Assign 5 RNA-Seq analysis using R DESeq2

Due Monday by 11:59pm **Points** 25 **Submitting** a file upload

Here is the R script we went over in class: **estrogen_MCF7_2.5M.R**. In this form it reads and analyses data sequenced at read depth 2.5 M. Modify the above file so it runs the data at sequencing depths 2.5M. 10M and 30M instead of only 2.5M.

1. read in counts from this **tab-delimited file**
(https://usfca.instructure.com/courses/1564374/files/63146010/download?wrap=1)
(https://usfca.instructure.com/courses/1564374/files/63146010/download?wrap=1)
(https://usfca.instructure.com/courses/1564374/files/63146010/download?wrap=1)
(https://usfca.instructure.com/courses/1564374/files/62360823/download?wrap=1)
(https://usfca.instructure.com/courses/1564374/files/62360823/download?wrap=1).

2. process the count data so as to create three DESeqDataSets: one for each of the three read depths.

3. create plots to visualize the results for all three and write one paragraph commenting on your results and comparing the results of the 3 read depths.

Upload your modified script (renamed to an appropriate name), your paragraph discussion, and your plots.

Here is the **tutorial** **(http://dwheelerau.com/2014/02/17/how-to-use-deseq2-to-analyse-rnaseq-data/)** we went over in class

Here is a **tutorial** **(http://www.sthda.com/english/wiki/rna-seq-differential-expression-work-flow-using-deseq2#other-comparisons)** that contains more statistical background

Here is the **DESeq2 vignette**
(https://usfca.instructure.com/courses/1564374/files/63146011/download?wrap=1)
(https://usfca.instructure.com/courses/1564374/files/63146011/download?wrap=1)
(https://usfca.instructure.com/courses/1564374/files/63146011/download?wrap=1) with relevant parts highlighted

Here is some information about the data:

#The data is for 3 DESeq2 experiments, each comparing the transcriptome of untreated MCF-7 tumor cells to those treated with estrogen.

There are 7 replicates for each of the conditions (untreated, treated) for each experiment (read depth = 2.5M, 10M, 30M)

These substrings of column names indicate the 7 replicates for control sample:

ControlCistrackID <-
c("2012.562","2012.563","2012.564","2012.565","2012.566","2012.568","2012.569")

# These substrings of column names indicate the 7 replicates for sample treated with estrogen:
E2CistrackID <- c("2012.570","2012.571","2012.572","2012.574","2012.575","2012.576","2012.577"
)

# These substrings of column names indicate the 3 different depth of reads for each of the replicates of both untreated and treated:
readDepth <- c("2.5M","10M","30M")