

Differences in gene expressions between adult and fetus brains

Mar Batlle

TASK 2

Getting the raw data

To start the Genomic Capstone Project, 6 datasets were downloaded from the European Nucleotide Archive (<https://www.ebi.ac.uk/ena/browser/home>) in fastq.gz format. Half of them corresponding to fetal samples: SRR1554537, SRR1554538, SRR1554541 and the other 3, were adult samples; SRR1554536, SRR1554535, SRR1554539.

Alignment to the reference genome

The galaxy server was used to align the sample to the reference genes, each of the 6 pair-ended samples was aligned to the built-in genome Human (Homo Sapiens) (b37) hg19 reference genome using the HISAT2 (Galaxy Version 2.1.0+galaxy5). As a result, each alignment generated one BAM file containing the results.

Alignment reads

After completing the alignments, it was determined that each sample had the following number of paired reads.

| Sample | Type | Number of Reads Paired |
|------------|-------|------------------------|
| SRR1554537 | Fetal | 55133946 |
| SRR1554538 | Fetal | 68026190 |
| SRR1554541 | Fetal | 69278357 |
| SRR1554536 | Adult | 21450348 |
| SRR1554535 | Adult | 38063721 |
| SRR1554539 | Adult | 33742728 |

Galaxy Workflow

In the following page you can see the workflow following using the galaxy site.

