## June 2017

# An Introduction to Deep-Sequencing Data Analysis

## Final Exercise

***Ester Feldmesser and Dena Leshkowitz***

### Introduction

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 was compared to gene expression of cells transfected with miR-23, miR-24 or miR-27.

The experiment was performed in mice cultured cells in quadruplicates, for the exercise we choose three of the samples in duplicates.

### Assignment Instructions

1. You need to submit a report by email (Dena.Leshkowitz@weizmann.ac.il and Ester.Feldmesser@weizmann.ac.il). The report should include the specified below.
2. Create a folder in your home directory for the final exercise.
3. The fastq files are located in your home directory in ~/course\_2017/Final\_RNAseq. There is one folder for each sample and inside the folder there is fastq file. There is also a file called pheno\_data.txt that relates between the fastq name and the biological sample information

Perform quality control for all the samples using FastQC as was done in exercise2. Make a new folder for the fastqc output under the final exercise folder. Look at the results and write in the **report** a short summary regarding the sequence quality.

1. Map the sequences , see commands used in Exercise 2 and notice that you should use mouse genome (index located in /shareDB/iGenomes/Mus\_musculus/UCSC/mm10/Sequence/STAR\_index/ and gtf file in /shareDB/iGenomes/Mus\_musculus/UCSC/mm10/Annotation/Genes/genes.gtf)
2. Next step will be to detect differentially expressed genes when comparing each of the miRNAs to the control (miRNA mimics). Open the WEXAC Rstudio through the web. We suggest loading the R script that was used for Arabidopsis and to make changes on it. You will need to change the working directory and the path to files. In addition, note that you have less samples than in the Arabidopsis exercise, therefore you will perform only two pairwise comparisons and you will need to change the column numbers in the K-means clustering. Prepare a report Excel file containing all the genes detected and their comparisons information. In addition create an Excel report for only the differentially expressed genes (threshold fdr<=0.05 and absolute logFC >=1). **Submit the excels with the report.**

Perform clustering using K-means in a similar way to the used in the clustering exercise (exercise 5) and create a heatmap. Try at least 2 and 3 clusters. Include the plots in the **report**.

1. Let’s assume that you are interested in the specific effect of miRNA24. Choose the cluster that best reveals this effect and perform an analysis in David or InterMine for mouse (MouseMine) using these genes. In David, use as background the default mouse background.
2. Summarise in the **report** what is the main specific biological effect/s of addition of miRNA24 to the cell culture.