

Lecture: Microarrays Bioinformatics

WS 2017/18

Assignment No. 6

(5 points)

Hand out: Thursday, December 7

Hand in due: Thursday, December 15, 10:00

Tutorial date: Tuesday, December 12, 10:15-11:45

Direct inquiries to: fabian.moertter@student.uni-tuebingen.de

Theoretical Assignments

1. Median Polish to summarize probe values

(5p)

The model to estimate a summarized expression value m_i of a probeset in the i -th array from its probe values x_{ij} is an additive linear model:

$$x_{ij} = m_i + a_j(+\epsilon)$$

Here a_j refers to the probe effect of the j th probe in a probeset (ϵ denotes a random error).

To estimate m_i and a_j Tukey's two-way median polish method is applied: here one iteratively subtracts row medians from each respective row value, then column medians from respective column values. This is continued until all row and column medians are zero or do no longer change. The final values of the median polish matrix are subtracted from the respective original matrix elements. These are now fitted probe values. The final summarized probeset values are derived by taking the median of all probes in each array.

Apply the median polish method to the following hypothetical expression values of just one probeset g consisting of 5 probes, measured in 4 arrays.

	A_1	A_2	A_3	A_4
p_1	8	13	15	9
p_2	11	7	6	15
p_3	8	5	7	12
p_4	14	16	16	13
p_5	4	7	6	5

Report intermediate steps and results of your computations, as well as the final values for g , which is a vector of length 4 (representing the 4 array values).

Practical Assignments

2. Lowess Normalization

Use the **same data** set as in the previous assignment (**affy-data.tsv**).

- (a) Using an **appropriate R package**, perform a **Lowess normalization on the MA values** that you calculated in the previous assignment.
- (b) Produce **diagnostics plots** of the normalized data including the Lowess function.
- (c) **Compare the two normalization methods** (linear regression from the previous assignment and Lowess normalization). Which one is **more appropriate** for the given data? Briefly discuss the results.

3. Invariant gene set algorithm and normalization

The **SingleExpressionCELfiles4A6.tsv** data set from the material folder contains **background corrected unnormalized expression** (\log_2) values of **two single-channel** (Affymetrix) microarray experiments.

- (a) Implement the **invariant gene set algorithm** as discussed in the lecture. Make sure to keep track of the invariant set (size). Apply your algorithm to the provided data.
- (b) Create a **plot showing the size** of the **invariant set** for each iteration.
- (c) **Apply** your **linear regression** function from the last assignment to
 - i. the **complete data** and **normalize**,
 - ii. the **invariant genes** only and **normalize then all genes**.
- (d) Think about **a visualization** how to **compare the two normalizations of (c)** and apply it to your results. Briefly **discuss** the results.

Please read the questions carefully. If there are any questions, you may ask them during the tutorial session or via e-mail to your tutor. You will usually get an answer in time, but late e-mails (e.g. on Thursday morning before class) might not be answered in time. Please upload your solutions in the Ilias system. Please pack your source code, the plots, as well as the theoretical part into **one single archive file (zip)**. Source code should compile correctly.