

Alignment of reads to a reference genome and Quality Control

- Alignment Tool used: HISAT2 (v 2.0)
- Reference Genome used: Grch38 p.12 (hg38)
- Total Number of samples: 3 fetal and 3 adult samples involving two runs per sample
- The first run involves paired and unpaired reads while the second run has 100% pairing

Sample	Run	Total Reads	Alignment Exactly once	Alignment more than once	Alignment zero times	Total Alignment Rate
SRX683794 (Adult)	SRR1554536	4070571 (Unpaired)	3626723 (89.10%)	421962 (10.37%)	21886 (0.54%)	99.46%
		21450348 (Paired)	20218602 (94.26%)	1014202 (4.73%)	217544 (1.01%)	98.99%
	SRR2071347	37025651 (Paired)	29473575 (79.60%)	3179721 (8.59%)	4372355 (11.81%)	88.19%
SRX683797 (Adult)	SRR1554539	9648705 (Unpaired)	9022038 (93.51%)	560019 (5.80%)	66648 (0.69%)	99.31%
		33742728 (Paired)	31805200 (94.26%)	1445879 (4.29%)	491649 (1.46%)	98.54%
	SRR2071350	49834056 (Paired)	35259549 (70.75%)	1928452 (3.87%)	12646055 (25.38%)	74.62%
SRX683792 (Adult)	SRR1554534	7208155 (Unpaired)	6639723 (92.11%)	511524 (7.10%)	56908 (0.79%)	99.21%
		28181772 (Paired)	26212392 (93.01%)	1564206 (5.55%)	405174 (1.44%)	98.56%
	SRR2071345	41133326 (Paired)	28977735 (70.45%)	2828591 (6.88%)	9327000 (22.68%)	77.32%
SRX683795 (Fetal)	SRR1554537	11724434 (Unpaired)	10677707 (91.07%)	963788 (8.22%)	82939 (0.71%)	99.29%
		55133946 (Paired)	51619039 (93.62%)	3133968 (5.68%)	380939 (0.69%)	99.31%
	SRR2071348	125129957 (Paired)	101478878 (81.10%)	7176814 (5.74%)	16474265 (13.17%)	86.83%
SRX683824 (Fetal)	SRR1554566	9962800 (Unpaired)	9043045 (90.77%)	842515 (8.46%)	77240 (0.78%)	99.22%
		53161501 (Paired)	49591395 (93.28%)	3267149 (6.15%)	302957 (0.57%)	99.43%

	SRR2071377	66177040 (Paired)	56347085 (85.15%)	4218713 (6.37%)	5611242 (8.48%)	91.52%
SRX683825 (Fetal)	SRR1554567	12466551 (Unpaired)	11423388 (91.63%)	957580 (7.68%)	85583 (0.69%)	99.31%
		61922935 (Paired)	58193574 (93.98%)	3325963 (5.37%)	403398 (0.65%)	99.35%
	SRR2071378	77670609 (Paired)	66088760 (85.09%)	4253961 (5.48%)	7327888 (9.43%)	90.57%

QUALITY CONTROL OF THE ALIGNMENT

- FASTQC (v 0.11.8) was used to do quality control on the BAM files

Sample	Runs	Per Base sequence quality	GC content (%)
SRX683794 (Adult)	SRR1554536	32.812	46.5
	SRR2071347	32.9409	47
SRX683797 (Adult)	SRR1554539	34.65495324	48.5
	SRR2071350	32.22108614	49
SRX683792 (Adult)	SRR1554534	34.69436006	52
	SRR2071345	32.54729672	51
SRX683795 (Fetal)	SRR1554537	31.47136605	49.5
	SRR2071348	31.01888734	49
SRX683824 (Fetal)	SRR1554566	32.27080346	48.5
	SRR2071377	32.43247485	49
SRX683825 (Fetal)	SRR1554567	32.70195493	47.5
	SRR2071378	32.63745564	47

- Adults:**

Values	Mean	Maximum value	Minimum Value	1 st Quartile	Median	3 rd Quartile
Per base sequence quality	33.31176603	34.69436006	32.22108614	32.46574408	32.87645	34.66480495
GC content	49	52	46.5	46.875	48.75	51.25

- Fetus:**

Values	Mean	Maximum value	Minimum Value	1 st Quartile	Median	3 rd Quartile
Per base sequence quality	32.08882371	32.70195493	31.01888734	31.35824637	32.35163915	32.65358046
GC content	48.41666667	49.5	47	47.375	48.75	49.125

- From the summary statistics above there is a similar per base sequence quality and alignment rates across both groups. Adult sample seem to have higher GC content although the difference is minor. There are no strange QC metrics for the given samples.