# Peer-graded Assignment: Alignment

[**Instructions**](https://www.coursera.org/learn/genomic-data-science-project/peer/daGIM/alignment)

Once you have understood the problem, the next step is to obtain the raw data so that you can perform your analysis. To complete this assessment you need to (a) find the raw data, (b) download and align the data for a specific set of samples, and (c) compile information on those samples that you will use in downstream analyses.

You will be comparing fetal brains to adult brains for this project. Accessing the data is described in detail in the project description.

You should download the data and align the data with a spliced aligner (some options include Tophat2, Hisat, STAR, but there are others) You may do this with command line tools or with Galaxy. If you do not have the computing resources to do this on your computer use the Galaxy Cloud servers to perform the alignment. In the paper they aligned the reads to version hg19 of the human genome, you may choose to align to any version of the human genome you think is appropriate.

The second component of this assessment is to collect information about the samples we will be analyzing. Go to the SRA and collect as much information as possible about the each sample. At minimum collect information on the age of the person who the brain came from and the RNA-quality (RIN) of the sample. Store the data in a tab-delimited text file where each column corresponds to a variable and each row corresponds to one samples values for those variables. This is the phenotype table we will use later on.

## **Review criteria**

**less**

Upload a pdf document describing your alignment strategy and how many reads aligned for each sample.

* Does it appear that an appropriate spliced aligner was used?
* Does the number of aligned reads seem appropriate given what you know about the data?

Upload a tab delimited text file containing the phenotype table as described above.

* Is the file tab-delimited?
* Are all of the required variables (age, RIN) in the data set?
* Are there measurements for each sample?

**PROMPT 1**

Upload a pdf document describing your alignment strategy and how many reads aligned for each sample.

Alignment Strategy

[Alignment Strategy](https://coursera-assessments.s3.amazonaws.com/assessments/1629567211687/6e9c2564-4de8-4707-8b61-237e2be2f46d/Alignment%20Strategy.pdf)

This file describes the alignment strategy and the number of reads that aligned for each sample

**PROMPT2**

Upload a tab delimited text file containing the phenotype table as described above.

Phenotype table

[Phenotype table](https://coursera-assessments.s3.amazonaws.com/assessments/1629557366092/f140c74c-0ed3-4db7-95f6-a9ef86a769e6/PhenotypeData.txt)

This is a tab delimited file with the phenotype table with required variables (age, RIN) and the measurements for each sample.