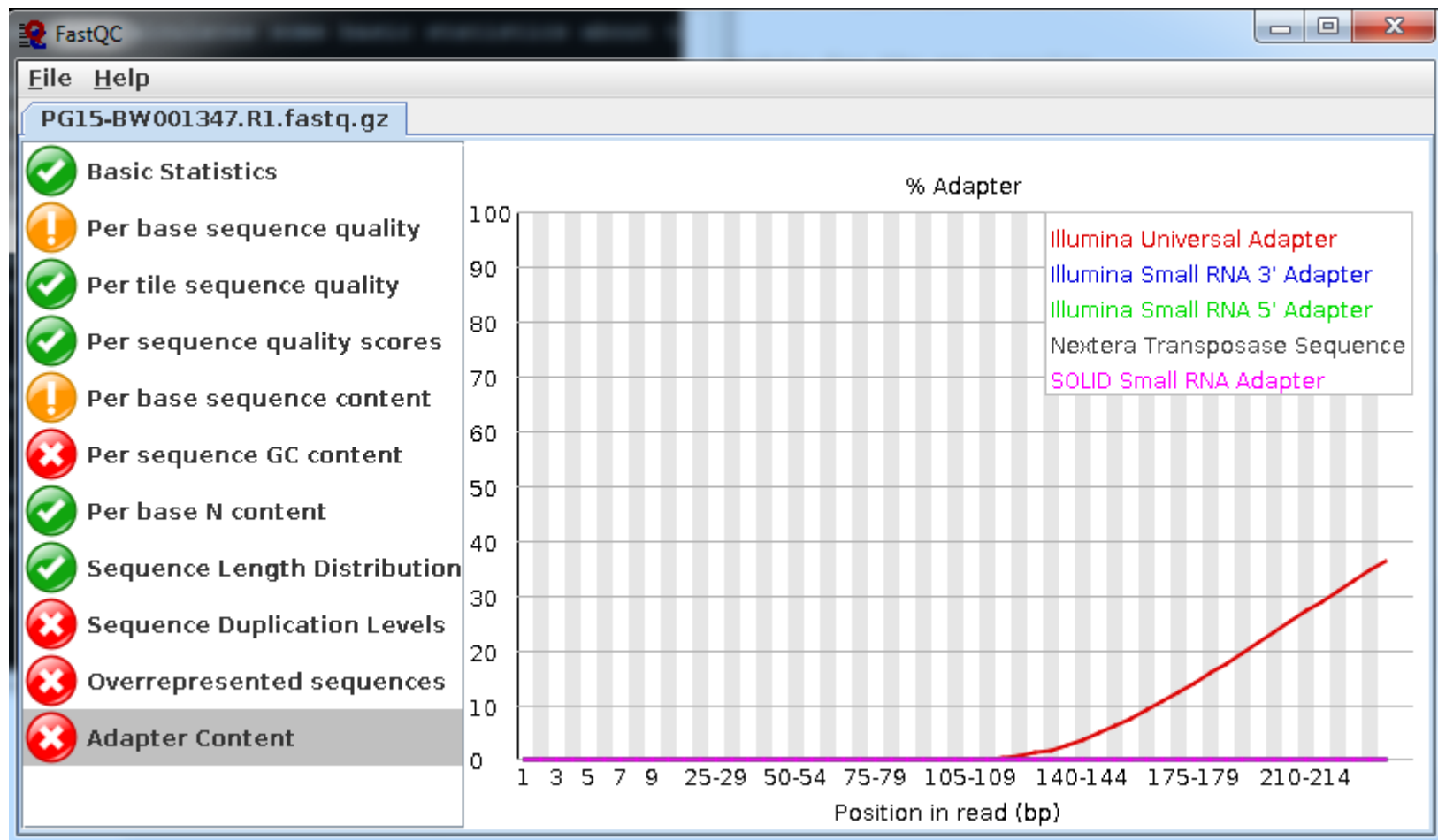
Duplicate reads



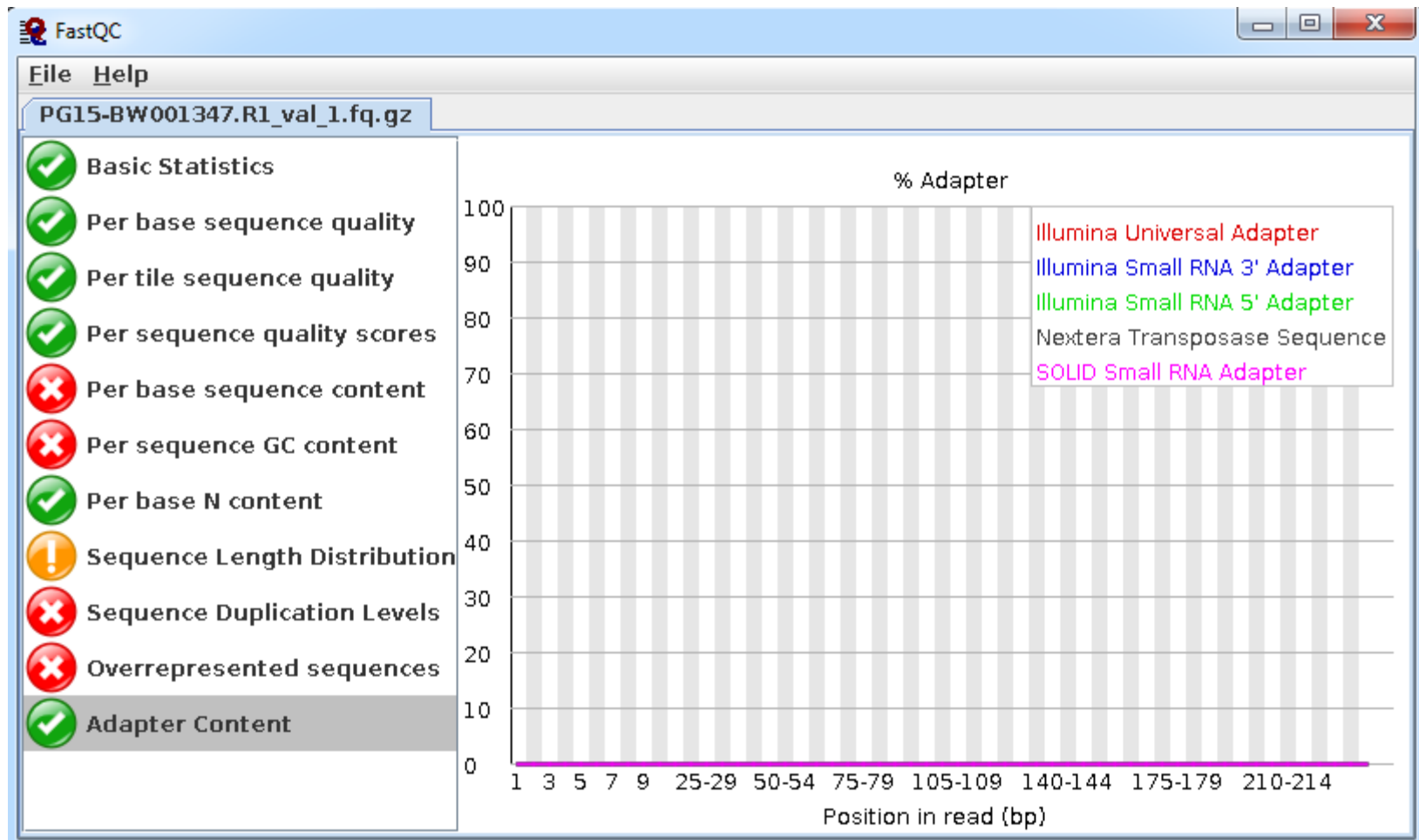
Adapter and Quality Trim – Trim Galore

- Removes adapters
- Trims low quality reads
- Removes read with ambiguous bases
- Removes short sequences
- Uses FastQ files as input

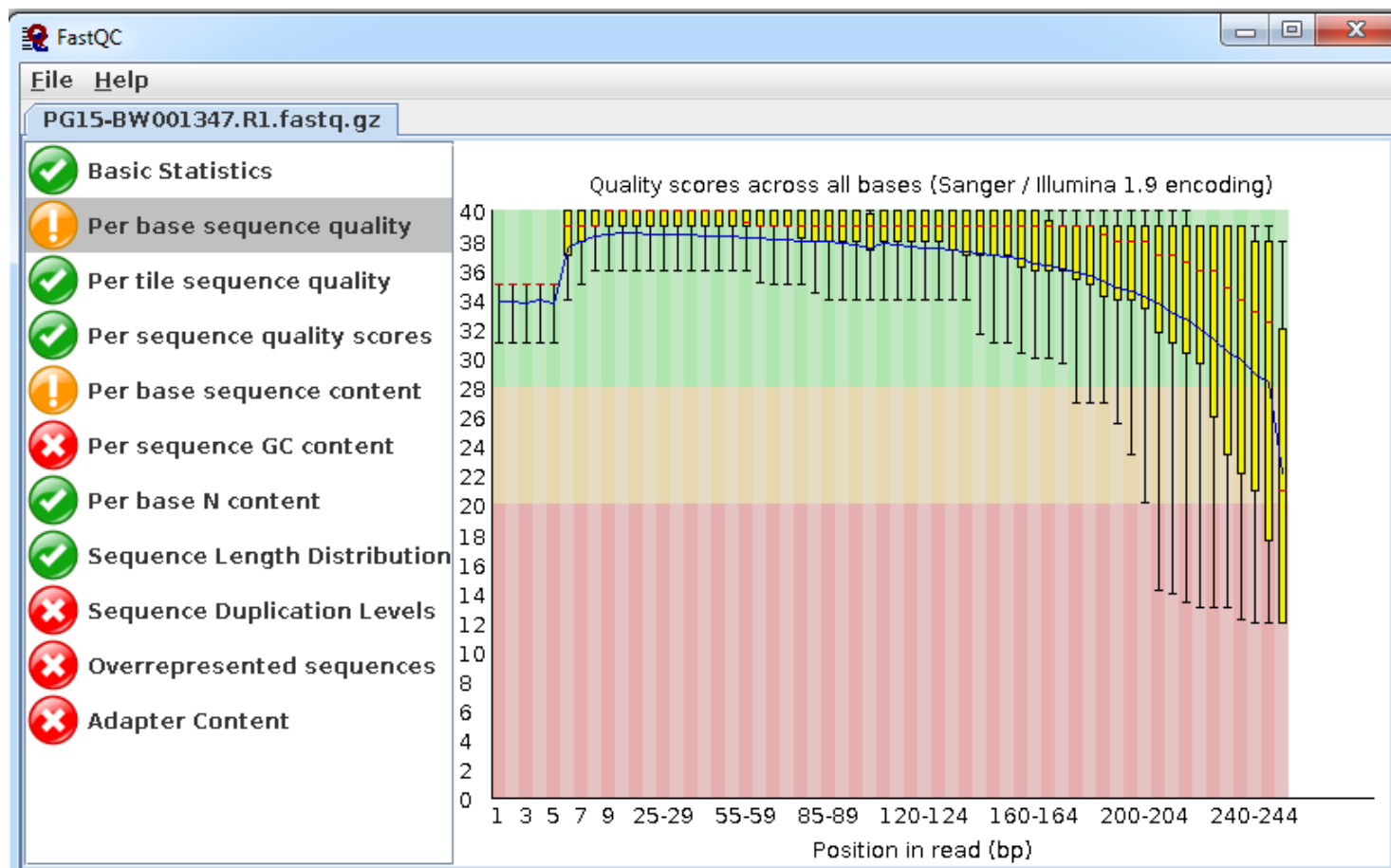
Adapter Trim



Adapter Trim



Quality Scores



Quality Scores

