Available online at www.sciencedirect.com
ScienceDirect
Biochimica et Biophysica Acta 1759 (2006) 319 – 327
BIOCHIMICA ET BIOPHYSICA ACTA
BBA
www.elsevier.com/locate/bbaexp

Review

Biological microarray interpretation: The rules of engagement

Rainer Breitling *

Groningen Bioinformatics Centre, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands

Received 8 May 2006; received in revised form 30 June 2006; accepted 30 June 2006
Available online 13 July 2006

Abstract

Gene expression microarrays are now established as a standard tool in biological and biochemical laboratories. Interpreting the masses of data generated by this technology poses a number of unusual new challenges. Over the past few years a consensus has begun to emerge concerning the most important pitfalls and the proper ways to avoid them. This review provides an overview of these ideas, beginning with relevant aspects of experimental design and normalization, but focusing in particular on the various tools and concepts that help to interpret microarray results. These new approaches make it much easier to extract biologically relevant and reliable hypotheses in an objective and reasonably unbiased fashion.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Microarray; Gene expression; Bioinformatics; Tool; Computational biology; System biology

1. Introduction

Microarrays are gaining popularity in biological laboratories by the day. In their standard application, they measure the gene expression status of a particular sample, by quantifying the mRNA levels of all genes in highly parallel fashion. A single array hybridization generates as many data as several classical Ph.D. theses taken together. This makes the technology very appealing as a shortcut towards data production, despite the large financial and technical demands. To turn these data into biological insight it is, however, necessary to interpret the collected data. After having come to grips with the manipulation of the large amount of collected data, the interpretation may seem almost trivial at first: there are many possible stories that a creative expert biologist can detect in the data, treating the results for each gene more or less like familiar Northern blot information. But in contrast to more traditional approaches the main challenge is not to come up with an explanation for the behavior of a single gene, but with one that is consistent and well supported in the context of complementary information on thousands of other genes.

New ideas on how to achieve this task are developed rapidly, new microarray tools and specialized microarray statistics are abounding. Making an informed choice between them may sometimes be daunting, but fortunately over the last few years a number of comparative evaluation studies and the accumulated experience from many thousands of microarray studies have lead to the emergence of a set of guidelines and informal standards that help in the process. In the following sections we will first outline the basic conditions that need to be fulfilled before a successful interpretation can start and describe how to detect differentially expressed genes and how to avoid the major pitfall associated with multiple testing. Then we proceed with a highly selective outline of methods to organize the differential expression information and to explore and annotate the results to obtain a biologically coherent interpretation. We conclude with an overview of higher-level tools that support the integration and extension of the array results with additional data sources. Fig. 1 summarizes the interpretation procedure.

2. Biological question and experimental design

A number of critical steps need to be taken before one can think of interpreting a microarray result successfully. One of these is the proper randomization during the experiment. It is

* Tel.: +31 50 363 8088; fax: +31 50 363 7976.
E-mail address: r.breitling@rug.nl.

0167-4781/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbaexp.2006.06.003

often jokingly said that the results of a laboratory experiment depend on the phase of the moon. For microarray analysis this is almost true. Gene expression responds very sensitively to changes in environmental conditions, even if they appear very minor. Consistent and statistically significant fluctuations in expression pattern have been reported in wild type yeast cultures incubated under constant standard growth conditions [1]. Even for genetically homogeneous inbred mouse strains reared under highly-controlled, pathogen-free laboratory conditions and matched for age and sex reproducible and statistically significant inter-individual variation in gene expression has been reported [2]. It was necessary to control many additional variables like social status, stress, and food intake to reduce this biological variation to a minimum. Even then, individual mice showed significant differential expression of some genes. Similar effects have been reported for *Arabidopsis*, where even simply touching the plants can lead to significant changes in gene expression [3]. Therefore, it is of utmost importance for the generation of interpretable microarray results to randomize or control all possible confounding factors. If two conditions are to be compared, then it is not sufficient to, for example, obtain the material at the same temperature and in the same medium, but it should preferably be grown in the same incubator, in random spatial arrangement and harvested in random order by the same person. The exact factors to be randomized will depend on the particular experimental set-up, but in any case it will be crucial to realize that the most unexpected subtle systematic difference between the sampling (and measurement) conditions will lead to biased results—making the data all but impossible to interpret.

To be able to interpret microarray data, it is also necessary to have a sufficient number of replicate measurements. Such replication is necessary to assess which results will have real predictive value, i.e. are expected to be verified in a new experiment, and which observations are only spurious. Statistical approaches allow one to assess the number of replicates necessary to reliably detect an expected effect of a certain size. For most applications, where financial and logistic constraints limit the number of hybridizations, this is merely an academic exercise. The more replicates the better. A real choice occurs when the type of replicates is chosen. Here it is important to realize that arrays show surprisingly little technical (labeling- or hybridization-specific) variation. Most of the observed expression variation is due to biological variability (day-to-day or interindividual variation; [4–7]). It is, therefore, most efficient and informative to perform as many *biological* replicates as possible. Rather than hybridizing a specific sample twice, isolate the same type of mRNA from a completely independent new sample. This will help to detect those expression changes that are reproducible and relevant. Detailed reviews of experimental design considerations, in particular for more complex experiments such as large time courses or multifactorial comparisons, are available in [8,9].

3. Hybridization

The choice of array platform and other technical details of the actual microarray experiment (sample preparation, hybridization, image processing) of course also influence the downstream analysis, but these will not be discussed in this paper. For

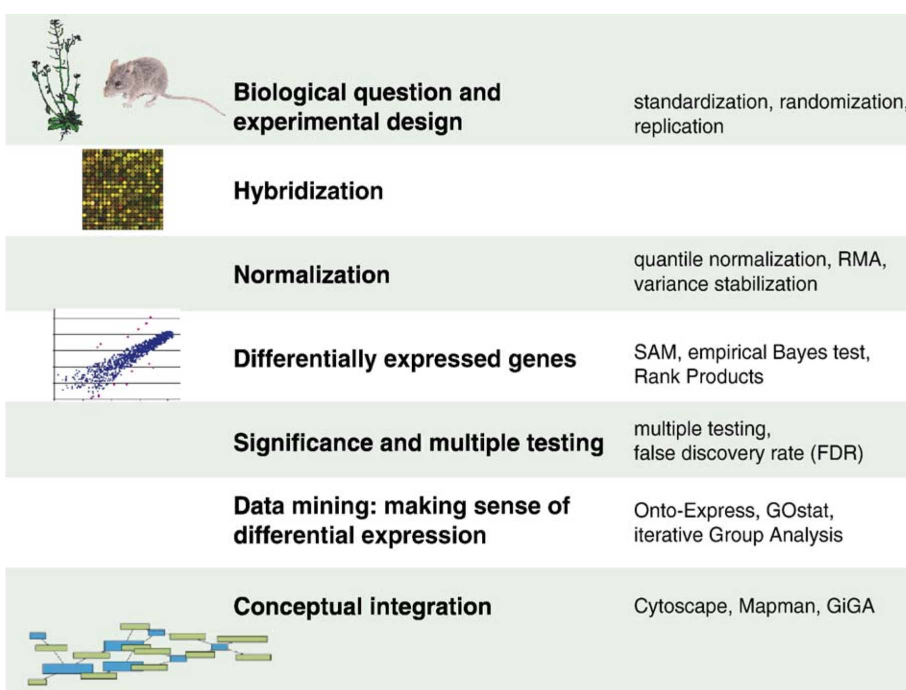

Fig. 1. Critical steps in microarray interpretation. The major steps of the usual interpretation process are outlined. Major pitfalls and some related analytical approaches are briefly indicated at the right. See main text for detailed discussion and the description of additional relevant tools and methods.

Download English Version:

https://daneshyari.com/en/article/1947211

Download Persian Version:

https://daneshyari.com/article/1947211

Daneshyari.com