

Final Exercise

by Ester Feldmesser, March 2016

The data for this exercise was downloaded from GEO (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE77105>). This is the summary of the experiment as it appears on the website: “MicroRNAs (miRNAs) are important regulators of cell fate decisions in immune responses. They act by coordinate repression of multiple target genes, a property that we exploited to uncover regulatory networks that govern T helper-2 (Th2) cells. A functional screen of individual miRNAs in primary T cells uncovered multiple miRNAs that inhibited Th2 cell differentiation. Among these were miR-24 and miR-27, which each functioned independently to limit interleukin-4 (IL-4) production”.

In this experiment, gene expression analysis of miRNA-deficient mouse CD4+ T cells transfected with miRNA mimics (control) was compared to gene expression of cells transfected with miR-23, miR-24 or miR-27.

The experiment was performed in mice cultured cells. Duplicates from each condition were chosen for the exercise. Quality control (FastQC) was performed previously and the quality was satisfactory. Fastq files were uploaded to Chipster and mapped to the mm10 genome using the star software. In order to save time, the mapping was already performed and saved in your home directory in a session named star_dup.zip. The session is located under the folder data/chipster_sessions.

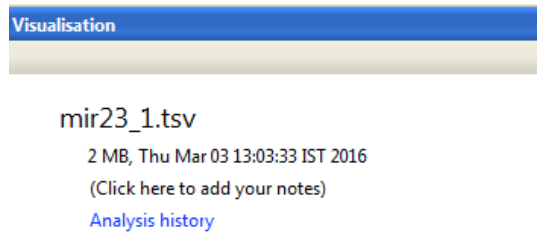
1. Open Chipster as you did in previous exercises and open the local session named star_dup.zip in data/chipster_sessions.
2. Click on Log_final.txt of the different samples to see what percentage of reads were mapped. Values above 80% are considered good.
3. Apply HTSeq to each of the bam files. Select the parameters as shown in the figure below. Mm10 is the same as Mus_musculuc.GRCm38.81.

| Analysis tools - RNA-seq - Count aligned reads per genes with HTSeq | |
|---|---------------------|
| Reference organism | Mus_musculus.G... |
| Chromosome names in the BAM file look like | chr1 |
| Does the BAM file contain paired-end data | no |
| Was the data produced with a strand-specific protocol | no |
| Mode to handle reads overlapping more than one feature | intersection-strict |
| Minimum alignment quality | 10 |
| Feature type to count | exon |
| Feature ID to use | gene_id |
| Add chromosomal coordinates to the count table | yes |

- Now until the run is finished we will move to the Variant detection lecture.

Continuation

- HTSeq creates two output files for each bam file: the tsv file that contains the counts and the txt file that contains a summary of the reads that do not align to exons.
- Rename the tsv files so it will be easier to keep track after the different samples. Right click on the tsv file and click on rename, now you can write a new name in the visualization pane.



According to the name of the original fastq file, rename the tsv as follows:

| Original fastq | New tsv name |
|------------------|--------------|
| SRR3112239.fastq | miR23_1.tsv |
| SRR3112240.fastq | miR23_2.tsv |
| SRR3112243.fastq | miR24_1.tsv |
| SRR3112244.fastq | miR24_2.tsv |
| SRR3112245.fastq | miR27_1.tsv |
| SRR3112246.fastq | miR27_2.tsv |
| SRR3112251.fastq | mimic_1.tsv |
| SRR3112252.fastq | mimic_2.tsv |

- Run DESeq2 to obtain pairwise comparisons between all the four experimental conditions (miR23, miR24, miR27 and mimic). A short reminder: Select all the tsv files and click the NGS tab, choose Utilities and select Define NGS experiment. The parameters should be as shown in the figure below. Insert in the phenodata file, in the group column a name for each condition. Run DESeq2.

Analysis tools - Utilities - Define NGS experiment

Type of experiment: RNA-seq

Does your data contain genomic coordinates: yes

Impute missing data: 0

Count column: count

8. Look at the de-list-deseq2.tsv file.

Q: How many genes are differentially expressed? Do you have all pairwise comparisons?

9. The table contains the raw counts for each gene. To cluster the differentially expressed genes, you will need to normalize and annotate them.

10. To normalize the counts, select the de-list-deseq2.tsv file and under the NGS tab choose Utilities and select the “Normalize count table”. Verify in the parameters window that the parameters are TMM for Calculate normalization factors and yes for Log transform. A new tsv file named ngs-data-table.tsv is created (do not confound it with a previous file of the same name).

11. Annotate the genes in the ngs-data-table.tsv file. Remember that we are working with mice.

12. Export the annotated.tsv file to your home directory.

13. Download it to your PC in the same way you did in the ChIP-Seq Exercise. In order to view the saved file look at the sftp pane in the MobaXterm, you might need to check in the “Follow terminal folder”. You can drag the files to your desktop.

14. Next step will be clustering the differentially expressed genes. Edit the downloaded file in Excel and save the result as txt to work with it in Expander.
Open the file and shift the tittles one column to the right. Take the symbol, description and the normalized reads (log2 transformed) columns. Delete unnecessary columns. The names of the relevant columns (normalized reads) are chip.sample#.tsv. Change them and give appropriate names to each sample. You can look at the phenodata.tsv in Chipster to know which sample is each one. Save the newly created file as txt.

15. Open Expander and upload the data. Remember that your data is log transformed.

The screenshot shows the 'Load Tabular Data' dialog box in the Expander software. The 'Organism' is set to 'mouse' and 'Expected gene IDs' is 'Entrez'. The 'Data name' is 'GE Data'. The 'Raw data file' is 'C:\Users\bsdena\Desktop\annotated.tsv'. The 'IDs conversion file' is empty. The 'Data type' is 'RNASeq counts' and the 'Data scale' is 'Log2'. The 'File contains detection calls (A, M, P flags)' checkbox is unchecked. The 'Set missing values to' radio button is selected with a value of '0.0'. The 'Estimate missing values with KNN' radio button is unselected. The 'OK', 'Cancel', and 'Advanced' buttons are at the bottom.

16. Click OK if you are asked to continue. Click No to download. Standardize and perform clustering using the click algorithm as done in the clustering exercise.

Q; How many clusters do you get?

17. Let's say that you are interested in the specific effect of miRNA27. Choose cluster 3 that has a relevant pattern and perform an analysis in David or [InterMine](#) for mouse (MouseMine). In David, use as background the default mouse background.
18. Write one sentence describing what in your opinion, is the main effect of addition of miRNA27 to the cell culture.