Published online February 13, 2004

Nucleic Acids Research, 2004, Vol. 32, No. 3 ***e29***
DOI: 10.1093/nar/gnh027

Reliability of gene expression ratios for cDNA microarrays in multiconditional experiments with a reference design

Rainer König*, Danila Baldessari[1], Nicolas Pollet[1], Christof Niehrs[1] and Roland Eils

Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, 69120 Heidelberg, Germany and [1]Division of Molecular Embryology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

Received November 24, 2003; Revised and Accepted January 10, 2004

ABSTRACT

In a typical gene expression profiling experiment with multiple conditions, a common reference sample is used for co-hybridization with the samples to yield expression ratios. Differential expression for any other sample pair can then be calculated by assembling the ratios from their hybridizations with the reference. In this study we test the validity of this approach. Differential expression of a sample pair (*i*, *j*) was obtained in two ways: directly, by hybridizations of sample *i* versus *j*, and indirectly, by multiplying the expression ratios for hybridizations of sample *i* versus pool and pool versus sample *j*. We performed gene expression profiling using amphibian embryos (*Xenopus laevis*). Every sample combination of four different stages and a pool was profiled. Direct and indirect values were compared and used as the quality criterion for the data. Based on this criterion, 82% of all ratios were found to be sufficiently accurate. To increase the reliability of the signals, several widely used filtering techniques were tested. Filtering by differences of repeated hybridizations was found to be the optimal filter. Finally, we compared microarray-based gene expression profiles with the corresponding expression patterns obtained by whole-mount *in situ* hybridizations, resulting in a 90% correspondence.

INTRODUCTION

Large-scale gene expression profiling with cDNA microarrays has emerged from specialized groups to mainstream laboratories in modern medical and biological research over a period of a few years (1). This technology allows us to explore a major subset or almost all genes for an organism. In multiconditional experiments a variety of conditions such as samples of several treatments, mutants, developmental stages or time points are examined. Thus, insights into gene expression and regulatory mechanisms have been achieved (2–4). However, compared with northern blot or RT–PCR analysis, gene expression values derived from cDNA microarrays are less reliable (5). Experimental biases in gene expression profiling occur due to local and global background, differing label incorporation, varying total amount of hybridized mRNA or bleaching effects of the dye. To reduce these biases and allow comparisons of hybridizations within multiconditional experiments, normalization methods have been developed (6–9). Jenssen and coworkers (10) suggested using the variability of double spots as a quality criterion for the expression signals. Furthermore, the mean and median of the pixel intensities of the spots were taken as a quality criterion: mean and median values are equal if the pixel intensities are normally distributed, which should be the case for good quality spots (11). Spot quality was also assessed in an extended examination considering and integrating spot size, signal-to-noise ratios, background uniformity and the saturation status (12). In another study it could be shown that expression signals may vary due to boundary effects of the slides (13). Statistical tests aim at estimating the significance of differential gene expression. Tibshirani and coworkers developed a significance test for each gene [significance analysis of microarrays (SAM)] (14,15). Their test was expanded to a robust classification scheme for the conditions (16). However, these tests depend on a sufficient number of repeated hybridizations.

In the present study, we assessed the reliability of a differential gene expression value in a multiconditional experiment when a limited number of repeats are available (in our case: one repeated colour-reversed hybridization). For this purpose we compared two experimental designs: the 'all-pairs design', considering all the possible combinations of samples, and the widely applied 'common reference design',

*To whom correspondence should be addressed. Tel: +49 6221 423600; Fax: +49 6221 423620; Email: r.koenig@dkfz.de
Present address:
Nicolas Pollet, Transgenèse et Génétique des Amphibiens, CNRS ESA8080, IBAIC Bat. 447, Université Paris-Sud, F-91405 Orsay, France

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

having a common reference for all hybridizations. While in the 'all-pairs design' samples are hybridized directly one against each other, within the 'common reference design' samples under investigation are hybridized with one reference sample, which may be a pool of samples, a sample at time point zero, a wild type or a non-malignant form of the tissue (2,3,15,17,18). Each of these experimental designs has advantages and limitations which are theoretically approached and extensively discussed by Yang and Speed (19).

In our experiment, we performed an all-against-all comparison of four samples (*Xenopus laevis* embryos at the four different developmental stages 10.5, 13, 19 and 38) and a pool, resulting in a total of 20 hybridizations. These four stages represent critical steps in development where we expect high variation in gene expression and for which comprehensive knowledge about gene expression and molecular mechanisms is available (20). Stage 10.5 represents the beginning of gastrulation, the process where specification of the three germ layers takes place. At stages 13 and 19, the beginning and end, respectively, of formation of the neural tube takes place. Stage 38 represents a tadpole stage in which most organs have developed. We experimentally tested if the comparison of gene expression profiles of samples through the 'common reference design' is as reliable as the 'direct' sample pair assessment. To improve the reliability of the data by discarding ambiguous signals, we evaluated four filtering methods including criteria such as differences between mean and median of the pixel intensities of a spot, differences of double spots in each hybridization and differences of repeated (colour-reversed) hybridizations. Finally, we confirmed the consistency of our microarray gene expression profiles with the corresponding gene expression patterns obtained by whole-mount *in situ* hybridizations (AxelDB) (21). We show that among the filter-validated clones with differential expression, 90% have a positive correlation between microarray and *in situ* hybridization results.

MATERIALS AND METHODS

Microarray slide manufacturing

cDNA clones were obtained from five different *X.laevis* libraries to represent a wide variability of sources. Half of the clones were from two libraries of stage 13 embryos and wild type (22) or LiCl treated (23). The other half of the clones were mixed from three libraries from liver, stage 24 and stage 10 embryos (Unigene dbest libraries 5540, 5539 and 5575, respectively) distributed by the IMAGE consortium and obtained from the RZPD (German Resource Center, Germany). Bacterial clones were grown in 96- or 384-well plates. The DNA inserts were amplified by PCR using vector primers. PCR products were purified with a Millipore MANU 030 system, resuspended in betaine buffer (24) and spotted on CMT-GAPS II Coated Slides (Corning, NY, USA). We used an OmniGrid spotter with 16 (4 × 4 configuration) Telechem SMP3 pins (Telechem International, CA). The average spot diameter and separation was 115 μm; 3840 spots were printed in non-adjacent duplicate with one spot in each half (considering the length) of the slide. About 10% of the spots were exogenous controls. Their corresponding synthetic mRNAs were spiked in the samples before labelling. Positive controls include 13 200-bp fragments of lambda genes spotted at three different concentrations and 10 *Arabidopsis* clones from the Stratagene 'SpotReport™ array Validation System'. Negative controls included human and mouse COT DNA, yeast tRNA, only-buffer-spots and empty positions. Note that, within our normalization scheme, we normalized due to the bulk of data and not due to (a much smaller number of) controls. The rationale for this is given by Beissbarth *et al*. (6). The expression signals of the lambda controls shown in Supplementary figure 2A support the validity of the chosen normalization approach. Excluding all controls, the number of 'endogenous' *Xenopus* clones spotted on the slides was 3121. Of these clones, 2580 could be assigned to 1914 defined Unigene contigs. Some contigs were redundantly spotted (1× 29, 1× 20, 1× 17, 1× 13, 1× 12, 3× 10, 1× 9, 3× 8, 2× 7 and less redundant).

Sample preparation

In vitro fertilization, embryo culture and staging of *X.laevis* embryos were carried out as described by Gawantka *et al*. (25). Samples were collected from embryos of the eight stages 10.5, 13, 15, 19, 24, 28, 38 and 47. Total RNA was isolated according to the protocol of Chomczynski and Sacchi (26), modified by adding a 4 M lithium chloride precipitation. A final on-column DNase digestion was performed using the Qiagen RNase free DNase set and RNeasy spin column. RNA was precipitated and resuspended in DEPC-treated water at a final concentration of 10 μg/μl. A pool was composed of an equal amount of all eight samples. The synthetic mRNAs of the external positive controls were obtained by *in vitro* transcription with the Ambion MEGAscript kit and added to the total RNA samples. The RT reaction was performed with 50 μg of total RNA, oligo(dT)20VN, Supercript II (GIBCO) and amino-allyl-dUTP (Sigma, Germany). Subsequently the monofunctional dye Cy3/Cy5 (Amersham, UK) was coupled to the cDNA and afterwards a quenching reaction with 4 M hydroxylamine was performed. cDNA for each corresponding sample was labelled with both dyes (separately) to allow repeated hybridizations with reversed colour labels. After each of the two reactions (RT and quenching), samples were purified using QIAquick PCR purification kit and Microcon YM-30 columns (Millipore, MA). Ten micrograms of polyA (Sigma, Germany) and 20 μg of yeast tRNA were added to avoid unspecific hybridizations. The final volume of the sample was 20 μl.

Hybridization and scanning

The slides were prehybridized with prehybridization buffer (SSC 5×, SDS 0.1%, BSA 1%) at 45°C for 50 min and then washed five times with water and once with 2-propanol. The sample was denaturated (95°C, then ice) and 20 μl of 2× hybridization buffer (formamide deionized 50%, SSC 9.6×, SDS 0.2%) was added. The hybridization was performed under a glass coverslip in Telechem hybridization chambers (45°C for 18–22 h). Slides were washed three times: with 1× SSC, 0.2% SDS solution at 45°C, and with 1× SSC, 0.2% SDS and 0.1% SSC solutions at room temperature. Slides were scanned after washing with the ScanArray Lite scanner (GSI Lumonics-Packard, CA). The images were analysed with GenePix Pro3 software (Axon Instruments, CA).

Normalization and data preparation

We applied a normalization technique described elsewhere (6). We used the median of the pixel intensities of a spot if not noted otherwise. After local background subtraction, the dataset of each hybridization was divided into four subsets: values for green of the first spot, green of the second spot, red of the first spot and red of the second spot of a clone, respectively. The 5% percentile of each of these subsets was subtracted from the signal values to correct for global background. Each value <1 was set to 1. A constant (constant = 10) was added to all values accounting for less variant ratios in the low signal range. Each signal value was divided by the median of its subset. For calculating this median, low signal values (one-third of all values) were not taken into account to reduce the influence of higher noise in this range. Each value was multiplied by a constant (median of the medians of all subsets of all hybridizations) to bring them into the range of the original raw values. The most reliable values from these datasets were extracted by selecting expression values with the minimal absolute difference between any of the two duplicated green and red intensities (6). Mean expression values were calculated by the arithmetic mean of the repeated (colour-reversed) hybridizations. From this, ratios were calculated for the hybridized sample pair and the logarithm with base 2 applied. These values will be referred to as 'log2-ratios' in the following and are indexed with the conditions, i.e. the developmental stages 10.5, 13, 19, 38 and pool, respectively (Fig. 1) (for interquartile ranges of all sample combinations after normalization see Supplementary fig. 3). Accordingly, $r_c^{i,j}$ is the log2-ratio for clone c of condition i and j, $i,j \in \{10.5, 13, 19, 38, \text{pool}\}$.

The quality criterion

For sample pair (i, j), the indirect log2-ratio is calculated by adding the log2-ratios of pairs (i, pool) and (pool, j):

$$ri_c^{i,j} := r_c^{i,\text{pool}} + r_c^{\text{pool},j}. \quad \mathbf{1}$$

This value was compared with the direct log2-ratio $r_c^{i,j}$. The log2-ratios are available at http://www.dkfz-heidelberg.de/tbi/people/koenig/Data/Xenopus1/index.html. To obtain a clear tendency, we categorized the differential expression values into three classes:

$$cat_c^{i,j} := \begin{cases} \text{up} & \text{if } r_c^{i,j} \geq 1, \\ \text{down} & \text{if } r_c^{i,j} \leq -1, \\ \text{not-changed} & \text{else.} \end{cases} \quad \mathbf{2}$$

The same categories were determined for $ri_c^{i,j}$. The quality for $ri_c^{i,j}$ was defined as 'good' if $ri_c^{i,j}$ and $r_c^{i,j}$ were of the same category. Note that log2-ratios of 1 and –1 denote twice over- and under-expression, respectively. This cut-off has been reported as a solid benchmark (27). Comparing different cut-off values in our study supports this view (Supplementary fig. 1). In addition to the qualitative comparison, we also performed a comparison based on the quantitative indirect and direct values (data not shown). The quantitative assessment, however, could not reveal any further insights.

Protocol of the algorithm

We implemented our algorithm in Matlab (www.mathworks.com). It can be easily ported to other common platforms, such as Excel (www.microsoft.com) or SPSS (www.sas.com). The general workflow is described in the following: (i) raw data of all hybridizations was uploaded and normalized as described above; (ii) log2-ratios of all sample pairs and sample–pool combinations were calculated (note that the ratios of the sample pairs are already the 'direct' ratios); (iii) indirect ratios were calculated by adding log2-ratios of the corresponding two sample–pool pairs (as described in equation 1); (iv) direct and indirect ratios were categorized (as described in expression 2); (v) categorized indirect ratios were compared with categorized direct ratios, and each match was counted and the sum divided by the number of all values to obtain the reliability value; (vi) to increase reliability, different filtering methods were applied on all data as described in the following section, and step (v) was then repeated on the filtered values.

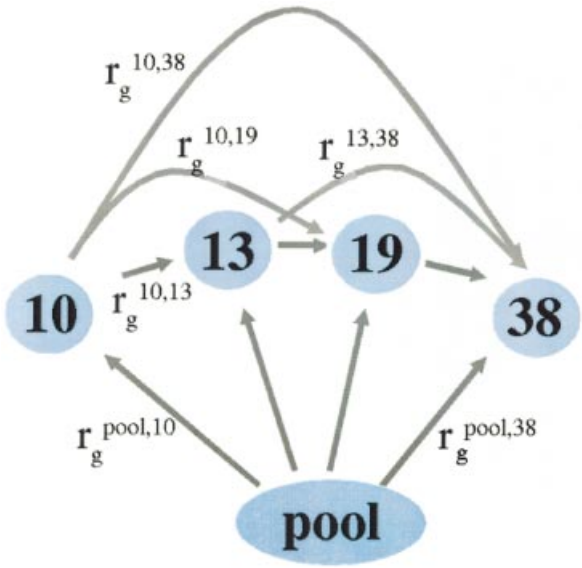

Figure 1. Experimental set-up. All possible pairs of conditions (developmental stages 10.5, 13, 19, 38 and a pool) were hybridized, yielding 10 different direct ratios $r^{10,19}$, $r^{13,38}$ etc. Each hybridization was repeated in colour-reverse mode.

Filtering

To increase the reliability of the data, signal validation techniques were applied for discarding ambiguous signal values. The signal validation techniques calculated a ranking for all values of all clones and sample pairs. Based on this ranking, values were discarded due to their low ranking at different stringency values between 0 (no values discarded) and 90 (90% discarded). We tested the following four validation criteria.

Validation criterion 1: spot intensity (si). The significance of a ratio increases with higher signal intensities (6, 28). We used signal intensities of the spots as our first validation criterion. The mean intensity of (normalized) green and red signal values for a clone c was calculated:

$$si_c^{i,j} := 1/16 \sum_{m=i,j} \sum_{k=1,2} ({}^k gprim_c^{m,\text{pool}} + {}^k gsec_c^{m,\text{pool}} + {}^k rprim_c^{m,\text{pool}} + {}^k rsec_c^{m,\text{pool}}) \quad i,j \in \{10.5, 13, 19, 38\}. \quad \mathbf{3}$$

where ${}^{k}gprim_{c}^{x,y}$, ${}^{k}gsec_{c}^{x,y}$, ${}^{k}rprim_{c}^{x,y}$, ${}^{k}rsec_{c}^{x,y}$ denote the normalized intensity values for clone c of sample pair (x, y) for primary and secondary spots, green and red, respectively. k denotes the hybridization (1: first, 2: second, colour reversed).

Validation criterion 2: difference of median and mean pixel intensities of a spot (dmm). The image processing software (GenePix Pro3, Axon Instruments, CA, USA) provides two intensity values for each spot and colour. It calculates the mean and median of the scanned pixel intensities of a spot. Mean and median values are equal if the pixel intensities are normally distributed, which should be the case for good quality spots (11). The absolute difference between mean and median intensities was taken as the quality criterion:

$$dmm_{c}^{i,j} := |ri_{c}^{i,j} - rim_{c}^{i,j}| \quad \mathbf{4}$$

where $ri_{c}^{i,j}$ denotes the indirect ratio as defined by equation 1 and $rim_{c}^{i,j}$ denotes the indirect ratio, calculated as in equation 1 only using mean instead of median values of the pixel intensities for a spot.

Validation criterion 3: difference of double spots (dds). The absolute differences between log2-ratios of primary and secondary spots were calculated:

$$dds_{c}^{i,j} := \Sigma_{m=i,j} \Sigma_{k=1,2} |{}^{k}raprim_{c}^{m,pool} - {}^{k}rasec_{c}^{m,pool}|$$
$$i, j \in \{10.5, 13, 19, 38\}. \quad \mathbf{5}$$

where ${}^{k}raprim_{c}^{x,y}$ and ${}^{k}rasec_{c}^{x,y}$ denote the log2-ratio of the first and the second spots for clone c of sample pair (x, y), respectively. k denotes the hybridization (1: first, 2: second, colour reversed).

Validation criterion 4: difference of the repeated hybridizations (hd). The log2-ratio differences of the first and the second hybridization (colour reversed) of each sample pair (i, j), $i, j \in \{10.5, 13, 19, 38, pool\}$ were calculated:

$$hd_{c}^{i,j} := 1/2 \, \Sigma_{m=i,j} |{}^{1}rh_{c}^{m,pool} - {}^{2}rh_{c}^{m,pool}| \quad \mathbf{6}$$

where ${}^{k}rh_{c}^{x,y}$ is the log2-ratio of hybridization k for clone c and sample pair (x, y).

RESULTS

Quality control without filtering

Indirect (hybridizations of sample i versus pool and pool versus sample j) and direct (sample i versus j) log2-ratios were compared for all sample pairs. Plots of indirect versus direct ratios show rather good correlation for pairs (10.5, 13), (19, 38), (13, 38) and (10.5, 38) (Fig. 2; for correlation coefficients see Table 1). Pairs (13, 19) and (10.5, 19) showed a lower correlation. Most of the hybridizations had a correlation coefficient above 0.8. As expected, when comparing gene expression of such diverse developmental stages, we observed highly scattered data in some cases. Both hybridizations for pair (10.5, 19), and one hybridization of pairs (10.5, 13), (pool, 10.5), (pool, 19), (10.5, 38), (19, 38) and (13, 38), respectively, had a correlation coefficient between 0.7 and 0.8. Pair (10.5, 38) had one hybridization with a correlation coefficient of 0.53. The scattering of both hybridizations for sample pair (10.5, 19) may be a reason for the low correlation coefficient of indirect versus direct values (0.69). Similarly, the scattering of hybridization for (pool, 19) may account for the low correlation coefficient of pair (13, 19). We compared the mean of indirect and direct log2-ratios for all sample pairs (Table 1). All sample pairs showed comparable values; (10.5, 19) and (10, 38) showed the largest difference of 0.3 which may be due to the larger scattering of their direct values. Note that our criterion of 1 for being differentially expressed was more than three times higher than this largest difference.

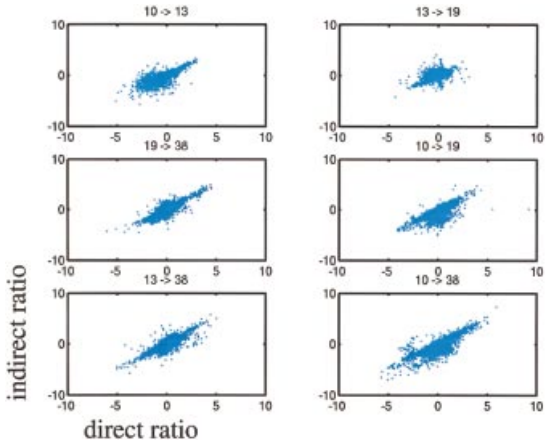

Figure 2. Indirect versus direct log2-ratios, plotted for each of the six possible sample pairs. Each spot represents a clone. Note that, under ideal conditions, indirect and direct ratios should be equal and all spots should be located on the 45° diagonal.

Table 1. Pearson correlation coefficients of direct and indirect ratios for all sample pairs and the mean of the direct and indirect ratios

Sample pair	Correlation coefficient	Mean of direct log2-ratios	Mean of indirect log2-ratios
10,13	0.80	–0.33	–0.29
13,19	0.44	–0.01	0.00
19,38	0.85	–0.08	–0.01
10,19	0.69	0.03	–0.29
13,38	0.83	0.01	–0.01
10,38	0.81	–0.02	–0.30

Differential expression values were categorized into 'up', 'down' and 'not-changed' (see Materials and Methods). Categories for indirect and direct values were compared for each clone and sample pair. Indirect values were indicated as 'good' if they were in the same category as the direct values. Accordingly, 82% of all values had good quality. The distribution of all values is shown in Figure 3. Note that for only very few of the non-matching cases (125 values, i.e. 0.7% of all values) indirect and direct values resulted in contradicting categories 'up' and 'down'.

Improved quality by filtering

To improve the quality of the data, we tested several signal validation techniques. Values were discarded with 10 different

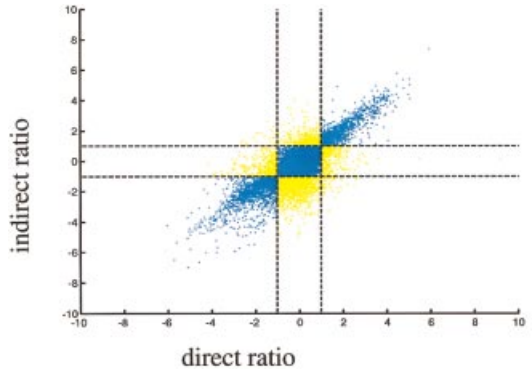

Figure 3. Indirect versus direct log2-ratios for all clones and sample pairs. Dark spots represent good quality values and light spots poor quality values according to our quality criterion (see Materials and Methods) Good quality values are grouped in three blocks corresponding to categories 'up' (upper, right), 'not-changed' (centred) and 'down' (lower, left).

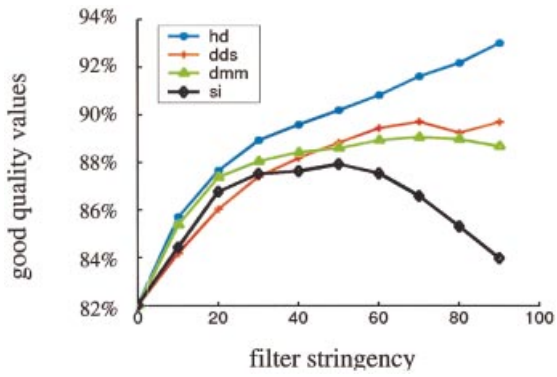

Figure 4. Percentages of good quality values are shown at different filtering stringencies. Here, a filter stringency of 0 corresponds to no filtering (i.e. all expression values are retained), and 80 corresponds to 80% of all values are discarded due to the applied signal validation technique. The signal validation technique that uses differences of repeated hybridizations (hd) performed best, as noted by a considerably steeper increase for stringencies larger than 30%. The other filters are based on differences of double spots (dds), differences between mean and median of the pixels for a spot (dmm) and signal intensities (si), respectively.

filtering stringencies, discarding 0%, i.e. no values filtered, up to 90% of all values (Fig. 4). Apparently, the percentage of good quality values increases similarly for all filters with a filter stringency up to 30%. For higher filter stringencies, however, the different filters show a distinctly different performance. The filter that uses differences of repeated hybridizations ('hd filter') performed best. It yielded 89% good quality values when discarding 30% of the values, and 93% good quality values when discarding 90% of all values. In contrast, the intensity filter ('si filter') showed a decreased performance when filtering >50%.

Apparently, this difference of filtering performance was caused by spots with saturated intensities. The intensity filter discarded only a few values of saturated spots. For all clones, 398 indirect ratios resulted from at least one saturated spot. The intensity filter eliminated only five of them at a filter stringency of 90%. In comparison, the hd filter (ddm and dds filter) reduced them to 35 (63 and 61, respectively) at the same filter stringency level.

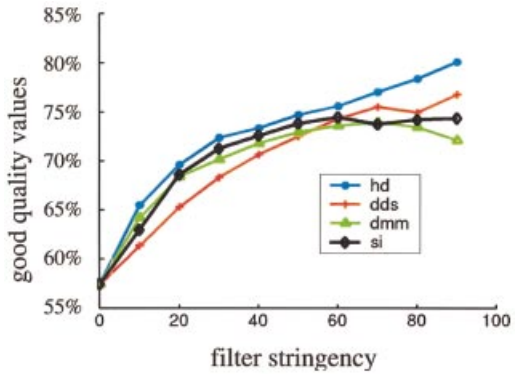

Figure 5. Performance of the filters when taking only expression values that showed changes in expression (only 'up' and 'down' categories, neglecting 'not-changed' values) Axes and labelling as in Figure 4.

Filtering differentially expressed genes during embryonic development

We were particularly interested in genes that *change* their expression during embryonic development. Therefore, we investigated how our filters performed for differential expression values of categories 'up' and 'down', disregarding the 'not-changed' category. The percentages of 'up' and 'down' good quality values were lower compared with all good quality values including the 'not-changed' category. However, filters could improve this ratio from 58% without filtering up to 82% when applying the hd filter at 90% stringency (Fig. 5). Interestingly, the intensity filter performed better for higher stringencies in the 'up' and 'down' categories compared with applying it to all values (compare Fig. 4). Apparently, the intensity filter discarded more values in the category 'not-changed', retaining more 'up' and 'down' values. At 90% stringency, the intensity filter retained 703 up/down values, in comparison with 352 for the hd filter. This higher amount of values compensated for ambiguous signals (e.g. of saturated spots).

Validation of results within the biological context

To assess our gene expression data with respect to biological context and relevance, we applied the best filter (hd filter) and discarded 30% of the values. Furthermore, we selected only those 261 clones that had good quality values for all sample pairs and showed differential expression for at least three sample pairs. These clones represent known genes as well as novel or uncharacterized genes. Among the known genes, some are key developmental regulators with a restricted differential expression pattern during embryonic development. Other clones showing a differential expression represented the elongation factor 1 alpha gene (*EF1α*). This gene is often termed 'housekeeping' because of its ubiquitous expression, but note that its transcription begins after mid-blastula transition (MBT, stage 8) and increases remarkably up to stage 12 (29). We validated our microarray results with gene expression patterns obtained by whole-mount *in situ* hybridizations. *In situ* hybridizations allow visualization of the spatial expression pattern of a gene. Qualitative changes of gene expression can be displayed among different embryo stages. The AxelDB database was used as the source of

Clone id	gene name	10 ->13	13 ->38
AGL_9G12	Xoct 25	-1	-1
726_19A22	*Cerberus*	-1	-1
AGL_25D01	*XmyoD*	1	0
AGL_16E03	*Xvent2*	0	-1
AGL_28F08	*Xpo*	1	-1
P76	*Chordin* 1	1	-1

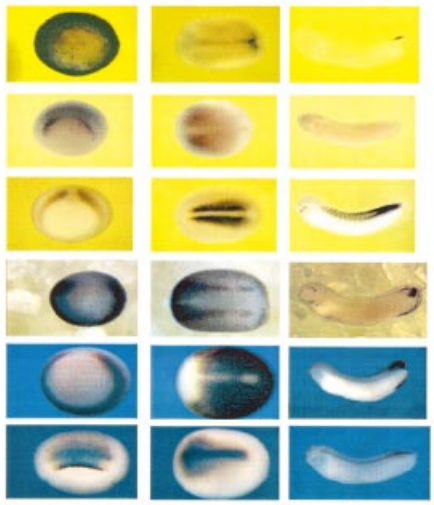

Figure 6. Comparison of differential expression obtained by our microarray study (columns 3 and 4) and *in situ* hybridizations from previous work (22), exemplified for six clones. Microarray-based differential expression is encoded by 1 for upregulated, –1 for downregulated and 0 for not-changed. For details see Results. Note that, for example, 10→13 specifies expressions of 13/10.

whole-mount *in situ* hybridization data (http://www.dkfz-heidelberg.de/molecular_embryology/axeldb.htm) (22). In AxelDB, embryos are presented at stages 10.5, 13 and 30, respectively. Stage 30 is early tailbud and stage 38 (used in our microarray experiment) is late tailbud stage. Between these two stages there are no major differences in expression of genes that are key regulators of early embryonic development.

The evaluation was done as exemplified in Figure 6. *In situ* hybridization images of clone AGL_9G12 show that the corresponding gene, *Xoct25*, is strongly expressed in almost the whole embryo at stage 10.5; the expression is reduced at stage 13 and even more at stage 30, when it is confined to the tip of the tailbud, a rather small area of the embryo. Our microarray data showed the same tendency: a decrease of expression (–1) from stage 10.5 to stage 13 (10.5→13) and a decrease (–1) from stage 13 to stage 38 (13→38). Note that the same direction is obtained considering all other combinations (Fig. 6): the expression level of *Xoct25* decreases (–1) from stage 13 to stage 19 (13→19) and remains 'not-changed' (0) from stage 19 to stage 38 (19→38). This corresponds to the fact that soon after stage 13, *Xoct25* expression drops to a minimum; therefore from stage 19 to stage 38 the expression remained 'not-changed'.

Figure 6 also shows the different behaviour of the two genes *Xoct25* and *Xoct91*. Interestingly, this is in perfect agreement with a previous analysis carried out with RNase protection, a technique considered to be one of the most sensitive for describing gene expression (30). The muscle marker gene *XmyoD* starts being expressed in muscle precursor cells at early gastrula (stage 10.5) and its expression increases with further muscle differentiation and development, remaining specifically expressed in the somites at late stages (stage 38). *Xpo* and *Chordin* expression increases from the beginning to the end of gastrulation and then it decreases in the tailbud stage. Also, *Cerberus* and *Xvent2* are expressed at tailbud stage at low levels but *Xvent2* is highly expressed in gastrula while *Cerberus* expression decreases following gastrulation.

For 174 (67%) out of 261 filtered clones in our study, we systematically compared their expression change during embryonic development with their expression as revealed by *in situ* images extracted from Axeldb. From these 174 clones, 62 (36%) were not suitable for our validation either because images for just one stage were found ($n = 39$) or because it was not possible to define the direction of their gene expression profile unambiguously ($n = 23$). Of the remaining 112 clones (representing 60 genes), 76 clones (35 genes) could be classified unequivocally (because all three embryonic stages were available in AxelDB), from which 71 clones (93%) were in agreement with the microarray data. Thirty clones were judged as 'most likely' confirming and six clones as 'most likely' not confirming the microarray results Here, 'most-likely' indicates that the evaluation was based on a comparison of *in situ* images where only two stages were available. In summary, of the 112 clones that were considered for validation, 101 (90%) confirmed our microarray results. Fifty of the 71 validated clones are listed in Table 2. Of the 261 filtered clones, 134 were singletons and the remaining 125 clones represented 34 genes. It can be seen that independent clones of the same gene typically show the same expression profile, further underlining the robustness of our filtering approach. For example, this was the case for the 29 *EF1*α clones.

DISCUSSION

Multiconditional experiments with microarrays have been rapidly established as a powerful tool for gene discovery, sample classification and studies on systems biology aspects.

Table 2. Clones that passed the hd filter (at 30% filter stringency) and quality criterion (same categories for indirect and direct ratios) were further selected either for being in agreement with known biological function or for matching expression patterns observed by whole-mount *in situ* hybridizations

Clone ID	Gene ID	Common name	Short description, expression	10→13	13→19	19→38	10→19	13→38	10→38
AGL_31C03	Xl743	Calreticulin CRP55 type	Epidermis	1	0	1	1	1	1
				1.08	0.24	1.3	1.75	1.42	2.79
				1.28	0.15	0.87	1.42	1.01	2.29
546_2L13	Xl91	Cardiac actin	Actin, alpha cardiac muscle (ALPHA 1)	0	1	1	1	1	1
				0.56	1.94	3.14	2.11	3.91	5.66
				0.85	1.93	2.57	2.78	4.49	5.35
726_19A22	Xl646	Cerberus	Head endoderm, organizer	–1	–1	0	–1	–1	–1
				–1.43	–1.42	–0.15	–3.66	–1.97	–2.92
				–1.61	–1.29	0.15	–2.9	–1.14	–2.75
726_21G12	Xl646	Cerberus	Head endoderm, organizer	–1	–1	0	–1	–1	–1
				–1.88	–1.69	0.02	–4.37	–2.01	–4.01
				–1.98	–2.01	–0.01	–3.99	–2.01	–3.99
726_19A22	Xl646	Cerberus	Head endoderm, organizer	–1	–1	0	–1	–1	–1
				–1.43	–1.42	–0.15	–3.66	–1.97	–2.92
				–1.61	–1.29	0.15	–2.9	–1.14	–2.75
P76	Xl391	Chordin	Organizer, notochord	1	0	–1	0	–1	–1
				1.17	–0.5	–1.99	0.53	–4.65	–2.57
				1.26	–0.88	–2.8	0.38	–3.69	–2.43
546A_1D01	Xl12	*Xgfpt*	Glutamine-fructose-6-phosphate transaminase	1	0	–1	1	–1	0
				2.05	–0.27	–1.99	1.72	–2.01	–0.17
				1.86	–0.26	–2.29	1.6	–2.55	–0.69
546_5D06	Xl159	Homologous to *btg1*	Somitogenic mesoderm	0	0	–1	0	–1	–1
				0.08	–0.03	–1.11	–0.08	–1.16	–1.03
				0.13	–0.08	–1.36	0.05	–1.44	–1.32
726_18I21	Xl502	Homologous to *btg2*	Somitogenic mesoderm	1	0	0	1	0	1
				1.05	0.84	–0.72	1.82	–0.01	1.03
				1.37	0.78	–0.94	2.15	–0.16	1.21
AGL_14C01	Xl502	Homologous to *btg2*	Somitogenic mesoderm	1	0	0	1	0	1
				1.78	0.57	–0.83	2.2	–0.09	1.9
				1.68	0.49	–0.41	2.17	0.09	1.77
546_5B12	Xl276	Homologous to low-affinity NGFR	Homologous to low-affinity nerve growth factor receptor	1	0	–1	1	–1	0
				1.76	0.02	–1.5	1.45	–1.03	–0.05
				1.38	–0.36	–1.22	1.02	–1.58	–0.2
546A_2I19	Xl54	Homologous to zonadhesin	Animal cap, epidermis	1	1	0	1	1	1
				3.11	1.16	0.95	3.01	1.94	4.14
				2.54	1.41	0.69	3.96	2.1	4.65
AGL_30E01	Xl684	Hyaluronan synthase related	Hyaluronan synthase related sequence protein	0	–1	–1	0	–1	–1
				0.77	–1.15	–2.51	–0.74	–3.95	–2.99
				0.49	–1.17	–3.18	–0.69	–4.35	–3.86
726_17L01	Xl713	Paraxial protocadherin	Organizer, paraxial mesoderm	1	0	–1	0	–1	–1
				1.28	0.08	–1.71	0.97	–1.79	–1.21
				1.01	–0.02	–2.04	1	–2.06	–1.04
AGL_30D02	Xl117	Pintallavis	Tailbud, organizer, chordoneural hinge	0	–1	–1	–1	–1	–1
				–0.42	–1.13	–1.09	–1.74	–2.47	–2.87
				–0.77	–1.06	–1.65	–1.83	–2.71	–3.47
546_4E20	Xl252	Retrotransposon-like element, 1A11	Retrotransposon-like element, 1A11,CCHC zinc finger type	0	0	–1	0	–1	–1
				–0.43	0.2	–1.01	–0.32	–1	–1.37
				–0.65	0.17	–1.22	–0.48	–1.05	–1.7
726_17J22	Xl709	Similar to Cerberus		–1	0	0	–1	0	–1
				–1.02	–0.21	0.7	–2.02	–0.17	–1.47
				–1.07	–0.51	0.5	–1.57	–0.01	–1.07
AGL_9A04	Xl488	*sox17a*	Transcription factor, endoderm, liver anlage	0	0	–1	0	–1	–1
				0.46	–0.12	–1.45	0.26	–2.46	–1.48
				0.48	–0.33	–1.46	0.15	–1.78	–1.31
726_21M07	Xl488	*sox17a*	Transcription factor, endoderm, liver anlage	0	0	–1	0	–1	–1
				0.82	–0.16	–1.98	0.4	–2.51	–1.33
				0.57	–0.2	–2.35	0.37	–2.55	–1.99
546_4M18	Xl6	*Xcad3*	Homeobox gene, posterior spinal cord	1	0	–1	0	–1	–1
				1.27	–0.24	–3.49	0.91	–2.39	–2.33
				1.02	–0.17	–2.96	0.84	–3.14	–2.12
546A_1B20	Xl6	*Xcad3*	Homeobox gene, posterior spinal cord	1	0	–1	0	–1	–1
				1.21	–0.14	–3.08	0.77	–2.08	–1.75
				1.1	–0.17	–2.51	0.93	–2.69	–1.59
24G2.1	Xcorglec	*Xcorglec*	Structural, epidermis	0	1	1	1	1	1
				0.31	1.98	2.28	3.35	4.47	5.09
				0.63	1.89	2.11	2.52	4	4.63
AGL_4C06	Xl407	*Xenf*	Early endodermal nuclear factor	–1	–1	0	–1	–1	–1
				–1.35	–1.49	–0.35	–3.35	–2.03	–4.09
				–1.52	–1.79	–0.35	–3.31	–2.14	–3.66

Table 2. *Continued*

Clone ID	Gene ID	Common name	Short description, expression	10→13	13→19	19→38	10→19	13→38	10→38
21G6.1	Xgs17	*Xgs17*	Unknown	–1	–1	0	–1	–1	–1
				–1.96	–1.3	–0.5	–2.75	–1.07	–3.99
				–2.23	–1.16	–0.8	–3.39	–1.97	–4.19
AGL_7C12	Xl37	*Xl37*	Unknown	–1	0	0	–1	0	–1
				–1.88	0.84	–0.5	–1.2	–0.05	–2.09
				–1.56	0.32	–0.44	–1.24	–0.13	–1.68
AGL_4C04	Xl405	*Xl405*	Unknown	0	0	–1	0	–1	–1
				0.25	–0.61	–2.62	–0.27	–3.23	–3.23
				0.21	–0.64	–2.71	–0.43	–3.35	–3.14
546_11P17	Xl405	*Xl405*	Unknown	0	0	–1	0	–1	–1
				0.46	–0.83	–2.15	–0.1	–3	–2.66
				0.36	–0.94	–2.45	–0.58	–3.39	–3.03
546_11G07	Xl571	*Xl571*	Unknown	0	0	–1	0	–1	–1
				0.94	–0.71	–1	0.08	–1.36	–1.3
				0.81	–0.56	–1.39	0.25	–1.95	–1.14
AGL_15E11	Xl786	*Xl786*	Unknown	0	0	1	1	1	1
				0.42	0.99	1.18	1.41	1.81	2.38
				0.77	0.48	1.29	1.24	1.76	2.53
726_21G14	Xl872	*Xl872*	Unknown	0	–1	0	–1	–1	–1
				–0.93	–1.94	–0.23	–2.37	–1.95	–3.04
				–0.99	–1.48	0.17	–2.47	–1.31	–2.3
AGL_25D01	Xl754	*XmyoD*	Myogenic factor, prospective	1	0	0	1	0	1
			somites of the early gastrula	1.3	0.41	–0.45	1.22	–0.46	1.12
				1.41	0.09	–0.19	1.49	–0.1	1.31
AGL_9G12	Xl408	*Xoct25*	Developmental regulator,	–1	–1	0	–1	–1	–1
			transcription factor, POU box	–1.25	–1.32	–0.42	–3.02	–2.42	–2.8
				–1.39	–1.97	–0.18	–3.37	–2.15	–3.54
AGL_1G10	Xl371	*Xoct91*	Developmental regulator,	0	–1	–1	0	–1	–1
			transcription factor, POU box	0.6	–1.15	–1.18	–0.67	–2.71	–2.14
				0.34	–1.23	–1.36	–0.89	–2.59	–2.25
546_9G10	Xl327	*Xom*	Developmental regulator,	0	0	–1	0	–1	–1
			homeobox gene, BMP4 group,	0	–0.22	–1.86	–0.32	–1.6	–2.46
			ventrolateral marginal zone	–0.42	–0.1	–1.81	–0.52	–1.92	–2.34
726_19B18	Xl327	*Xom*	Developmental regulator,	0	0	–1	0	–1	–1
			homeobox gene, BMP4 group,	–0.29	–0.35	–2.85	–0.31	–2.66	–3.71
			ventrolateral marginal zone	–0.01	–0.36	–2.81	–0.38	–3.17	–3.18
AGL_16E03	Xl798	*Xvent2*	Developmental regulator,	0	0	–1	0	–1	–1
			homeobox gene, BMP4 group,	–0.22	–0.09	–1.53	–0.41	–2.44	–2.27
			ventrolateral marginal zone	–0.44	–0.37	–1.89	–0.81	–2.26	–2.7
AGL_28F08	Xl52	*Xpo*	*Xenopus* zinc finger CCHC	1	0	–1	1	–1	–1
			posterior gene, ventrolateral	2.74	0.32	–3.57	2.59	–3.43	–1.31
			marginal zone, tailbud	2.47	0.43	–4.67	2.9	–4.24	–1.77
AGL_8F05	Xl52	*Xpo*	*Xenopus* zinc finger CCHC	1	0	–1	1	–1	–1
			posterior gene, ventrolateral	2.65	0.13	–4.44	3.42	–4.09	–1.64
			marginal zone, tailbud	2.63	0.32	–4.98	2.95	–4.67	–2.03
546A_2G21	Xl52	*Xpo*	*Xenopus* zinc finger CCHC	1	0	–1	1	–1	0
			posterior gene, ventrolateral	1.64	0.09	–1.54	1.34	–1.49	0.16
			marginal zone, tailbud	1.47	0.38	–2.26	1.85	–1.88	–0.41
726_21O16	Xl52	*Xpo*	*Xenopus* zinc finger CCHC	1	0	–1	1	–1	–1
			posterior gene, ventrolateral	2.71	0.43	–3.97	2.83	–3.15	–1.24
			marginal zone, tailbud	2.38	0.37	–3.96	2.75	–3.59	–1.22
726_17J15	Xl382	*EF1α*	Elongation factor 1 alpha chain	1	0	0	1	0	1
				1.95	0.23	0.33	2.66	0.57	2.73
				1.47	0.52	–0.01	1.99	0.51	1.98
726_21D21	Xl382	*EF1α*	Elongation factor 1 alpha chain	1	0	0	1	0	1
				1.96	0.18	0.47	2.35	0.56	2.85
				2.02	0.08	0.35	2.1	0.43	2.45
726_21E17	Xl382	*EF1α*	Elongation factor 1 alpha chain	1	0	0	1	0	1
				2.15	0.19	0.41	2.36	0.3	2.73
				2.04	0.12	0.19	2.16	0.31	2.35
726_21E24	Xl382	*EF1α*	Elongation factor 1 alpha chain	1	0	0	1	0	1
				2.13	0.32	0.37	2.51	0.7	2.73
				1.88	0.35	–0.09	2.23	0.26	2.14
726_21 H15	Xl382	*EF1α*	Elongation factor 1 alpha chain	1	0	0	1	0	1
				2.02	0.13	0.4	2.51	0.57	2.75
				1.68	0.19	0.22	1.88	0.42	2.1
726_21K24	Xl382	*EF1α*	Elongation factor 1 alpha chain	1	0	0	1	0	1
				1.91	0.39	0.32	2.5	0.54	2.77
				1.7	0.16	0.12	1.86	0.28	1.98

Table 2. *Continued*

Clone ID	Gene ID	Common name	Short description, expression	10→13	13→19	19→38	10→19	13→38	10→38
726_21N13	Xl382	*EF1α*	Elongation factor 1 alpha chain	1	0	0	1	0	1
				2.09	0.33	0.41	2.38	0.29	2.79
				1.91	0.1	0.02	2.01	0.12	2.03
726_21O12	Xl382	*EF1α*	Elongation factor 1 alpha chain	1	0	0	1	0	1
				2.13	0.35	0.46	2.26	0.63	2.78
				2.12	0.24	0.02	2.36	0.25	2.37
AGL_27C06	Xl382	*EF1α*	Elongation factor 1 alpha chain	1	0	0	1	0	1
				1.74	0.12	0.36	2.31	0.45	2.67
				1.42	0.38	–0.08	1.8	0.3	1.72

Of 71 clones, 29 represent the EF1α gene, of which nine are represented. The columns denote internal ID (AxelDB entries) (20), gene name, gene description and expression of sample pair (10.5, 13), (13, 19), (19, 38), (13, 38) and (10.5, 38). The rows show three values for each clone: the qualitative differential expression of the values (1, upregulated; –1, downregulated; 0, not-changed), and the direct and indirect quantitative log2-ratios. Note that, for example, 10→13 specifies expressions of 13/10.

Still, this high-throughput technology requires precise accuracy estimations. In a typical set-up, the analysis of multiconditional experiments relies on signal comparisons in more than one sample pair. A signal of a gene in a given sample pair may have to be compared with a signal of a gene in another sample pair. Such interconditional comparisons can only be performed after appropriate normalization. However, even after normalization not every interconditional comparison is reliable. A major source of noise is a considerable fluctuation of the measured signals. We addressed this problem by testing different quality criteria for expression ratios. These tests were applied after categorizing differential expression values into three different classes, namely upregulated, downregulated and not-changed. Although information on the exact quantitative expression values is thus lost, the further analysis and comparison of qualitative values is more robust and straightforward. Note that our data were taken from the rather complex system of a developing embryo, where many genes are expected to be differentially expressed. The reliability of a signal was defined by a triangle comparison: indirect ratios had to be in the same category as their direct counterparts to be regarded as good quality values. This approach can be regarded as an application of the transitivity condition (triangle inequality relation) for metric spaces: the error of an indirect path is greater than or equal to the error of a direct path.

We systematically assessed the accuracy and validation of experimental data in an all-pairs set-up, where we performed an all-against-all comparison of four conditions using different embryonic stages of *X.laevis*. Prior to filtering, 82% of all values fulfilled our quality criterion, i.e. roughly four-fifths of our expression profiles showed equal direct and indirect values when regarding the categories upregulated, downregulated and not-changed. After filtering with different stringencies, the percentage of good quality values could be increased up to 93% when discarding 90% of the data. This high stringency reflects the difficulty of filtering more than 90% of good quality values. Discarding only 30% increased the percentage of good quality clones up to 89%.

The signal validation technique using differences of the hybridizations (hd filtering) performed best. The increase in reliability was maximal when discarding one-third of all values.

We propose hd filtering for signal validation if triangle comparisons are not possible (non-all-pairs set-up). The hd filtering method requires repeated hybridizations. Our study suggests that if cost demands limiting the experiment to single hybridizations, the method using differences of mean and median signals should be used for filtering.

This filter also showed robust results. However, the use of this filter is restricted to only one hybridization and cannot correct for fluctuations due to differences between the slides that may additionally arise from varying hybridization conditions. Interestingly, the mean–median filter performed better for lower stringencies than filtering by differences of differential expression of double spots. Notably, the latter requires duplicate spots leading to an unfavourable less compact chip design. Compared with the si filter, these three filters performed better. This may be due to the fact that they effectively discarded values being influenced by saturation effects.

A critical question remains as to whether the expression profiles for those clones that passed the statistically motivated tests would correspond to their known biological function during embryogenesis. After hd filtering and selecting only those 261 clones that had good quality values for all sample pairs and showed differential expression, 90% of the microarray-based expression profiles were confirmed by whole-mount in situ hybridization data. This high percentage underlines the validity of our filtering approach.

Reliability of gene expression values depends on the applied normalization method. We used an elaborated method that was originally set up for one-channel hybridizations (6) and applied for radio-active dyes on membrane filters (31). More recently, it has also been successfully applied to two-dye cDNA microarray data (32). For comparison, within initial trials we tested the lo(w)ess-normalization as another commonly used normalization method (8, 9) (for implementation of the limma package, see www.bioconductor.org). Notably, our reliability criterion provides an elegant tool for comparing different normalization methods, by the difference of indirect and direct expression values. For the lo(w)ess normalization, the percentage of good values was 67% without filtering, rising to 71% when filtering 30% with the best filter (hd filter) and 74% when filtering 90%. Notably, the performance of the si filter was quite stable: it stayed between 69% and 70% for

stringencies of 30–90%. According to these findings, we did not obtain better results with lo(w)ess normalization than with the normalization we used. In this study, we only considered cDNA microarrays with two-sample hybridizations. However, the triangle principle may also be applied to one-sample set-ups such as Affymetrix oligochips or membrane filters with radioactive dyes if repeated hybridizations had been performed. In this case, differential expression ratios for possible cross-combinations may be calculated. These ratios may serve as input for the quality criterion by comparison of direct and indirect ratios.

In our study, direct and indirect values were similar on a qualitative basis. Accordingly, expression data of large-scale profiling experiments may be collected efficiently via a common reference to obtain robust information about the qualitative expression behaviour of an organism. It remains to be investigated in future studies whether quantitative expression ratios may further increase the reliability and accuracy of differential gene expression profiles.

In summary, our quality criterion assigned a reliability measure to gene expression ratios. This approach is particularly suited to the analysis of biological questions in developmental biology. The categorizing approach is straightforward and can be used in situations where clustering techniques cannot be directly applied as in very well defined cellular systems. By effectively discarding unreliable genes, genes with expression profiles reflecting their real biological function are retained that can be used for further analysis, such as classification and clustering.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We thank Daniel Göttel for his generous help with the micro-robot devices, and Gunnar Wrobel and Frank Diehl for suggestions and materials. We are grateful to Ursula Fenger for technical help and to Suresh Swaminathan and Rajeeb Swain for their suggestions during the preparation of the manuscript. Benedikt Brors was very helpful with critical comments on an earlier version of the manuscript. The work was funded by the Nationales Genom-Forschungsnetz (NGFN) and the Deutsche Forschungsgemeinschaft.

REFERENCES

1. Shoemaker,D.D. and Linsley,P.S. (2002) Recent developments in DNA microarrays. *Curr. Opin. Microbiol.*, **5**, 334–337.
2. DeRisi,J.L., Iyer,V.R. and Brown,P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**, 680–686.
3. Spellman,P.T., Sherlock,G., Zhang,M.Q., Iyer,V.R. Anders,K., Eisen,M.B., Brown,P.O., Botstein,D. and Futcher,B. (1998) Comprehensive identification of cell-cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell*, **9**, 3273–3297.
4. Gasch,A.P., Spellman,P.T., Kao,C.M., Carmel-Harel,O., Eisen,M.B., Storz,G., Botstein,D. and Brown,P.O. (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell*, **11**, 4241–4257.
5. Mangalathu,S.R., Vernon,S.D., Taysavang,N. and Unger,E.R. (2001) Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR. *J. Mol. Diagn.*, **3**, 26–31.
6. Beissbarth,T., Fellenberg,K., Brors,B., Arris-Prat,R., Boer,J.M., Hauser,N.C., Scheideler,M., Hoheisel,J.D., Schuetz,G., Poustka,A. *et al.* (2000) Processing and quality control of DNA array hybridization data. *Bioinformatics*, **16**, 1014–1022.
7. Zien,A., Aigner,T., Zimmer,R. and Lengauer,T. (2001) Centralization: a new method for the normalization of gene expression data. *Bioinformatics*, **17** (Suppl. 1), S323–S331.
8. Smyth,G.K., Yang,Y.-H. and Speed,T.P. (2003) Statistical issues in microarray data analysis. *Methods Mol. Biol.*, **224**, 111–136.
9. Yang,Y.H. and Speed,T.P. (2002) Design and analysis of comparative microarray experiments. In Speed,T.P. (ed.), *Statistical Analysis of Gene Expression Microarray Data*. CRC Press, Boca Raton, FL.
10. Jenssen,T., Langaas, Kuo,W.P., Smith-Sorensen,B., Myklebost,O. and Hovig,E. (2002) Analysis of repeatability in spotted cDNA microarrays. *Nucleic Acids Res.*, **30**, 3235–3244.
11. Tran,P.H., Peiffer D.A., Shin Y., Meek L.M., Brody J.P. and Cho K.W. (2002) Microarray optimizations: increasing spot accuracy and automated identification of true microarray signals. *Nucleic Acids Res.*, **30**, e54.
12. Wang,X., Ghosh,S. and Guo,S.W. (2001) Quantitative quality control in microarray image processing and data acquisition. *Nucleic Acids Res.*, **29**, e75.
13. Raffelsberger,W., Dembele,D., Neubeuer,M.G., Gottardis,M.M. and Gronemayer,H. (2002) Quality indicators increase the reliability of microarray data. *Genomics*, **80**, 385–394.
14. Tusher,V., Tibshirani,R. and Chu,G. (2002) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA*, **98**, 5116–5121.
15. Chu,G., Narasimhan,B., Tibshirani,R. and Tusher,V. (2002) Significance analysis of microarrays. Users guide and technical document. http://www.stanford.edu/~wanjen/sam.pdf.
16. Tibshirani,R., Hastie,T., Narasimhan,B. and Chu,G. (2002) Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc. Natl Acad. Sci. USA*, **99**, 6567–6572.
17. van't Veer,L.J., Dai,H., van de Vijver,M.J., He,Y.D., Hart,A.A.M., Mao,M., Peterse,H.L., van der Kooy,K., Marton,M.J., Witteveen,A.T. *et al.* (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, **415**, 530–536.
18. Alizadeh,A.A., Eisen,M.B., Davis,R.E., Ma,C., Lossos,I.S., Rosenwald,A., Boldrick,J.C., Sabet,H., Tran,T., Yu,X. *et al.* (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, **403**, 503–511.
19. Yang,Y.H. and Speed,T. (2002) Design issues for cDNA microarray experiments. *Nature Rev. Genet.*, **3**, 579–588.
20. Gilbert,S.F. (2003) *Developmental Biology* (7th edn). Sinauer Associates, Sunderland, MA.
21. Pollet,N., Schmidt,H.A., Gawantka,V., Vingron,M. and Niehrs,C. (2000) Axeldb: a *Xenopus laevis* database focusing on gene expression. *Nucleic Acids Res.*, **28**, 139–140.
22. Gawantka,V., Pollet,N., Delius,H., Vingron,M., Pfister,R., Nitsch,R. Blumenstock,C. and Niehrs,C. (1998) Gene expression screening in *Xenopus* identifies molecular pathways, predicts gene function and provides a global view of embryonic patterning. *Mech. Dev.*, **77**, 95–141.
23. Glinka,A., Delius,H., Blumenstock,C. and Niehrs,C. (1996) Combinatorial signaling by Xwnt-11 and Xnr3 in the organizer epithelium. *Mech. Dev.*, **60**, 221–231.
24. Diehl,F., Grahlmann,S., Beier,M. and Hoheisel,J.D (2001) Manufacturing DNA microarrays of high spot homogeneity and reduced background signal, *Nucleic Acids Res.*, **29**, e38.
25. Gawantka,V., Delius,H., Hirschfeld,K., Blumenstock,C. and Niehrs,C. (1995) Antagonizing the Spemann organizer: role of the homeobox gene Xvent-1. *EMBO J.*, **14**, 6268–6279.
26. Chomczynski,P. and Sacchi,N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.*, **162**, 156–159.
27. VanBogelen,R.A., Greis,K.D., Blumenthal,R.M., Tani,T.H. and Matthews,R.G. (1999) Mapping regulatory networks in microbial cells. *Trends Microbiol.*, **7**, 320–328.
28. Huber,W., von Heydebreck,A., Sültmann,H., Poustka,A. and Vingron,M (2002) Variance stabilization applied to microarray data calibration and

to the quantification of differential expression. *Bioinformatics*, **18** (Suppl. 1), S96–S104.
29. Johnson,AD and Krieg,PA. (1995) A *Xenopus laevis* gene encoding EF-1 alpha S, the somatic form of elongation factor 1 alpha: sequence, structure and identification of regulatory elements required for embryonic transcription. *Dev. Genet.*, **17**, 280–290.
30. Hinkley,C.S., Martin,J.F., Leibham,D. and Perry,M. (1992) Sequential expression of multiple POU proteins during amphibian early development. *Mol. Cell. Biol.*, **12**, 638–649.
31. Fellenberg,K., Hauser,N.C., Brors,B., Neutzner,A., Hoheisel,J.D. and Vingron,M. (2001) Correspondence analysis applied to microarray data. *Proc. Natl Acad. Sci. USA*, **98**, 10781–10786.
32. Wittig,R., Nessling,M., Will,R.D., Mollenhauer,J., Salowsky,R., Münstermann,E., Schick,M, Helmbach,H., Gschwendt,B., Korn,B. *et al*. (2002) Candidate genes for cross-resistance against DNA-damaging drugs. *Cancer Res.*, **62**, 6698–6705.