

## Bioinformatics III

Prof. Dr. Volkhard Helms

Thorsten Will, **Maryam Nazarieh**, Duy Nguyen, Daria Gaidar  
Winter Semester 2016/2017

Saarland University  
Chair for Computational Biology

### Exercise Sheet 10

**Due: January 30, 2017 13:15**

Submit your solutions on paper, hand-written or printed at the *beginning* of the lecture or in building E2 1, Room 3.03. Alternatively you may send an email with a single PDF attachment. Additionally hand in all source code via mail to [maryam.nazarieh@bioinformatik.uni-saarland.de](mailto:maryam.nazarieh@bioinformatik.uni-saarland.de).

## Gene Expression and DNA Methylation

### Exercise 10.1: Read data and Normalization (25 points)

The data given in the supplementary comprises of two already normalized data sets for expression and methylation of 100 genes in blood and skin tissues with 19 samples, where (HSC, MPP1, MPP2, CLP, CMP, GMP, MEP, CD4, CD8, B\_cell, Eryth, Granu and Mono) are different blood cells and (TBSC,ABSC,MTAC,CLDC,EPro,EDif) are for skin tissue:

- Read the data into a data matrix where the rows correspond to the set of genes in each sample and columns correspond to the different samples.
- Filter the data, for both expression and methylation data, by removing entries with empty/NA expression and methylation values. If there are several entries with the same gene name, take the average mean expression and methylation value for each gene in every sample.
- Submit the final matrices as your solution.

### Exercise 10.2: Network Construction (50 points)

In this exercise, determine if the expression data at gene level has a normal distribution (You can use Shapiro test).

Construct a co-expressed network with three methods (Pearson, Spearman, Kendall) with threshold = 0.75.

Visualize the network using Cytoscape tool which is an open source software platform for visualizing complex networks and integrating these with any type of attribute data.

- Explain and discuss your results.

### Exercise 10.3: Clustering (25 points)

Apply a bottom-up hierarchical clustering on both expression and methylation data. Rotate the data such that samples are the rows of the data and genes are the columns of data.

- what is the similarity between the two sorts of hierarchical clustering with respect to expression and methylation?
- can hierarchical clustering cluster the samples into tissues (skin and blood?)

Have fun!