Homework #1: Module 1 (RNA-seq)

Create a shell script to submit a job which uses featureCounts:

For your homework, you will be tasked with using featureCounts (part of the Subread package) to generate a counts matrix to be used for DESeq2 analysis. For this assignment two things will be very useful:

- Reading through the featureCounts documentation in the read summarization step.
- Reading through technical details of the experiment we analyzing (hint: details can be found on the same link we used to find the fastq files)

You will need to write your own shell script and submit it to the hotel queue. This is not designed be a copy-and-paste job, so be sure to read the manual and think carefully about your command (4 pts).

General questions:

1. Describe the main steps of a full differential expression RNA-seq analysis from the type of files you would need, up to and including performing differential gene analysis. For each of the main steps, give an example of the tool you would use (if relevant) and most importantly the purpose of what is done computationally. (6 pts)

2. What are the similarities and differences between RPKM and TPM (describe the differences between how the methods work)? Which one might be considered preferable and why? (2 pts)

3. What kind of normalization should you perform to your count matrix if you choose to use DESeq2 for your differential gene expression analysis? (1 pt)

13 points total