

## ***Genomic Data Science Capstone: Week 4 (Alignment)***

*For this assignment I performed an analysis of RNA-seq experiments obtained by Jaffe et al (2015). To complete this project, I compared fetal and adult brains by accessing raw data and aligning specific set of samples.*

*The alignment strategy I used:*

### **Step 1: To obtain SRA files:**

- a) Log in within the website [www.usegalaxy.org](http://www.usegalaxy.org)*
- b) Go to “Get Data”*
- c) Select “EBI SRA”*
- d) Search FASTQ files*
- e) Download the following samples:*

<i>Adult brain</i>	<i>Fetal brain</i>
SRR1554539	SRR1554567
SRR1554534	SRR1554541
SRR1554536	SRR1554537

### **Step 2: To run HISAT2 for alignment:**

- a) Select “hg19” (to use built-in reference genome)*
- b) Select “paired end”*
- c) In summary options select:*

*Output alignment summary in a more machine-friendly style > yes*

*Print alignment summary to a file > yes*

### **Step 3: To calculate aligned reads:**

- *Total pairs = paired × sets*
- *Total alignment reads = (aligned reads 1 × time) + (aligned concordant > 1 × time) + (unpaired aligned concordant 1 × time)*

**Alignment result**

<i>AGE</i>	<i>ID</i>	<i>UNPAIRED</i>	<i>ALIGNED 0 TIMES</i>	<i>ALIGNED 1 TIME</i>	<i>ALIGNED &gt;1 TIME</i>
<b><i>ADULT</i></b>	SRR1554539	9 33742728 (100.00%)	502791 (1.49%)	32422514 (96.09%)	817423 (2.42%)
	SRR1554534	28181772 (100.00%)	415707 (1.48%)	26719314 (94.81%)	1046751 (3.71%)
	SRR1554536	21450348 (100.00%)	141926 (0.66%)	20543060 (95.77%)	765362 (3.57%)
<b><i>FETAL</i></b>	SRR1554541	69278357 (100.00%)	561499 (0.81%)	66406174 (95.85%)	2310684 (3.34%)
	SRR1554537	55133946 (100.00%)	474170 (0.86%)	52684816 (95.56%)	1974960 (3.58%)
	SRR1554567	61922935 (100.00%)	488761 (0.79%)	59351545 (95.85%)	2082629 (3.36%)