Exercises 10

Ewa Szczurek
szczurek@mimuw.edu.pl

Exercise 1
RNA-seq data analysis on the Galaxy server (http://centromere:8080/root):

1. Upload to your history
    a. SharedData/Data Libraries/GTF 4 Cufflink (the reference Annotation GTF files for Cufflink)
    b. SharedData/Data Libraries/workshop_sample_data/RNA-seq

2. Map with Tophat
3. Find splice junctions using Cufflinks
4. Merge the transcripts using Cuffmerge
5. Perform DE analysis using Cuffdiff

    Analyze the input/output files and the program parameters.

Homework
Analyze the input/ output files in this analysis in your Galaxy session.

1. Cufflinks
    a. -If you have not yet done so, run Cufflinks with the options:
        i. **Use Reference Annotation:** Use Reference Annotation, setting **Reference Annotation:** dm3_genes.gtf
            - How many transcripts have been found for cond1 and how many for cond2 ?
        ii. **Use Reference Annotation:** No
            - How many transcripts have been found for cond1 and how many for cond2 ?

    -What is the difference between Cufflinks results when **Use Reference Annotation** option is on and when it is off (between i. and ii.)?

2. Inspect the results of Cuffmerge. What is the number of significantly differentially expressed genes?