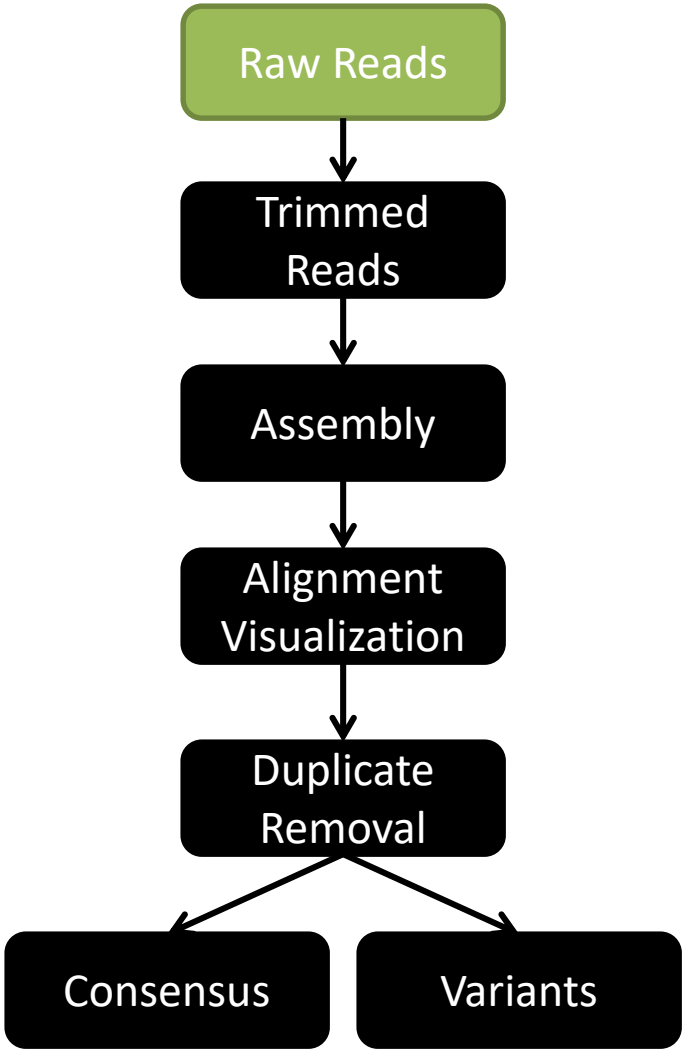
Quality Control and Visualization

Quality Check

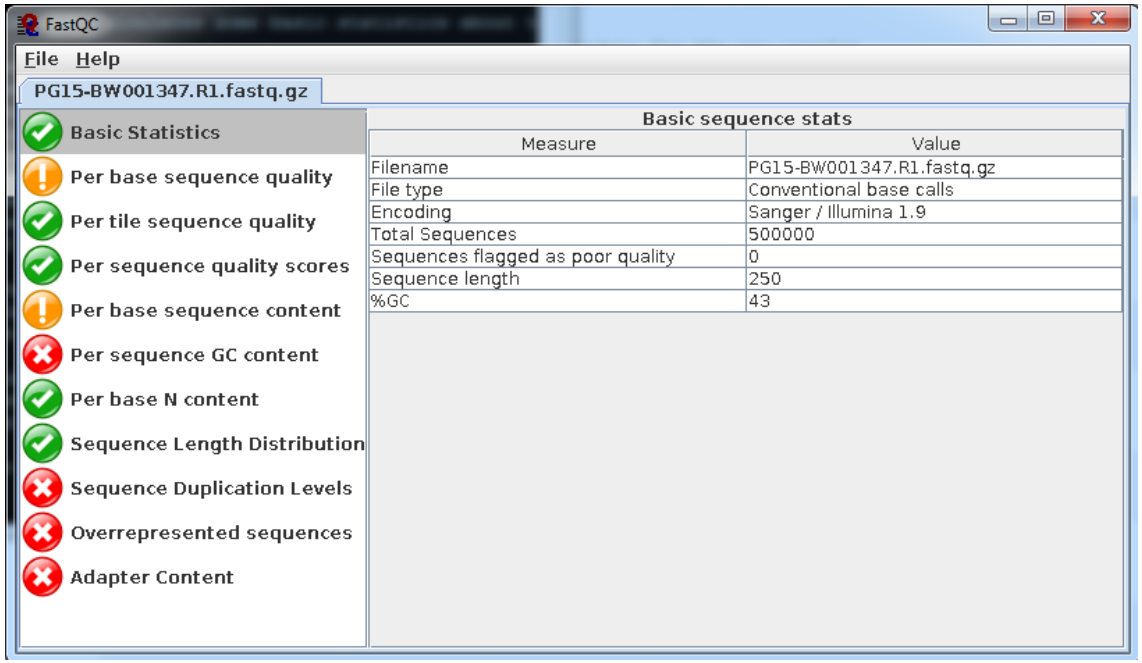
- Garbage in Garbage out
- Low quality data generated towards the end of reads.
- High level of duplication and low quality reads can effect variant calling.
- To identify read through into adapters.
- Basic statistics of a run.

Where do we start?

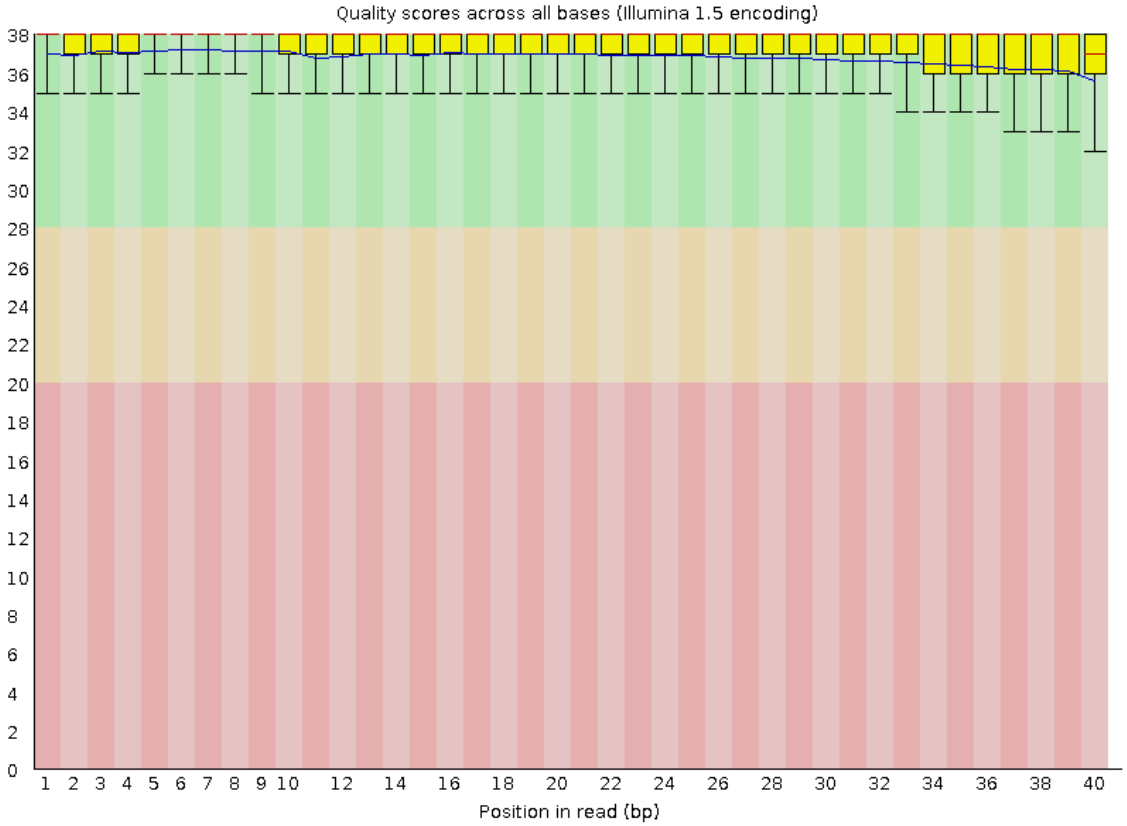


Quality Check - FASTQC

- 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

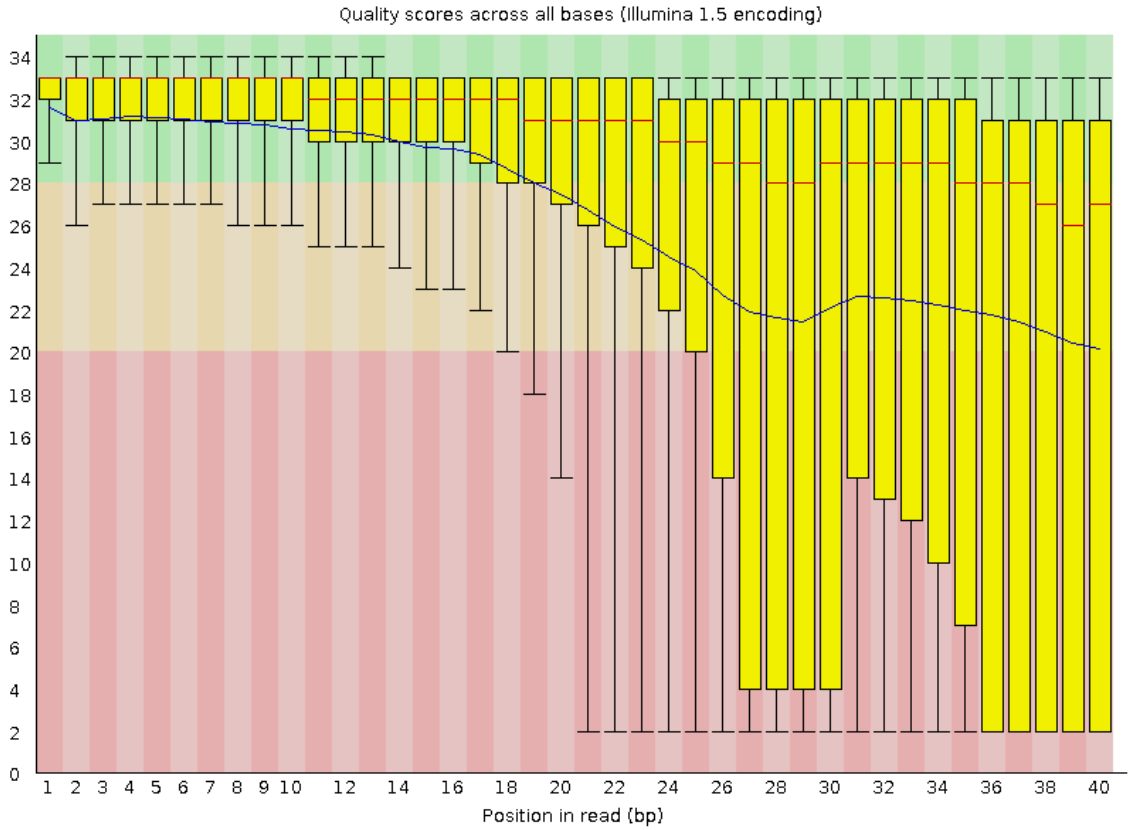


Quality Scores



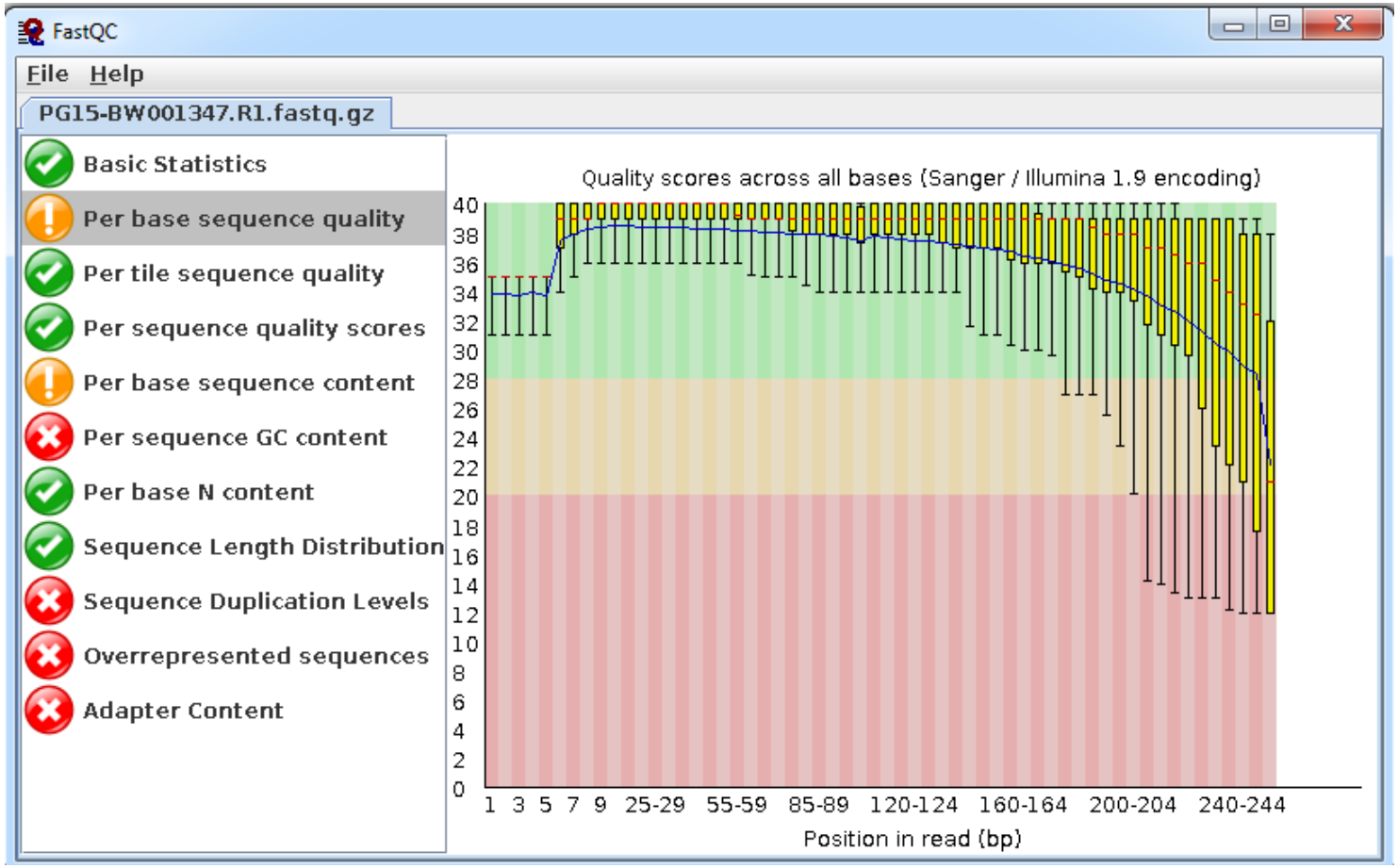
Q40 = 1 in 10,000
Q30 = 1 in 1,000
Q20 = 1 in 100
Q10 = 1 in 1

Quality Scores

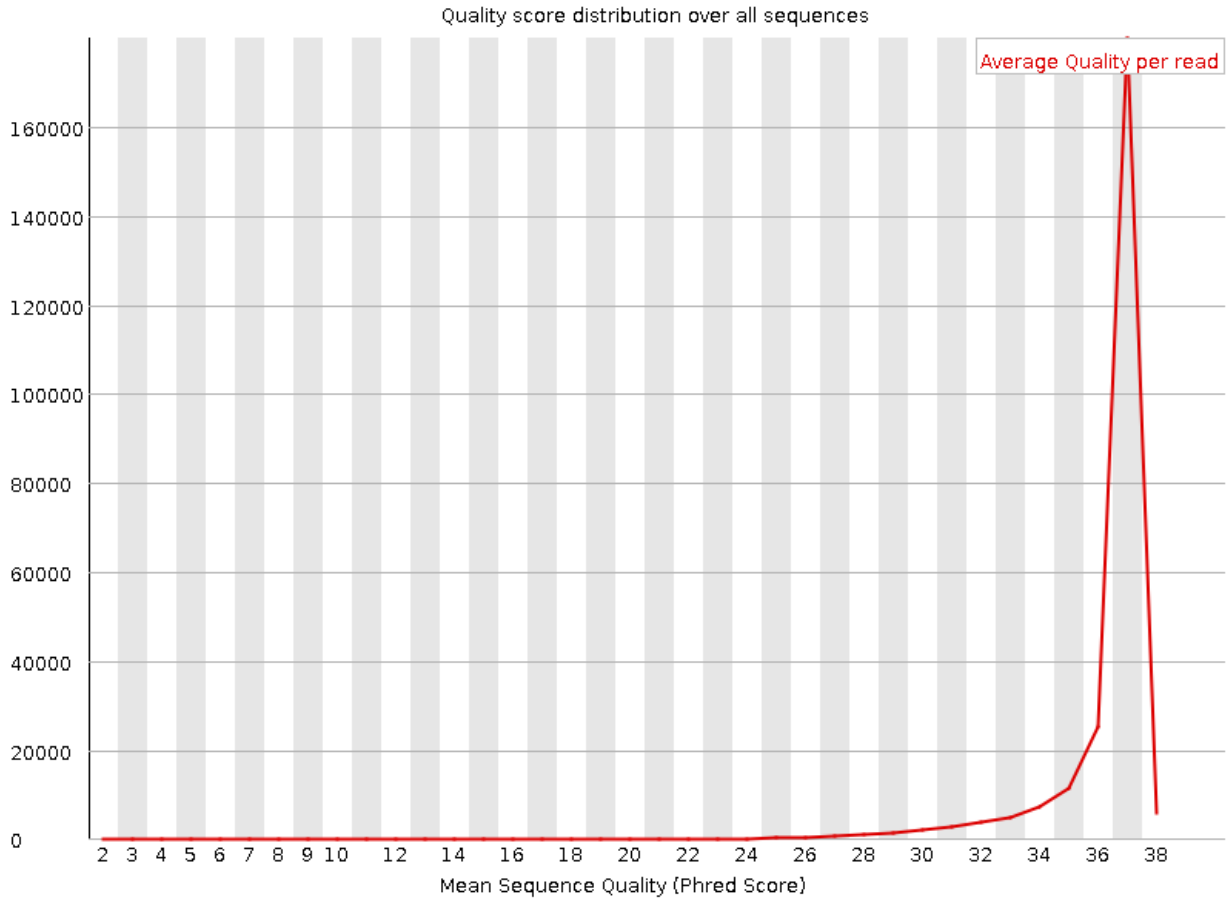


Q40 = 1 in 10,000
Q30 = 1 in 1,000
Q20 = 1 in 100
Q10 = 1 in 1

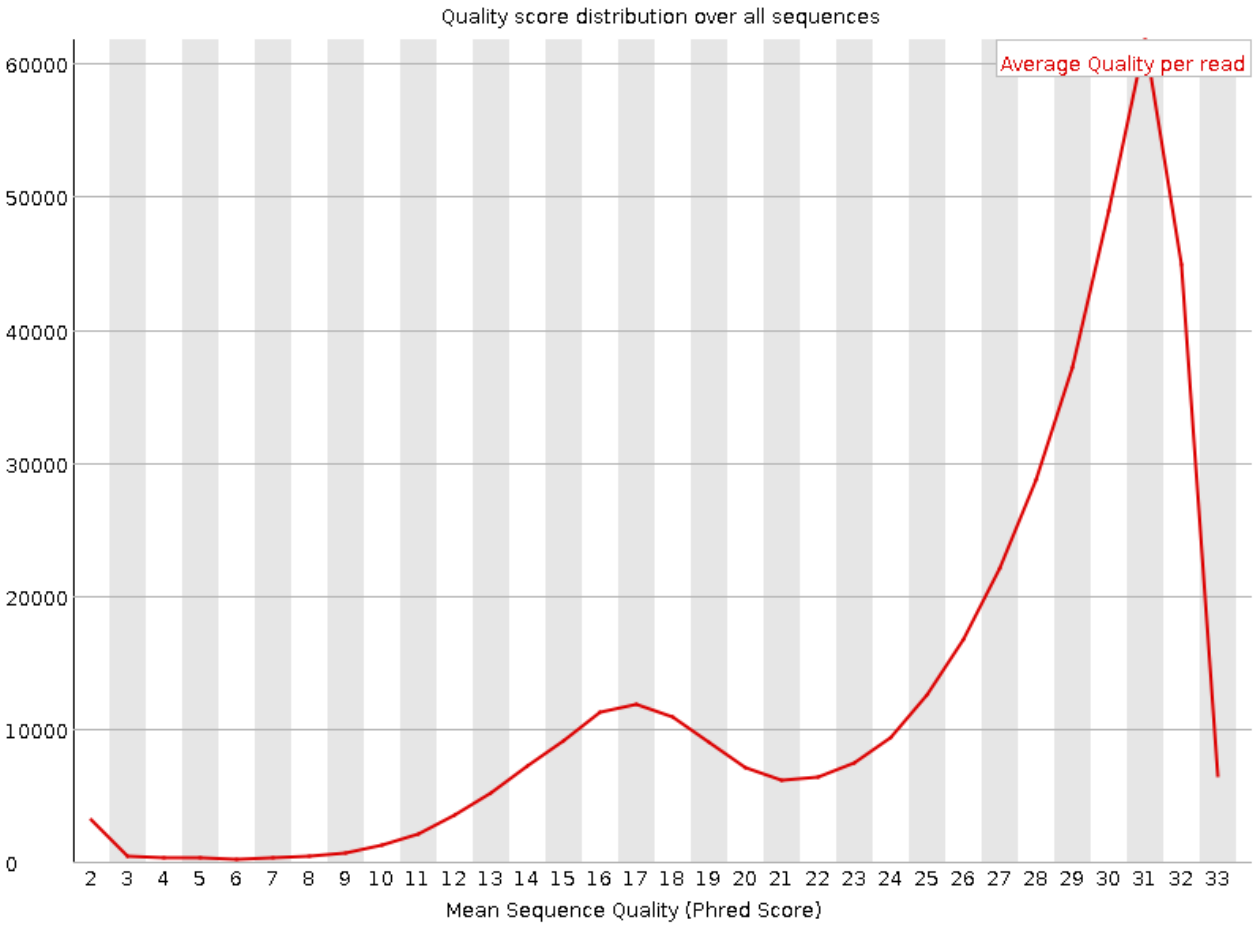
Quality Scores



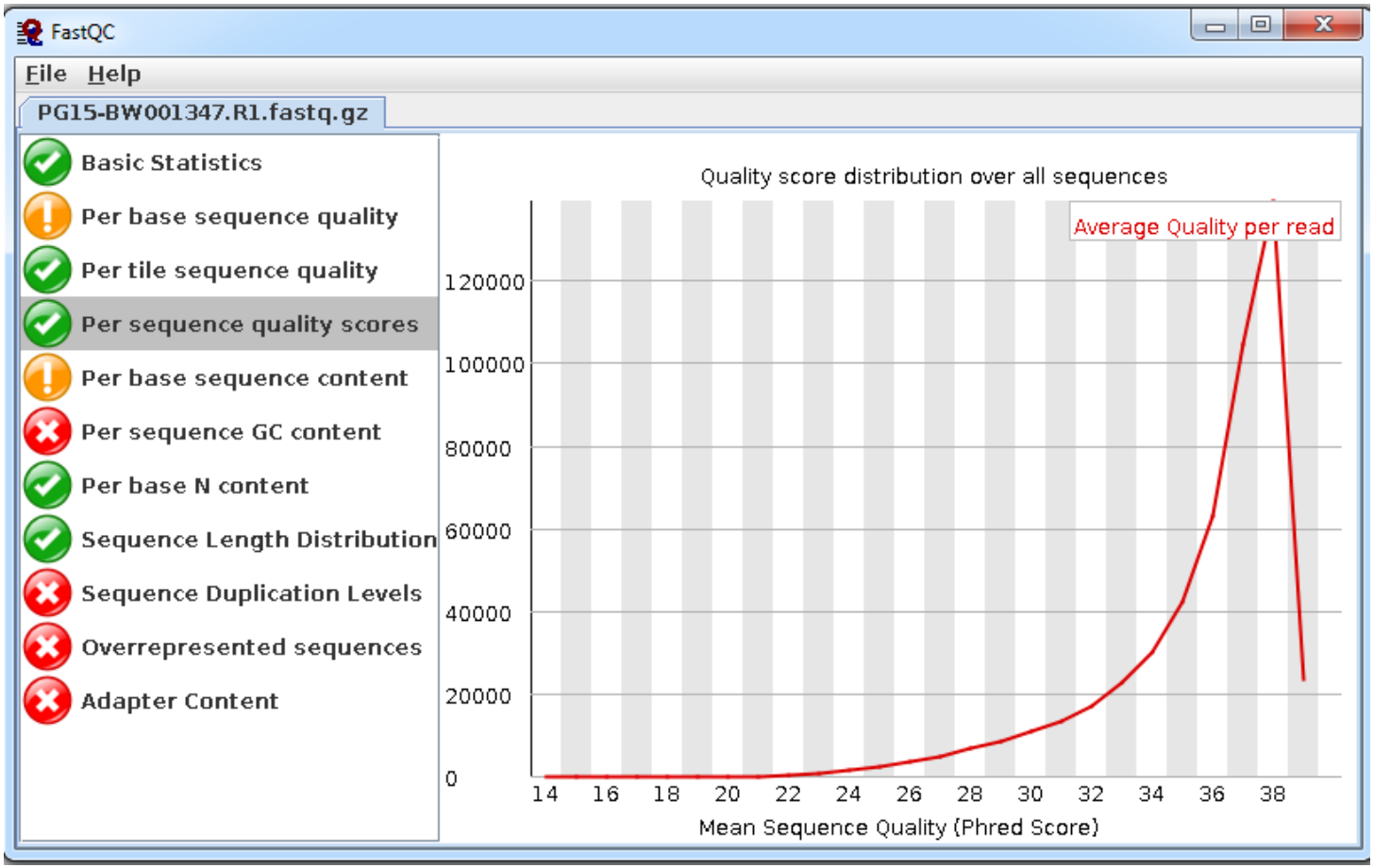
Average Quality



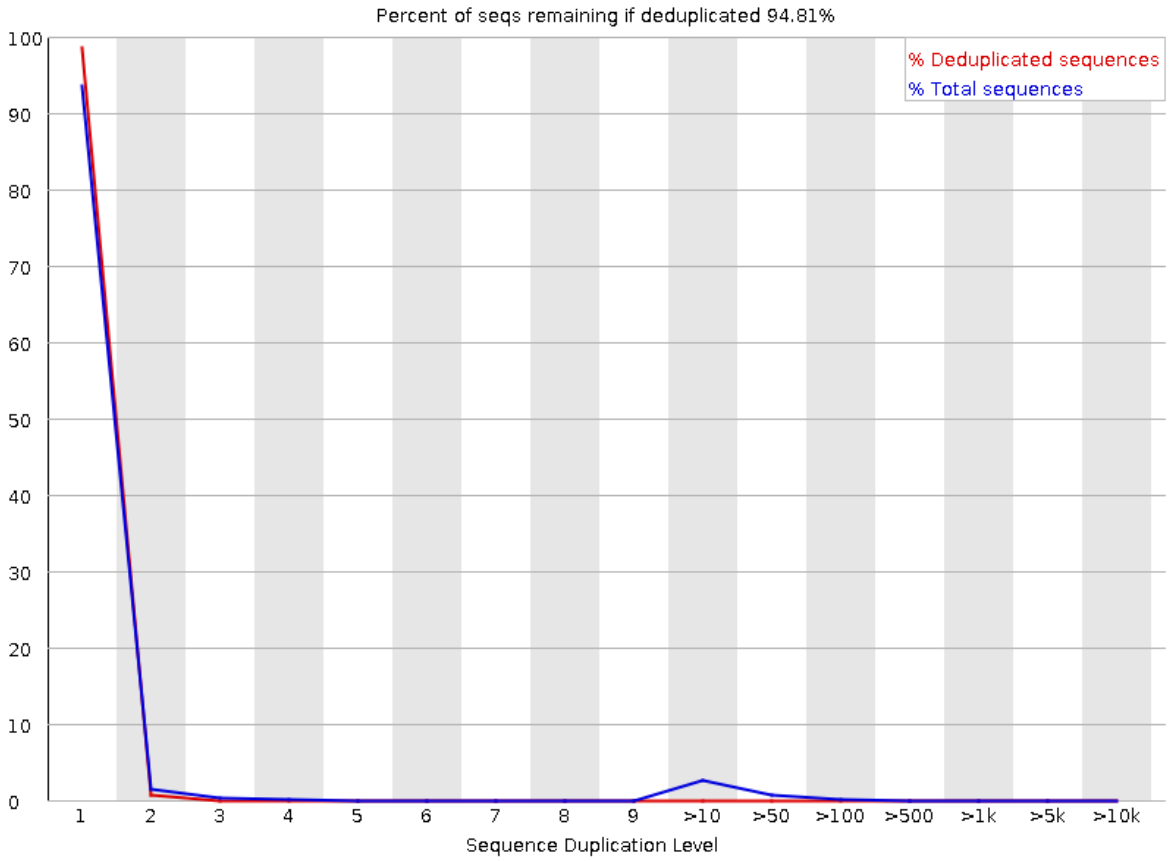
Average Quality



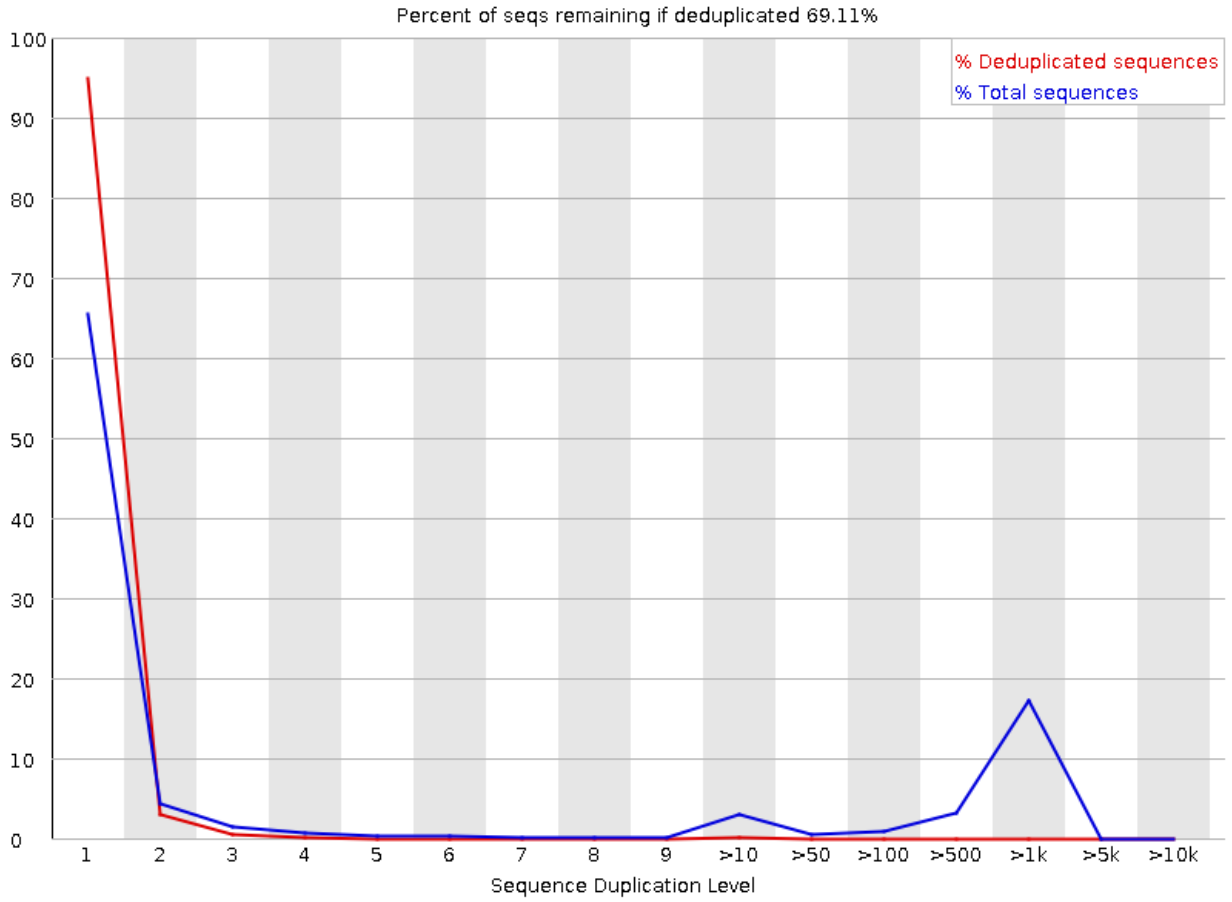
Average Quality



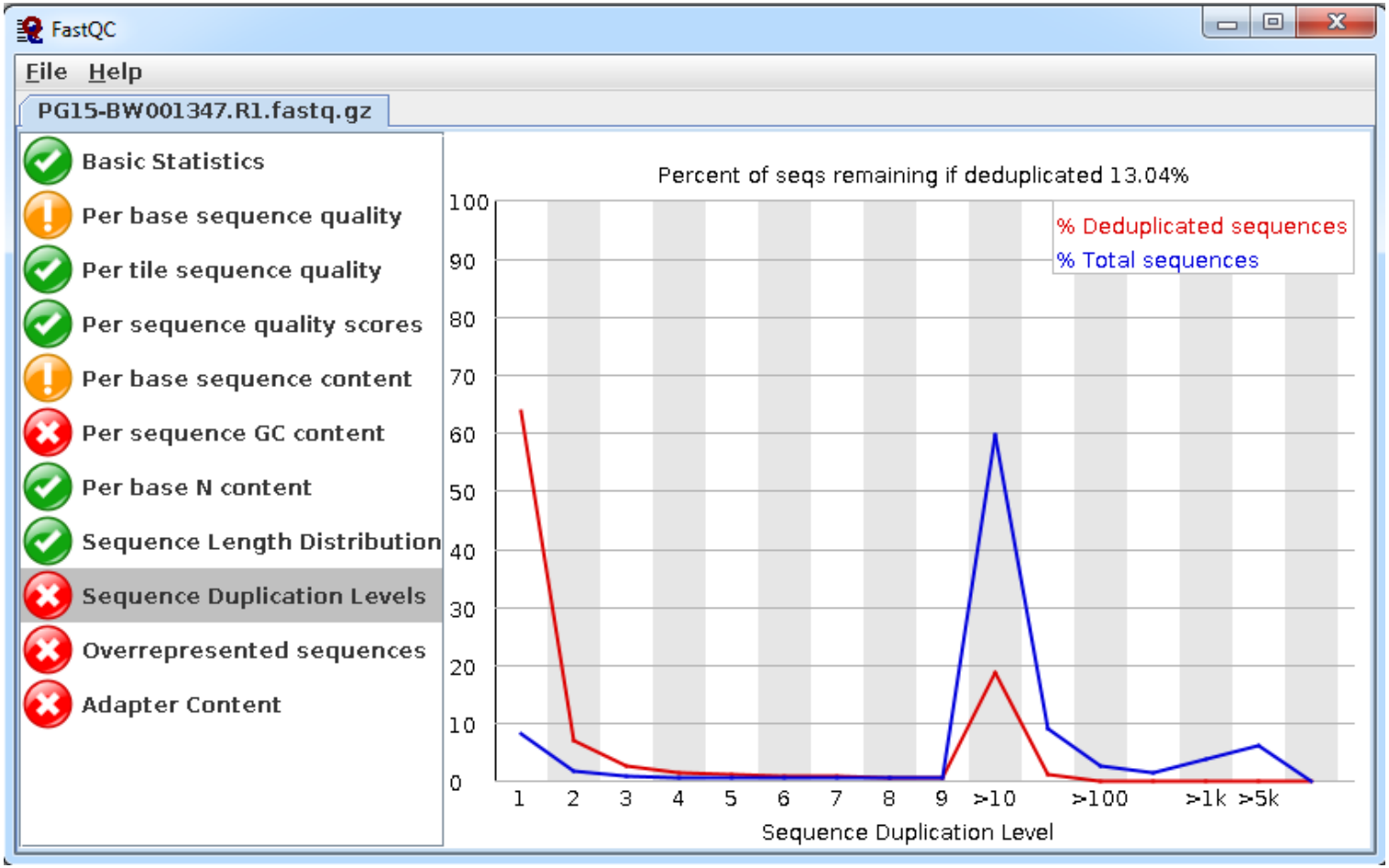
Duplicate reads



Duplicate reads



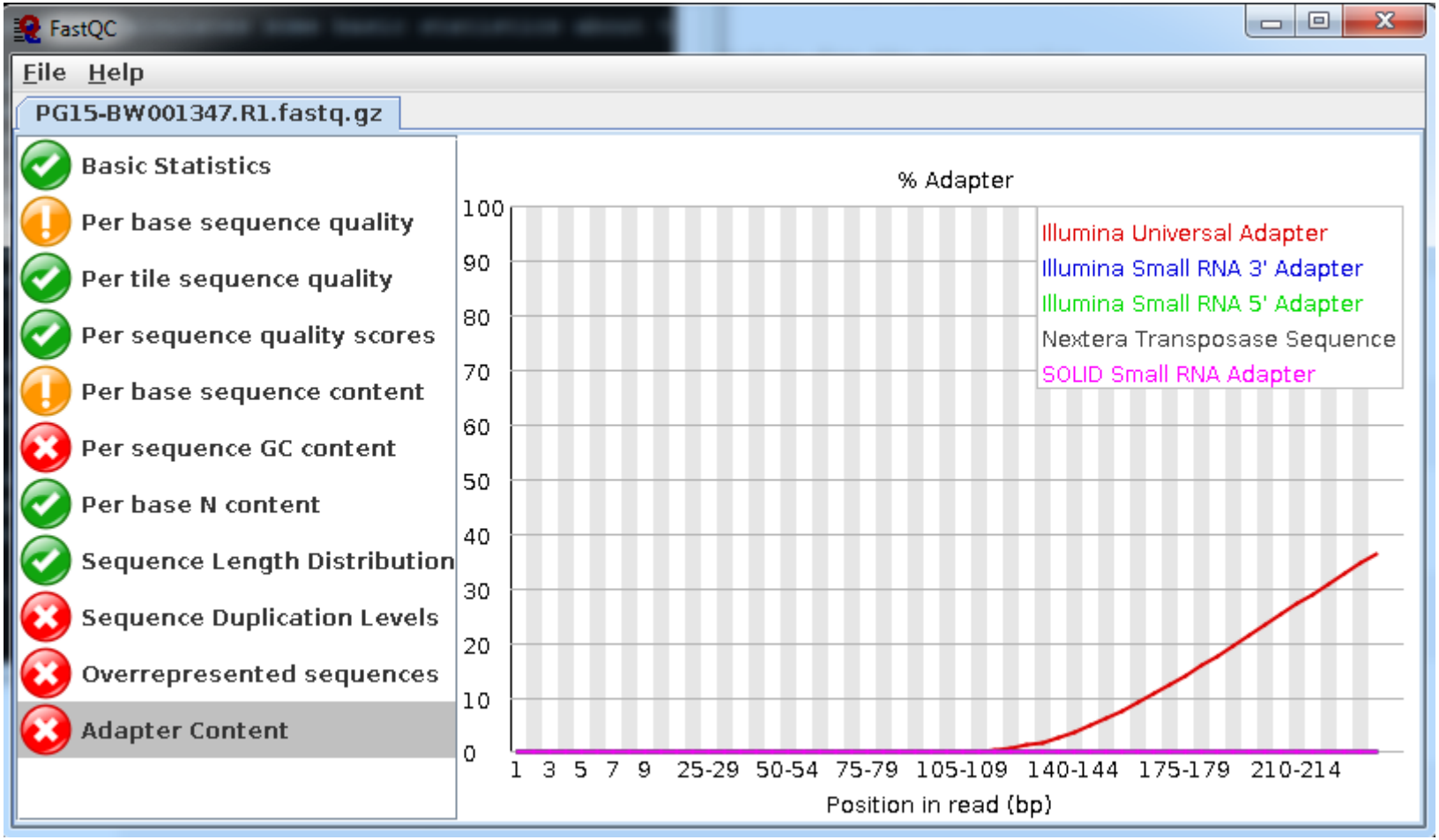
Duplicate reads



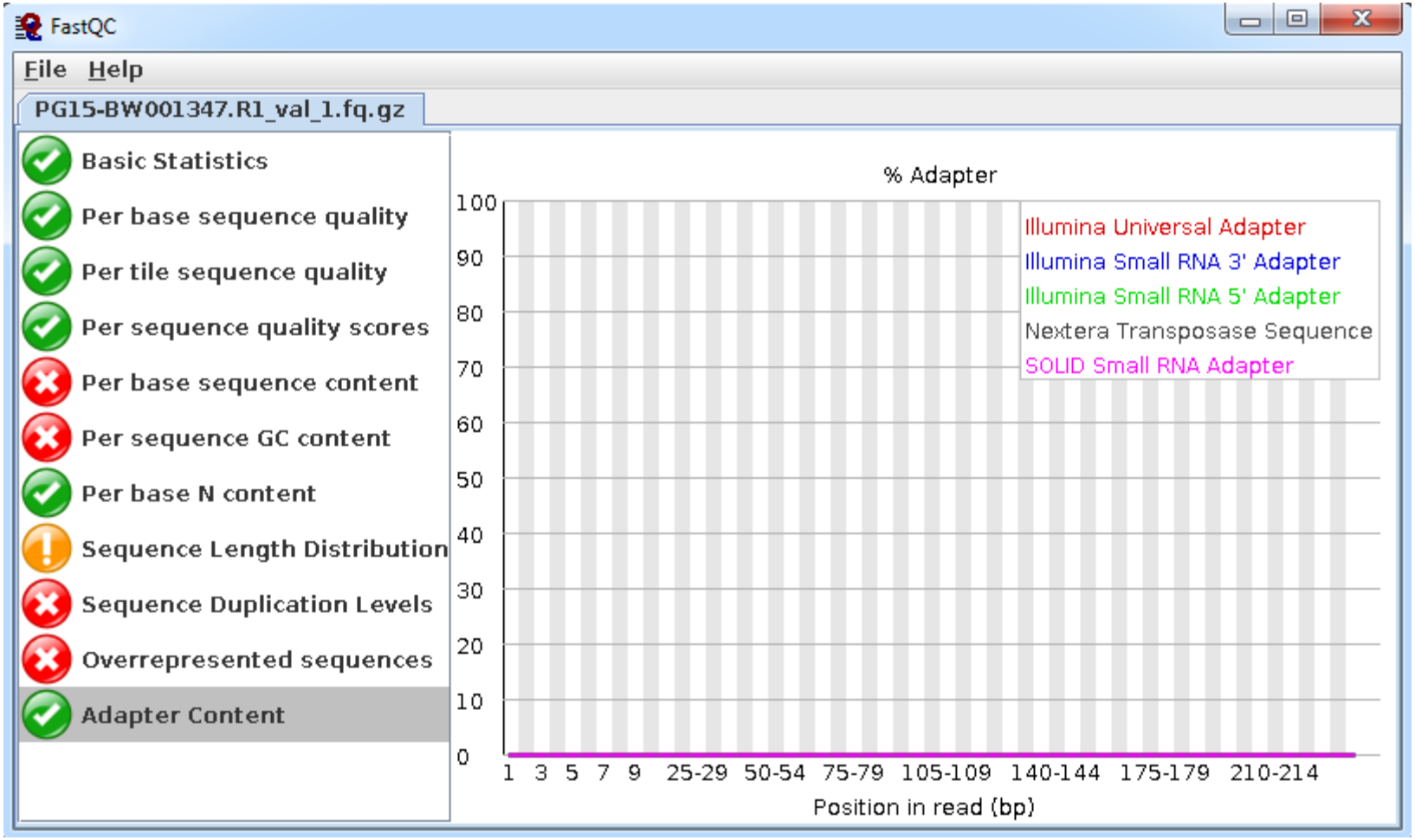
Adapter and Quality Trim – Trim Galore

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

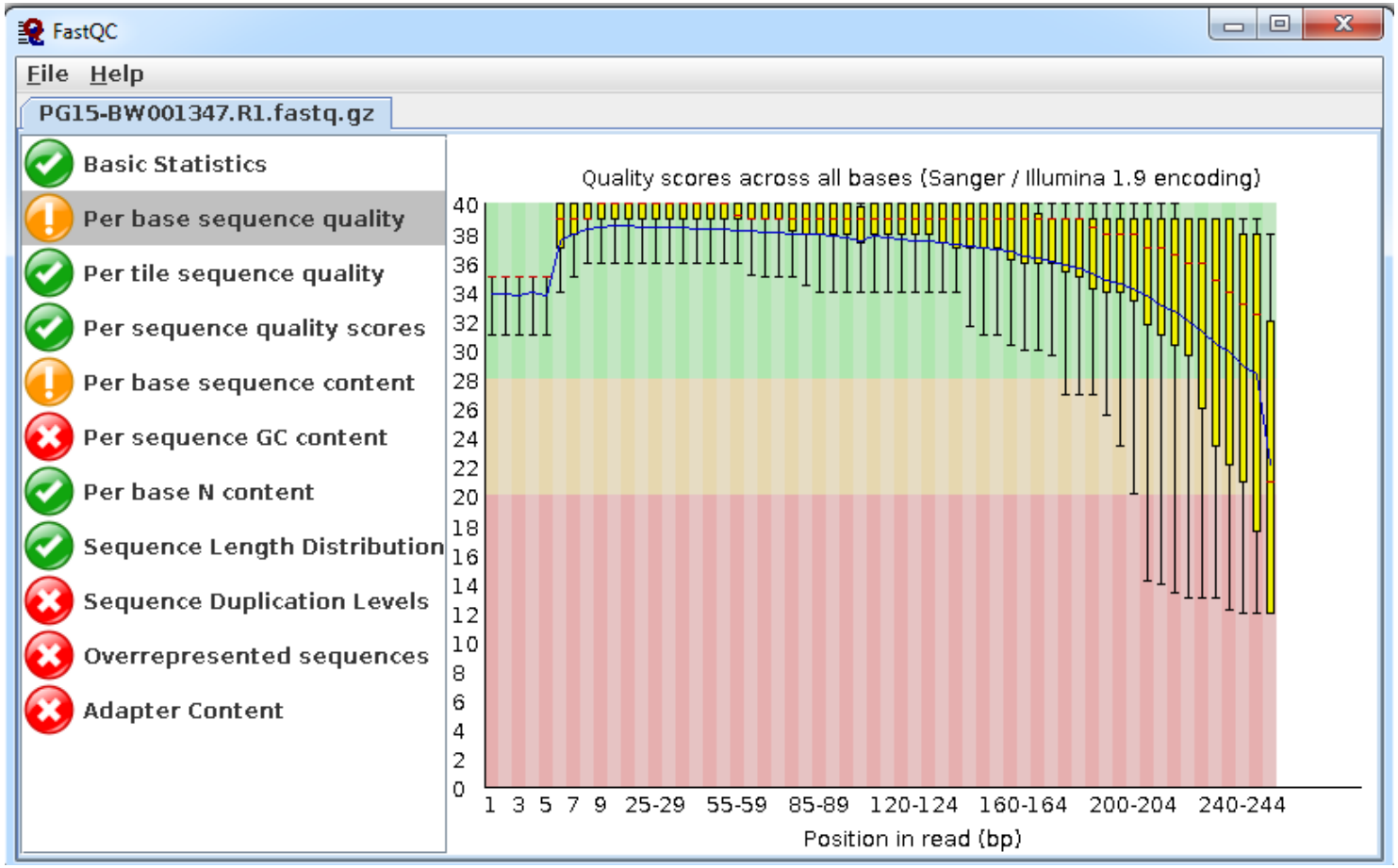
Adapter Trim



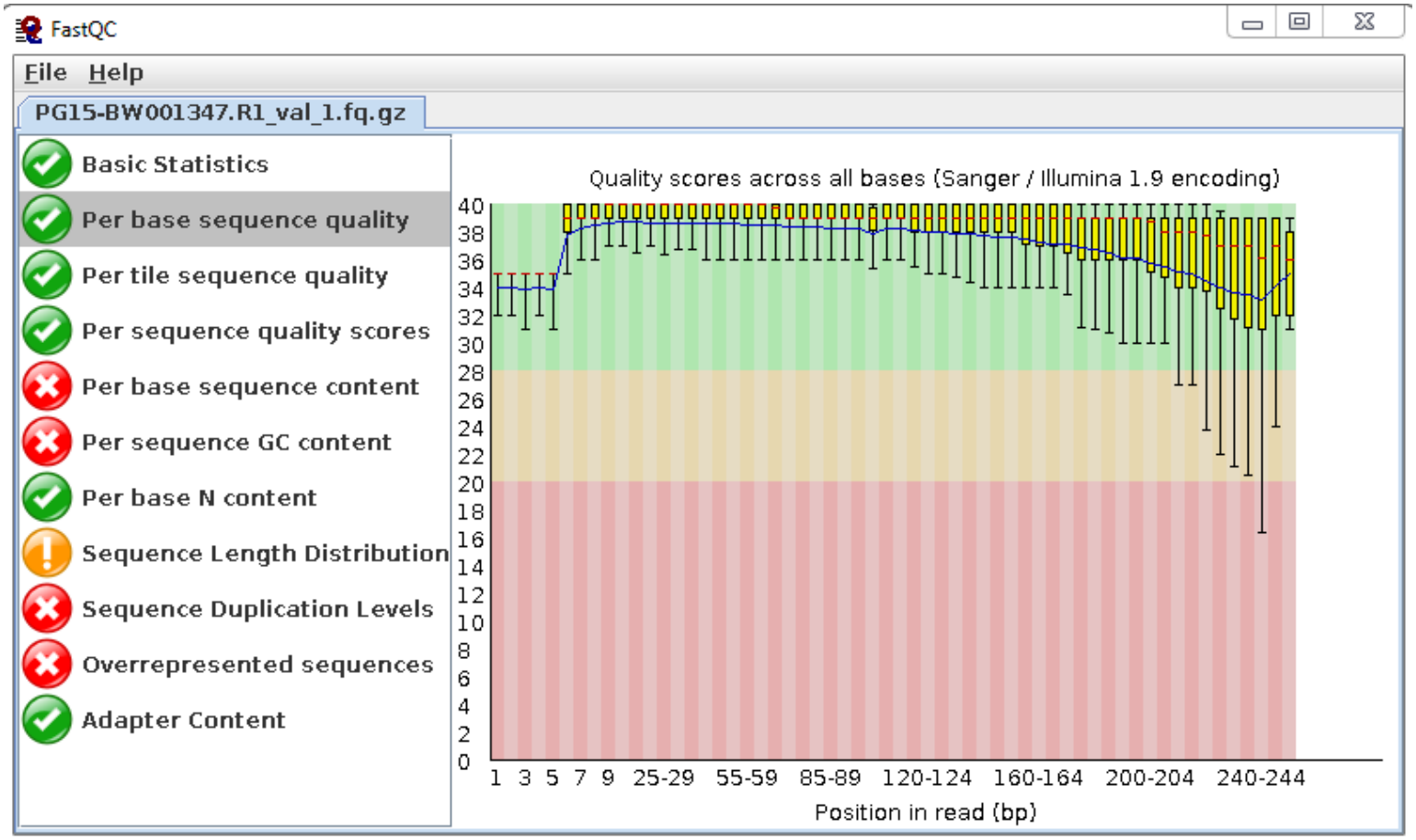
Adapter Trim

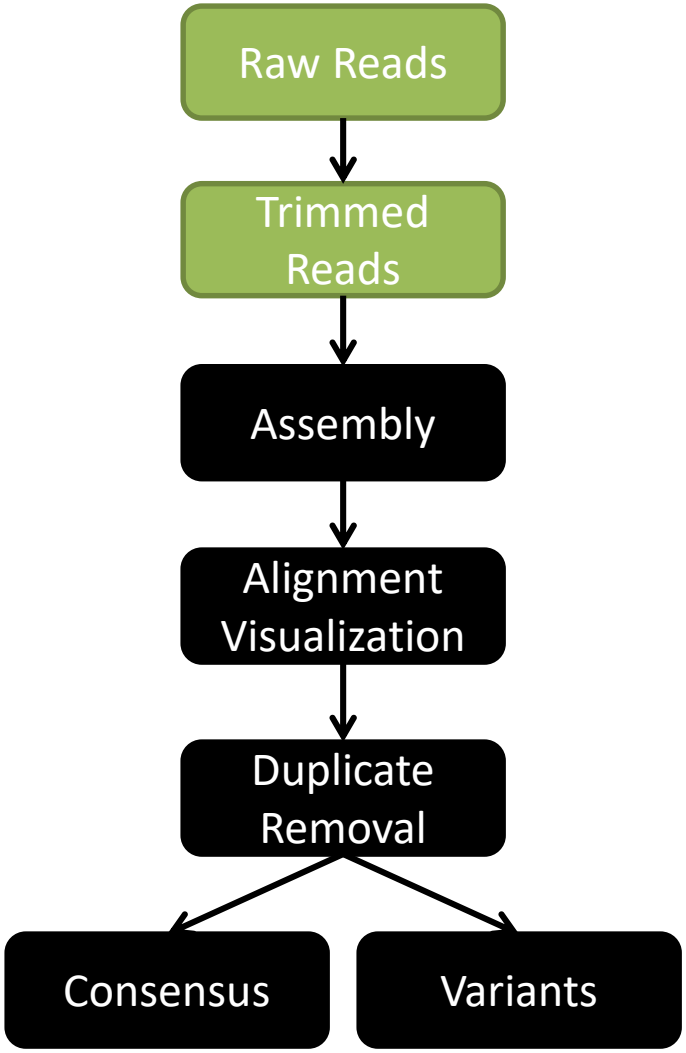


Quality Scores



Quality Scores




Raw Reads
Trimmed Reads
Assembly
Alignment Visualization
Duplicate Removal
Consensus
Variants