Available online at www.sciencedirect.com
SCIENCE DIRECT®
Analytical Biochemistry 346 (2005) 217–224
ANALYTICAL BIOCHEMISTRY
www.elsevier.com/locate/yabio

A quality-controlled microarray method for gene expression profiling

Thomas Degenkolbe [a], Matthew A. Hannah [a], Susanne Freund [a], Dirk K. Hincha [a,*], Arnd G. Heyer [b], Karin I. Köhl [a]

[a] *Max-Planck-Institut für Molekulare Pflanzenphysiologie, D-14424 Potsdam, Germany*
[b] *Biologisches Institut, Abt. Botanik, Universität Stuttgart, Pfaffenwaldring 57, D-70550 Stuttgart, Germany*

Received 31 May 2005
Available online 19 September 2005

Abstract

Gene expression profiling on microarrays is widely used to measure the expression of large numbers of genes in a single experiment. Because of the high cost of this method, feasible numbers of replicates are limited, thus impairing the power of statistical analysis. As a step toward reducing technically induced variation, we developed a procedure of sample preparation and analysis that minimizes the number of sample manipulation steps, introduces quality control before array hybridization, and allows recovery of the prepared mRNA for independent validation of results. Sample preparation is based on mRNA separation using oligo(dT) magnetic beads, which are subsequently used for first-strand cDNA synthesis on the beads. cDNA covalently bound to the magnetic beads is used as template for second-strand cDNA synthesis, leaving the intact mRNA in solution for further analysis. The quality of the synthesized cDNA can be assessed by quantitative polymerase chain reaction using 3′- and 5′-specific primer pairs for housekeeping genes such as glyceraldehyde-3-phosphate dehydrogenase. Second-strand cDNA is chemically labeled with fluorescent dyes to avoid dye bias in enzymatic labeling reactions. After hybridization of two differently labeled samples to microarray slides, arrays are scanned and images analyzed automatically with high reproducibility. Quantile-normalized data from five biological replica display a coefficient of variation $\leqslant 45\%$ for 90% of profiled genes, allowing detection of twofold changes with false positive and false negative rates of 10% each. We demonstrate successful application of the procedure for expression profiling in plant leaf tissue. However, the method could be easily adapted for samples from animal including human or from microbial origin.

Keywords: cDNA microarray; Transcript profiling; cDNA synthesis and labelling; Reduced technical variability; Statistical analysis of expression profiling data

The hybridization of nucleic acid microarrays has become a standard method to generate gene expression profiles. The major advantage over classical methods such as Northern blotting is the multiparallel analysis of expression for a large number of genes in one experiment (see [1–3] for reviews). Expression profiles are used to identify phenotypically influential candidate genes or to map metabolic or regulatory networks. In a candidate gene approach, samples from different treatments or different genotypes are compared to identify differentially expressed genes. Various strategies for candidate gene identification have been described such as the use of fixed fold change thresholds or statistical tests such as general-linear-model-based *t* tests [4–7]. The statistical test generates a *p* value that indicates the probability to falsely declare a significant change in gene expression (false positives; type I error). However, with increasing stringency of the tests ($p < 0.05$), the probability to overlook existing changes (false negatives; Type II error) increases to unacceptably high values. This problem results from the high variability of replicate data, reaching 100% for expression levels of a substantial portion of genes in independent replicates [8], and the low number of experimental replica usually applied. Theoretical power analysis for *t* tests indicates that more than 70% of the genes showing two fold up-regulation will be overlooked when the selection criterion $p < 0.05$ is employed on a dataset with a sample size of five and a CV of 100% [9].

* Corresponding author. Fax: +49 331 567 8250.
E-mail address: hincha@mpimp-golm.mpg.de (D.K. Hincha).

0003-2697/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2005.08.027

In many studies, the number of replicates is even lower than five (Stanford microarray database, TIGR Solanaceae database), further increasing the type II error rate. This weakens the applicability of microarray expression profiling especially in cases where the false negative rate is as critical as the false positive rate. This is the case, e.g., for network mapping, where the statement "no connection" cannot be made with confidence when the type II error is large.

One possibility to solve this problem is the application of statistical tests that are less sensitive to high noise and low replication than the *t* test, such as the PRIM method [9] or the widely used significance analyses of microarrays method [7]. In addition, experimental reproducibility can be optimized to reduce variation between replicates.

The variability of replicate microarray data is the sum of biological, technical, and so-called residual variability [10]. Biological variability of replicate samples originates from differences in gene expression in different biological samples, e.g., due to subtle differences in growth conditions. Technical variability is introduced during slide production, RNA isolation, cDNA synthesis and labeling, and image analysis. Mixed-model analysis of variance showed that technical variance can be as large as or even larger than the biological variance for microarray data [11,12]. Therefore, reducing technical variability has a major impact on the number of replicates required to detect a given change of gene expression.

The aim of the method presented in this paper is to reduce technical variability during the preparation of labeled cDNAs for hybridization by introducing quality control steps and by optimizing image analysis. We could decrease the sum of technical and biological variance to less than 45% in normalized data for 90% of the genes represented on the commercial tomato cDNA chip that we used for expression profiling.

Feasibility of the method is demonstrated by the identification of drought-regulated genes in potato leaves. In addition, expression profiling using tomato and tobacco leaf samples hybridized to the same microarray slides also yielded useful results (data not shown). For application of the method to animal, including human, or microbial samples, adaptation of the tissue homogenization and lysis procedures may be required. All following biochemical reactions would be the same and can be directly applied.

Another important aspect of this method is that the isolated mRNA is not lost during cDNA synthesis but can be recovered and used again for cDNA synthesis or for an independent confirmation of the results of expression profiling. In the case of very small or unique biological samples, this may be the only possibility for an independent confirmation of results.

Materials and methods

cDNA synthesis and labeling

Approximately 200 mg of potato leaf tissue was transferred into a 2-ml microcentrifuge tube and snap-frozen in liquid N_2. The sample was homogenized in a ball-mill (Retsch, Haan, Germany) for 90 s at 25 Hz. After adding 1.5 ml lysis buffer (100 mM Tris–HCl, 500 mM LiCl, 10 mM EDTA, 1% (w/v) lithium-dodecylsulfate (LiDS),[1] pH 8.0), the sample was homogenized for another 60 s. After centrifugation, mRNA was isolated from 750 µl of the supernatant using Dynabeads Oligo$(dT)_{25}$ (Dynal, Oslo, Norway) that had been equilibrated in lysis buffer. Samples were incubated with the beads for 5 min at room temperature with gentle mixing. The beads were separated from solution on a magnetic sample holder and washed twice with 1 ml wash buffer A (10 mM Tris–HCl, 150 mM LiCl, 1 mM EDTA, 0.1% LiDS, 0.05% Tween 20, pH 8.0). Beads were transferred into fresh microcentrifuge tubes using a magnetic PickPen (BioNobile, Turku, Finland) and then washed twice with 500 µl of wash buffer B (10 mM Tris–HCl, 150 mM LiCl, 1 mM EDTA, 0.05% Tween 20, pH 8.0) and once with 250 µl SuperScript II buffer (Invitrogen). Reverse transcriptase was added to the samples and they were incubated for 1 h at 42 °C. The beads were separated on a magnet and the supernatant was discarded. The beads were washed twice with 250 µl SuperScript II buffer with 0.05% Tween 20 (Invitrogen) with transfer of the beads into a fresh microcentrifuge tube using the PickPen. Subsequently, the mRNA was eluted with 20 µl of 10 mM Tris–HCl (pH 8.5) by incubation for 2 min at 95 °C. The beads were separated on a magnet and the supernatant containing the mRNA was transferred to a fresh tube. After repeating the elution procedure, the beads were washed twice with 250 µl TE (pH 8.5).

As a quality control of the synthesized cDNA we determined the 3′/5′ ratio by quantitative PCR using primers for GAPDH (GAPDH 5′ forward: 5′-AAGGACAAGGCTG CTGCTCAC; reverse: 5′-AACTCTGGCTTGTATTCAT TCTCG; GAPDH 3′ forward: 5′-TTCAACATCATCCCT AGCAGCACT; reverse: 5′-TAAGGTCGACAACAGAA ACATCAG). Two microliters of well-homogenized beads was used in a 25-µl volume PCR in a 5700 Sequence Detection System from Gene Amp, using the QuantiTect SYBR Green PCR kit (Qiagen, Hilden, Germany).

Second-strand cDNA synthesis was performed on the beads for 2.5 h at 37 °C using a hexanucleotide mix and Klenow polymerase from Roche. Subsequently, the beads were washed twice with 200 µl 2× SSC and the cDNA was eluted with 30 µl 10 mM Tris–HCl (pH 8.5) by incubation for 2 min at 95 °C. The supernatant containing the cDNA was transferred to a fresh tube and the elution was repeated once with 30 µl of fresh buffer. The beads could then be stored in TE containing 0.02% NaN_3 at 4 °C.

The cDNA from the combined elutions was purified using the Qiaquick PCR Purification Kit (Qiagen) following the manufacturer's instructions. The cDNA was eluted with labeling buffer (Ulysis Kit, Invitrogen). cDNA concentration was determined spectrophotometrically.

[1] *Abberviations used:* LiDS, lithium-dodecylsulfate; SSC, standard saline citrate; TE, Tris-EDTA; GAPDH, glyceraldehyde-3-phosphate dehydrogenase.

The volume of the samples was then reduced to 20 µl, using Microcon YM-30 columns (Millipore). cDNAs were labeled using either the Ulysis Alexa-Fluor 532 or the Ulysis Alexa-Fluor 647 Nucleic Acid Labeling Kit (Invitrogen) following the manufacturer's instructions. Samples from control and water-deficit-treated plants, labeled with different dyes, were combined, purified with the Qiaquick PCR Purification Kit, and concentrated to 4–5 µl using a Microcon YM-30 column.

Chip hybridization

For hybridization experiments, we used microarrays with cDNAs from tomato spotted on glass slides, obtained from the Center for Gene Expression Profiling (Boyce Thompson Institute, Cornell University, Ithaca, NY; http://bti.cornell.edu/CGEP/CGEP.html). The arrays contain approximately 12,000 cDNAs corresponding to 8500 different genes [13]. Slides were prehybridized for 45 min at 42 °C in hybridization buffer (0.1% SDS, 25% formamide, 5× SSC, 10 mg/ml bovine serum albumin fraction V, 40 mM NaP (pH 6.8)) and then washed successively in 2× SSC, 0.2× SSC, and distilled water at room temperature. Slides were dried by centrifugation at 1500 rpm for 3 min.

Concentrated samples were mixed with 50 µl of hybridization buffer at 56 °C. After addition of 2.5 µl Liquid Block (Amersham Bioscience) samples were denatured for 5 min at 95 °C. A dry, prehybridized slide was put into a hybridization chamber (Scienion, Berlin, Germany) and the sample was applied to the slide under a cover slip (LifterSlip; Erie Scientific, NH). The closed hybridization chamber was incubated overnight at 42 °C in the dark.

After hybridization, the slides were washed for 10 min each at 42 °C with 2× SSC, 0.2% SDS and at room temperature with 2× SSC and 0.2× SSC. The dried slides were scanned using an FLA-8000 laser scanner (Fuji) with excitation at 532 nm and emission at 570 nm for Alexa-Fluor-532-labeled cDNAs and excitation at 635 nm and emission at 675 nm for Alexa-Fluor-647-labeled cDNAs.

Image and data analysis

Image analysis was performed with the GeneSpotter software version 2.3 (MicroDiscovery, Berlin, Germany). To determine the reproducibility of image analysis, single image files were processed four times independently. For each spot of an individual image, the coefficient of variation (CV) for the estimated mean signal intensity for both dyes was determined.

The data were normalized either by division by the mean signal for the respective slide and dye (means normalization) or by quantile normalization [14] over all slides and both dyes. To determine the variance of expression in individual plants grown under the same condition in a single experiment, the coefficients of variation were calculated for the mean signal intensity of each individual spot on the slide for the raw, the means-normalized, and the quantile-normalized data. Ranks were calculated using the procedure Rank in SAS 8.2 (SAS institute, Cary, NC) with settings "percent" and "ties = high." The procedure assigns ranks to the values of the variable CV according to the size of the value. These values are then divided by the total number of observations. Ties are converted to the highest rank assigned to the tied numbers. Thus, the 592 lowest values (CV = 0%) are converted to the rank 0.12 and the highest value (CV = 59%) is converted to 1. Furthermore, induction factors for genes between drought-stressed and control conditions were calculated as mean expression ratios from the three replicate hybridizations based on raw, means-normalized or quantile- normalized data.

The effect of sample size on the power of statistical tests for different CVs was estimated using the sample size estimate procedure for a two-sample t test in SAS 8.2 with $\alpha = 0.1$, group 1 mean = 1, and group 2 mean = 1.5 or 2. The effect of CV on the minimally detectable expression ratio (threshold expression ratio) was iteratively estimated for a power of 0.9 and $\alpha = 0.1$ using the sample size estimate procedure for a two-sample t test. The number of genes in the first dataset (three replicates, one experiment) that had an induction or repression ratio larger than the threshold expression ratio for each CV was determined and normalized to the total number of genes on the slide.

For seven selected genes, expression values measured by slide hybridization were compared to those measured by quantitative RT-PCR (see above) on the same mRNA that was used for array probe synthesis by calculation of the Pearson correlation coefficients (SAS 8.2) for raw expression values from slide hybridization and C_t values from quantitative RT-PCR.

Plant material

Plantlets of *Solanum tuberosum* cv. Désirée were transplanted from axenic culture to pots filled with fertilized sand and placed in a climate chamber (16 h light, 150 µmol m^{-2} s^{-1}, 25 °C, 30% r.h.). For the first 7 days after planting, all plants were kept well watered. Then, for half of the plants the water supply was limited to impose a water-deficit that resulted in 50% growth reduction. After 4 weeks, leaf samples were taken from upper, fully developed leaves and immediately frozen in liquid nitrogen. Three samples from individual control and drought-stressed plants from one experiment were used for the profiling experiments to estimate expression ratios and coefficients of variation. For the test case experiment, samples from individual drought-stressed and control plants from five independent experiments were used for expression profiling.

Results and discussion

Probe synthesis and labeling procedure

The probe synthesis procedure is based on a method developed by Chen et al. [15] that allows mRNA isolation

directly from a crude homogenate (Fig. 1). This is much simpler and faster than conventional methods starting from the isolation of total RNA. Since the mRNA and, at later stages in the procedure, the first-strand cDNA is bound to magnetic beads, the washing and separation steps are easily performed using magnetic separators, thus reducing sample loss or variation introduced by liquid transfer steps.

Quality control for cDNA synthesis was performed by an analysis of the 3′–5′ ratio of the cDNA corresponding to GAPDH by quantitative PCR on an aliquot of the bead-bound cDNA (Fig. 1), because poor quality of cDNA can severely distort microarray results [16]. High quality of the cDNA synthesis is shown by a 3′–5′ ratio close to unity (Table 1), indicating full length of the cDNA strands. Although we never obtained 3′–5′ ratios larger than 1.3, this would have led to exclusion from further processing.

Following first-strand cDNA synthesis on the beads, mRNA was recovered after thermal denaturation and stored for later validation experiments. We tested the quality of a second round of cDNA synthesis from this mRNA by determining the 3′–5′ ratio of the cDNA corresponding to GAPDH by quantitative PCR as described above and found ratios almost identical to those reported in Table 1 (data not shown). Also, we have used this mRNA successfully in several cases for an independent confirmation of array data by quantitative RT-PCR (see below). We conclude from these results that cDNA quality is not compromized in a repeated cDNA synthesis step.

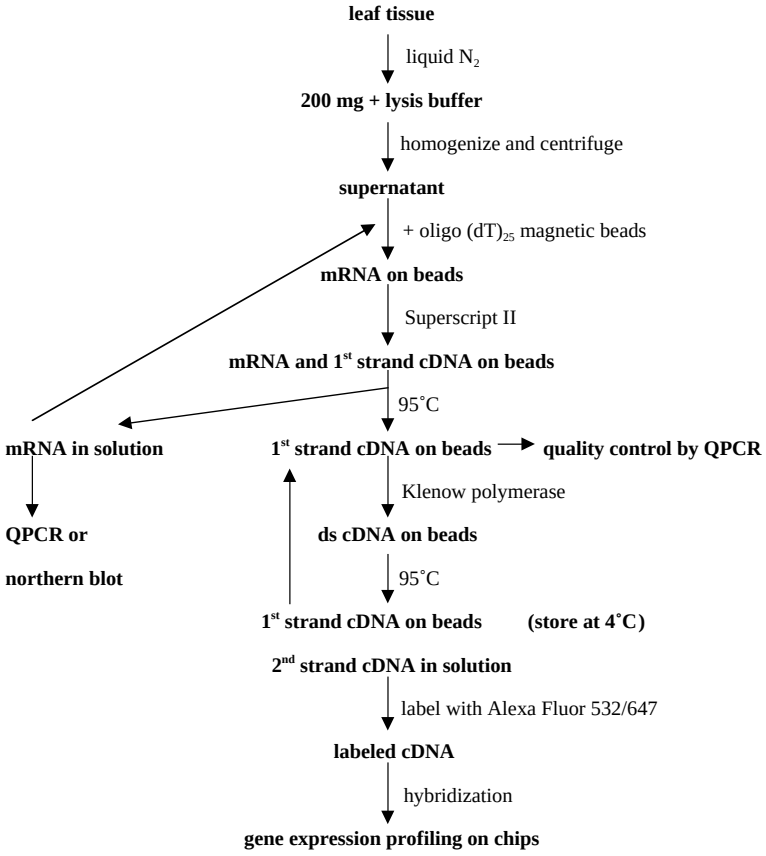

Fig. 1. Flow chart for sample processing. The figure summarizes the work flow from tissue sampling to chip hybridization. In addition to the linear work flow, possible loops, and control points are indicated. See the text for experimental details.

Table 1
Quality control of first-strand cDNA synthesis by RT-qPCR

Plant treatment	C_t value from 3′-specific primers	C_t value from 5′-specific primers	3′/5′ ratio
Control	18.97 ± 0.77	18.86 ± 1.04	1.01
	18.32 ± 0.87	18.66 ± 1.04	0.98
	18.68 ± 0.89	18.88 ± 0.90	0.99
Drought stress	18.49 ± 0.87	18.77 ± 1.06	0.99
	18.73 ± 1.03	19.06 ± 1.19	0.98
	18.84 ± 0.74	18.70 ± 1.25	1.01

Primers specific for the 3′ and 5′ ends of the GAPDH gene, were used to assess completeness of cDNA synthesis. Samples were generated from three independent control and water-deficit- treated plants each, and four parallel measurements were taken from each sample. Data are the mean ± SD from these four measurements. Ratios were calculated from the mean values.

First-strand cDNA covalently bound to the oligos$(dT)_{25}$ magnetic beads was used as template for second-strand synthesis (Fig. 1). The amount of second-strand cDNA needed for labeling reactions ranged between 0.7 and 1.5 μg. When this amount is not reached in one reaction, second-strand synthesis can be repeated on the same beads without need for additional RNA or biological sample material. While we showed in separate experiments that this is possible (data not shown), we did not use this option in the experiments reported below, as cDNA yield was always sufficient with the first reaction.

After second-strand cDNA synthesis, first-strand cDNA on beads was stored at 4 °C and could be used in a new second-strand synthesis at a later time in case the hybridization failed (data not shown).

The second-strand cDNA was chemically labeled with fluorescent dyes for hybridization on microarrays. We used a labeling procedure that did not involve dye-coupled nucleotides but rather used postsynthesis labeling of the cDNA to avoid dye bias introduced by preferential incorporation of differentially modified nucleotides during enzymatic reactions [17]. Additionally, we used Alexa dyes instead of the more commonly used Cy dyes because of their superior physical properties [18], especially their higher photostability [19].

Reproducibility of the image analysis

Image analysis identifies the location of each spot on the image of the hybridized slide, sets a circular detection area around the respective spot, and converts the pixel intensities within this area into a number by calculating the mean or median signal intensity. All three steps are potential sources of error. To estimate the magnitude of technical variation introduced by image analysis, the mean signal intensities for individual spots of a Tomato CGEP array were calculated repeatedly with the GeneSpotter software. This software identifies the spot location automatically and estimates the signal intensities for fixed spot diameters.

The software assigns quality measures to each spot that indicate whether the signal is likely to be "below background," "artifact," "overshined," or "poorly shaped." On the slide image chosen for the repeated analysis, 59% of the spots were estimated to be above background level in all four image analyses ($p = 0.05$), whereas 34% were estimated to be below background level for all readings. The remaining 7% were flagged to be absent in one to three of the four image analyses. This distribution was similar to other images of expression profiles for potato leaf material. We thus conclude that the chosen slide was representative, and the results could be generalized to all our experiments.

We calculated CV for the mean signal intensities for four independent image analyses of the same slide and plotted these values against the relative rank of the CV to estimate the reproducibility of all measurements from individual spots with more than two readings (Fig. 2). This differs from the commonly used method of plotting intensities of two samples against each other and performing a correlation analysis. Depicting CV against its percentage rank has the advantage that data from more than two samples can be analyzed together and variance parameters needed for sample size estimation can be derived easily (see below). The CV of rank 50% is the median, and the CV of rank 90% is the 90th percentile (P90) of the variance distribution. In our image analysis, the signal intensities for 50% of all cDNAs on the slide had a CV lower than 1 and 90% had a CV lower than 6% (Fig. 2). For 12% of all spots, no variation (CV = 0%, rank = 0.12) was detected. Spots that were always estimated "above background" or "below background" had even lower median (0.4% in both cases) and P90 (4.4 and 2.7%, respectively) values than the overall distribution. This indicates that high reproducibility of image analysis is independent of the signal-to-noise ratio.

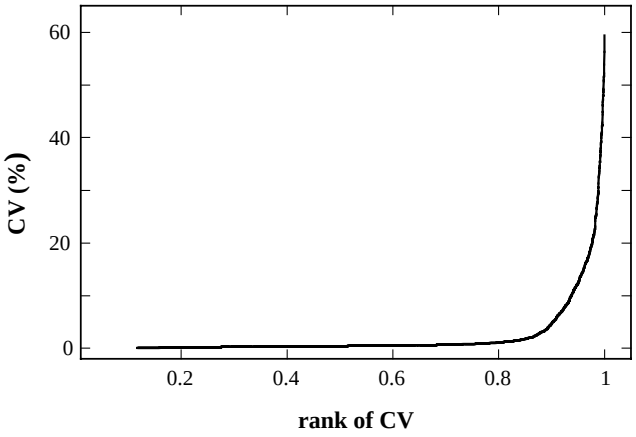

Fig. 2. Reproducibility of image analysis. The coefficient of variation (CV) of the mean signal intensity for each spot estimated by the image analysis program GeneSpotter was determined for four readings of an individual slide and plotted against the rank of the CV.

The GeneSpotter software uses a default setting of the spot diameter of 7.69 pixels. In addition, we used a value of 5.96 pixels, to estimate the influence of spot diameter on CV on an independent slide. The lower spot diameter reduced the median of the mean signal CV from 1 to 0.1% and the P90 from 7.6 to 6.4% for this slide. Also, the overall percentage of spots flagged as low quality was reduced when the smaller spot diameter setting was used.

We conclude from our tests that fluorescence intensities for most spots can be determined with high confidence. Nevertheless, a small fraction of the spots showed a large technical variance from the image analysis. This indicates that the contribution of the image analysis procedure to the overall CV of expression profiling data can be significant and should always be analyzed and optimized for any experimental protocol.

Reproducibility of biological replicates

Variability between expression profiles derived from samples belonging to the same treatment group can originate from unintended genetic and environmental

differences between plants and from technical variation as discussed above. The sum of these components of variation determines the number of replicates needed to detect a given difference between treatment groups. Thus, the magnitude of variation needs to be known to properly design experiments. To reduce experimental costs, the amount of variation should be reduced as far as possible.

To determine the necessary sample size to detect a two-fold change in gene expression, we assessed variance by processing three samples from independent, genetically identical plants from a potato clone cultivated either under standard or water-deficit conditions in one experiment. To show the variance distribution, the CV of signal intensities of the individual spots on the arrays is depicted against the relative rank of the CV for raw and quantile-normalized data (Fig. 3). To facilitate comparison, the median and P90 of the distribution for raw and normalized data are given in Table 2.

Without normalization, signal intensities from 90% of the spots on the arrays could be determined with a CV of less than 48% in samples from control plants and less than 59% in drought-stressed plants. Quantile normalization resulted in a P90 of 45 and 46% for control and drought-stressed plants, respectively, thus decreasing variance and increasing homogeneity of variance. Means-normalized data showed an intermediate distribution with higher median but similar P90 in comparison to the quantile-normalized data.

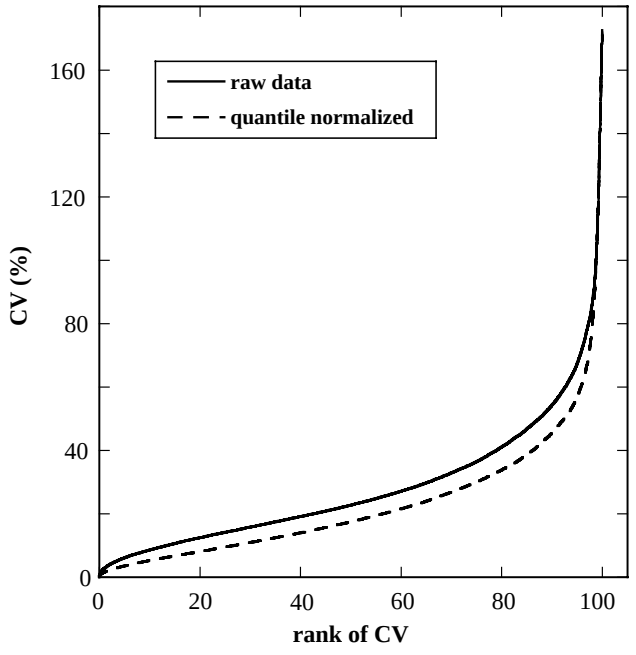

Fig. 3. Reproducibility of biological replicates. Coefficient of variation (CV) for all cDNAs spotted on the array based on raw data or quantile-normalized mean fluorescence values plotted against the relative rank of the CV. CV were estimated from raw or quantile- normalized data derived from three independent, individual plant samples cultivated under water-stress conditions and from data derived from plants cultivated under control conditions. CV data from both treatments were then joined in one dataset before calculating the ranks.

Table 2
Median and P90 of the coefficients of variation (CV) of the fluorescence signal intensity

Normalization	Treatment	Median	P90
None	Control	19.6	48.4
	Drought	26.1	58.8
Mean	Control	20.2	47.0
	Drought	26.2	46.8
Quantile	Control	17.7	45.0
	Drought	16.9	46.0

Leaf samples from control ($n = 3$) or from drought-stressed ($n = 3$) *Solanum tuberosum* plants were hybridized to CGEP tomato cDNA arrays. The CV was calculated for each spotted cDNA and each treatment based on the raw data after image analysis (normalization = none), after means normalization (mean), or after quantile normalization.

Based on this variance estimate, a power analysis was performed for a two-sided *t* test (Fig. 4). This analysis revealed that a sample size of five would allow acceptable control of both type I and type II errors. With a CV of 45% a sample size of five allows the detection of a 2.0-fold change in gene expression with a power of 90%, and a type I

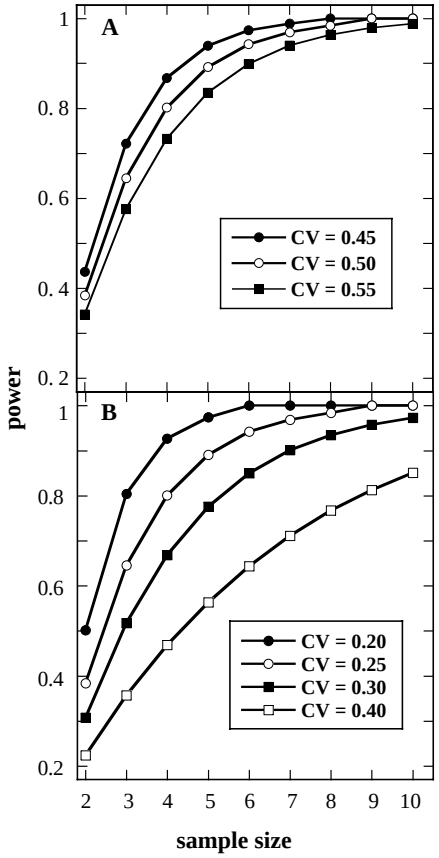

Fig. 4. Theoretical power analysis for a two-sided *t* test to detect a 2.0-fold (A) or a 1.5-fold (B) change in gene expression as a function of sample size. Indicated coefficients of variation (CV) and a type I error (false positive rate) of 0.1 were used as input values to determine the fraction of changes in gene expression that would be detected at a given number of replica. A power of 1.0 denotes a type II error (false negative rate) of zero, i.e., 100% of all changes that occurred are detected.

Table 3
Threshold expression ratio for the given coefficient of variation with type I and type II errors of 0.1 in a two-sided *t* test

CV (%)	Minimum detectable expression ratio	% of all genes
40	1.82	1.99
45	1.92	1.40
50	2.03	1.04
55	2.13	0.86
60	2.22	0.72

The percentage of all genes that exceed the minimum expression ratio is given for a dataset from leaves of drought-stressed and control plants grown in a single experiment.

Table 4
Number of genes that showed at least the indicated change in expression at a *p* value of less than 0.1 or 0.05 when gene expression levels of drought-stressed and control plants were compared

Change in gene expression	Number of genes at $p < 0.1$	Number of genes at $p < 0.05$
1.5-fold	179	111
1.7-fold	67	39
2.0-fold	16	9
2.2-fold	13	7

error rate of 10%. The detection of 1.5-fold changes with a type I error of 0.1, a power of 90%, and a sample size of five requires the CV to be less than 21%. Our optimized method resulted in a CV smaller than 45 or 21% in 90 or 59% of the genes, respectively. For these genes, a 2.0-fold (1.5-fold) induction or repression could be detected with a likelihood of more than 90% from a sample size of five, which means a type II error of less then 10%. In the remaining 10% (41%) of the genes, the CV is too high to allow detection of a 2.0-fold (1.5-fold) change in expression based on five replica with this error rate.

The reduction of the CV by optimization of the hybridization method has a marked effect on the number of genes that can be detected as significantly different by ANOVA based methods (Table 3). A reduction of the CV from 60 to 45% decreases the detection limit from 2.2-fold to 1.9-fold changes. In our dataset on leaf RNA from moderately drought- stressed and control potato plants grown in a single experiment, the number of genes exceeding the 1.9-fold threshold is almost twice as high as the number of genes exceeding the 2.2-fold threshold.

Effect of drought stress on gene expression in potato leaves as a test case

Using the method characterized above, we searched for genes that are induced by a moderate water-deficit treatment in potato. Based on the calculations described above, we used data from five leaf samples each from drought-stressed and control plants that were grown in five independent experiments. This dataset was completely independent of the dataset that was used to estimate CV. We filtered for potential candidates using three criteria, namely quality score of the spots on the slide, significance level in an *F* test, and induction or repression factor. Genes represented by spots which received a bad quality flag in more than three of five replica were excluded from the analysis. The importance of including spot quality information in the data evaluation has been shown previously [20]. Since, we used a mild water-stress, we did not expect to see massive changes in gene expression. Therefore, a cutoff of 1.5-fold induction or repression was used for filtering the data. Among the genes that passed the quality control, 179 showed an at least 1.5-fold change in expression and a *p* value of less than 0.1 when gene expression levels of drought-stressed and control plants were compared. A higher cutoff of at least 2.0-fold change decreases the number of detected genes by more than 60% (Table 4). Within the group of significantly induced genes were several that had previously been described as stress inducible. One of these was the dehydrin TAS 14, which belongs to the group of LEA proteins and is induced in tomato by salt and osmotic stress [21]. TAS 14 was induced 1.7-fold ($p = 0.014$) in leaves of drought-stressed potato plants and would thus have been overlooked in a 2.0-fold change cutoff approach of data analysis. For seven selected genes, expression measured by the slide hybridization method was followed by quantitative RT-PCR measurements using the mRNA samples generated by our probe synthesis procedure (Fig. 1). In five of seven genes, expression values of both methods correlated significantly ($p \leqslant 0.05$).

Conclusion

We have demonstrated an improved experimental procedure for the preparation of labeled cDNAs for array hybridization. We believe that this procedure is superior to currently used protocols, because it minimizes technical variability and offers additional safety and control options not available in other protocols. In addition, we have analyzed technical variability due to the scanning procedure and shown how to minimize it by adjusting the spot diameter. Using data obtained from potato leaf tissue samples, we present evidence for the crucial importance of achieving low CV values in array hybridization experiments. We provide quantitative data on the effect of CV on gene expression profiling results that support our approach of minimizing data variability at all experimental levels to optimize the yield of useful data from such expensive experiments.

Acknowledgments

We thank Mirjam Müller for skillful technical assistance. This research was financially supported by the BMBF (German Federal Ministry of Education and Research) with Grant No. 032632C. The responsibility for the content of this publication lies with the authors.

References

[1] D. Baldwin, V. Crane, D. Rice, A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants, Curr. Opin. Plant Biol. 2 (1999) 96–103.

[2] P.O. Brown, D. Botstein, Exploring the new world of the genome with DNA microarrays, Nat. Genet. 21 (1999) 33–37.

[3] D.J. Duggan, M. Bittner, Y. Chen, P. Meltzer, J.M. Trent, Expression profiling using cDNA microarrays, Nat. Genet. 21 (1999) 10–14.

[4] D. Finkelstein, R. Ewing, J. Gollub, F. Sterky, J. Cherry, S. Somerville, Microarray data quality analysis, lessons from the AFGC project, Plant Mol. Biol. 48 (2002) 119–131.

[5] W. Jin, R.M. Riley, R.D. Wolfinger, K.P. White, G. Passador, G.G. Gibson, The contribution of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*, Nat. Genet. 29 (2001) 389–395.

[6] J. Quackenbush, Microarray data normalization and transformation, Nat. Genet. 32 (2002) 496–501.

[7] V.R. Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing radiation response, Proc. Natl. Acad. Sci. USA 98 (2001) 5116–5121.

[8] B. Winer, Statistical Priciples in Experimental Design, McGraw-Hill, New York, 1971.

[9] S.W. Cole, G. Zoran, J.A. Zack, Controlling false-negative errors in microarray differential expression analysis: A PRIM approach, Bioinformatics 19 (2003) 1808–1816.

[10] G.A. Churchill, Fundamentals of experimental design for cDNA microarrays, Nat. Genet. 32S (2002) 490–495.

[11] J.J. Chen, R.R. Delongchamp, C.A. Tsai, F. Sistare, H.M. Hsueh, K.L. Thompson, V.G. Desai, J.C. Fuscoe, Analysis of variance components in gene expression data, Bioinformatics 20 (2004) 1436–1446.

[12] X. Cui, G.A. Churchill, How many mice and how many arrays? Replication in mouse cDNA microarray experiments, CAMDA 2005 (2003).

[13] R. Alba, Z. Fei, P. Payton, Y. Liu, S.L. Moore, P. Debbie, J. Cohn, M. D'Ascenzo, J.S. Gordon, J.K.C. Rose, G. Martin, S.D. Tanksley, M. Bouzayen, M.M. Jahn, J. Giovannoni, ESTs, cDNA microarrays, and gene expression profiling: Tools for dissecting plant physiology and development, Plant J. 39 (2004) 697–714.

[14] B.M. Bolstad, R.A. Irizarry, M. Astrand, T.P. Speed, A comparison of normalization methods for high density oligonucleotide array data based on variance and bias, Bioinformatics 19 (2003) 185–193.

[15] I. Chen, U. Haehnel, L. Altschmied, I. Schubert, H. Puchta, The transcriptional response of *Arabidopsis* to genotoxic stress - a high-density colony array study (HDCA), Plant J. 35 (2003) 771–786.

[16] H. Auer, S. Lyianarachchi, D. Newsom, M.I. Klisovic, G. Marcucci, K. Kornacker, Chipping away at the chip bias: RNA degradation in microarray analysis, Nat. Genet. 35 (2003) 292–293.

[17] Y.H. Yang, S. Dudoit, P. Luu, D.M. Lin, V. Peng, J. Ngai, T.P. Speed, Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation, Nucleic Acids Res. 30 (2002) e15.

[18] W.G. Cox, M.P. Beaudet, J.Y. Agnew, J.L. Ruth, Possible sources of dye-related signal correlation bias in two-color DNA microarray assays, Anal. Biochem. 331 (2004) 243–254.

[19] J.E. Berlier, A. Rothe, G. Buller, J. Bradford, D.R. Gray, B.J. Filanoski, W.G. Telford, S. Yue, J. Liu, C.-Y. Cheung, W. Chang, J.D. Hirsch, J.M. Beechem, R.P. Haugland, R.P. Haugland, Quantitative comparison of long-wavelength Alexa Fluor dyes to Cy dyes: Fluorescence of the dyes and their bioconjugates, J. Histochem. Cytochem. 51 (2003) 1699–1712.

[20] X. Wang, S. Ghosh, S.W. Guo, Quantitative quality control in microarray image processing and data acquisition, Nucleic Acids Res. 29 (2001) e75.

[21] J.M. Godoy, D.A. Waner, T. Horie, S.L. Li, R. Horie, K.B. Abid, J.I. Schroeder, A tomato cDNA inducible by salt stress and abscisic acid: Nucleotide sequence and expression pattern, Plant Mol. Biol. 15 (1990) 695–705.