INF-BIO9121/5121
Home/Oral exam 2017
RNA-seq

Files necessary for this exam can be found at */data/exam/RNAseq_home_exam*. Copy the folder to your VM before analysing it.

Scripts used during the course can be found in http://inf-biox121.readthedocs.io/en/2017/Lectures/. Modify them accordingly for the home exam.

Question 1: Quality control and trimming

*Perform this step on *sample1_R1/2.fastq.gz* and *sample2_R1/2.fastq.gz* files found in *fastq_files* folder.

Perform FastQC on the raw (untrimmed) reads.

- Present and explain key plots from the output.

Perform adapter and quality trimming with Trimmomatic.

You will find trimmomatic tool at */share/inf-biox121/trimmomatic/Trimmomatic-0.36/trimmomatic-0.36.jar*. Make sure you use *java -jar* to invoke this tool. Appropriate adapter file can be found at */share/inf-biox121/trimmomatic/Trimmomatic-0.36/adapters/TruSeq3-PE-2.fa.*

- Present the trimming statistics.

Perform FastQC on the trimmed reads.

- Present and discuss the changes in quality post trim.

Question 2: Mapping using tophat2

*Perform this step on sample1 and sample2 from above after quality trimming.

Map the reads using Tophat2. Use the bowtie2 index in folder *reference/gadMor2_ena*. There is no transcriptome-index for this data. Use the above index as the genome reference.

- How many reads were initially discarded by tophat2? Why were they discarded?
- For each sample, present the percentage of reads that were mapped (left, right, and overall) and the concordant pair alignment rate. Explain the difference between these statistics.

Question 3: Differential expression calculation

You are not required to perform these steps due to computational demands.

- Briefly, explain what each of Cufflinks, Cuffmerge, and Cuffdiff do, including the type(s) of information that are used as input and what is the output.

Question 4: Differential expression assessment and visualization

*The remaining steps are performed using the Cuffdiff output folder that was generated for you *cuffdiff_out*. It contains the full set of 6 samples, sample1 and sample2 with three replicates each.

Visually assess your data using CummeRbund in R/Rstudio.

- Present an overall assessment of your samples with a dispersion plot, a PCA or MDS plot, a CV^2 plot, a boxplot, a dendrogram, and a heatmap. You will be expected to explain these plots during the examination. Remember to use "replicates=TRUE" parameter is some of the above plots.
- Perform the pairwise comparison; present a volcano plot, the numbers of differentially expressed genes (FDR<0.05), and the names of the most 6 differentially expressed genes.

Step 7: Functional annotation

For each of the top differentially expressed genes identified in the previous step, identify the name and function of the genes from www.ensembl.org.

- Present and discuss the most 6 differentially expressed genes, whether they were up- or down-regulated, and by how much they were differentially expressed.
- Find the orthologues for these genes in Human or an organism of your choice.

Please make sure you have your commands available during the oral exam.

THE BEST OF LUCK!

If you run into technical issues or have difficulties interpreting the assignment you can contact Arvind Sundaram (arvind.sundaram@medisin.uio.no). (Remember that this is an exam - some interpretations and assumptions we expect you to do based on what we have covered during the course).