Bioinformatics Workshop on S. aureus-related genomics

Prof. Dr. Volkhard Helms
Ruslan Akulenko
Summer Semester 2014

Saarland University
Chair for Computational Biology

Exercise Sheet 4

August 28, 2014

Data processing and Statistical tests applied to *Staphylococcus aureus* data

In this assignment you will practically learn how to work with data in R, perform data preprocessing and visualization of your data, run statistical tests which will help to reveal additional knowledge from S. aureus MicroArray data.
In the first part you will study basic R functions and data types, learn how to load data and parse it. Some basic types of plotting will be also covered.
In the second part of the assignment, you will clean the data and predict missing values.
In the final part, the following statistical tests will be covered:
Student's t-test, Kolmogorov-Smirnov test and Wilcoxon signed-rank test.
These are the most widely used approaches which you will learn with later visualization of your results.

Exercise 1.
Run RStudio which is already installed on your PC. Get familiar with the user interface. Open existing script using File → Open File and select **Exercise4.R**. Make sure that the data file is located in your working directory. To switch directory use Session → Set Working Directory → Choose Directory.
This file contains complete code used in the current assignment. Above every line there is a commented description starting with the sign '**#**'. You can execute every line separately by placing cursor and clicking “Run”. However, the code is ordered according to the data processing steps.

Exercise 2.
Preprocessing stage starts with the loading data and cleaning part: separating data into several matrices, adjust names and other properties. You can have first impression about your data by plotting it. Then apply LFM to predict missing (ambiguous) entries. This procedure is computationally heavy and is expected to be executed around 8 min.

Exrcise 3.
Perform Student's t-test, Kolmogorov-Smirnov test and Wilcoxon signed-rank test. Plot and analyze the results.
Save completed data matrix to the file. It will be used later in the assignment 5.