

Assignment 2: QC The Alignment

For this assignment we take the alignment files (*.sam files) generated from the previous assignment and run FastQC on them.

We used the fastqc command line tool in interactive mode for the assignment. The reports were saved for further analysis.

Percentage of Mapped Reads for each sample

Sample	Type	Reads Mapped	Percent Mapped
SRR1554537	fetal	118101807	99.75%
SRR1554538	fetal	145893064	99.77%
SRR1554541	fetal	147568454	99.75%
SRR1554566	fetal	115275781	99.77%
SRR1554567	fetal	132306889	99.77%
SRR1554568	fetal	103235507	99.76%
SRR1554535	adult	80203609	99.74%
SRR1554536	adult	45114554	99.87%
SRR1554539	adult	70483001	99.73%
SRR1554556	adult	104055748	99.79%
SRR1554561	adult	82851562	99.71%
SRR1554534	adult	60071030	99.73%

fetal_avg_mapping_rate 99.76%

adult_avg_mapping_rate 99.76%

The overall percentage of reads mapped is same for both the fetal and adult samples. Although as we can see above the adult sample SRR1554536 has the highest mapping rate (99.87%) and the adult sample SRR1554561 has the lowest mapping rate (99.71%), it does not make a difference in the overall mapping rate.

Average Quality Score of Mapped Reads

Sample	Type	Avg_qual_score	per_b_seq_quality	Total Sequences	Basic
SRR1554537	fetal	34.28822447	good	1.18E+08	pass
SRR1554538	fetal	34.79538539	good	1.46E+08	pass
SRR1554541	fetal	33.61501326	low	1.48E+08	pass
SRR1554566	fetal	34.00849666	good	1.60E+08	pass
SRR1554567	fetal	34.27971447	good	1.33E+08	pass
SRR1554568	fetal	33.97647128	good	1.03E+08	pass
SRR1554535	adult	33.4352196	low	80414830	pass
SRR1554536	adult	34.22883518	good	45172764	pass
SRR1554539	adult	35.84105791	good	70670921	pass
SRR1554556	adult	34.79055322	good	1.04E+08	pass
SRR1554561	adult	35.00277462	good	83093446	pass
SRR1554534	adult	35.8043798	good	60234971	pass
<hr/>					
Avg_qual_score_fetal		34.16055092			
Avg_qual_score_adult		34.85047005	_____		

The average quality scores for the fetal and adult samples are very similar. As we can see above the average fetal quality score is only slightly lower than that of the adult samples.

There is a similar trend in the average quality scores of all mapped reads (fetal and adult samples). The average quality of the beginning base pairs starts above 30 phred score and then it goes up for a while till around 20 bp of the run and then it starts to go down. Even at the very end of the run the phred score is mostly above ~28 so there is no problem with the quality of the reads.