

Tim Tran

- Does the document appear to have an appropriate QC?

Yes the document have very good QC for both the fetal samples and the Adult samples.

FastQC was used to QC the samples. See the FASTQC results below:

- Is the mapping rates similar for fetal and adult samples?

The mapping rates are similar for both fetal and adult samples. Both have mapping rate of about > 95%

- Is there a trend in the average quality score of mapped reads?

Yes, the average quality score of mapped reads are between 34-38 Phred's score for all the samples. One would think that the adult samples will have lower quality scores because of the lower RIN score indicating poor quality of the RNA. However, their adult samples have about the same quality scores falling in the range above. All mean quality Phred's Score are in the range of 34-40, indicating good quality.

Here are the FASTQC data for Feta Samples:

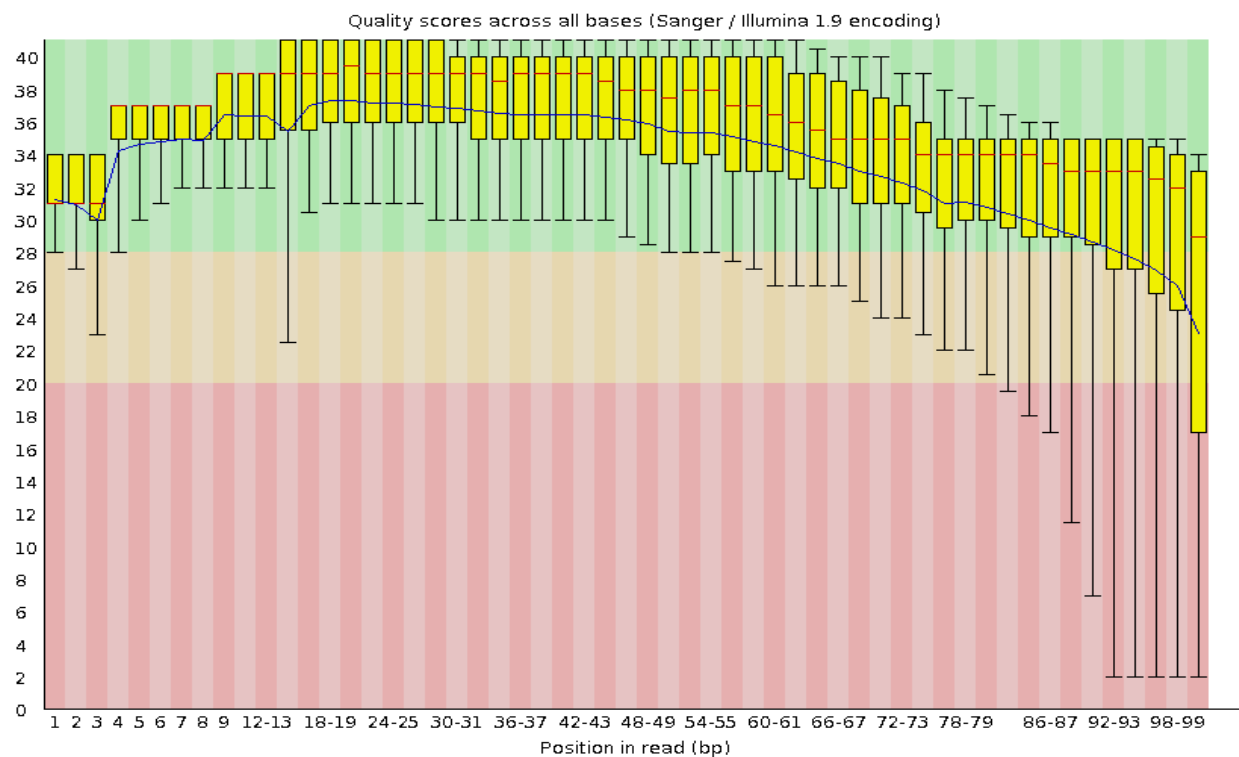
Fetal Samples:

R3452_DLPFC_polyA_RNAseq_total; SRX683795; SRR1554537

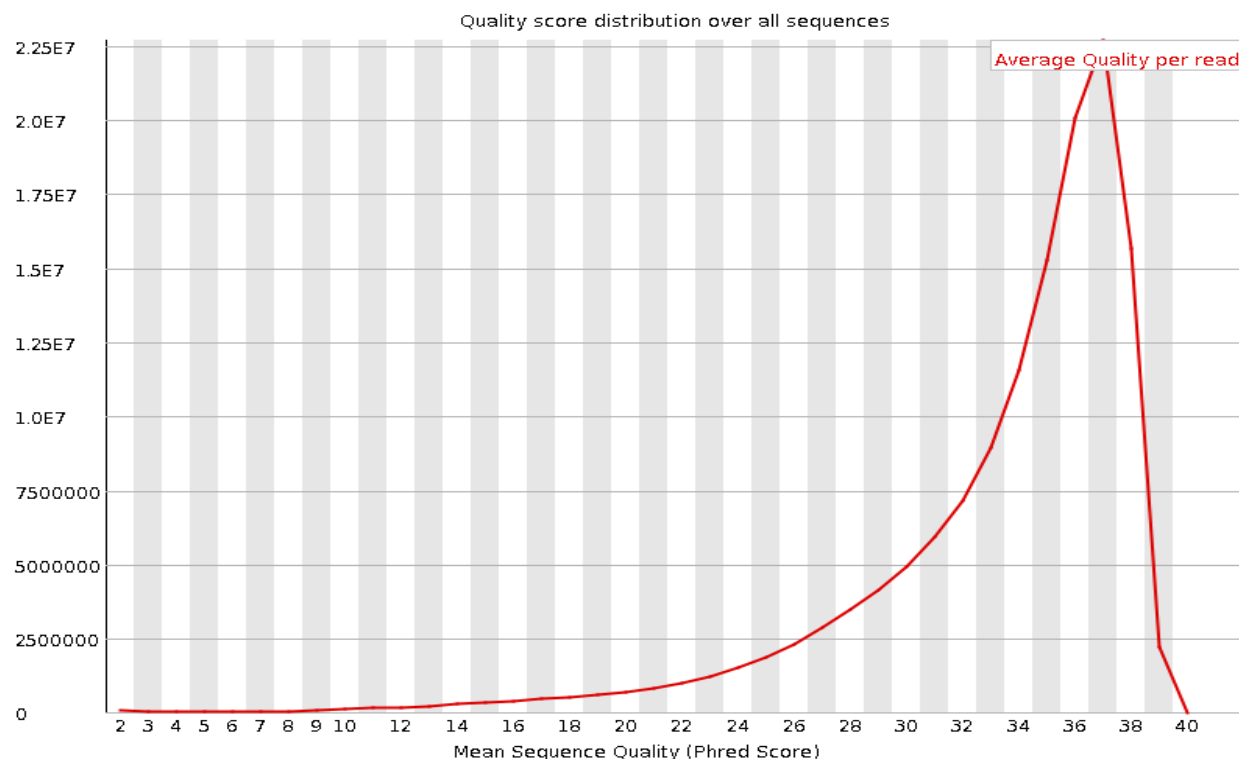
The quality score was mostly in the green areas (range 28- 40) indicating excellent quality score for this sample. The score decreased at the end of the sequence to about 18-26 score at around base 92-99 position in read-base indicating that the sequencing got worse toward the end of the sequencing. The total sequence is **139,023,385 kilobase**. The sequence length is 100 bp. The mean sequence quality score (Phred Score) is around 36.

Basic Statistics

Measure	Value
Filename	SRR1554537_1.fastq.gz_accepted_hits.bam
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	139023385
Sequences flagged as poor quality	0
Sequence length	100
%GC	48



The quality score was mostly in the green areas (range 28- 40) indicating excellent quality score for this sample, but decreased at the end of the sequencing at position 98-99 bp.



The mean sequence quality score (Phred Score) is around 36.

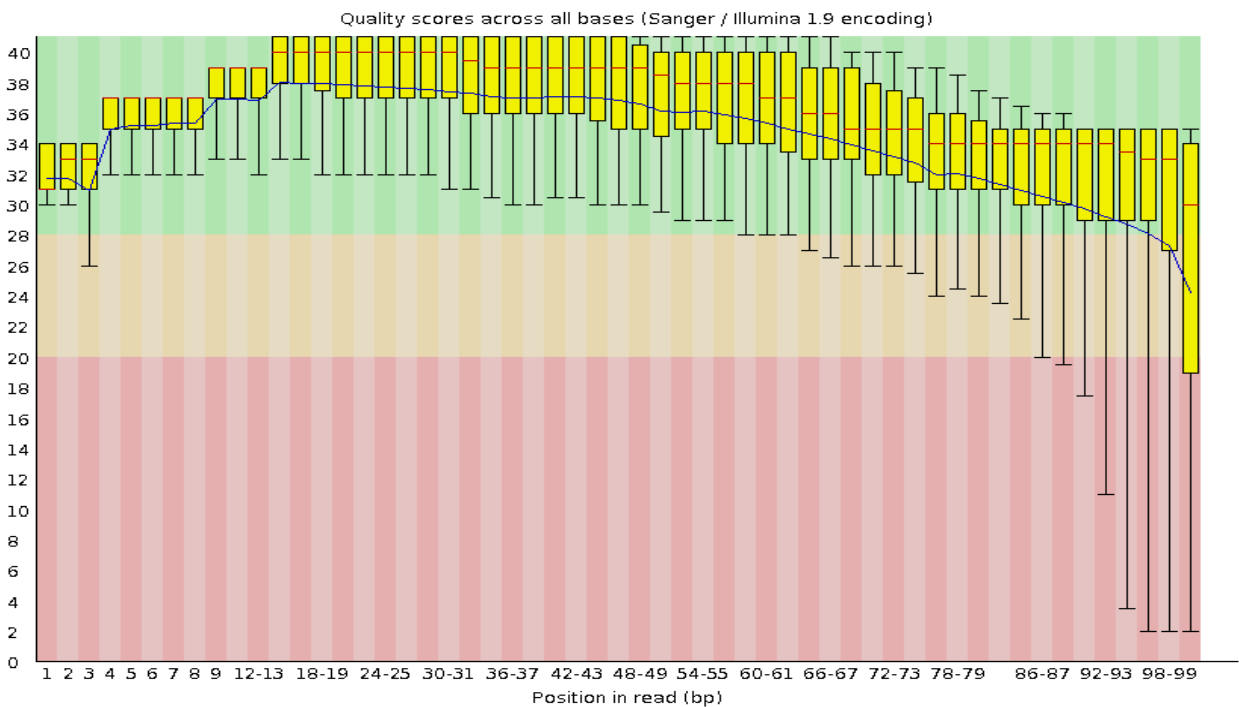
R3462_DLPFC_polyA_RNAseq_total; SRX683796; SRR1554538

The quality score was mostly in the green areas (range 28- 40) indicating excellent quality score for this sample. The score decreased at the end of the sequence to about 18-28 score at around base 98-99 position in read-base indicating that the sequencing got worse toward the end of the sequencing. The total sequence is **168,599,101** kilobase. The sequence length is 100 bp. The mean sequence quality score (Phred Score) is around 36-38.

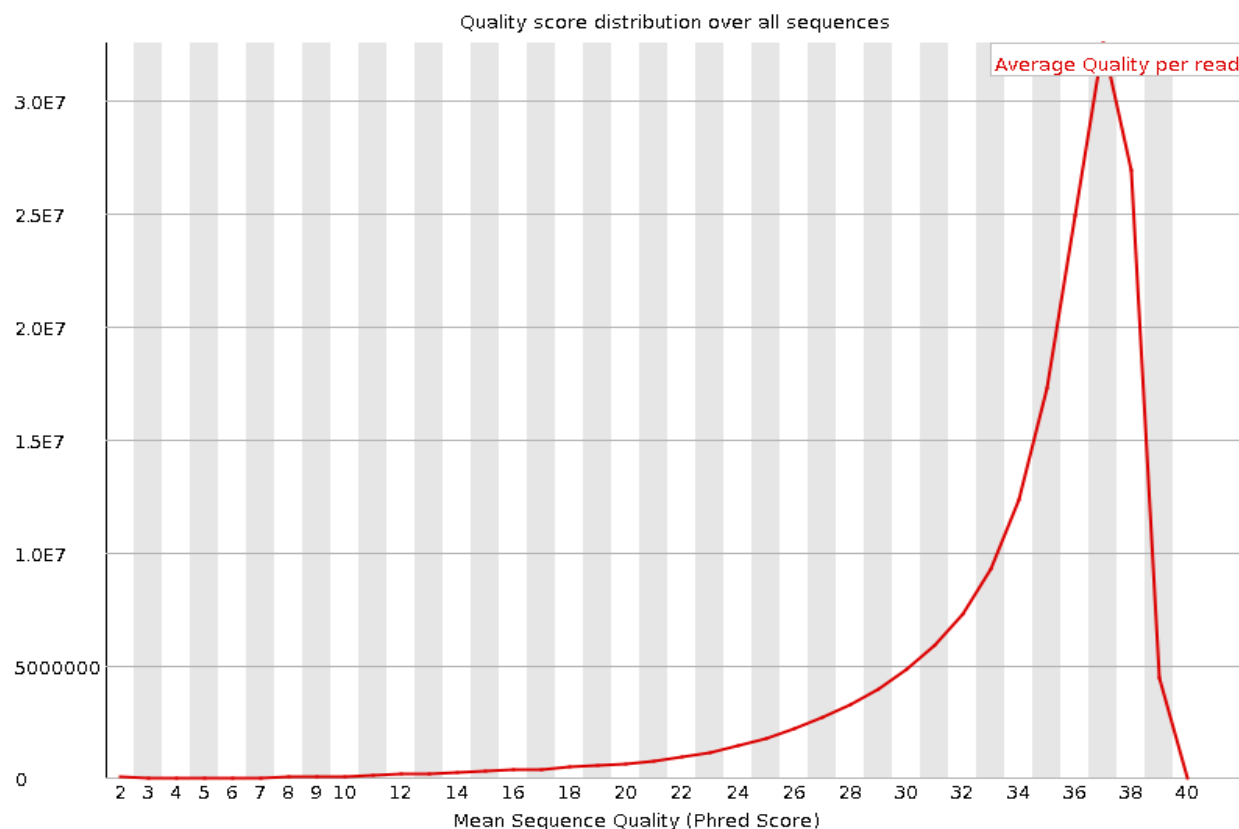
Basic Statistics

Measure	Value
Filename	SRR1554538_1.fastq.gz_accepted_hits.bam
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	168599101

Measure	Value
Sequences flagged as poor quality	0
Sequence length	100
%GC	47



The quality score was mostly in the green areas (range 28- 40) indicating excellent quality score for this sample, but decreased at the end of the sequencing at base 98- 99 with score below 28.



The mean sequence quality score (Phred Score) is around 36-38.

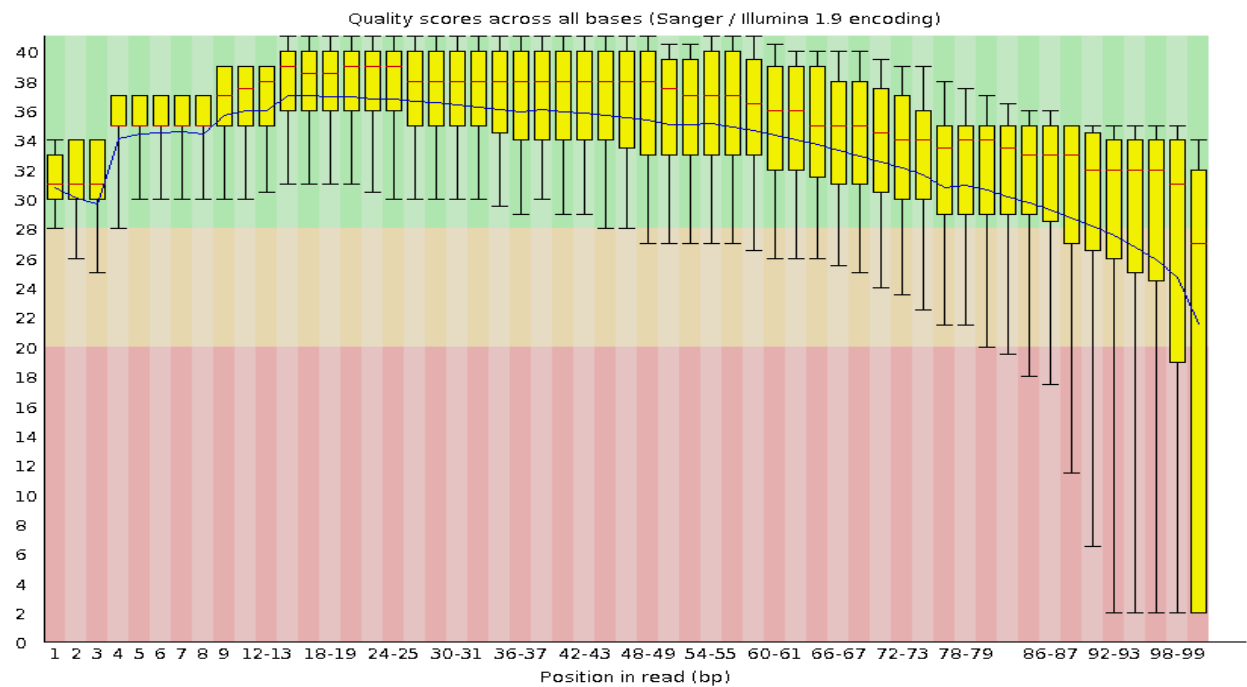
R3485_DLPFC_polyA_RNAseq_total; SRX683799; SRR1554541

The quality score was mostly in the green areas (range 28- 40) indicating excellent quality score for this sample. The score decreased at the end of the sequence to about 28 to 2 score at around base 86-99 position in read-base indicating that the sequencing got worse toward the end of the sequencing. The total sequence is 182,825,541 kilobase. The sequence length is 100 bp. The mean sequence quality score (Phred Score) is around 36.

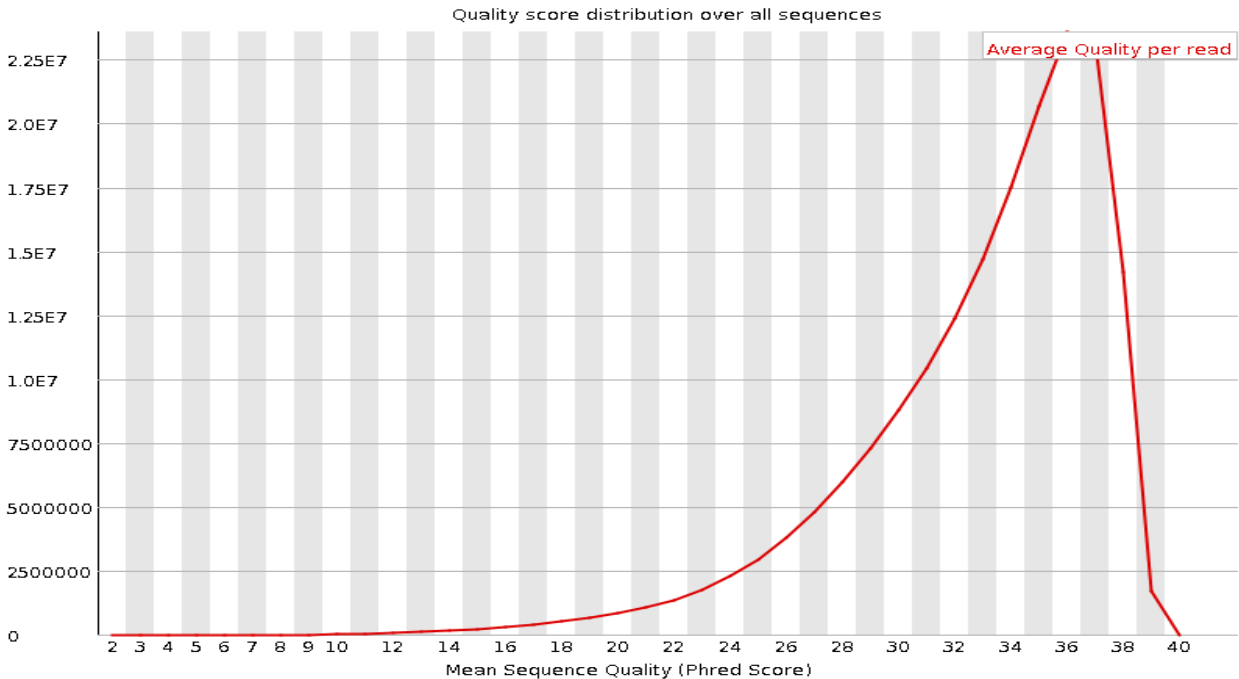
Basic Statistics

Measure	Value
Filename	SRR1554541_1.fastq.gz_accepted_hits.bam
File type	Conventional base calls

Measure	Value
Encoding	Sanger / Illumina 1.9
Total Sequences	182825541
Sequences flagged as poor quality	0
Sequence length	100
%GC	47



The quality scores decreased to around 28 down to 2 at base position 98-99, indicating the sequencing is poor toward the end of the sequencing run.



The mean sequence quality score (Phred Score) is around 36.

Here are FASTQC data for Adult samples

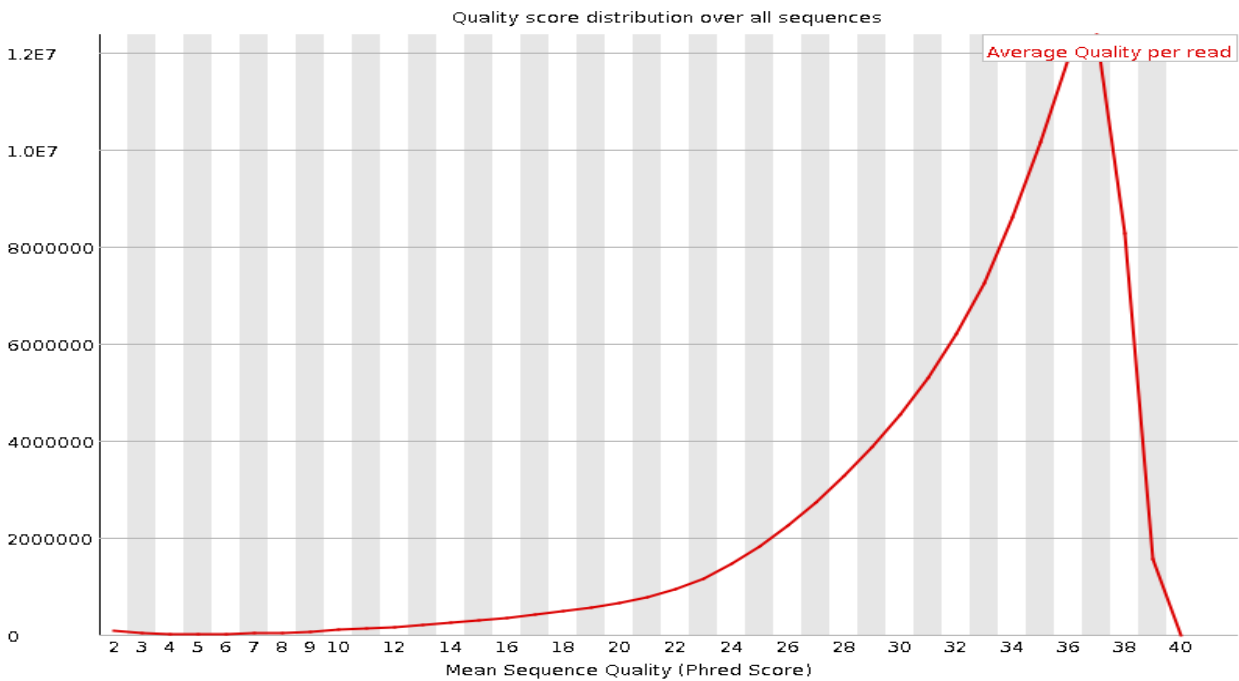
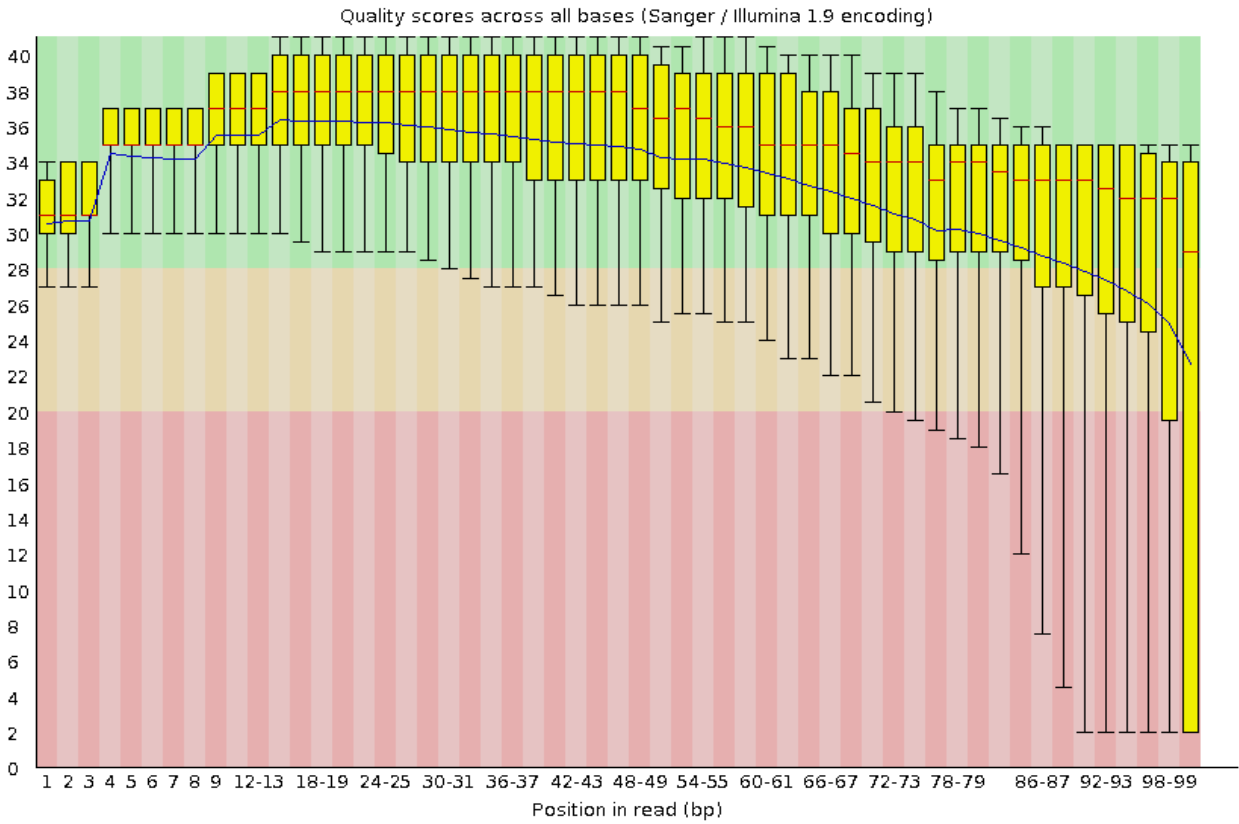
Adult samples:

R2869_DLPFC_polyA_RNAseq_total; SRX683793; SRR1554535

The quality scores are in the range of (28 – 40) indicating all good sequencing bases. In position 87-99 bases, the quality score plummeted to below 28, indicating the quality of sequencing decrease at the end of sequencing. At position 99 bases, the quality score decrease to 2 indicating very poor sequencing at the end of the reads. The ranges are between 2-20 toward the end of the reads. The total sequences for this adult sample is **98,964,659** which is shorter compared to the Fetal samples. The mean sequence quality score (Phred's score) is about 36.

Basic Statistics

Measure	Value
Filename	SRR1554535.fastq.gz_accepted_hits.bam
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	98964659
Sequences flagged as poor quality	0
Sequence length	100
%GC	48



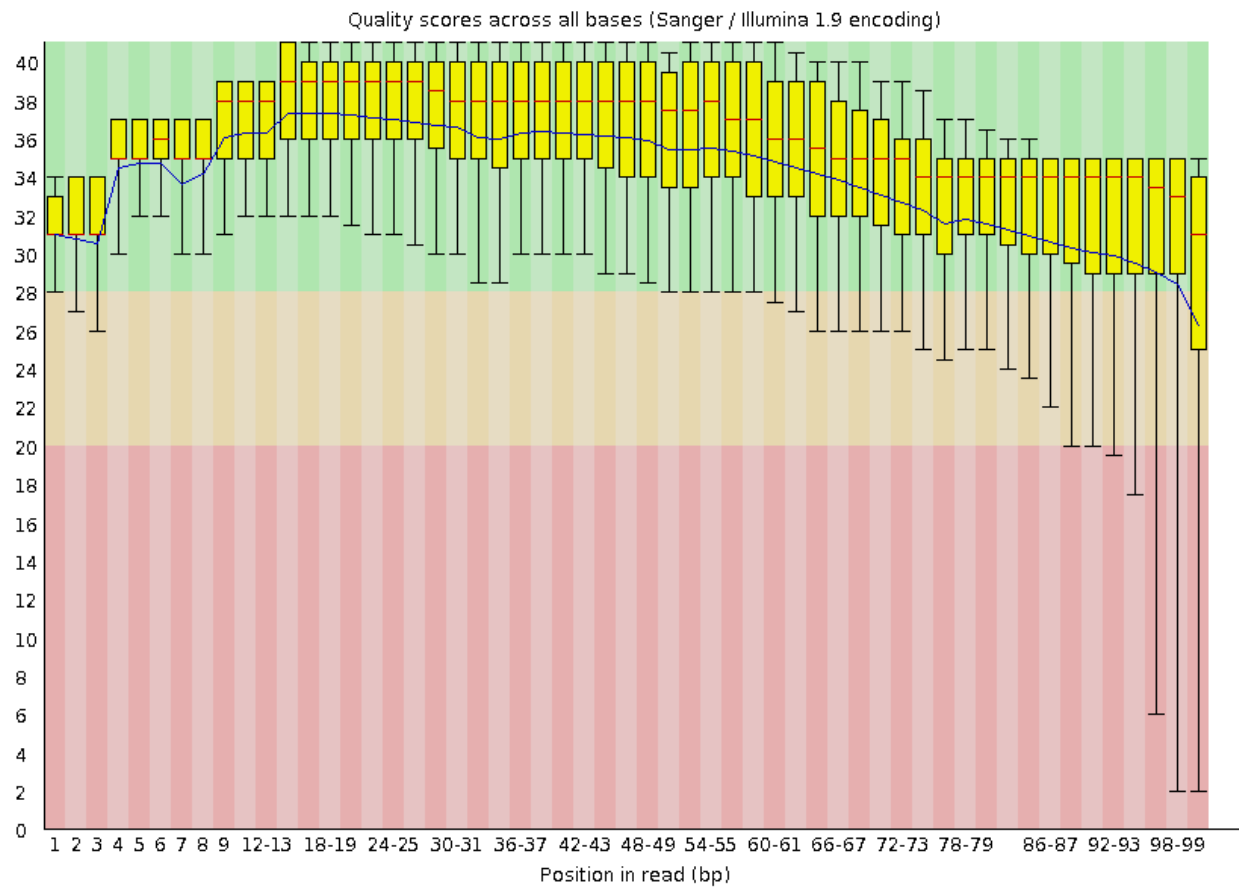
The mean sequence quality (Phred Score) is about 36.

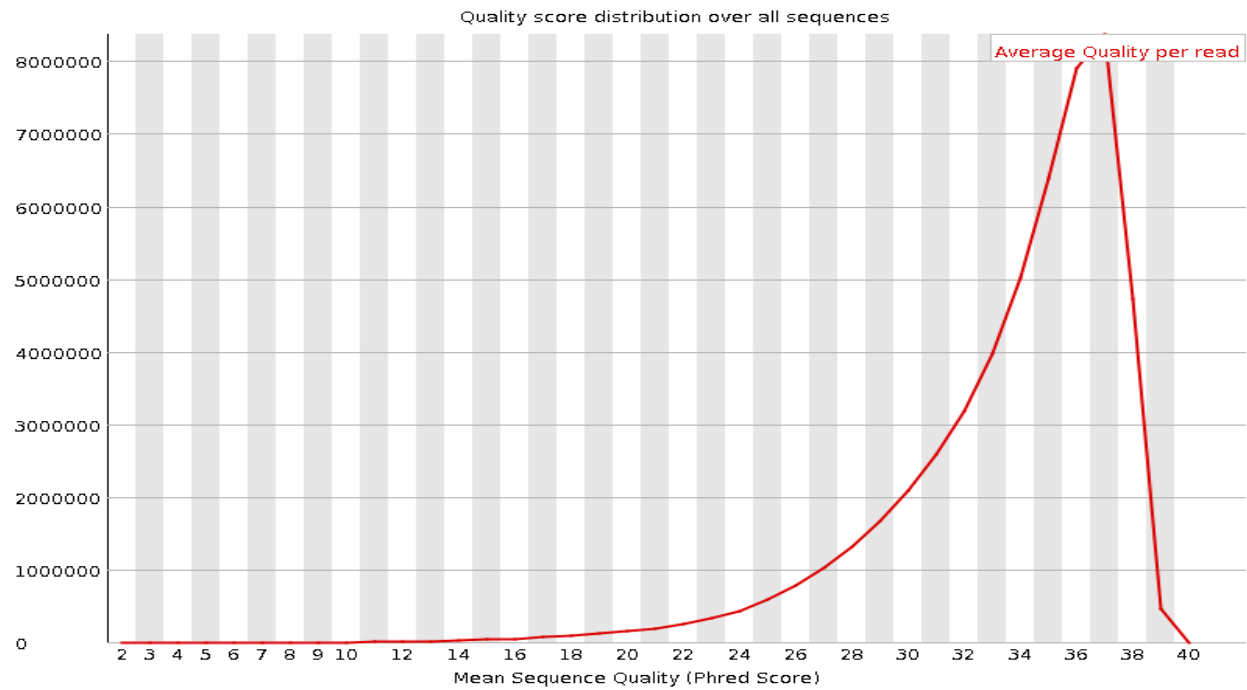
R3098_DLPFC_polyA_RNAseq_total; SRX683794; SRR1554536

The quality scores are in the range of (28 – 40) indicating all good sequencing bases. In position 100 bases, the quality score plummeted to below 28 , indicating the quality of sequencing decrease at the end of sequencing. At position 99 bases, the quality score decrease to under 28-32 indicating poor sequencing at the end of the reads. The total sequences for this adult sample is 52,207,634 which is shorter compared to the Fetal samples. The mean sequence quality score (Phred's score) is about 36.

Basic Statistics

Measure	Value
Filename	Galaxy29-SRR1554536_accepted_hits.bam
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	52207634
Sequences flagged as poor quality	0
Sequence length	100
%GC	46





The mean quality score (Phred Score) is around 36 indicating good quality.

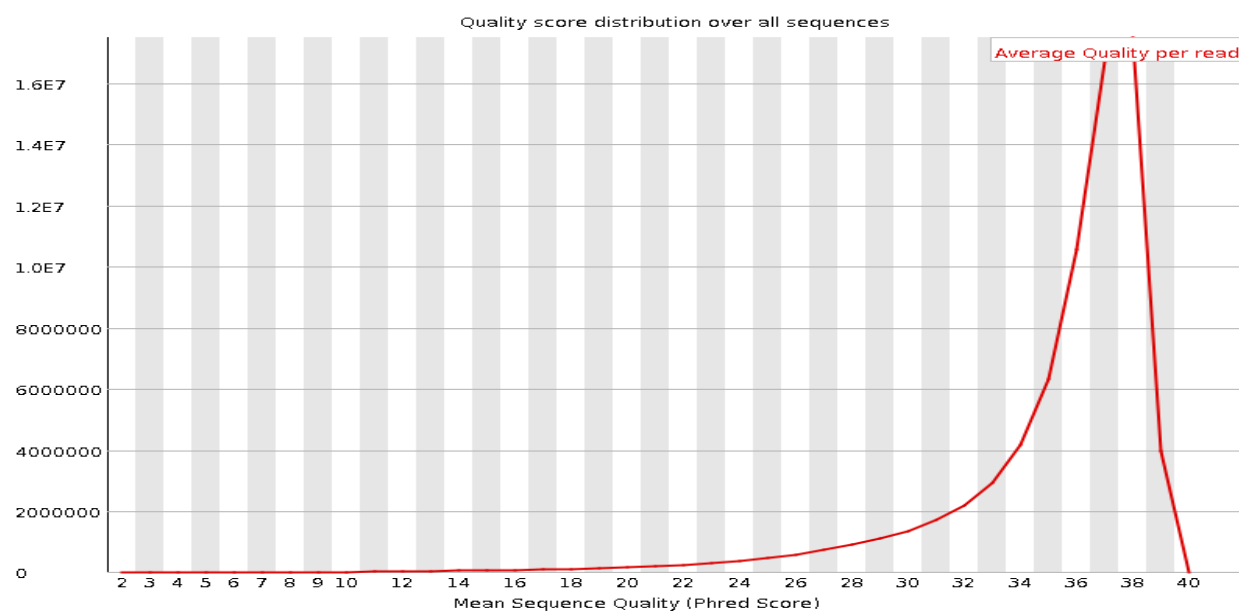
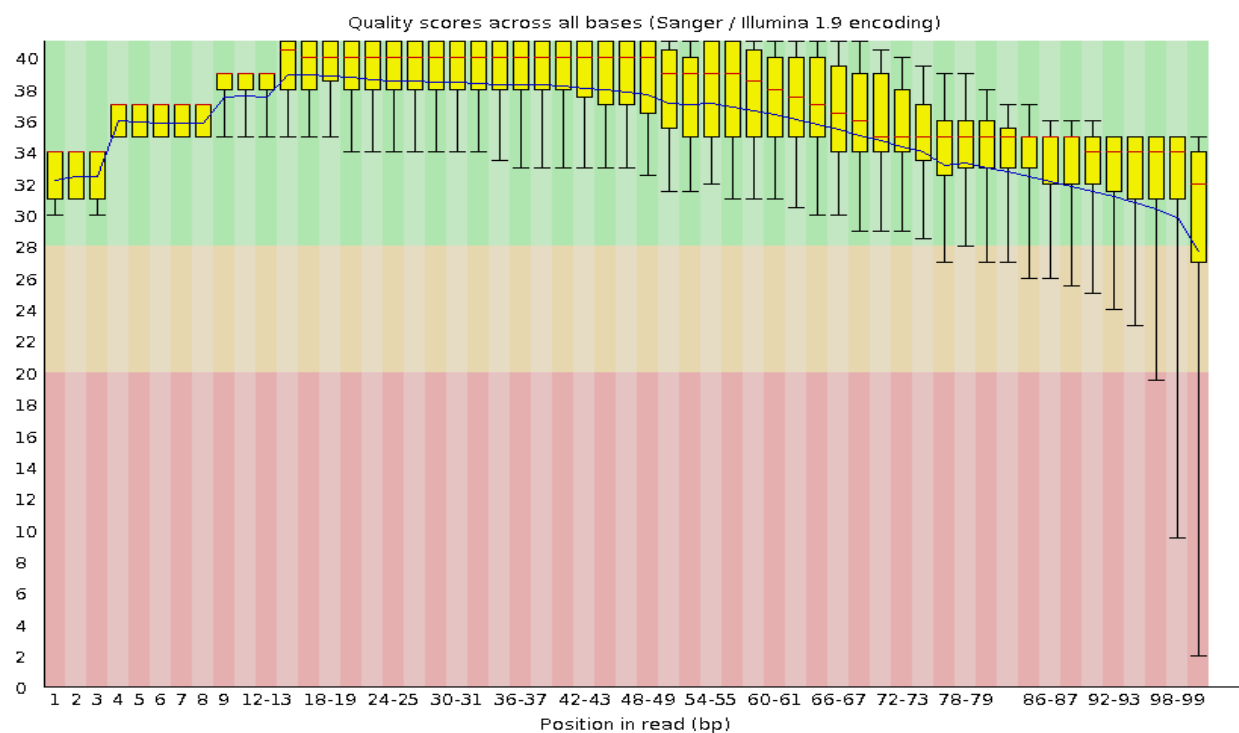
SRR1554539

R3467_DLPFC_polyA_RNAseq_total; SRX683797; SRR1554539

The quality score was mostly in the green areas (range 28- 40) indicating excellent quality score for this sample, but decreased at the end of the sequencing. In the position 99 bases, the quality score plummeted to below 28 , indicating the quality of sequencing decrease at the end of sequencing. The total sequences for this adult sample is 73,705,800 which is shorter compared to the Fetal samples. The mean sequence quality (Phred's score) is about 38.

Basic Statistics

Measure	Value
Filename	SRR1554539_accepted_hits_sorted.bam
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	73705800
Sequences flagged as poor quality	0
Sequence length	100
%GC	48



The mean quality score (Phred Score) is around 38 indicating good quality.

