Week02_Thursday

Perform the following steps to embark on the RNA-seq odyssey:

Step 0: Practice using sra-tools to convert an SRA file into a set of fastq files.

1. Download SRR13075764.1 from Canvas.
2. Use FileZilla or scp to move it to your supercomputer account.
3. Using **less** to view it results in a warning that it might be a binary file.
4. The tool to decode this file is in the **ncbi-sra** package.
5. This is a paired-end sequencing file, so use **fastq-dump** to split it into two files, one containing the forward reads and the other containing the reverse reads.
6. The command is **fastq-dump**, with the arguments **--split-e** and **--split-spot**.
7. Use less to examine the resulting files. This is fastq format.

Step 1: Assemble into groups with 5-7 people per group.

Step 2: Think of an idea for an RNA-seq project. Remember, RNA-seq can be used to compare genome-wide levels of gene expression across two or more tissue samples. Consider the following points as you think up your project:

1. Try to limit your comparison to two tissue types, so we can analyze it easily.
2. The organism involved should have a sequenced genome. See this database for a list of publicly available complete genome sequences: https://www.ncbi.nlm.nih.gov/genome/browse#!/overview/.
3. Choose a topic for which the short-read data are already available. Search the NCBI SRA to see if there are sequences available for your species. The data need to fulfill the following criteria:
    a. The dataset should be paired-end reads.
    b. The sequencing platform should be Illumina.
    c. The data should be replicated, with at least 3 samples per condition.

Step 3: Have each person in your group download one short-read file. Each person will process their single file, and then we will combine them into a single analysis later. Each person should download a different file.

Step 4: Convert the file to fastq format. Make sure the file is split into two files, one for the forward read and one for the reverse read. This step might need to be homework.

Step 5: Each person write up a small description of the planned project (work separately on this), including species name and the accession numbers for the data. Submit this as your answer to the Canvas assignment, which is due Monday.