Normalization strategies for cDNA microarrays

Johannes Schuchhardt*, Dieter Beule, Arif Malik¹, Eryc Wolski¹, Holger Eickhoff¹, Hans Lehrach¹ and Hanspeter Herzel

Institute for Theoretical Biology, Humboldt-Universität zu Berlin, Invalidenstrasse 43, D-10115 Berlin, Germany and
¹Max Planck Institute of Molecular Genetics, Ihnestrasse 73, D-14195 Berlin, Germany

Received as resubmission February 16, 2000; Revised March 13, 2000; Accepted March 20, 2000

ABSTRACT
Multiple Arabidopsis thaliana clones from an experimental series of cDNA microarrays are evaluated in order to identify essential sources of noise in the spotting and hybridization process. Theoretical and experimental strategies for an improved quantitative evaluation of cDNA microarrays are proposed and tested on a series of differently diluted control clones. Several sources of noise are identified from the data. Systematic and stochastic fluctuations in the spotting process are reduced by control spots and statistical techniques. The reliability of slide to slide comparison is critically assessed within the statistical framework of pattern matching and classification.

INTRODUCTION
Large areas of medical research and biotechnological development will be transformed by the evolution of high throughput techniques (1–3). Miniaturization and automatization enables the concurrent performance of many thousands or even millions of small-scale experiments on oligonucleotide chips (4,5) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8). Manufacturing processes and labeling techniques will lead to different performances (9,10) or spotted microarrays (6–8).

© 2000 Oxford University Press
MATERIALS AND METHODS

Array preparation

A complex probe from several mouse tissues was purified and reverse transcribed with radioactively labeled cDNA. Arabidopsis thaliana cDNA (GenBank accession nos AF104328 and U29785) was spiked in a fixed amount for normalization purposes (18). Clones were amplified by PCR reaction, 5′-aminomodified for attachment to glass slides, and purified (19). Prior to spotting, glass slides were cleaned and derivatized for covalent attachment of cDNA. A 384 pin gridding head (X5251; Genetix, Christchurch, UK) was used for spotting a grid of 384 blocks, each containing 36 spots. All clones were spotted twice within a block (double spotting).

Details of the spotting pattern of library and control clones are explained in Figure 1. Altogether nine slides with an identical spotting pattern were produced.

The radioactively labeled probe was hybridized on the cDNA array for 10 h at 42°C. For details on spotting technique and hybridization procedures see Eickhoff et al. (20).

Scanning and image processing

Arrays were exposed for 16 h to a Fuji BAS-SR 2025 intensifying screen (Raytest, Germany) and scanned at 25 μm resolution with a Fuji BAS 5000 phosphorimager (Raytest). The image was converted into a table of signal intensities using proprietary software.

Data processing

Intensity data were ordered in a table, each column corresponding to a slide and each row to a spot on the slide. The following normalization procedures were tested for their efficiency:

- no normalization, averaging over k slides;
- normalization by average intensity of control spots (slide-wise normalization) and averaging over k slides;
- division by the intensity of the two constant spots and averaging over k slides (pin-wise normalization);
- slide-wise normalization of the diluted and constant signals, averaging of the dilution and control signals over several slides, then quotient formation (average pin-wise normalization).

RESULTS

Non-specific background and overshining

The level of background noise and the influence of neighboring signal intensities is illustrated in Figure 2. The intensity of background spots is plotted versus the average signal intensity of the four next neighbor spots.

The y-axis intercept of the linear regression gives an estimation of the non-specific background. The small background intensity indicates that there are only weak overshining effects for the 6×6 spotting pattern. The regression can be used for correction of the systematic part of these errors. The radius used to quantify spots was varied systematically: for the given spotting density only weak changes are observed if the scanning radius is kept in a reasonable range of about half the spotting distance (data not shown). The magnitude of the background and overshining effects is substantially smaller than fluctuations induced by spotting variabilities quantified below.

Assessment of spotting variabilities

In order to facilitate interpretation of the experimental data we neglect all non-linearities from image processing and assume that hybridization reactions reach mass action equilibrium. Due to the fact that different spots of a dilution series compete for the same probe the amount of probe bound in each spot is proportional to the amount of target cDNA present in the spot. The observed signal intensity then reflects the amount of spotted cDNA. Fluctuations in spot size and in the hybridization
process can now be inferred from fluctuations in the signal gained from control spots. Scatter plots for typical experiments are presented in Figure 3 and give insight into the magnitude of fluctuations and correlations between double spotted signals.

In Figure 3a signal intensities of the dilution series are shown. The 384 points represent double spotted pairs from the same block of one slide. Ideally the spots would form a straight line along the diagonal. Deviations from this ideal behavior essentially reflect random fluctuations in target volume (item 6 listed in the Introduction). Points are increasingly dense close to the origin owing to the construction of the dilution series: dilution is stepwise increased by a factor of two, resulting in an exponentially growing number of points in the region of low signal intensities. The correlation value of $C = 0.90$ gives an impression of the signal reproducibility within one block.

In Figure 3b intensity values of two slides are compared. The correlation is substantially smaller ($C = 0.76$) and the slope of 0.6 is significantly different from 1. The smaller correlation value indicates that additional sources of noise complicate the comparison of different slides. According to our reasoning above, the main sources of additional noise will be inhomogeneities in hybridization (items 8 and 9) as well as non-linear transmission and saturation effects in scanning and image processing (item 12). The fact that the cumulative intensity of all spots of the dilution series should be the same in all nine slides implies a simple normalization procedure: divide the signal intensity of each spot by the average intensity of all control spots on the slide. This intuitive normalization procedure will be referred to as ‘slide-wise normalization’. Applying slide-wise normalization to the example given in Figure 3b would result in a corrected slope close to 1. The values from different slides are now directly comparable, which is an obvious advantage of this normalization procedure.

Due to the stochastic character of the above-mentioned fluctuations, averaging is a sensible method to reduce the noise level. We demonstrate this by normalizing nine slides as
strongly overlap, as seen from projecting the points to the only the dilution signal is known the different classes will respond to 64 different pins (see Materials and Methods). If each dilution level is represented by 64 spots corresponding to the constant control signal. The intensity of the averaged diluted control signal is deviation induced by different pin characteristics.

Figure 4. Intensity of diluted control signal versus intensity of the constant control signal. Values are obtained by averaging over eight slides after normalization by slide average. The six different dilution classes are marked by six different symbols. Classes are separable to a fair degree in two dimensions, owing to correlations between the constant and diluted signals. Class boundaries as defined by the nearest mean classifier are indicated by lines.

described above and plotting the averaged normalized intensities (Fig. 3c). Scattering of the signal is considerably reduced, as reflected in the substantially higher correlation coefficient of \( C = 0.99 \). Despite the small deviations of the averaged signal from the diagonal, the six classes of the dilution series cannot be distinguished in the scatter plot in Figure 3c. This hints at the presence of another source of noise not reduced by this averaging procedure.

This variability stems from systematic variations in pin geometry, as clearly shown in Figure 3d. Here, only the points belonging to the second dilution level are plotted. If only stochastic sources of noise were present in the signal, a narrow circular distribution would be expected. In contrast, a correlation coefficient of \( C = 0.97 \) and a wide signal distribution indicate correlated variations in the signal that are present despite averaging over different slides. On the other hand, the strong correlations can be exploited for correction of the systematic deviations induced by different pin characteristics.

Figure 4 illustrates how the presence of a constant control signal can help to distinguish the different classes of the dilution series. The intensity of the averaged diluted control signal is plotted versus the intensity of the averaged constant control signal. Each dilution level is represented by 64 spots corresponding to 64 different pins (see Materials and Methods). If only the dilution signal is known the different classes will strongly overlap, as seen from projecting the points to the y-axis. In two dimensions, however, the six classes are reasonably well separated. Basically this is due to the fact that the reference signal helps to decide whether an observed strong signal intensity is due to high concentration in the target clone or rather a consequence of systematically excessive volume spotted by the corresponding pin. The most direct way to achieve normalization by the constant control signal is by calculating the ratio of the dilution signal and the constant control signal. In geometrical terms this corresponds to a projection of the data point along a ray from the origin onto the vertical axis. In the next section we investigate how suitable ways of averaging and normalization improve the separability of the dilution classes.

**Classification of the dilution levels**

Separation of the dilution levels based on signal strengths can be regarded as a classification task: given a randomly chosen position on the slide, infer the correct dilution class from the signal intensity observed at that point. Classification is perfect when all spots are assigned to the correct class. Of more practical importance than reassignment is prediction of signal intensity patterns of unknown class. This problem is comparable to the task of inferring an mRNA level from the signal intensity observed in a spot, where the probe rather than the target dilution is of interest. Classification and prediction are done with a nearest mean classifier (21). This classifier is represented by a number of center vectors, each vector defined as the center of one class. Any pattern classified by this method is associated with the class defined by the closest center. Two issues will be investigated:

- improvement of classification and prediction performance by averaging multiple experiments;
- improvement of classification and prediction performance by suitable normalization.

Improvement by averaging is assessed by dividing a set of eight slides into training and test sets of varying size. In four steps an increasing number of slides \( k \) is used to infer the dilution class in a given position of the slide. For \( k = 1 \) the classification task is to recover the correct dilution class of a given spot using the nearest mean classifier trained on the same slide. The prediction task is to predict the correct dilution class of a given spot using the nearest mean classifier trained on a different slide. The quality of classification is measured by the average intensity observed in a spot, where the probe rather than the target dilution is of interest. Classification and prediction are done with a nearest mean classifier (21). This classifier is represented by a number of center vectors, each vector defined as the center of one class. Any pattern classified by this method is associated with the class defined by the closest center. Two issues will be investigated:

- improvement of classification and prediction performance by averaging multiple experiments;
- improvement of classification and prediction performance by suitable normalization.

Improvement by averaging is assessed by dividing a set of eight slides into training and test sets of varying size. In four steps an increasing number of slides \( k \) is used to infer the dilution class in a given position of the slide. For \( k = 1 \) the classification task is to recover the correct dilution class of a given spot using the nearest mean classifier trained on the same slide. The prediction task is to predict the correct dilution class of a given spot using the nearest mean classifier trained on a different slide. The quality of classification is measured by the average intensity observed in a spot, where the probe rather than the target dilution is of interest. Classification and prediction are done with a nearest mean classifier (21). This classifier is represented by a number of center vectors, each vector defined as the center of one class. Any pattern classified by this method is associated with the class defined by the closest center. Two issues will be investigated:

- improvement of classification and prediction performance by averaging multiple experiments;
- improvement of classification and prediction performance by suitable normalization.

Improvement by averaging is assessed by dividing a set of eight slides into training and test sets of varying size. In four steps an increasing number of slides \( k \) is used to infer the dilution class in a given position of the slide. For \( k = 1 \) the classification task is to recover the correct dilution class of a given spot using the nearest mean classifier trained on the same slide. The prediction task is to predict the correct dilution class of a given spot using the nearest mean classifier trained on a different slide. The quality of classification is measured by the average intensity observed in a spot, where the probe rather than the target dilution is of interest. Classification and prediction are done with a nearest mean classifier (21). This classifier is represented by a number of center vectors, each vector defined as the center of one class. Any pattern classified by this method is associated with the class defined by the closest center. Two issues will be investigated:

- improvement of classification and prediction performance by averaging multiple experiments;
- improvement of classification and prediction performance by suitable normalization.

**DISCUSSION**

The evaluation performed for multi-spotted *A. thaliana* control clones allowed the identification of essential sources of noise resulting from the spotting and hybridization processes. Systematic signal fluctuations in target transmission during the spotting process can be reduced by monitoring control spots. Stochastic fluctuations can be reduced by averaging intensity
arrays using glass slides and radioactively marked probes, the probe in the two-color labeling strategy may be a disadvantage implicit by a reference probe. The unavoidable use of a reference either explicit by a defined set of normalization clones or of the actual spot size. Generally, the comparison of different two-color labeling strategy, since ratios should be independent comparability of microarray data, irrespective of where the solution could be the initial step towards an enhanced spotted array and spiking of these clones into the hybridization and labeling technique employed.

Comparing the four normalization schemes proposed, for our experiment the average pin-wise strategy (D) seems most appropriate. Training and test performance are persistently higher than in all other procedures. An explanation for this fact is given by the following consideration. By first calculating an average over several slides and then calculating the ratio of the averaged quantities, this strategy tends to avoid a strongly fluctuating denominator. This makes strategy (D) superior to strategy (C), where first the ratio of individual quantities is calculated and than an average is taken. Of course, the optimal strategy will depend on the relative amount of noise contributed by different sources and will very much depend on the array and labeling technique employed.

We envisage that the inclusion of control clones in every spotted array and spiking of these clones into the hybridization solution could be the initial step towards an enhanced comparability of microarray data, irrespective of where the array is produced. Normalization may be less important for the two-color labeling strategy, since ratios should be independent of the actual spot size. Generally, the comparison of different experiments (slides) will require some kind of normalization: either explicit by a defined set of normalization clones or implicit by a reference probe. The unavoidable use of a reference probe in the two-color labeling strategy may be a disadvantage if results from different laboratories are to be compared.

Although our considerations have concentrated on cDNA arrays using glass slides and radioactively marked probes, the principal sources of systematic and stochastic signal variability are expected to be similar in other microarray technologies involving spotting techniques. A proper normalization of array data will be a prerequisite to successfully meet future challenges like identification of SNPs, identification of expression profiles and reverse engineering of regulatory genetic networks.

**REFERENCES**