Prof. K. Nieselt, Faculty of Science

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Lecture: Microarrays Bioinformatics WS 2017/18

Assignment No. 1 (5 points)

Hand out:	Thursday, October 26
Hand in due:	Thursday, November 9, 10:00
Tutorial date:	Tuesday, Nov 7, 10:00-11:30
Direct inquiries to:	fabian.moertter@student.uni-tuebingen.de

Theoretical Assignments

1. Application of Microarrays (4p)

The design of a microarray experiment/study always depends on the subject that the researcher wants to investigate. Think about an experimental design or approach using microarrays to address either one of (**choose only one of a) OR b**) the following questions:

(a) As you learnt in 'Grundlagen der Bioinformatik', after sequencing a genome typically genes (mostly protein-coding) are predicted using sequence-based approaches such as HMMs (e.g. `GenScan`). One possibility to validate the predictions is via expression experiments (note: this only addresses the true positives, not the false positives. Why?). How can microarrays (including design of the array) be used to confirm the result of gene prediction?

(b) Describe an experiment that could be used to identify core genes that are active / expressed in every tissue in humans. Which experiments do you suggest that address this question? What characteristics (in general) does a gene have to show in your experiment in order to be 'defined' as a core gene?

Hand in a written text (in English or German) of up to one page.

2. Sensitivity and Specificity of a Microarray (1p)

As explained in the lecture, an important aspect in many high-throughput studies are the terms 'sensitivity' and 'specificity' of a genome-wide expression microarray.

Please briefly explain and illustrate (e.g. by producing a figure) these two terms for the example of a whole human genome microarray.

Practical Assignments

1. **Specificity of oligo probes**

   In the following exercises on this and the following assignments you will learn to read in data sets and to access and plot your data in R. It is useful to consult the help for the `read.table` function before starting. The material necessary for this assignment (`A01.zip`) is available from the course website within `ILIAS`.

   The following task addresses the topic of oligo design. In this real-world data set, two so-called capture microarrays were designed for the bacterium *Treponema pallidum*. A capture microarray has the purpose to represent the whole genome (and not only the genes themselves). The designed probes are of length 60 and 100, respectively. Try to make yourself familiar with the tasks on the file `OligoHitsums1.csv` as the second file is quite large.

   (a) Read in the files `OligoHitsums1.csv` and `OligoHitsums2.csv`. The files contain a row for each probe of the *Treponema pallidum* microarray. The probe name is in the first column. In each of the following columns, the number of 'hits' (i.e., alignment of the probe sequence with a region of the same length as the probe in the genome) a probe has with exactly i mismatches, where i goes from 0 to the length of the probe.
   Column 1: Probe name
   Column 2: Number of hits with no mismatch
   Column 3: Number of hits with one mismatch
   Column 4: Number of hits with two mismatches
   ...

   (b) Produce a visualization using the boxplot. You may have to split the data in order to allow for better interpretation. Try to interpret the plots: is there a visible threshold for the number of mismatches that give meaningful results. Is this also a meaningful threshold from the biological point of view?

   (c) How many probes have exactly one perfect match, how many have more than one perfect match? Visualize the results. You can either use again a boxplot, or find out about other useful plots in R for this data.

   (d) Discuss briefly your results: just for the aspect analysed here, how do you judge the specificity of the oligo probes on this array? Which other aspects should be analysed to be able to answer 'how good is this array'?

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.