

Research Report

Conditions to Ensure Competitive Hybridization in Two-Color Microarray: A Theoretical and Experimental Analysis

BioTechniques 32:1342-1346 (June 2002)

Youmin Wang, Xujing Wang,
Sun-Wei Guo, and Soumitra
Ghosh

Medical College of Wisconsin,
Milwaukee, WI, USA

ABSTRACT

We derived a theoretical model that explains certain biases observed in the two-color microarray hybridization experiments reported in the literature. We show that true competition is achieved only when the hybridization kinetics of the two differentially labeled probes are the same. If the hybridization kinetics of the two differentially labeled probes is different, which can occur when the labeling and hybridization conditions for the two probes are dissimilar, then differential expression observed becomes a function of the amount of the target (*i.e.*, DNA spotted on the slide). We use this model to validate the microarray methodology by determining the differential expression of four select Arabidopsis genes and two human genes (β -actin and GAPDH) as a function of the amount of target arrayed. We show through both modeling and experiments that the rate constants for Cy5- and Cy3-labeled probes are the same under our experimental conditions. Therefore, the target concentrations need not greatly exceed the probe concentration. It is obvious from the data presented that a simple treatment of an individual hybridization rate calculation does not fully describe what is occurring in today's complex, multispecies experiments. The method of validation is easily implemented to ensure data reliability by two-color microarray.

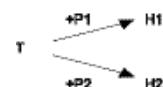
INTRODUCTION

Despite the growing popularity of two-color cDNA microarrays, little data are available in the literature regarding the potential biases of the two-color method. The biases can arise from unequal hybridization kinetics of the two differentially labeled probes and can produce false differential expression results. Indeed, a recent publication reports that the differential expression of three yeast gene fragments deviated from expected values as much as 40-fold, depending on target concentrations (21). This observation casts serious doubt on the assumption that the kinetics of the two differentially labeled probes is equal because the underlying sequences are identical. It also raises a question as to whether the two-color method is truly competitive in every microarray experiment. To understand the theoretical basis of the biases that have been observed, we derived a kinetic model that can explain the bias. Although the kinetics of nucleic acid hybridization in the conventional single-color methods, such as reverse dot blots, has been well described (4,8,9), the theory cannot be applied directly to the two-color methodology because there is a competitive component. Using our derived model, we also have developed a simple method to validate the microarray methodology.

KINETIC MODELING

The two-color competitive hybridization reaction can be described

by the following scheme:



where T denotes the target, P1 and P2 are differentially labeled probes from test and reference samples, respectively, and H1 and H2 are the respective hybridization products of P1 and P2. The rate of product formation is given in Reference 4:

$$\frac{dC_{H1}}{dt} = k_1 C_T C_{P1} = k_1 (C_T^0 - C_{H1} - C_{H2})(C_{P1}^0 - C_{H1}) \quad [Eq. 1]$$

$$\frac{dC_{H2}}{dt} = k_2 C_T C_{P2} = k_2 (C_T^0 - C_{H1} - C_{H2})(C_{P2}^0 - C_{H2}) \quad [Eq. 2]$$

where k_1 and k_2 are the rate constants, C_T^0 , C_{P1}^0 , and C_{P2}^0 are the initial concentrations of target and two probes, respectively, C_{H1} and C_{H2} are the concentrations of H1 and H2 at time t , respectively. The ratio of dC_{H1} and dC_{H2} is thus

$$\frac{dC_{H1}}{dC_{H2}} = \frac{k_1}{k_2} \frac{(C_{P1}^0 - C_{H1})}{(C_{P2}^0 - C_{H2})} \quad [Eq. 3]$$

Solving Equation 3, we obtain

$$(C_{P1}^0 - C_{H1})^{k_2} = Q(C_{P2}^0 - C_{H2})^{k_1} \quad [Eq. 4]$$

where $Q = \left(\frac{C_{P1}^0}{C_{P2}^0}\right)^{\frac{k_2}{k_1}}$, which was obtained

using initial conditions ($t = 0$, $C_{H1} = C_{H2} = 0$). C_{H1} and C_{H2} cannot be solved explicitly due to the nonlinearity of Equation 4. However, if we assume that hybridization reaction is generally carried out until equilibrium or

$$\frac{dC_{H1}}{dt} = \frac{dC_{H2}}{dt} = 0 \quad [Eq. 5]$$

then this leads to a significant simplification of Equation 4.

There are several possible scenarios which we discuss individually.

(i) When $C_T^0 > C_{P1}^0 + C_{P2}^0$, then

$$\frac{C_{H1}}{C_{H2}} = \frac{C_{P1}^0}{C_{P2}^0}$$

because $C_{P1}^0 - C_{H1} = C_{P2}^0 - C_{H2}$ and $C_T^0 - C_{H1} - C_{H2} \neq 0$.

That is, the ratio of the concentrations of the products directly reflects the ratio of the concentrations of the probes.

(ii) When $C_T^0 < C_{P1}^0 + C_{P2}^0$, then

$$(C_{P1}^0 + C_{H2} - C_T^0)^{k_2} = \left(\frac{C_{P1}^0}{C_{P2}^0}\right)^{k_1} (C_{P2}^0 - C_{H2})^{k_1}$$

[Eq. 6]

because, to satisfy Equations 1, 2, and 5, the following must be true:

$$C_T^0 - C_{H1} - C_{H2} = 0 \quad [Eq. 7]$$

Equation 6 can be further simplified under the following two possible scenarios.

(iia) When $k_1 = k_2$, then Equation 6 becomes

$$(C_{P1}^0 + C_{H2} - C_T^0) = \left(\frac{C_{P1}^0}{C_{P2}^0}\right) (C_{P2}^0 - C_{H2}) \quad [Eq. 8]$$

Solve Equation 8 for C_{H2} , and we have

$$C_{H2} = C_T^0 \frac{C_{P2}^0}{C_{P1}^0 + C_{P2}^0} \quad [Eq. 9]$$

Substitute Equation 9 into Equation 7, and we obtain

$$C_{H1} = C_T^0 \frac{C_{P1}^0}{C_{P1}^0 + C_{P2}^0} \quad [Eq. 10]$$

$$\therefore \frac{C_{H1}}{C_{H2}} = \frac{C_{P1}^0}{C_{P2}^0}$$

Therefore, the concentration ratio of the products is equal to the initial concentration ratio of the probes.

(iib) When $k_1 \neq k_2$, then Equation 4 becomes

$$(C_{P1}^0 - C_{H1}) = C_{P1}^0 (C_{P2}^0)^{-a} (C_{P2}^0 - C_{H2})^a$$

[Eq. 11]

where $a = k_1/k_2$. Equation 11 is constrained by Equation 7 and can only be solved numerically.

To visualize how k_1 and k_2 affect differential expression (i.e., the ratio C_{H1}/C_{H2}), we plotted Figure 1. We can see that when $k_1 \neq k_2$, the differential expression decreases or increases as the target amount is increased. Only at a sufficiently high target amount does the differential expression become independent of target amount, as is the case when $k_1 = k_2$.

Based on the theoretical analysis above, we can see that, to achieve reliable differential expression results without regard to hybridization kinetics, we could use a large excess of targets. Indeed, that is what has often been assumed (20). However, to the best of our knowledge, no data have been re-

ported that this assumption is satisfied. To the contrary, our calculation demonstrates that this assumption is not always true. Consider, for example, a spotting solution containing 100 ng/ μ L DNA (7) and a spotting volume of 5 nL (7,17): the maximum amount of DNA arrayed is 0.5 ng. The efficiency of target DNA binding to poly-L-lysine-coated glass surface is 20%–50% (unpublished data). If we assume 50% binding and 100% availability for hybridization (both of which are definitely overestimates), then 0.5 ng cDNA corresponds to approximately 2.5×10^8 molecules/spot (5). For probes, most microarray experiments use 1–5 μ g mRNA or 50–200 μ g total RNA (6,7,10,13,15,16), which corresponds to $1–5 \times 10^6$ cells (12) or about 7.5×10^7 copies of low abundant mRNAs (1). The total amount of one mRNA probe in two-color hybridization is thus 1.5×10^8 copies. For intermediate and high abundant mRNA probes, the number is even higher, ranging from 3×10^9 to 1.2×10^{11} copies (1). Assuming good mixing during hybridization, these calculations show that the target amount is not, in fact, much in excess compared to the amount of probe. According to our model, data obtained under this condition would not be reliable if $k_1 \neq k_2$, which can occur as shown in Reference 21. How do we determine whether $k_1 = k_2$? One method suggested by our model is to determine the differential expression of known genes as a function of the target amount (Figure 1).

MATERIALS AND METHODS

Arrays

The microarrays were constructed by arraying PCR-amplified cDNA clones on poly-L-lysine-coated glass microscopic slides. Human cDNA clones were purchased from Research Genetics (ResGen™; Invitrogen, Carlsbad, CA, USA), and *Arabidopsis* clones were obtained from Incyte Genomics (Palo Alto, CA, USA). We selected four *Arabidopsis* clones and two human clones [β -actin and glyceraldehyde-3-phosphate dehydrogenase (GAPDH)] as control targets. The four *Arabidopsis* clones represent chloro-

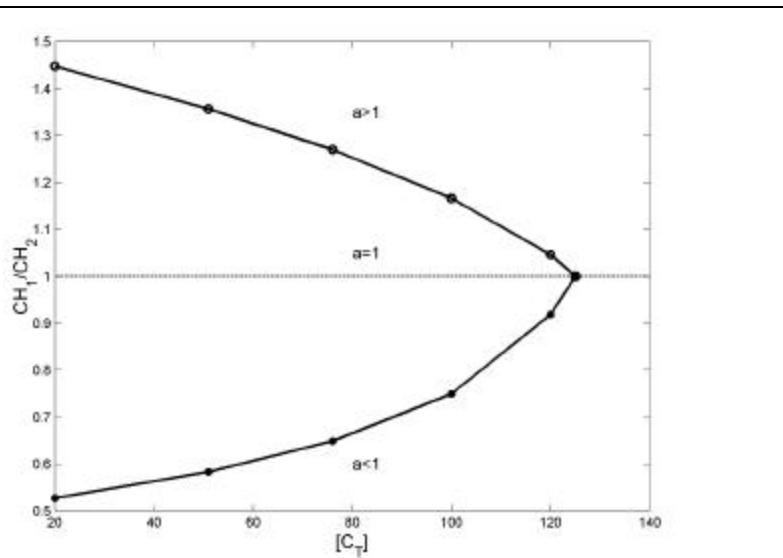


Figure 1. Theoretical prediction of the effect of the target amount on differential expression (ratio of hybridization products) at various a values ($a > 1$, $k_1 > k_2$; $a = 1$, $k_1 = k_2$; and $a < 1$, $k_1 < k_2$).

phyl a/b binding protein, ribulose-1,5-bisphosphate, and cellulose synthase *Arabidopsis thaliana* genes. There is no significant homology among their sequences and to human gene sequences, as determined by screening the sequences against the nonredundant database using the BLAST 2.0 sequence alignment program (2). No cross hybridization of the human cDNA and *Arabidopsis* clones was further confirmed experimentally (data not shown). Spotting solutions were made from six 2-fold dilution series of the stock solutions (approximately 0.2 $\mu\text{g}/\mu\text{L}$ for the *Arabidopsis* clones and 0.8 $\mu\text{g}/\mu\text{L}$ for the two human clones). The amount of target arrayed ranged from 0.031 to 1 ng for the *Arabidopsis* clones (or on a molar basis, 1×10^{-4} to 3.6×10^{-3} pmol for Figure 2, #1, 2×10^{-4} to 6.3×10^{-3} pmol for #2, 1×10^{-4} to 3.5×10^{-3} pmol for #3, and 1×10^{-4} to 3.2×10^{-3} pmol for #4) and from 0.125 to 4 ng for human clones. The clones were spotted onto the slide (approximately 5 nL), using a GMS 417 Arrayer (Genetic MicroSystems, Woburn, MA, USA), in at least triplicate for each of the four quadrants. The preparation of the cDNA clones, printing, and post-processing of slides was carried out using the protocol similar to the one given in Reference 16.

Probe Samples, Hybridization, and Detection

Arabidopsis probes were made as follows. The clone inserts were amplified by PCR, cloned into a pSP64 poly(A) expression vector, transformed into JM 109 Competent Cells (Promega, Madison, WI, USA), and transcribed in vitro using RiboMaxTM SP6 RNA polymerase (Promega). The length of the *Arabidopsis* probes ranged from 480 to 911 nucleotides, and the GC content ranged from 41% to 50%. In addition, two sources of human RNA were used. One was isolated from a melanoma cell line UACC-903 and one was from skeletal muscle. The purified RNA was quantified using the spectroscopic method. Complex mRNA (2 μg) or total RNA (100 μg) was used, into which *Arabidopsis* probes were spiked. The amounts spiked were 1 ng (3.6×10^{-3} pmol) for #1; 6 ng (3.8×10^{-2}

pmol) for #2; 20 ng (7.0×10^{-2} pmol) for #3; and 24 ng (7.6×10^{-2} pmol) for #4 (Figure 2). The sample was then split equally into two, one labeled with Cy5 and the other with Cy3. Labeling was done using SUPERSCRIPTTM II First-Strand Synthesis System for RT-PCR (Invitrogen) and Cy5-dUTP and Cy3-dUTP (Amersham Biosciences, Piscataway, NJ, USA). Labeled Cy5 and Cy3 probes were combined and purified with a PCR purification kit (Qiagen, Valencia, CA, USA) and concentrated using an Amicon[®] 30 kit (Millipore,

Bedford, MA, USA). Hybridization was performed in 11.25 μL hybridization solution containing 20 μg poly(dA), 12 μg yeast tRNA, 0.3% SDS, 3.4x SSC, and 20 μg Cot I DNA at 65°C for 18 h. After hybridization, the slides were washed, dried, and scanned using a ScanArray[®] 5000 (GSI Lumonics, Billerica, MA, USA).

Image and Data Analysis

Images were processed using Mataray image-quality assurance software

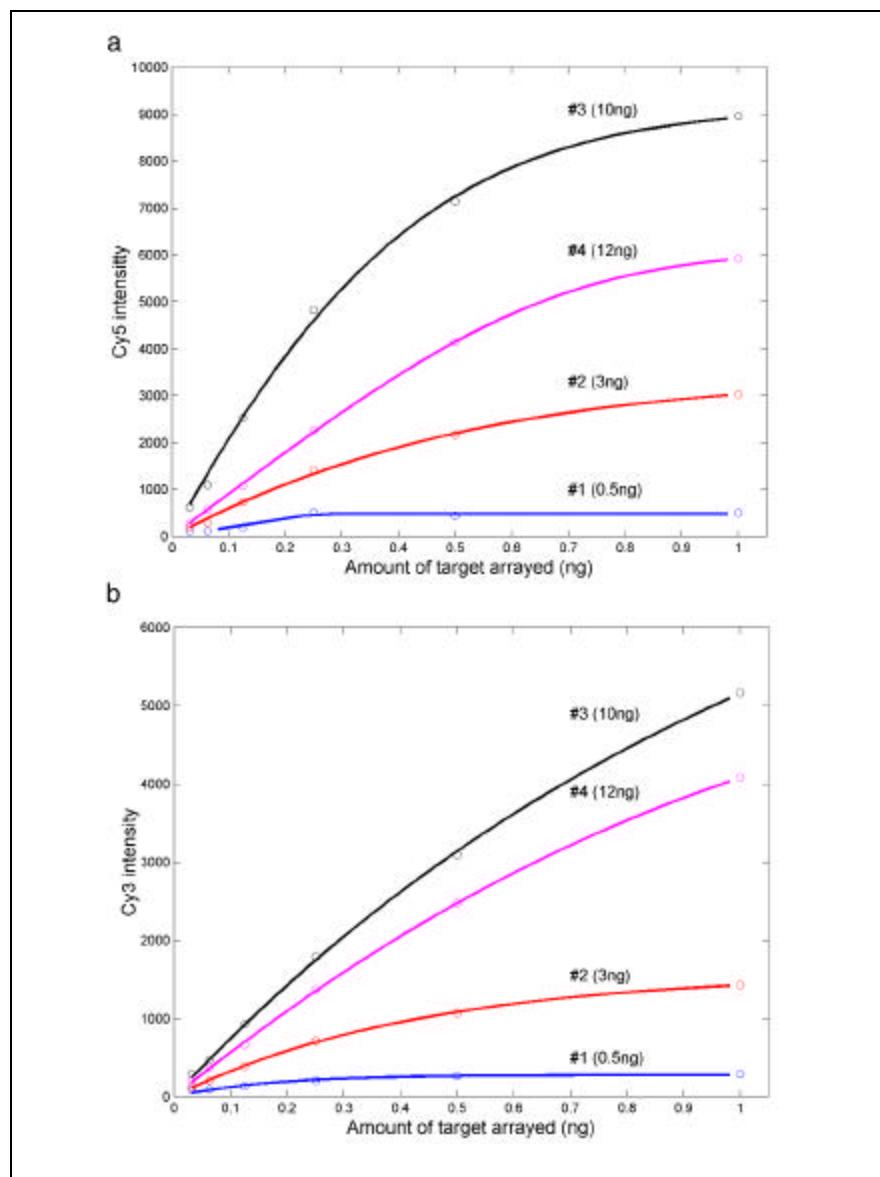


Figure 2. Spot intensities of four *Arabidopsis* clones as a function of the amount of target arrayed. Circles are the experimental data (each data point is the average of 12 spots). The fitted curves are the results of the analysis of the experimental data using Equations 1, 2, and 11. (a) Cy5 intensity. (b) Cy3 intensity.

developed in this laboratory (18). Signal and background values for Cy5 and Cy3 dyes and a composite quality score were obtained for each spot. All spots presented here passed our stringent filters. The mean signal intensity of the spots was used to calculate the global mean for each clone, which was represented by at least 12 spots on each slide. The global mean was used in all plots.

RESULTS

Determination of the Relationship between k_1 and k_2

The relationship between rate constants of hybridization for Cy5- and Cy3-labeled probes (i.e., whether $k_1 = k_2$) was determined by using the intensity ratio as a function of target amount. The observed Cy5/Cy3 intensity ratio for the four *Arabidopsis* clones and two human clones (β -actin and GAPDH) is nearly constant over the range of target amount (data not shown), suggesting $k_1 = k_2$ for these clones.

Figure 2, a and b, shows the individual intensities of Cy5 and Cy3 as a function of target amount. The solid

lines are a least-square fit to the experimental data using our model (Equations 1, 2, and 11), by minimizing $\sum_i (C_{H1,i}^{\text{exp}} - C_{H1,i}^{\text{theory}})^2$ and $\sum_i (C_{H2,i}^{\text{exp}} - C_{H2,i}^{\text{theory}})^2$ simultaneously, where $C_{H1,i}^{\text{exp}}$ and $C_{H2,i}^{\text{exp}}$

are the experimentally measured Cy5 and Cy3 intensity values, respectively. The fitted $a (=k_1/k_2)$ values for each of the clones are 2.5, 2.71, 3.27, and 3.33, respectively, with a mean value of 2.95 ± 0.41 . Note that since the fitted a was estimated from the intensity rather than the concentration and that intensity values are different for Cy5 and Cy3 because of different quantum yield and extinction maximum, the estimated a should be close to the ratio of the product of quantum yield and the extinction maximum for Cy5 and Cy3, which is around 3 (3,19). Therefore, our estimates are in excellent agreement with the theoretical prediction. It should also be noted from Figure 2, a and b, that clone #3 had a much stronger intensity than #4, although #4 was spiked at a higher amount. This could be due to differences in the sequence of the clones that lead to differences in the labeling and/or hybridization efficiency.

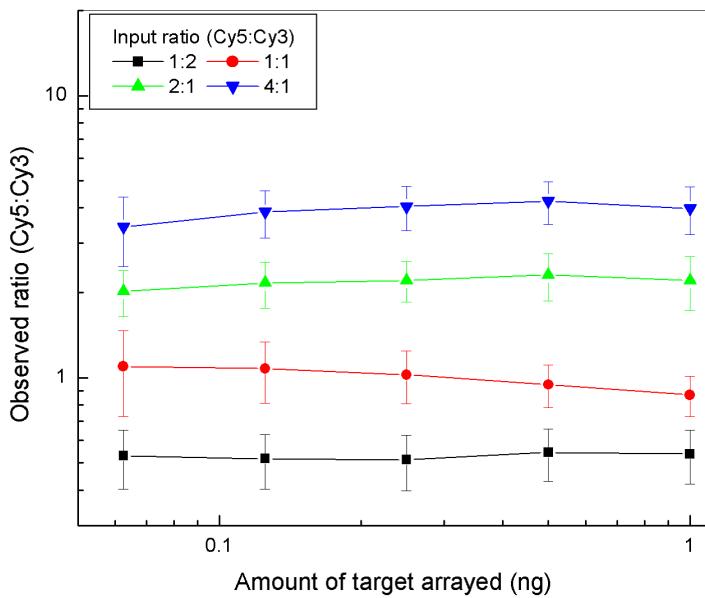


Figure 3. The observed intensity ratio as a function of the target amount at various input ratios. The slides contained 320 spots of a single clone (β -actin gene). Skeletal muscle mRNA was labeled and combined at ratios shown. Each data point is the average of 64 spots \pm SD.

Effect of Relative Abundance on Differential Expression

To determine the impact of the relative abundance of one dye-labeled probe over the other on hybridization kinetics, we varied the input ratios of Cy5:Cy3 labeled probes. Skeletal muscle mRNA was quantified, divided, and labeled with either Cy5 or Cy3. The labeled probes were combined at predetermined Cy5:Cy3 ratios (4:1, 2:1, 1:1, and 1:2, respectively), while the Cy3-labeled probe was kept constant at 0.25 µg and hybridized to four arrays containing 320 spots of a single cDNA clone (β -actin) with five 2-fold dilution series. There is a good correlation between the observed intensity ratio and the input ratio, with a Pearson correlation coefficient of 0.998. Moreover, the observed ratio is independent of the amount of target arrayed in all four input ratios (Figure 3).

DISCUSSION

Unlike high-density synthetic oligonucleotide arrays in which only one labeled probe is used for each slide (14), cDNA microarrays typically involve the co-hybridization of two differentially labeled probes that are both competing for the cognate target element. The two-color strategy has added more power to the microarray methodology because, by comparing two samples in a single hybridization, the distracting effect of many variables such as the labeling method, hybridization conditions, target features, and the sequence of the gene can be minimized or even eliminated (17). However, the reliability of the data is highly dependent on the validity of the probe competition. Our kinetic model (see Figure 1) indicates that to achieve a bona fide competition, the two probes should ideally be indistinguishable kinetically (i.e., $k_1 = k_2$). If $k_1 \neq k_2$, the amount of target should be in large excess over the amount of probe used. We have shown both through modeling and experiments that the rate constants for Cy5- and Cy3-labeled probes are the same under our experimental conditions. Therefore, the target concentrations need not greatly exceed the probe concentration. The validation method

proposed should give us more confidence than the single-point method, in which one spikes several control genes at predetermined ratios and determines differential expression at a single-target concentration (11).

ACKNOWLEDGMENTS

The cell line UACC-903 was kindly provided by Dr. Michael L. Bittner (National Institutes of Health, Bethesda, MD, USA). We thank Kate Frederick, Obrad Kokanovic, Dr. Steve Nye, and Jill Waukau for performing microarray and in vitro transcription experiments. Thanks also to the Children Hospital of Wisconsin Foundation for the funding.

REFERENCES

- Alberts, B. 1994. Molecular Biology of the Cell, 3rd ed. Garland Publishing, New York.
- Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
- Amersham Pharmacia Biotech. Product specification.
- Anderson, M.L.M. and B.D. Young. 1985. Quantitative filter hybridization, p.73-109. In B.D. Hames and S.J. Higgins (Eds.), *Nucleic Acid Hybridization—A Practical Approach*. IRL Press Ltd., Oxford and Washington, DC.
- Bertucci, F., K. Bernard, B. Loriod, Y.C. Chang, S. Granjeaud, D. Birnbaum, C. Nguyen, K. Peck, et al. 1999. Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples. *Hum. Mol. Genet.* 8:1715-1722.
- Duggan, D.J., M. Bittner, Y. Chen, P. Meltzer, and J.M. Trent. 1999. Expression profiling using cDNA microarrays. *Nat. Genet.* 21(1 Suppl):10-14.
- Eisen, M.B. and P.O. Brown. 1999. DNA arrays for analysis of gene expression. *Methods Enzymol.* 303:179-205.
- Flavell, R.A., E.J. Birfelder, J.P. Sanders, and P. Borst. 1974. DNA-DNA hybridization on nitrocellulose filters. 1. General considerations and non-ideal kinetics. *Eur. J. Biochem.* 47:535-543.
- Flavell, R.A., P. Borst, and E.J. Birfelder. 1974. DNA-DNA hybridization on nitrocellulose filters. 2. Concatenation effects. *Eur. J. Biochem.* 47:545-548.
- Hegde, P., R. Qi, K. Abernathy, C. Gay, S. Dharap, R. Gaspard, J.E. Hughes, E. Snestrud, et al. 2000. A concise guide to cDNA microarray analysis. *BioTechniques* 29:548-556.
- Heller, R.A., M. Schena, A. Chai, D. Shalon, T. Bedilion, J. Gilmore, D.E. Woolley, and R.W. Davis. 1997. Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc. Natl. Acad. Sci. USA* 94:2150-2155.
- Kane, M.D., T.A. Jatkoe, C.R. Stumpf, J. Lu, J.D. Thomas, and S.J. Madore. 2000. Assessment of the sensitivity and specificity of oligonucleotide (50-mer) microarrays. *Nucleic Acids Res.* 28:4552-4557.
- Khan, J., R. Simon, M. Bittner, Y. Chen, S.B. Leighton, T. Pohida, P.D. Smith, Y. Jiang, et al. 1998. Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. *Cancer Res.* 58:5009-5013.
- Lockhart, D.J., H. Dong, M.C. Byrne, M.T. Follettie, M.V. Gallo, M.S. Chee, M. Mittmann, C. Wang, et al. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* 14:1675-1680.
- Ross, D.T., U. Scherf, M.B. Eisen, C.M. Perou, C. Rees, P. Spellman, V. Iyer, S.S. Jeffrey, et al. 2000. Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* 24:227-235.
- Schena, M., D. Shalon, R. Heller, A. Chai, P.O. Brown, and R.W. Davis. 1996. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. USA* 93:10614-10619.
- Shalon, D., S.J. Smith, and P.O. Brown. 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.* 6:639-645.
- Wang, X., S. Ghosh, and S.W. Guo. 2001. Quantitative quality control in microarray image processing and data acquisition. *Nucleic Acids Res.* 29:e75.
- Willard, H.H., L.L. Merritt, Jr., J.A. Dean, and F.A. Settle, Jr. 1988. *Instrumental Methods of Analysis*. Wadsworth Publishing, Belmont, CA.
- Winzeler, E.A., M. Schena, and R.W. Davis. 1999. Fluorescence-based expression monitoring using microarrays. *Methods Enzymol.* 306:3-18.
- Yue, H., P.S. Eastman, B.B. Wang, J. Minor, M.H. Doctolero, R.L. Nuttall, R. Stack, J.W. Becker, et al. 2001. An evaluation of the performance of cDNA microarrays for detecting changes in global mRNA expression. *Nucleic Acids Res.* 29:e41-1.

Received 3 January 2002; accepted 16 April 2002.

Address correspondence to:

Dr. Youmin Wang
Department of Pediatrics
Max McGee National Research Center for Juvenile Diabetes
Medical College of Wisconsin
Children's Hospital of Wisconsin
8701 Watertown Plank Road
Milwaukee, WI 53226, USA
e-mail: ymwang@mcw.edu