

Genomic data Science Capstone – Week 4 Submission

Gulnaz Zaidi

Downloading the data

FastQ files for 4 sets of Fetal and adult samples from the SRR155* series were downloaded from SRA page of NCBI, using the “**Faster Download and Extract Reads in FASTQ**” format from NCBI SRA” tool on Galaxymain. To distinguish the fetal and adult samples, the fetal samples are labeled as F1, F2, F3, F4. The adult samples are labeled as A1, A2, A3, A4. The downloaded samples are –

Fetal		Adult	
F1	SRR1554537	A1	SRR1554535
F2	SRR1554566	A2	SRR1554539
F3	SRR1554567	A3	SRR1554556
F4	SRR1554568	A4	SRR1554561

Alignment

As the samples are RNA sequencing data, the FastQ files were aligned using HISAT2 tool on Galaxymain. HISAT is a fast and sensitive splice-aware alignment program, which matches the reads exclusively with the exons (gene-coding regions) and ignores the intronic (non-coding) regions. HISAT2 creates output files in BAM format. Hg19 was used as the reference genome.

The Alignment summaries for each sample are –

F1 - SRR1554537

55133946 reads; of these:

55133946 (100.00%) were paired; of these:

476666 (0.86%) aligned concordantly 0 times

52688941 (95.57%) aligned concordantly exactly 1 time

1968339 (3.57%) aligned concordantly >1 times

F2 - SRR1554566

53161501 reads; of these:

53161501 (100.00%) were paired; of these:

388090 (0.73%) aligned concordantly 0 times

50650544 (95.28%) aligned concordantly exactly 1 time

2122867 (3.99%) aligned concordantly >1 time

F3 - SRR1554567

61922935 reads; of these:

61922935 (100.00%) were paired; of these:

488761 (0.79%) aligned concordantly 0 times

59351545 (95.85%) aligned concordantly exactly 1 time

2082629 (3.36%) aligned concordantly >1 times

F4 - SRR1554568

48184702 reads; of these:

48184702 (100.00%) were paired; of these:

396509 (0.82%) aligned concordantly 0 times

46133743 (95.74%) aligned concordantly exactly 1 time

1654450 (3.43%) aligned concordantly >1 times

A1 - SRR1554535

38063721 reads; of these:

38063721 (100.00%) were paired; of these:

533477 (1.40%) aligned concordantly 0 times

36345262 (95.49%) aligned concordantly exactly 1 time

1184982 (3.11%) aligned concordantly >1 times

A2 - SRR1554539

33742728 reads; of these:

33742728 (100.00%) were paired; of these:

503288 (1.49%) aligned concordantly 0 times

32420443 (96.08%) aligned concordantly exactly 1 time

818997 (2.43%) aligned concordantly >1 times

A3 - SRR1554556

49480779 reads; of these:

49480779 (100.00%) were paired; of these:

433552 (0.88%) aligned concordantly 0 times

47652724 (96.31%) aligned concordantly exactly 1 time

1394503 (2.82%) aligned concordantly >1 times

A4 - SRR1554561

39272751 reads; of these:

39272751 (100.00%) were paired; of these:

593770 (1.51%) aligned concordantly 0 times

37506556 (95.50%) aligned concordantly exactly 1 time

1172425 (2.99%) aligned concordantly >1 times