Genomic data Science Capstone – Quality Check - Week 5

Gulnaz Zaidi

Quality control on alignments:

Quality check on Alignments was done by using the FastQC tool on Galaxy, to determine the sequencing accuracy. The mapping summary from Alignment step was used to determine the mapping rates. The samples were mapped against hg19 using HISAT2 on Galaxy.

The Fetal samples are named F1, F2, F3 and F4. The Adult samples are named A1, A2, A3 and A4.

Inferences:

1. The mapping rates of alignments for the fetal and adult samples are excellent. However, the mapping rates are marginally higher in fetal samples(99.2%) as compared to adult samples(98.68%), as seen in Table1.
2. The average quality scores for all samples are in very good, as seen in Table 2 and Table3. This indicates very high sequencing accuracy in all samples. None of the selected samples show problems during Quality check.
3. Even though, the quality scores of all sample are good, 3 adult samples - SRR1554539, SRR155456 and SRR1554561 show better quality scores than the other samples. On an average, the adult samples show slightly higher Quality scores than the fetal samples.

Results:

	Fetal	% Mapped reads		Adult	% Mapped reads
F1	SRR1554537	99.14	A1	SRR1554535	98.6
F2	SRR1554566	99.27	A2	SRR1554539	98.51
F3	SRR1554567	99.21	A3	SRR1554556	99.12
F4	SRR1554568	99.18	A4	SRR1554561	98.49
	Average	**99.2**		Average	**98.68**

Table1 – Percentage of mapped reads in fetal and adult samples

	F1 - SRR1554537	F2 - SRR1554566	F3 - SRR1554567	F4 - SRR1554568
FETAL				
	A1 - SRR1554535	A2 - SRR1554539	A3 - SRR155456	A4 - SRR1554561
ADULT				
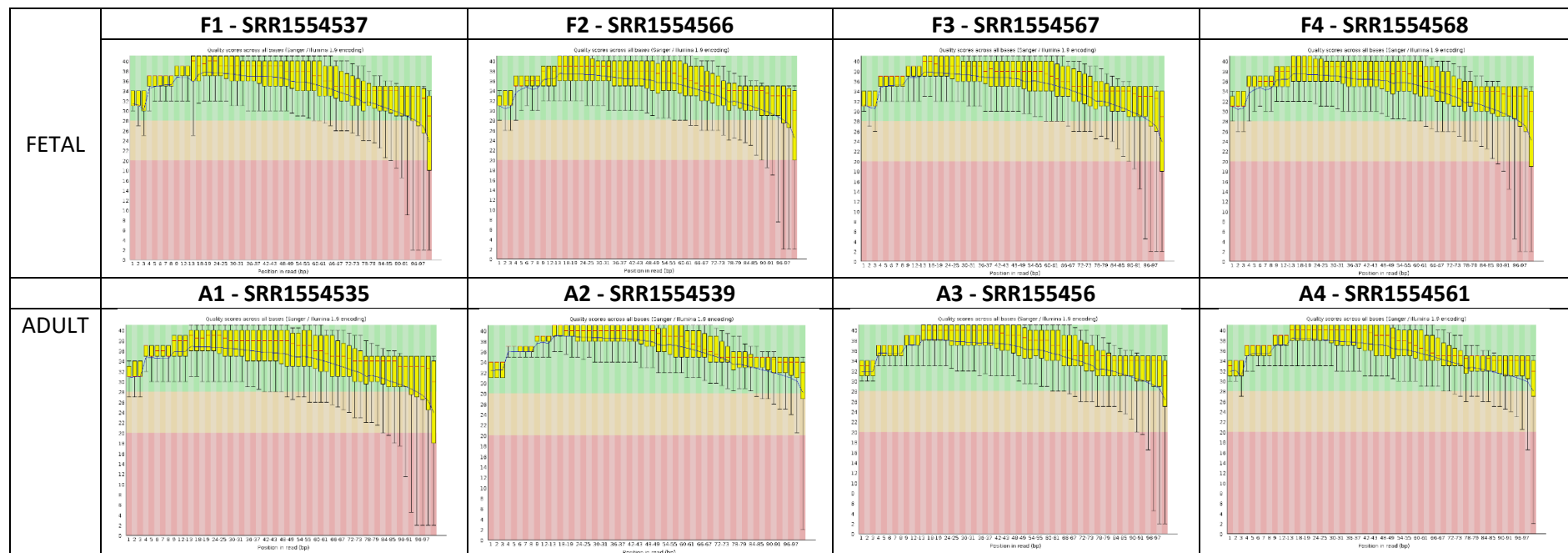
Table 2 – Comparison of 'Per Base Sequence Quality' of fetal and adult samples

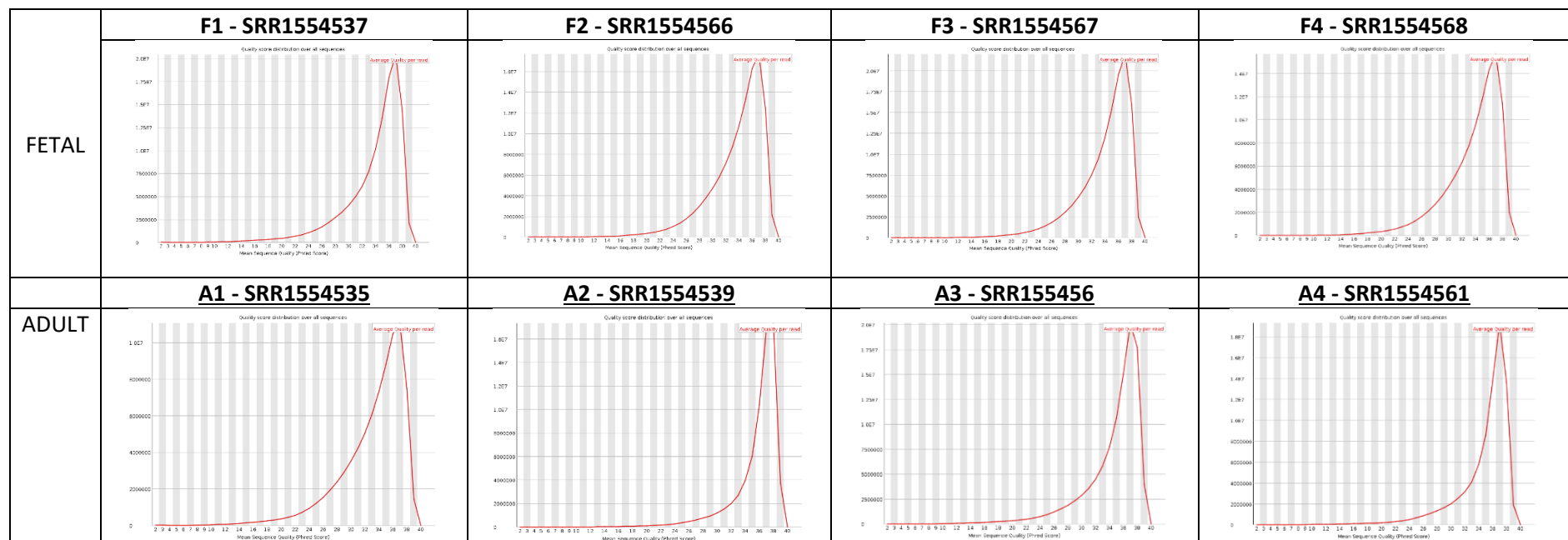|	F1 - SRR1554537	F2 - SRR1554566	F3 - SRR1554567	F4 - SRR1554568
FETAL				
	A1 - SRR1554535	A2 - SRR1554539	A3 - SRR155456	A4 - SRR1554561
ADULT				

Table 3 – Comparison of 'Per sequence quality scores' of Fetal and Adult samples