

Intensity-based analysis of two-colour microarrays enables efficient and flexible hybridization designs

Peter A. C. 't Hoen^{1,*}, Rolf Turk¹, Judith M. Boer¹, Ellen Sterrenburg¹,
Renée X. de Menezes³, Gert-Jan B. van Ommen¹ and Johan T. den Dunnen^{1,2}

¹Center for Human and Clinical Genetics, ²Leiden Genome Technology Center and ³Department of Medical Statistics, Leiden University Medical Center, Wassenaarseweg 72, 2333 AL Leiden, The Netherlands

Received December 17, 2003; Revised January 30, 2004; Accepted January 31, 2004

ABSTRACT

In two-colour microarrays, the ratio of signal intensities of two co-hybridized samples is used as a relative measure of gene expression. Ratio-based analysis becomes complicated and inefficient in multi-class comparisons. We therefore investigated the validity of an intensity-based analysis procedure. To this end, two different cRNA targets were hybridized together, separately, with a common reference and in a self-self fashion on spotted 65mer oligonucleotide microarrays. We found that the signal intensity of the cRNA targets was not influenced by the presence of a target labelled in the opposite colour. This indicates that targets do not compete for binding sites on the array, which is essential for intensity-based analysis. It is demonstrated that, for good-quality arrays, the correlation of signal intensity measurements between the different hybridization designs is high ($R > 0.9$). Furthermore, ratio calculations from ratio- and intensity-based analyses correlated well ($R > 0.8$). Based on these results, we advocate the use of separate intensities rather than ratios in the analysis of two-colour long-oligonucleotide microarrays. Intensity-based analysis makes microarray experiments more efficient and more flexible: It allows for direct comparisons between all hybridized samples, while circumventing the need for a reference sample that occupies half of the hybridization capacity.

INTRODUCTION

DNA microarrays are widely used to measure genome-wide changes in mRNA expression levels across conditions such as developmental stages, disease states, drug treatment and gene disruption (1–5). Affymetrix GeneChips, prepared by photolithography, and spotted cDNA and 50–70mer oligonucleotide microarrays are currently the most frequently used platforms. The GeneChip is a one-colour system based on the immunofluorescent detection of biotinylated nucleic acids. The

difference in perfect and mismatch probe intensities is used for gene expression measurements (6). Spotted microarrays are commonly hybridized with two samples labelled with two different fluorophores. For these arrays, the ratio of the signal intensities in the two channels is a relative measure of gene expression.

Normalization is essential to remove systematic biases in microarray data. For two-colour arrays, normalization algorithms can be applied to (log-transformed) ratios (7) (e.g. using a LOWESS algorithm). Alternatively, ANOVA models that account for array, dye and spot effects can be applied to the individual signal intensities on all the arrays (8). In both cases, after normalization, the ratio of the co-hybridized samples is usually calculated to minimize the influence of spatial variation in spot morphology and hybridization efficiency on the experimental outcome. Furthermore, some suggest that ratio-based analysis is important because of possible competitive hybridization of the two targets due to saturation of binding sites on the array (9).

Ratio-based analysis can be applied to experiments with a reference or loop design (10,11). A disadvantage of the reference design is that half of the acquired data represent only one sample that is often not biologically relevant, thereby doubling the number of arrays required (10,11). A loop design has other disadvantages (11). The calculated ratios have variable levels of precision since some samples are more directly related than others, and the set of hybridizations cannot be extended. This has important implications for studies in which not all samples become available at the same time; new samples could only be included in the experiment via forming new subloops, and only if biological material from the earlier samples is still available.

An intensity-based analysis in which the signal intensities in the two channels are kept separately, also after normalization, would allow for hybridization designs that are more efficient than the reference design and more flexible than the loop design. We designed a set of experiments to determine whether an intensity-based analysis would be justified for our spotted long-oligonucleotide microarrays. Our aims are 2-fold: first, to investigate whether hybridization patterns are sufficiently uniform across arrays; secondly, to verify if there is evidence for competition between targets for binding sites

*To whom correspondence should be addressed. Tel: +31 71 527 6611; Fax: +31 71 527 6075; Email: p.a.c.hoen@lumc.nl

on the array. We run two parallel statistical analyses, one ratio-based and the other one intensity-based, and compare their results.

MATERIALS AND METHODS

Microarray and target preparation

Murine oligonucleotide microarrays were produced in the Leiden Genome Technology Center by spotting the Sigma-Genosys mouse 7.5K oligonucleotide library (v. 1.0) (65mer, 20 μ l in 50% DMSO) in duplicate on poly-L-lysine-coated slides (12). RNA was isolated from hind limb muscles of 20-week-old (sample A) and 8-week-old (sample B) C57Bl/10ScSn-DMD^{mdx/J} (*mdx*) mice (R. Turk, E. Sterrenberg, E. de Meijer, G.J.B. van Oumen, J.T. den Dunnen and P.A.C. 't Hoen, in preparation). A reference RNA was created by pooling RNA from the following mouse tissues: muscle, testis, kidney, liver, brain, heart, spleen, ovary, uterus and whole embryos. The RNA was amplified by T7-polymerase-driven linear amplification and labelled through incorporation of aminoallyl-UTP and subsequent coupling to Cy3 or Cy5 monoreactive dyes, as described previously (12). Microarrays were hybridized with 1.5 μ g of the indicated (Table 1) cRNA targets. Hybridization and washing was done in a GeneTAC hybstation (Genomic Solutions) (12).

Feature extraction and data analysis

Feature extraction was performed with GenePix 3.0 software (Axon Instruments Inc.). Spots with intensities lower than background or aberrant spot shape were flagged by the software and checked manually. Only spots that were not flagged on any of the analysed arrays were taken into account in further analyses, leaving 2224 data points per array. Local background-subtracted median signal intensities were used as intensity measures. Scaled gene expression ratios in samples A and B were calculated after transformation (natural logarithm) of the background-corrected intensities and subtraction of the average of the LN-transformed intensities (linear scaling).

Statistical analysis

SPSS (version 10.0.7) was used for standard statistical tests. To determine the significance level in comparisons of groups of Pearson correlation coefficients (r), the r values were first transformed using the formula

$$r(N-1)^{1/2}/(1-r^2)^{1/2}.$$

This was done to normalize the distribution of the correlation coefficients (13). The significance level used in all tests was 0.05.

Data analysis with MAANOVA (v2.0) was performed using Matlab (6.5R13, The MathWorks Inc). Using the MAANOVA package (www.jax.org/staff/churchill/labsite/software/anova/), we have fitted linear fixed-effects models, taking dye and average gene effects into account. The hyb-design was used as the test variable. The array effect could not be fitted because it is confounded with the hyb-design. Two models were compared: the null model where no effect due to hyb-design is assumed, and the alternative model in which an effect of hyb-design is expected. Hyb-design-specific p values were obtained via F-statistics for the individual features of

Table 1. Overview of used Hyb-designs and ratio calculations

Hyb Design	Array	Cy3	Cy5	Ratio
CoHyb	1	A	B	R1
	2	B	A	R2
ComRef	3	A	REF	R3
	4	REF	A	R4
	5	B	REF	R5
	6	REF	B	R6
OneColour	7	A	–	R7
	8	–	B	
	9	B	–	R8
	10	–	A	
SelfSelf	11	A	A	
	12	B	B	

R1–R8 are calculated from scaled LN-transformed background-subtracted intensities.

Average ratios are then calculated according to:

$$\text{LN}(\text{RatioCoHyb}): 0.5 * [\text{LN}(R1) - \text{LN}(R2)]$$

$$\text{LN}(\text{RatioComRef}): 0.5 * [\text{LN}(R5) - \text{LN}(R6)] - 0.5 * [\text{LN}(R3) - \text{LN}(R4)]$$

$$\text{LN}(\text{RatioOneColour}): 0.5 * [\text{LN}(R8) - \text{LN}(R7)].$$

each individual target (A and B) separately. F-statistics and corresponding p values are based upon the F2 statistic available in the MAANOVA package, which is a shrunk version of the classic F-statistic. To avoid distributional assumptions, the package offers the possibility of computing p values for hypothesis tests via permutation methods. We have chosen to perform the F-test with restricted residual shuffling and 1000 iterations. For more technical details about the model-fitting procedure, we refer to the package documentation.

This approach yielded a list of gene-specific p values relating to the hyb-design effect. In order to test if there was an overall hyb-design effect across all genes, we compared the proportion of computed p values of <0.05 with the expected one under no overall effect, 5%, via a conventional normal hypothesis test.

RESULTS

We performed a set of experiments to assess the influence of co-hybridization on microarray gene expression measurements. Two different mouse RNA samples and a RNA reference were amplified separately in the presence of aminoallyl-UTP (12). Each cRNA target was independently labelled with Cy3 and Cy5. Aliquots were hybridized to murine 65mer oligonucleotide microarrays, according to the scheme in Table 1. This yielded the following hyb-designs: co-hybridization of sample A and B (CoHyb), hybridization of A and B against the common reference (ComRef), hybridization of A and B separately (OneColour) and self-self hybridization (SelfSelf). Dye swaps were performed for each hyb-design.

Analysis of raw background-corrected signal intensities for sample A and B suggests that the distribution of intensity measurements is not influenced by the hyb-design (Fig. 1). To confirm this, we fitted a linear mixed-effects model to the data using MAANOVA (8) and calculated p values corresponding to the hyb-design effect per feature. Since no more p values of <0.05 were found than would be expected by chance (2.2%

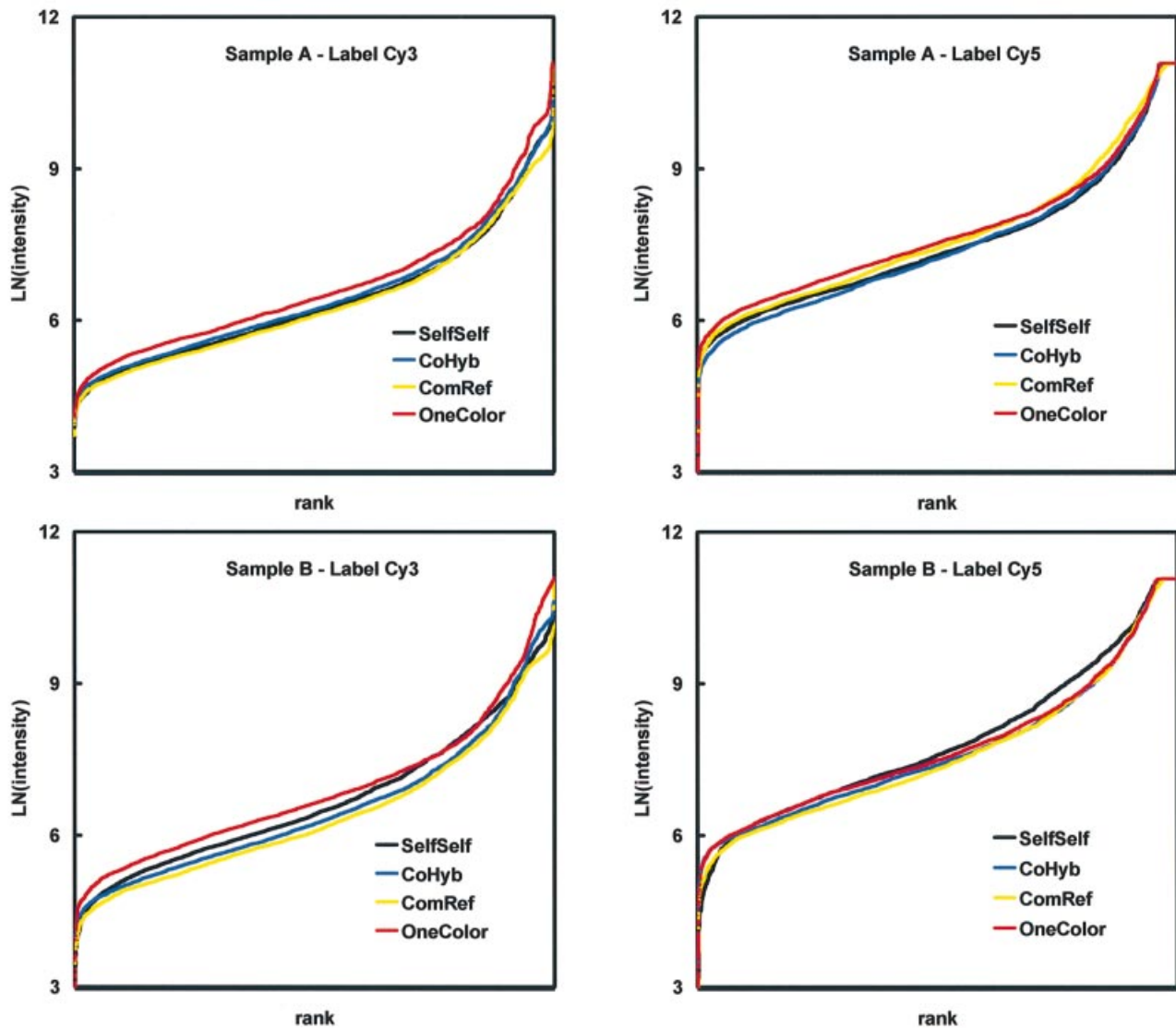


Figure 1. Effect of hybridization design on signal intensity distributions. The LN-transformed background-corrected Cy3 and Cy5 signal intensities from samples A and B, observed in the co-hybridizations (CoHyb, arrays 1 and 2, pink), hybridizations with the common reference (ComRef, arrays 3–6, yellow), one-colour hybridizations (OneColour, arrays 7–10, red) and self–self hybridizations (SelfSelf, arrays 11 and 12, blue), were ranked in rising order and plotted. The Cy5 signal levels off at high intensities due to scanner saturation.

and 5.3% for samples A and B, respectively), we concluded that the hybridization design did not significantly affect the measured signal intensity. We searched for indications of competition effects that would specifically affect measurements of highly abundant RNA species. To this end, we ranked the features according to average LN-transformed signal intensity on all arrays and made groups of 25 subsequent features. Then we performed a pairwise comparison of the signal intensities of the individual features in the different hybridization designs with Student's *t*-test. We found no statistically significant differences in intensity values between the one- and two-colour hybridizations in either of the signal intensity ranges. This indicates that, on spotted oligonucleotide microarrays, there is, even for high abundant RNA targets, no competition for binding sites. The presence of a co-hybridized target labelled in a different colour will, therefore, not affect signal intensities.

The intensity measurements on the arrays with different hybridization designs were highly consistent. To show this, the Pearson correlation coefficients of the separated Cy3 and Cy5 background-corrected intensities were calculated for each hybridization (Tables 2 and 3). Correlation coefficients for intensity measurements of a sample labelled in the same colour and hybridized to different arrays ranged between 0.88 and 0.98. We found that the correlations of signal intensities on one array (self–self hybridizations) were significantly stronger than the correlations of signal intensities from the same sample hybridized to different arrays ($p = 0.002$, Student's *t*-test). This confirms earlier observations that gene expression comparisons on the same array are more accurate than comparisons between different arrays (11). In addition, the signal intensities from targets labelled in the same colour are correlated significantly more strongly than the signal intensities from targets labelled in opposite colours ($p = 0.023$ and p

Table 2. Pearson correlation coefficients of raw Cy3 or Cy5 signal intensities from sample A, hybridized using different designs

CoHyb-3								
CoHyb-5	0.951							
ComRef-3	0.909	0.864						
ComRef-5	0.928	0.919	0.882					
OneColour-3	0.933	0.893	0.879	0.856				
OneColour-5	0.958	0.977	0.873	0.923	0.893			
SelfSelf-3	0.964	0.939	0.932	0.907	0.954	0.946		
SelfSelf-5	0.951	0.976	0.872	0.920	0.905	0.980	0.967	
Sample A	CoHyb-3	CoHyb-5	ComRef-3	ComRef-5	OneColour-3	OneColour-5	SelfSelf-3	SelfSelf-5

Pearson correlation coefficients are coloured according to strength: dark grey, $R > 0.95$; light grey, $0.90 < R < 0.95$; white, $R < 0.90$.

Table 3. Pearson correlation coefficients of raw Cy3 or Cy5 signal intensities from sample B, hybridized using different designs

CoHyb-3								
CoHyb-5	0.951							
ComRef-3	0.929	0.904						
ComRef-5	0.915	0.946	0.901					
OneColour-3	0.950	0.919	0.879	0.891				
OneColour-5	0.949	0.981	0.909	0.946	0.919			
SelfSelf-3	0.750	0.817	0.771	0.802	0.730	0.825		
SelfSelf-5	0.747	0.822	0.758	0.798	0.709	0.827	0.966	
Sample B	CoHyb-3	CoHyb-5	ComRef-3	ComRef-5	OneColour-3	OneColour-5	SelfSelf-3	SelfSelf-5

Pearson correlation coefficients are coloured according to strength: dark grey, $R > 0.95$; light grey, $0.90 < R < 0.95$; white, $R < 0.90$.

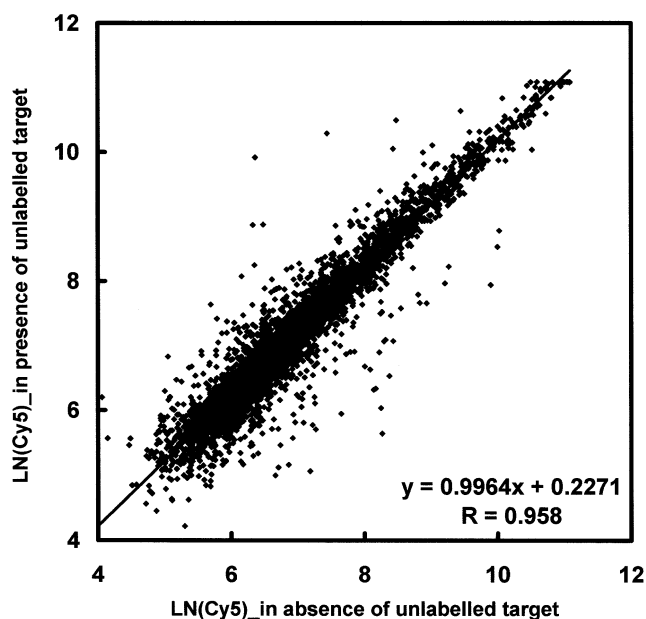


Figure 2. Effect of addition of unlabelled template on signal intensity measurements. Cy5-labelled cRNA was hybridized in the absence [x-axis, array 17 (table S1)] and presence [y-axis, array 18 (table S1)] of an equal amount of unlabelled cRNA. LN-transformed background-corrected signal intensities are plotted, together with the calculated regression line and the Pearson correlation coefficient.

= 0.037 for samples A and B, respectively, paired Wilcoxon test). This is indicative of a small dye effect and stresses the importance of dye swaps and balanced designs.

The experiment was repeated in a confined fashion with similar results (Supplementary tables S1 and S2 and fig. S1). In this second experiment, we also investigated the effect of co-hybridization with an identical unlabelled target. Again, we found a strong correlation of the intensity readings in

hybridizations with and without the unlabelled target and only a very slight effect on absolute intensity values (Fig. 2).

To detect influences of the hyb-design on ratio calculations, we determined the individual gene expression ratios between sample A and B in three different ways: (i) averaging of the scaled ratios in the two co-hybridization experiments; (ii) calculating the ratios of the individual samples over the common reference and subsequently eliminating the common reference to obtain the ratio of A to B; (iii) calculating the ratios based on the intensity readings in the OneColour hybridizations (Table 1). We observed a high degree of correlation in the three types of ratio determinations ($R \geq 0.8$) (Fig. 3).

DISCUSSION

The main conclusion from our study is that co-hybridized targets do not influence each other's hybridization to spotted 65mer oligonucleotide microarrays. Even for highly abundant RNAs, we found no evidence for competition of the two targets for binding sites on the array. From an approximate calculation, we can see that indeed the number of binding sites generally exceeds the number of target molecules that can reach their complementary binding sites. According to the manufacturer, our pins deposit 0.7 nl of the 20 μ M oligonucleotide solution onto the glass surface which amounts to 8×10^9 probe molecules, of which 20–50% is suggested to be available for binding (9). When we apply $2 \times 1.5 \mu$ g of RNA and take into account that the cRNA is on average 1 kb in length, and that an abundant RNA species may comprise up to 1% of the RNA applied, there are $\sim 5 \times 10^{10}$ copies of this specific RNA molecule in solution. Owing to limiting diffusion distances of the molecules over the glass surface, only a fraction of these will reach the spot on the array. For PCR products spotted at a concentration of 0.5 μ g/ μ l, the amount of binding sites is estimated to be only 10–20% of that

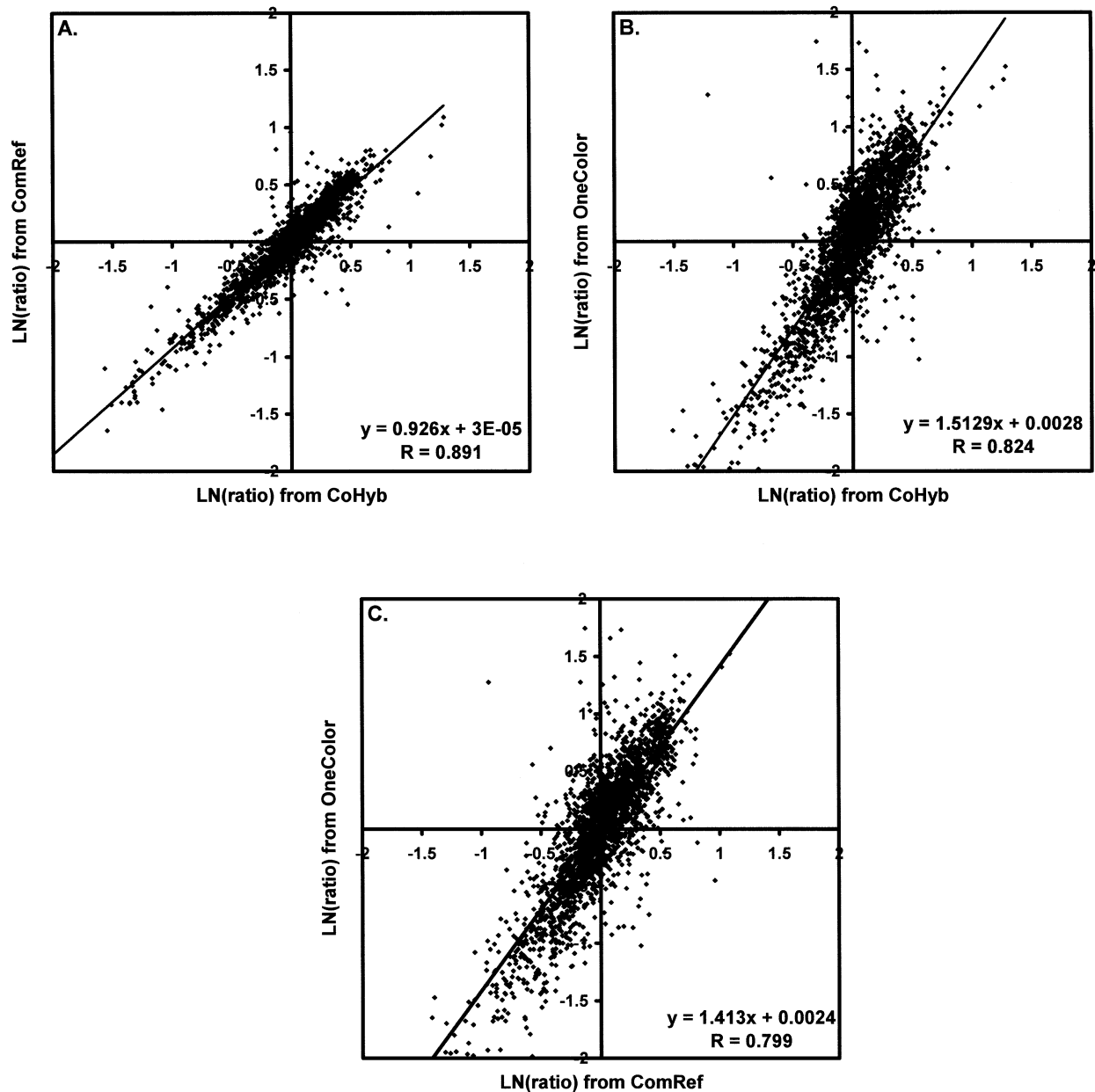


Figure 3. Effect of hybridization design on ratio calculation. After LN-transformation, background-corrected intensity values were linearly scaled by subtracting the mean LN-signal intensity in the separate channels of the arrays. Subsequently, gene expression ratios for samples A and B were calculated from the co-hybridized targets (CoHyb, arrays 1 and 2), targets hybridized with the common reference (ComRef, arrays 3–6) and the one-colour hybridizations (arrays 7–10), as described in the footnote to Table 1.

on the oligonucleotide arrays, indicating that competition may play a role, but only for higher abundant RNAs on cDNA arrays (14). Furthermore, the variation in concentration of spotted PCR products is larger than for oligonucleotide arrays, and therefore the influence of competition on the signal may be spot dependent, partly explaining the probe-concentration-dependent accuracy of cDNA arrays (15). The hybridization kinetics/thermodynamics on spotted microarrays are different from Affymetrix GeneChips, where saturation of binding sites is clearly observed and hybridization is adequately described by the Langmuir adsorption formula (16–18). This is probably due to the higher amount of target molecules applied (~15 μ g

cRNA), the substantially lower amount of binding sites [~ 10^7 per feature (19)] and the extensive mixing.

The independence of the measurements in the two channels of the microarray indicates that signal intensities from the two channels can be taken into subsequent analyses separately. After simultaneous normalization of the separated intensities on all arrays in the experiment, expression ratios can be calculated across all experimental groups. This is supported by our observation that the resulting gene expression ratios of sample A to sample B are independent of the hybridization design and the associated method of calculation. Obviously, this procedure critically depends on the uniformity of the arrays; spatial

hybridization effects should be absent and the spotting should be highly reproducible between different arrays. Judged from the high correlation of signal intensity measurements across different arrays, this uniform hybridization seems achievable.

In multi-class experiments, it is difficult to decide which samples should be co-hybridized. Since this and other studies show that differential expression measurements from co-hybridized samples are more accurate than from samples on different arrays, it has been suggested that the comparisons in which the experimenter is most interested should be hybridized to the same array (11,20). Co-hybridization effects may be included in mixed ANOVA models (21). Since this will be at the expense of a large reduction in the degrees of freedom in the experiment, it is best to avoid a co-hybridization bias. This can be achieved by co-hybridizing each group as many times to one group as to all the other groups (blocking of the effect), but this might not be possible when there are many different experimental groups. We suggest, therefore, that samples are randomly assigned from the different experimental groups to the arrays. This attributes an extra level of randomization in microarray designs above those suggested by Churchill (20), i.e. randomization of treatment and sampling, dye assignments, slides in a batch and spots on the array. To eliminate the dye effect that we and others (12,22,23) have demonstrated, it is essential to keep the design balanced with respect to dye, i.e. to label samples from a certain group as many times in one colour as in the other.

In summary, we show that the intensity measurements on spotted oligonucleotide arrays are not affected by co-hybridized targets and demonstrate the validity of separate analysis of the signal intensity readings from the different channels of the array. An intensity-based type of analysis has considerable advantages over ratio-based analysis in experiments that include many different groups. First, it enables comparisons between samples which were not hybridized to the same array without the need to relate the samples to a common reference or other samples in the same series. Secondly, the set of hybridizations is extendable, as long as temporal effects in array and target preparation and hybridization remain negligible. In future, multi-colour hybridizations (15), applying more than two targets on the same array, may become more general. For these multi-colour experiments, an intensity-based analysis would be even more favourable than for two-colour arrays, since a ratio-type of analysis will be complicated by the possible calculation of many different ratios from one array.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

REFERENCES

1. Van 't Veer,L.J., Dai,H., van de Vijver,M.J., He,Y.D., Hart,A.A., Mao,M., Peterse,H.L., van der,K.K., Marton,M.J., Witteveen,A.T. *et al.* (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, **415**, 530–536.
2. Arbeitman,M.N., Furlong,E.E., Imam,F., Johnson,E., Null,B.H., Baker,B.S., Krasnow,M.A., Scott,M.P., Davis,R.W. and White,K.P. (2002) Gene expression during the life cycle of *Drosophila melanogaster*. *Science*, **297**, 2270–2275.
3. Boer,J.M., de Meijer,E.J., Mank,E.M., van Ommen,G.B. and den Dunnen,J.T. (2002) Expression profiling in stably regenerating skeletal muscle of dystrophin-deficient mdx mice. *Neuromusc. Disord.*, **12**, S118–S124.
4. Giaever,G., Chu,A.M., Ni,L., Connelly,C., Riles,L., Veronneau,S., Dow,S., Lucau-Danila,A., Anderson,K., Andre,B. *et al.* (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, **418**, 387–391.
5. Ross,D.T., Scherf,U., Eisen,M.B., Perou,C.M., Rees,C., Spellman,P., Iyer,V., Jeffrey,S.S., Van de,R.M., Waltham,M. *et al.* (2000) Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genet.*, **24**, 227–235.
6. Lockhart,D.J., Dong,H., Byrne,M.C., Follettie,M.T., Gallo,M.V., Chee,M.S., Mittmann,M., Wang,C., Kobayashi,M., Horton,H. *et al.* (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.*, **14**, 1675–1680.
7. Quackenbush,J. (2002) Microarray data normalization and transformation. *Nature Genet.*, **32** (Suppl.), 496–501.
8. Kerr,M.K., Martin,M. and Churchill,G.A. (2000) Analysis of variance for gene expression microarray data. *J. Comput. Biol.*, **7**, 819–837.
9. Wang,Y., Wang,X., Guo,S.W. and Ghosh,S. (2002) Conditions to ensure competitive hybridization in two-color microarray: a theoretical and experimental analysis. *Biotechniques*, **32**, 1342–1346.
10. Dobbin,K., Shih,J.H. and Simon,R. (2003) Statistical design of reverse dye microarrays. *Bioinformatics*, **19**, 803–810.
11. Yang,Y.H. and Speed,T. (2002) Design issues for cDNA microarray experiments. *Nature Rev. Genet.*, **3**, 579–588.
12. 't Hoen,P.A., de Kort,F., Van Ommen,G.J. and den Dunnen,J.T. (2003) Fluorescent labelling of cRNA for microarray applications. *Nucleic Acids Res.*, **31**, e20.
13. Cox,D.R. and Hinkley,D.V. (1974) *Theoretical Statistics*. Chapman & Hall, London, UK.
14. Stillman,B.A. and Tonkinson,J.L. (2001) Expression microarray hybridization kinetics depend on length of the immobilized DNA but are independent of immobilization substrate. *Anal. Biochem.*, **295**, 149–157.
15. Hessner,M.J., Wang,X., Khan,S., Meyer,L., Schlicht,M., Tackes,J., Datta,M.W., Jacob,H.J. and Ghosh,S. (2003) Use of a three-color cDNA microarray platform to measure and control support-bound probe for improved data quality and reproducibility. *Nucleic Acids Res.*, **31**, e60.
16. Hekstra,D., Taussig,A.R., Magnasco,M. and Naef,F. (2003) Absolute mRNA concentrations from sequence-specific calibration of oligonucleotide arrays. *Nucleic Acids Res.*, **31**, 1962–1968.
17. Held,G.A., Grinstein,G. and Tu,Y. (2003) Modeling of DNA microarray data by using physical properties of hybridization. *Proc. Natl Acad. Sci. USA*, **100**, 7575–7580.
18. Chudin,E., Walker,R., Kosaka,A., Wu,S.X., Rabert,D., Chang,T.K. and Kreder,D.E. (2002) Assessment of the relationship between signal intensities and transcript concentration for Affymetrix GeneChip arrays. *Genome Biol.*, **3**, RESEARCH0005.
19. Lockhart,D.J. and Winzeler,E.A. (2000) Genomics, gene expression and DNA arrays. *Nature*, **405**, 827–836.
20. Churchill,G.A. (2002) Fundamentals of experimental design for cDNA microarrays. *Nature Genet.*, **32** (Suppl.), 490–495.
21. Cui,X. and Churchill,G.A. (2003) Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol.*, **4**, 210.
22. Tseng,G.C., Oh,M.K., Rohlin,L., Liao,J.C. and Wong,W.H. (2001) Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res.*, **29**, 2549–2557.
23. Yang,Y.H., Dudoit,S., Luu,P., Lin,D.M., Peng,V., Ngai,J. and Speed,T.P. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.*, **30**, e15.