

## Dye-Bias Correction in Dual-Labeled cDNA Microarray Gene Expression Measurements

Barry A. Rosenzweig,<sup>1</sup> P. Scott Pine,<sup>1</sup> Olen E. Domon,<sup>2</sup> Suzanne M. Morris,<sup>2</sup> James J. Chen,<sup>2</sup> and Frank D. Sistare<sup>1</sup>

<sup>1</sup>Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Laurel, Maryland, USA; <sup>2</sup>National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, Arkansas, USA

A significant limitation to the analytical accuracy and precision of dual-labeled spotted cDNA microarrays is the signal error due to dye bias. Transcript-dependent dye bias may be due to gene-specific differences of incorporation of two distinctly different chemical dyes and the resultant differential hybridization efficiencies of these two chemically different targets for the same probe. Several approaches were used to assess and minimize the effects of dye bias on fluorescent hybridization signals and maximize the experimental design efficiency of a cell culture experiment. Dye bias was measured at the individual transcript level within each batch of simultaneously processed arrays by replicate dual-labeled split-control sample hybridizations and accounted for a significant component of fluorescent signal differences. This transcript-dependent dye bias alone could introduce unacceptably high numbers of both false-positive and false-negative signals. We found that within a given set of concurrently processed hybridizations, the bias is remarkably consistent and therefore measurable and correctable. The additional microarrays and reagents required for paired technical replicate dye-swap corrections commonly performed to control for dye bias could be costly to end users. Incorporating split-control microarrays within a set of concurrently processed hybridizations to specifically measure dye bias can eliminate the need for technical dye swap replicates and reduce microarray and reagent costs while maintaining experimental accuracy and technical precision. These data support a practical and more efficient experimental design to measure and mathematically correct for dye bias. **Key words:** cDNA, dye bias, dye swap, genomics, microarray. *Environ Health Perspect* 112:480–487 (2004). doi:10.1289/txg.6694 available via <http://dx.doi.org/> [Online 15 January 2004]

A frequent goal of genome-scale gene expression experiments is to identify significant alterations in transcript levels resulting from the exposure of a living system to a test agent at a given dose and time. For animal experimentation, biological replication is achieved through strategically paired dosings of individual animals. A study design involving, for example, a control group and three different test agent dose groups, each at two time points, can achieve three clear biological replicate measurements of each transcript on an array at each dose and time point using 24 animals. For a similar gene expression microarray experiment using cells in culture rather than animals, the designation of what constitutes a biological replicate is less clear. Does a biological replicate require a complete second experiment with cells expanded at some separate time, or could a replicate be considered a parallel set of additional culture-flask incubations arbitrarily assigned to a control or treatment group and treated and processed that same day? The volume of cultured cells needed to generate sufficient quantities of RNA for microarray gene expression analyses practically limits implementation of such an experiment to the pooling of multiple large flasks of cultured cells from each of the studied dose and time points. Eight samples of RNA would be collected

for microarray analyses from such an experiment. To achieve the same minimum of three biological replicate measurements and therefore a similar level of confidence in biological accuracy, generally and practically speaking, this entire cell culture dosing experiment would be repeated on two additional separate occasions.

Technical replicate measurements of each of the individual biological replicates are generally incorporated at the discretion of the experimenter and may depend on factors including the amount of sample available, the budget of the experimenter, and whether the design of the specific microarray platform incorporates replicate probes. Replicate probes designed into a microarray platform to measure abundance of the same target gene transcript in a sample provide one approach to enhance confidence in the technical accuracy of relative transcript-level measurements. Repeat hybridizations of the same sample set using additional microarrays represent an additional level of technical replication to enhance confidence in the accuracy of each measurement of gene expression change. The precision and accuracy of replicated biological measurements is optimized when the technical replicate measurements are accurate and no confounding systematic error is introduced during sample processing.

Dual-labeled microarray hybridization protocols can introduce a systematic dye-bias error that could confound identification of true biological effects distinct from technical artifact (Goryachev et al. 2001; Ideker et al. 2000; Kerr et al. 2001; Tseng et al. 2001; Wang et al. 2001). Differences observed between red and green channel fluorescence intensities for a given transcript may be due to either a true biological difference resulting from the exposure of test agent to the cells or to a systematic bias resulting from individual transcript-dependent differences in efficiencies of dye incorporation and sample hybridizations.

Gene expression microarray data normalization typically corrects for systematic trends in dye incorporation and hybridization that affect all the transcripts similarly based on either array location or fluorescence intensity (Quackenbush 2002; Tseng et al. 2001; Yang et al. 2002). Some transcripts, however, behave differently from the global population (Tseng et al. 2001; Zhou et al. 2002), and this is the focus of the current investigation.

Confidence in data accuracy has been gained through a study design approach that eliminates dye-bias artifacts by adding a technical replicate involving a dye reversal or dye-swap labeling of the same two paired samples hybridized to an additional microarray (see "Terminology" in "Materials and Methods"). A logical approach to improve accuracy is to simply increase the number of such technical dye swap replicates of the same sample. The practical cost and sample limitations associated with microarray gene expression measurements require investigators to make difficult choices and carefully consider the minimum number of arrays to use while not unduly compromising accuracy and precision. Such genome-scale gene expression

This article is part of the mini-monograph "Application of Genomics to Mechanism-Based Risk Assessment."

Address correspondence to B.A. Rosenzweig, Division of Applied Pharmacology Research (HFD-910), Center for Drug Evaluation and Research, U.S. FDA, 10903 New Hampshire Ave., Life Sciences Building 64, Silver Spring, MD 20993 USA. Telephone: (301) 796-0125. Fax: (301) 796-9818. E-mail: rosenzweig@cder.fda.gov

The authors declare they have no competing financial interests.

Received 22 August 2003; accepted 12 January 2004.

comparisons can quickly become costly. The goal of this study was to demonstrate that transcript-dependent dye bias is consistent and measurable on a given day and can therefore be corrected mathematically using a pair of split-control hybridizations (see "Terminology" in "Materials and Methods") to achieve a similar high level of technical accuracy. The data show that these split-control hybridizations must be performed concurrently in the same batch with corresponding treatment and control array labelings and hybridizations. Careful measurement of dye-bias and mathematical correction to eliminate this systematically introduced error result in accurate sets of microarray data attained with greater efficiency. For example, for the eight samples of RNA derived from the single typical cell culture experiment described above, we describe an approach to measure gene expression changes, achieving with 10 microarrays a level of technical accuracy equivalent to that of a balanced replicate dye swap design involving 16 microarrays. The biological interpretations of the effects of (+/-)-*anti*-benzo[*a*]pyrene-*trans*-7,8-dihydrodiol-9,10-epoxide (BPDE) on the consistent gene expression changes observed in thymidine kinase (TK) 6 cells derived from the data described in this publication are published separately (Akerman et al. In press).

## Materials and Methods

### RNA Isolation

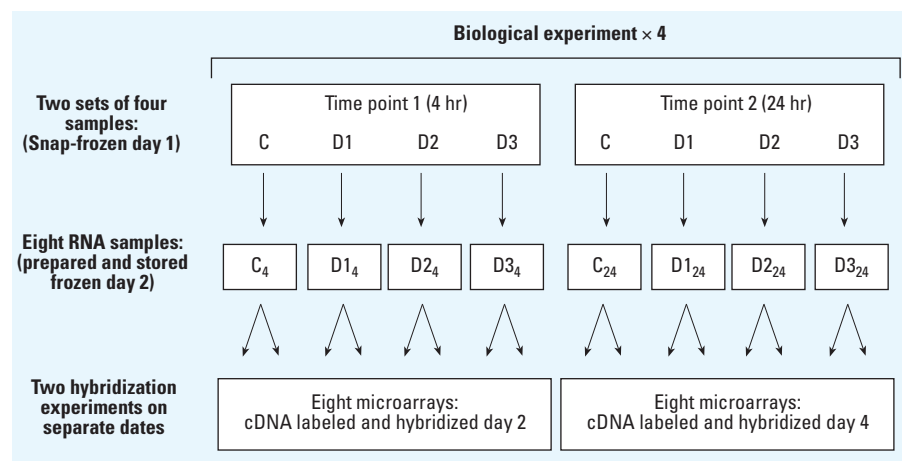
Human TK6 cells in suspension culture were harvested and collected after treatment with either vehicle or with one of three concentrations (0.01, 0.1, 1  $\mu$ M) of BPDE for 4 or 24 hr. BPDE is a member of the polycyclic aromatic hydrocarbon family and these substances are extremely hazardous. Special precautions are to be taken when working with these compounds. Cell pellets were snap-frozen at  $-70^{\circ}\text{C}$ . RNA was isolated using the Qiagen RNeasy Midi kit protocol (Qiagen 1999). Cell pellets were thawed in RNA lysis buffer (RLT; Qiagen, Valencia, CA), homogenized using a VirTis Tempest rotor-stator homogenizer (VirTis, Gardiner, NY), and further processed according to the RNeasy handbook protocol. After final column elution the samples were precipitated with lithium chloride, washed with 70% ethanol, and hydrated in sterile water treated with diethyl pyrocarbonate. Quantitative and qualitative analyses were performed on each sample using the RNA 6000 Assay on an Agilent 2100 bioanalyzer (Agilent Technologies, Palo Alto, CA). RNA samples were aliquoted and frozen at  $-70^{\circ}\text{C}$  until labeled and hybridized within 90 days.

**Microarray processing.** Hybridizations were performed using the Human-350 microarray, a glass slide with 350 spotted human cDNA probes (PHASE-1 Molecular Toxicology, Inc., Santa Fe, NM). Each cDNA probe was spotted in quadruplicate on each slide for a total of 1,400 spots per microarray. Sample cDNA labeling and microarray hybridizations and washes were performed according to the supplied protocol. Twenty micrograms of each RNA was reverse transcribed, and their corresponding cDNAs were independently labeled by incorporation of either Cyanine (Cy)3-dCTP or Cy5-dCTP. Dye incorporation was measured using a Cytofluor multiwell plate reader (PerSeptive Biosystems, Inc., Framingham, MA). The two labeled samples to be compared were purified, combined, and hybridized to the same microarray. Each microarray was scanned using a ScanArray 4000 microarray scanner (PerkinElmer, Inc., Wellesley, MA) with dual lasers. Low-resolution horizontal line prescans were performed on each microarray before higher resolution scanning to balance the overall fluorescence intensity of the entire microarray between the two dyes. The laser power and photomultiplier tube gain settings were assessed and slightly adjusted for each individual microarray to achieve optimal balance with the least amount of postscanning processing and normalization. QuantArray (PerkinElmer, Inc.) software was used to quantitate the relative transcript level for each spot of the microarray from the ScanArray output

TIF file. Local fluorescent background was subtracted and  $\log_2$  transformed, the resulting data locally weighted scatterplot smoothing (LOWESS) was normalized, and  $\log_2$  ratios were calculated. LOWESS normalization, also known as locally weighted least-squares regression, uses a smoothing curve to normalize a data set (Cleveland et al. 1992; Yang et al. 2001). Gene transcript signals beneath the mean signal strength of the four plant and one bacterial control genes were excluded from analyses. All statistical analyses and the level of statistical significance for each gene were determined using the Student unpaired, two-tailed *t* test, assuming unequal variance. The complete data set is currently being submitted to ArrayExpress (EMBL–European Bioinformatics Institute, Hinxton, UK; <http://www.ebi.ac.uk/arrayexpress>) and will be available for public download by the second quarter of 2004. Accession numbers referencing this data set will be available on the International Life Sciences Health and Environmental Sciences Institute website (<http://hesi.ils.org/index.cfm?pubentityid=120>).

### Terminology

**Dye-swap hybridization pair.** The control cDNA sample labeled with Cy5 and the treated sample cDNA labeled with Cy3 were combined and hybridized on the first microarray. Conversely, the control sample cDNA labeled with Cy3 and the treated sample cDNA labeled with Cy5 are combined and hybridized on the second microarray of the pair.



**Figure 1.** Experimental design for sample treatment, processing, and data generation. Four separate biological experiments were run on four separate occasions. Each biological experiment represented two time points, at 4 hr posttreatment and at 24 hr posttreatment; each was treated with either vehicle or a low, medium, or high dose of BPDE for a total of 16 samples per experiment. cDNA synthesis, hybridization, washing, and scanning were performed for sets of eight microarrays at a time on separate dates for each time point. Two microarrays were performed with a dye reversal on each pair of RNA samples, resulting in eight microarrays per time point and a total of eight sets of microarray hybridizations performed on eight separate dates.

**Split-control hybridization.** The same cDNA control sample was split and labeled separately with either Cy3 or Cy5. The split Cy3- and Cy5-labeled samples were then combined and hybridized to the same microarray. Two such microarrays processed concurrently represent a split-control hybridization pair.

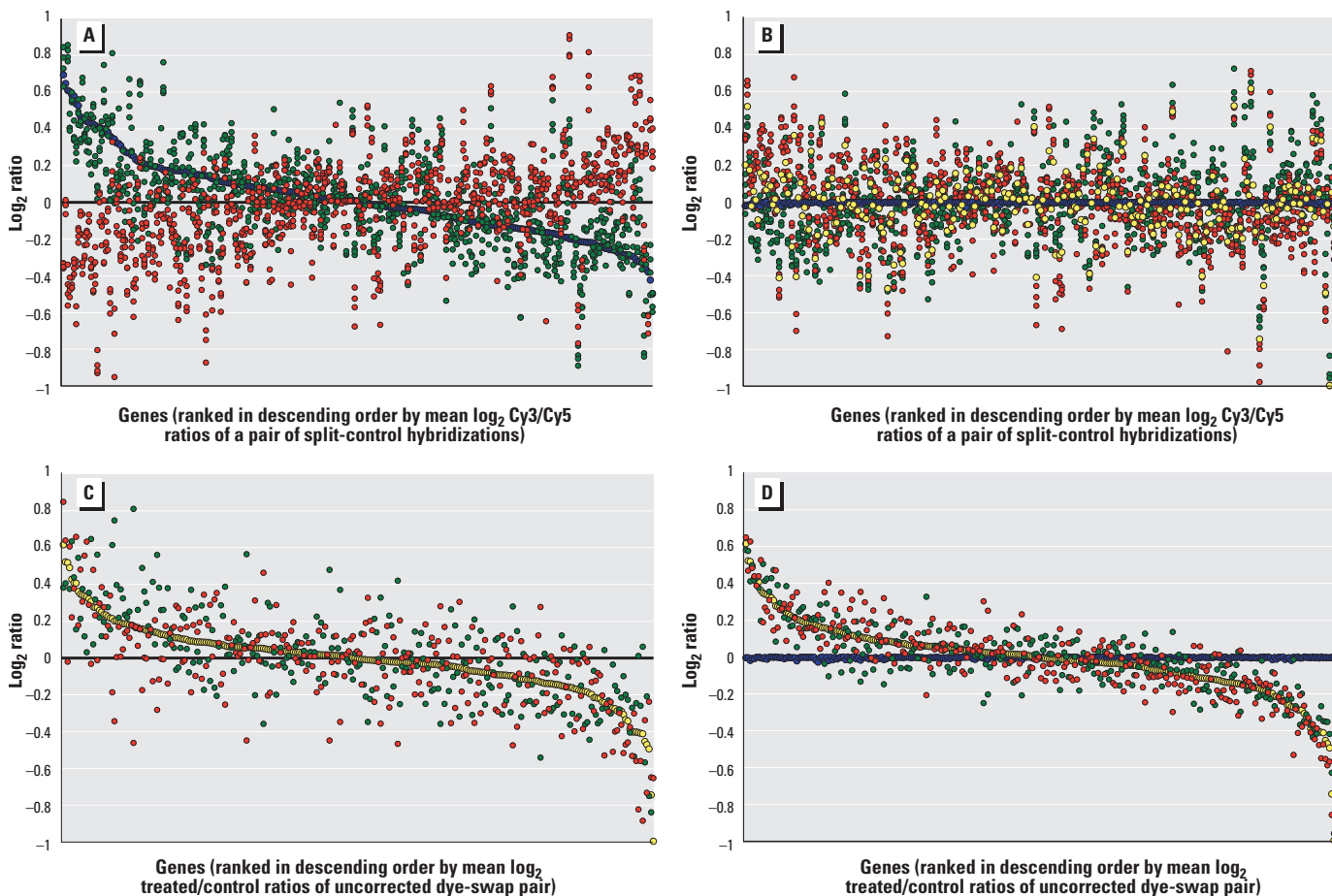
**Dye-bias correction.** The magnitude and direction of the apparent dye-bias effect on the spot intensity values of each dye-swap hybridization pair are systematically mitigated by the dye-bias correction factors

derived for each probe using the split-control cDNA microarray hybridization data values calculated for each probe. To correct for dye bias, the treated Cy3/control Cy5 values for each probe were divided by the mean of the split-control Cy3/Cy5 ratios for that same probe, or the treated Cy5/control Cy3 values for each probe were multiplied by the mean of the split-control Cy3/Cy5 ratios.

## Results

To demonstrate, accurately measure, and compensate for transcript-dependent dye

bias, two split-control cDNA microarrays were dedicated for each set of hybridizations performed on any given day. Figure 1 is a description of how the eight sets of hybridizations were processed separately to generate these data. In Figure 2 an example of a subset of LOWESS-normalized data from four cDNA microarray hybridizations are presented: one pair of split-control hybridizations, a concurrent treatment versus control hybridization, and the corresponding concurrent reverse-labeled and dye-swap treatment versus control



**Figure 2.** Dye-bias effect before and after correction in a representative split-control hybridization experiment. (A) A sample of total RNA from control cells was split; half was labeled with Cy3 and the other half with Cy5. Both halves were hybridized to a single cDNA microarray (split-control hybridization). The data were first LOWESS normalized to correct for differences between samples in dye incorporation and inherent fluorescent intensities between dyes. The log<sub>2</sub> Cy3/Cy5 ratios for a pair of split-control hybridizations were calculated and the means were ranked in descending order along the x-axis (blue circles). The individual spot fluorescent ratio data from two other arrays shown (green and red) are from a dye-swap pair of treated versus control hybridizations performed concurrently on the same day as two split-control arrays. One set of cDNA microarray data points (green) represents a hybridization in which the treated cDNA was labeled with Cy3 and the control cDNA was labeled with Cy5. The other set of data points (red) represents a hybridization in which the treated cDNA was labeled with Cy5 and the control cDNA was labeled with Cy3. For both of these microarrays, the identity of each spot is linked by the x-axis to the same gene of the split-control hybridization. (B) The magnitude and direction of the apparent dye-bias effect on the individual spot Cy3/Cy5 ratios are reduced by the dye-bias correction factors derived for each probe using the split-control cDNA microarray hybridization data values. The four corrected individual spot fluorescent Cy3/Cy5 ratio values for each probe are plotted from each of the pair of dye-swap treated versus control hybridizations (green and red) performed concurrently with the two split-control arrays. The simple means of the eight dye-swap-treated/control ratios are plotted (yellow) for each transcript linked by the x-axis to the same genes ranked as in panel A. (C) One set of data points represents the mean treated/control gene expression ratios of four values of each probe for the microarray of a dye-swap pair for which the treated sample cDNA was labeled with Cy3 and the control sample cDNA with Cy5 (green). The other set of data points (red) represents the mean-treated/control gene expression ratios of four values of each probe for the microarray of a dye-swap pair for which the treated sample cDNA was labeled with Cy5 and the control sample cDNA with Cy3. Neither data set was first corrected for dye bias. The genes are listed in descending order according to the mean of the eight log<sub>2</sub> treated/control ratios from the uncorrected dye-swap pair of microarrays (yellow). (D) Same data as in panel C, except the values of the individual treated/control expression ratios for each probe on each microarray were first dye-bias corrected in the same fashion as the Cy3/Cy5 ratios of the split-control hybridizations processed concurrently (blue).



hybridization. These data represent the control and corresponding 1  $\mu$ M BPDE-treated cultured TK6 cells at the 4-hr time point of experiment 3 of Table 1. The highly significant contribution of only the dye bias to the relative ratios between red and green fluorescence signals of a typical dual color cDNA microarray hybridization involving BPDE-treated cultured TK6 cells is seen in Figure 2A. Ranking in descending order the Cy3 to Cy5 log ratios for each probe of the split-control hybridization pair clearly shows a pattern to the data conveyed by the dye-bias effect. Some transcripts in this single experiment show much greater dye bias than others. For some, there is an apparent Cy3-labeled transcript bias, whereas for others the bias is greater for Cy5. These data show that dye bias alone contributes a strong signal to some transcripts (as high as 1.5-fold in the data of Figure 2). The data show that this artifact

**Table 1.** Ninety-nine percent confidence limits of  $\log_2$  Cy3/Cy5 ratios of eight pairs of split-control hybridizations before and after dye-bias corrections.<sup>a</sup>

Experiment no.	Before correction		After correction	
	Upper	Lower	Upper	Lower
1	0.60	-0.47	0.25	-0.25
2	0.65	-0.51	0.30	-0.34
3	0.69	-0.49	0.28	-0.27
4	0.49	-0.40	0.32	-0.36
5	0.41	-0.47	0.30	-0.32
6	0.59	-0.51	0.40	-0.38
7	0.53	-0.56	0.34	-0.38
8	0.70	-0.51	0.25	-0.23

<sup>a</sup>Each experiment was performed on separate dates using eight different control RNA samples collected from eight unique biological samples. See Figure 1 for a more detailed explanation of the experimental design. Each experiment represents the combined data from duplicate split-control hybridizations depicting a total of 2,800 ratio values (four replicates of each gene spotted on each array containing 350 genes for a total of 1,400 spots per array on each of two arrays).

could make difficult the ability to distinguish systematic experimental noise from a true biological effect of the test agent on relative transcript abundance if neither a split-control hybridization nor an experimental dye-swap replicate were performed. For any given set of experimental microarray hybridizations performed on the same day, we observed a similar trend in the data consistent with the split-control array when the treated sample was labeled with Cy3 and the control with Cy5. Conversely, an inverse trend was observed in the data from the dye swap hybridization array when the labeling was reversed and the treated sample was labeled with Cy5. Figure 2B shows the same data after mathematical correction for this dye bias on both of the experimental microarray hybridization experiments. These data were corrected for this dye bias by dividing the individual Cy3 (treated):Cy5 (control) microarray ratios by the mean split-control microarray Cy3/Cy5 ratio for each particular transcript, or conversely, by multiplying the individual Cy5 (treated):Cy3 (control) dye-swap microarray values by the mean split-control microarray Cy3/Cy5 ratio of each transcript. The pattern to the data seen in Figure 2A is clearly mitigated by this process. In Figure 2C and D, the genes are rank ordered, not by dye-bias strength measured with the split controls as in Figure 2A and B but rather by the mean ratio values of the eight determinations from the two dye-swap experimental arrays, comparing a treated with a control sample (in this case RNA from TK6 cells treated with 1  $\mu$ M BPDE compared with the control sample, both from the 4-hr time point). These data show that when only a single microarray is used and the treatment effect is calculated by first using the concurrent split-control data to dye-bias correct either the treated Cy3/control Cy5 or the

treated Cy5/control Cy3 microarray data, the results from each microarray agree well with the means calculated from the more conventional replicate dye-swap microarrays (Figure 2D). When the four individual data points on each single microarray are not first corrected for dye bias using the concurrent split-control hybridization array data, the agreement with the dye swap pair is poor (Figure 2C) by comparison. The linear correlation coefficients for comparisons between the means of the dye-swap replicate expression ratios with those of the array with the treated sample labeled with Cy5 is 0.70 before dye-bias correction but improves to 0.91 after dye-bias correction. For the array with the treated sample labeled with Cy3, the linear correlation coefficients improve from 0.75 before dye-bias correction to 0.89 after correction. If the results from each single microarray are not first dye-bias corrected, the data, which would reflect both the treatment effect as well as any dye-bias effects, show poorer agreement with the dye-swap replicate mean expression ratios, which should reflect more accurately only the true treatment effect.

The split-control hybridization data used to generate a portion of the data presented in Figure 2 are represented in Tables 1 and 2 as experiment 3. Table 1 lists the 99% confidence intervals of the  $\log_2$  Cy3/Cy5 ratios for all eight split-control hybridization pairs performed across eight separate dates. The 99% confidence interval varied substantially from day to day, indicating that the magnitude of the dye-bias effect can be variable. The spread in the data (Table 1) of the split-control hybridization pairs from experiments 3 and 8, for example, is quite broad compared with experiments 4 and 5. As might be expected, the  $\log_2$  confidence intervals are reduced for each experiment after dye-bias

**Table 2.** Correlations between data from eight pairs of split-control hybridizations performed on different dates.<sup>a</sup>

Experiment no.	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b	7a	7b	8a	8b
1a	1.00	0.80*	0.78	0.67	0.64	0.70	0.33	0.12	-0.14	-0.47	0.29	0.49	0.29	0.45	0.60	0.62
1b		1.00	0.61	0.81	0.69	0.79	0.06	0.21	-0.17	-0.15	0.49	0.56	0.19	0.52	0.66	0.66
2a			1.00	0.62*	0.61	0.63	0.43	0.20	-0.07	-0.39	0.22	0.21	0.39	0.46	0.62	0.54
2b				1.00	0.63	0.83	-0.06	0.52	-0.34	-0.10	0.38	0.40	0.02	0.65	0.67	0.59
3a					1.00	0.71*	0.04	0.13	-0.06	-0.19	0.41	0.33	0.40	0.48	0.56	0.58
3b						1.00	0.02	0.45	-0.28	-0.18	0.29	0.36	0.10	0.57	0.58	0.61
4a							1.00	0.00*	0.25	-0.28	-0.12	-0.10	0.40	-0.07	0.13	0.12
4b								1.00	-0.36	0.09	0.03	-0.12	-0.26	0.35	0.28	0.18
5a									1.00	0.40*	0.37	-0.13	0.48	0.03	0.04	-0.05
5b										1.00	0.32	-0.25	-0.01	0.15	0.01	-0.12
6a											1.00	0.29*	0.34	0.56	0.49	0.42
6b												1.00	-0.04	0.21	0.29	0.33
7a													1.00	0.34*	0.49	0.46
7b														1.00	0.75	0.73
8a															1.00	0.78*
8b																1.00

<sup>a</sup>The linear correlation coefficients of the  $\log_2$  Cy3/Cy5 ratio means ( $n = 4$  spots each) for 350 genes is listed for each pairing of the 16 microarrays processed either on the same day within a given experiment (indicated by asterisk) or on different dates across seven other experiments.

correction. Dye-bias correction of the data eliminates a large component of the variance, thus giving greater confidence in the control data for comparison with the corresponding experimental treatment arrays for testing statistical significance. The spread in the data of experiments 3 and 8 is substantially reduced after dye-bias correction to spreads similar to those of experiments 4 and 5 and all the other experiments listed in Table 1. The data in the table suggest that after dye-bias correction, few false positives would be identified at altered gene expression ratios greater than 1.3-fold with a significance of  $p < 0.01$ . The table indicates, generally, that without the inclusion of a split control to accurately measure and correct for dye bias, technical replicate measurements of 1.6-fold are needed to reach the same level of statistical confidence.

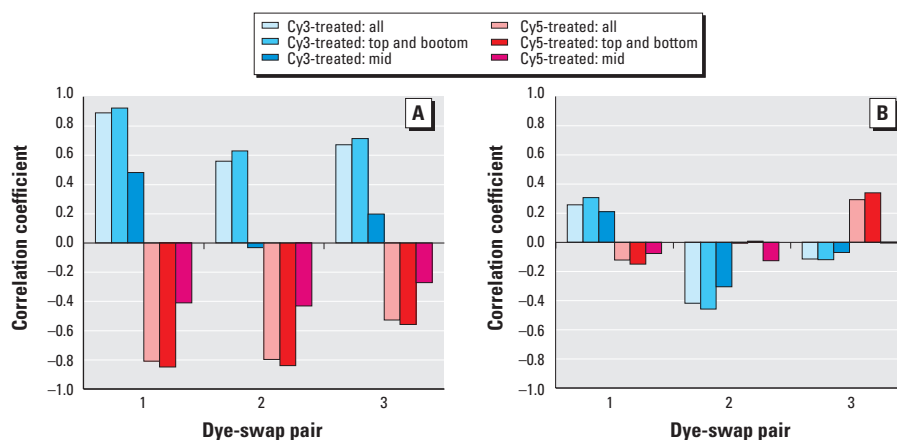
Table 2 is a list of the linear correlation coefficients for the individual Cy3/Cy5 ratios across experimental pairs for the same eight split-control hybridization experiments. Four of the eight hybridization pairs show a weak correlation among themselves within a given set and a weak correlation with the remaining experiments. These four experiments showing the weakest correlation with their corresponding partner of the same day also displayed the smallest 99% confidence intervals before the data were dye-bias corrected and therefore apparently the least amount of dye bias. Experiments 4, 5, 6, and 7 (Table 1) are the four experiments with the least spread in the data, as evidenced by the upper limit of the  $\log_2$  99% confidence intervals reaching 0.49, 0.41, 0.59, and 0.53, respectively. Each set of split-control hybridization Cy3/Cy5 values from these same experiments 4, 5, 6, and 7 shows poor correlation with their hybridization pair among all the concurrent pairs listed in Table 2, with linear correlation coefficients of 0.00, 0.40, 0.29, and 0.34, respectively. The Cy3/Cy5 ratios from the individual arrays of experiments 4, 5, 6, and 7 also correlate poorly with the ratios of any of the arrays of experiments 1, 2, 3, and 8. Conversely, the remaining four of the eight hybridization pairs show strong correlation with their corresponding partner of the pair and generally less correlation with those of the remaining experiments. Experiments 1, 2, 3, and 8 with the upper limit of the 99% confidence intervals reaching 0.60, 0.65, 0.69, and 0.70, respectively, show higher linear correlation coefficients between hybridization pairs with values of 0.80, 0.62, 0.71, and 0.78, respectively. Furthermore, the Cy3/Cy5 ratios from each of the individual arrays of experiments 1, 2, 3, and 8 generally

correlate better with all the others among these four experiments than with those of experiments 4, 5, 6 and 7. This suggests that when dye bias occurs, it may be seen among the same probes.

In Figure 3, the linear correlation coefficients were determined for three separate experimental BPDE treatments in dye-swap experiments performed on the same day with their corresponding split-control hybridization pair (experiment 3 of Tables 1 and 2). Before dye-bias correction (Figure 3A), a strong positive correlation existed between the split-control Cy3/Cy5 ratios and the experimental treatment dye swap hybridization Cy3/Cy5 ratios when the treated sample was labeled with Cy3. A strong negative correlation was seen when the treated sample was labeled with Cy5. When the genes were ranked as explained in Figure 2, then compartmentalized into thirds, the strongest correlation was seen within the top and bottom two-thirds of the entire gene set. After dye-bias correction (Figure 3B), the strong correlation between the treated and control fluorescence ratios and the split-control Cy3/Cy5 ratios was mitigated, resulting in a dye-bias corrected set of data essentially devoid of this systematically introduced experimental artifact.

To further investigate the consistency of the dye-bias effect for individual probes across different hybridization dates, the individual Cy3/Cy5 ratios were calculated for each transcript across eight separate

split-control hybridization experiments, and the probes were again ranked in descending order according to these Cy3/Cy5 ratios. Figure 4A demonstrates both a consistent trend in a dye-bias effect of individual probes across these eight split-control hybridizations as well as the variation that can be seen across experiments, as alluded to in Tables 1 and 2 where experiments with both large and small dye-bias effects have been noted. Some genes tended to be slightly biased in the same direction each time across experiment dates. Figure 4A shows that the same trend described in the single experiment of Figure 2 tended to persist across certain transcripts. The same set of genes from the same set of split-control hybridizations was ranked in Figure 4B in ascending order by the magnitude of their standard deviations. The genes identified in split-control hybridization experiments with the least amount of associated dye bias tended to have slightly lower variance than the genes showing the greatest amount of dye bias (Figure 4B). The genes with the lower variances, at the left of the figure, tended to have means that aligned closer to the  $\log_2$  zero value, whereas the means of those to the right showed greater dye bias by deviating further from the zero value. Figure 4C presents data for the 10 genes with the smallest and largest variance as well as the 10 genes with the greatest Cy3 or Cy5 bias. In some experiments the genes displaying the most consistent and



**Figure 3.** Linear correlation coefficients of individual treated versus control dye-swap hybridization pairs with corresponding split-control hybridization pairs. (A) Before dye bias. (B) After dye bias. The correlations of the mean  $\log_2$  ratios for 350 genes across three representative data sets are presented. Each of the three sets of microarrays was processed on the same day. The light blue bar represents the linear correlation coefficients of the eight  $\log_2$  Cy3/Cy5 ratios ( $n = 4$  spots/gene; 350 genes  $\times$  2 arrays) for each gene from two split-control microarray hybridizations with the  $\log_2$  treated/control ratio values where treated samples were labeled with Cy3. The pink bar represents the correlation coefficients of the same  $\log_2$  Cy3/Cy5 split-control ratios with the treated/control ratio values where treated sample was labeled with Cy5. Correlation coefficients were derived for three subsets of data from each microarray. The first bar represents the correlation coefficient across the entire set of genes (all). The second bar represents only the top third and the bottom third of the genes ranked in descending order by mean  $\log_2$  Cy3/Cy5 split-control ratios. The third bar represents the middle third of the ranked genes (mid).

strongest dye-bias effect nevertheless showed large variation across experiments. This figure shows that the magnitude of the dye-bias effect depends on both the particular experiment and on the relationship between the individual microarray probe and its target transcript. Note that the 10 Cy3-labeled gene transcripts and the 10 Cy5-labeled transcripts identified as showing the most consistent experimentally introduced dye bias tended to display the greatest effects in experiments 1, 2, 3, and 8, the same experiments similarly identified in Tables 1 and 2.

The graphs in Figure 5 demonstrate important similarities and distinctions between data derived from dye-swap technical replicate hybridizations and data derived from single microarray hybridizations without a dye swap that have been dye-bias corrected using concurrent split-control hybridization data. Figure 5A shows data from experiment 2 of Tables 1 and 2. These data derive from control cells and from cells harvested 24 hr after being treated 4 hr with 1  $\mu$ M BPDE. Two separate sets of data were calculated using each of the individual microarrays comprising the dye-swap pair after correction of each data point for dye bias using the data from the two concurrent split-control microarray

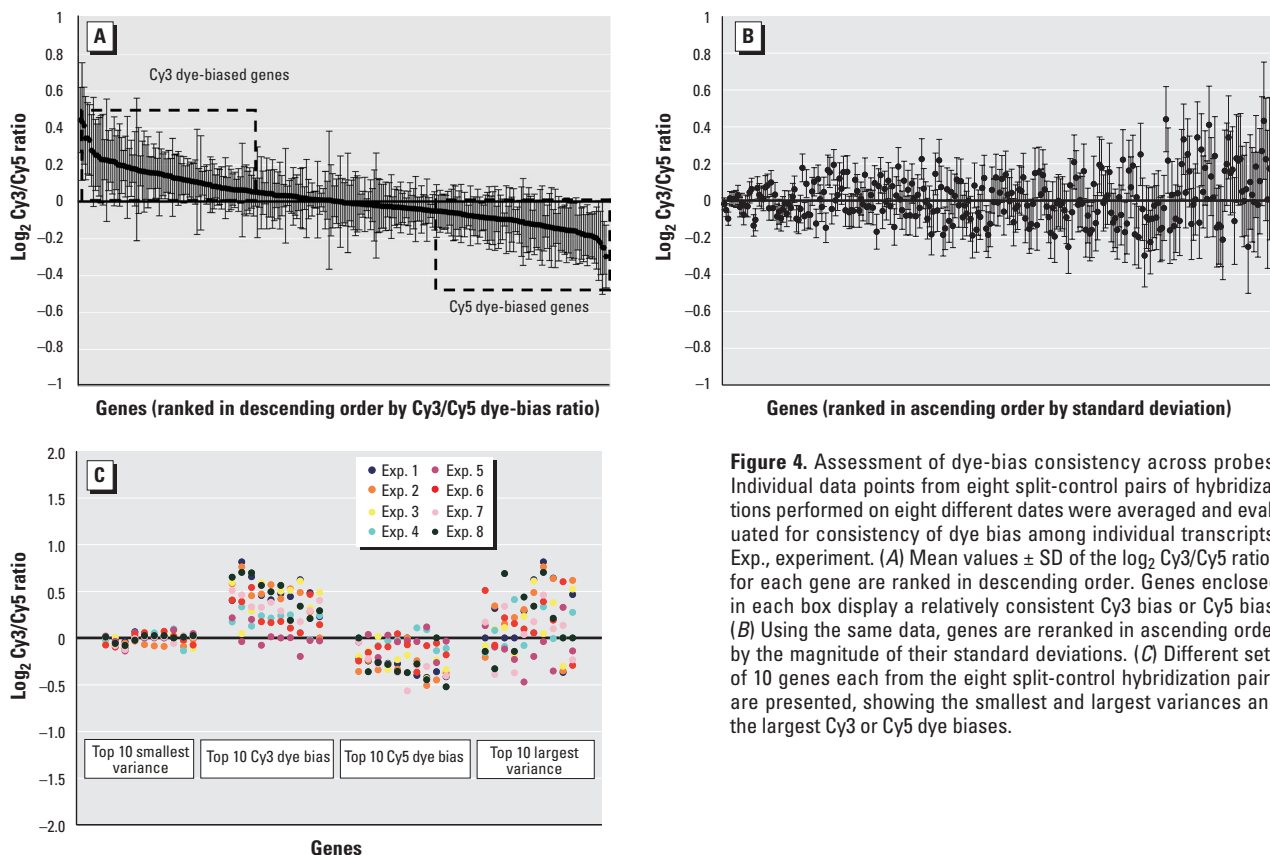
hybridizations. These two sets of results were compared with a third data set calculated simply as the mean and standard deviation of the eight measurements of treated/control expression ratios from the dye-swap pair (with no dye-bias correction). Little deviation from the dye-swap pair mean  $\log_2$  ratios was seen for each gene of the individual arrays when a dye-bias correction was first applied. Furthermore, when the data were displayed such that their mean values were brought to a common zero value (for presentation purposes only), but the actual variation around the means for each transcript was maintained, general reduction in standard deviations was seen when dye bias was first corrected (Figure 5A inset).

Figure 5B presents a subset of the same data from one of the microarrays in the dye-swap pair. When the mean  $\log_2$  ratios of the four gene expression ratios of this microarray were plotted against the mean  $\log_2$  ratios of the eight values from the two dye-swap microarrays, a stronger linear relationship was seen if the single microarray data were first corrected for dye bias. For all genes the linear correlation coefficient for this microarray improved from 0.924 to 0.953 after dye-bias correction. For those dye-bias corrected genes statistically different from

control at  $p < 0.001$ , the linear correlation coefficient improved further to 0.983. These results indicate that the data from a single dual-color experimental hybridization comparing a treated and control sample, when dye-bias corrected using concurrently processed split-control dye-bias sample information, will provide data that are comparably accurate to those of a replicate dye swap experiment.

## Discussion

Individual experimenters seek approaches to optimize accuracy of gene expression ratios from genome-scale microarray experiments as free as possible of experimental artifact introduced during sample processing. It has been shown previously, and we have confirmed in this study, that two-color directly labeled cDNA microarray hybridization protocols can systematically introduce a dye-bias error. Removal of this variable dye bias is important to obtain a more accurate set of statistically significant microarray data to enhance confidence in interpretation of the biological effects of the test agent. Many labs perform dye reversal or dye-swap experiments in which two microarray hybridizations are performed for each given dose and time data point, alternating the labeling of each treated and control sample



**Figure 4.** Assessment of dye-bias consistency across probes. Individual data points from eight split-control pairs of hybridizations performed on eight different dates were averaged and evaluated for consistency of dye bias among individual transcripts. Exp., experiment. (A) Mean values  $\pm$  SD of the  $\log_2$  Cy3/Cy5 ratios for each gene are ranked in descending order. Genes enclosed in each box display a relatively consistent Cy3 bias or Cy5 bias. (B) Using the same data, genes are reranked in ascending order by the magnitude of their standard deviations. (C) Different sets of 10 genes each from the eight split-control hybridization pairs are presented, showing the smallest and largest variances and the largest Cy3 or Cy5 dye biases.

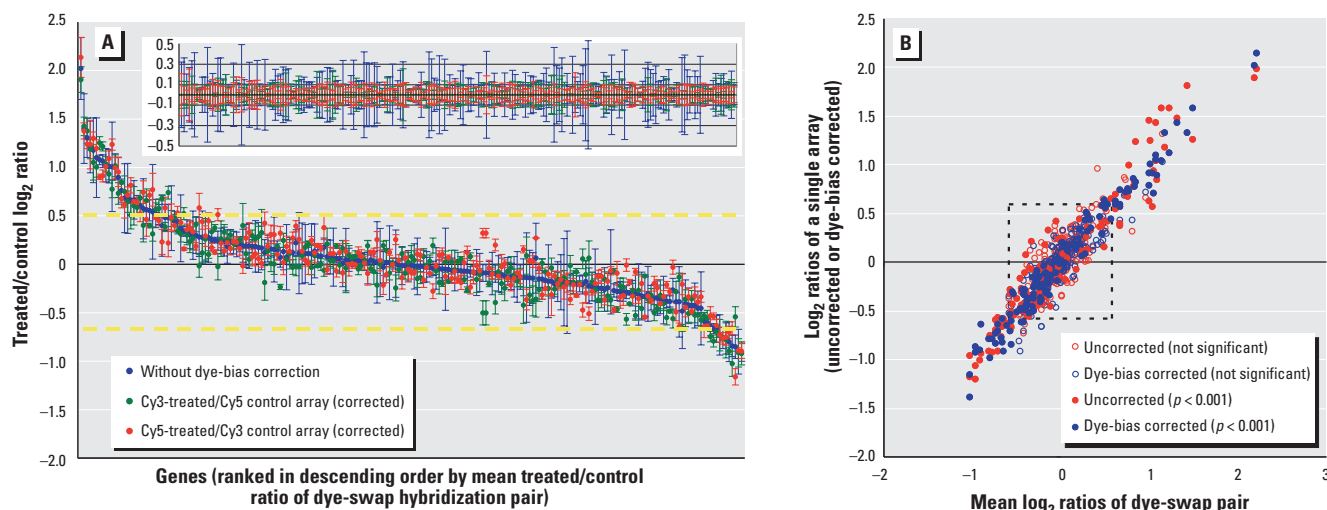
with the Cy3 and Cy5 dyes (Dobbin et al. 2003; Yang et al. 2001). The microarray results are then averaged between the dye-swap pair, eliminating the dye-bias artifact. We propose an alternate approach that allows the use of only one microarray per dose or time point within a set of microarray hybridizations and maintains the accurate detection of significantly altered genes. Although the dye-swap approach to dye-bias correction is acceptable and widely practiced, it can be costly. Furthermore, the proposed alternate approach offers improved technical replicate precision, which may be an important consideration, depending on whether the precision of technical replicates is considered by researchers in selecting genes for further investigation of biological significance. Because of the fairly large numbers of cell culture flasks that must be harvested for sufficient cellular material for RNA analyses, there is a tendency to rely on technical replicate measurements to help filter out nonresponding genes and identify those altered transcripts that can survive confidence-testing thresholds. We reasoned that biologically relevant alterations devoid of technical artifact could be identified to then evaluate for robustness of biological replication. In common practice, investigators tend to consider both a fold-change cutoff value as well as statistical confidence  $p$ -values to identify genes of interest for further biological evaluation and investigation (Akerman et al. In press; Wolfinger et al. 2001). From this perspective, accurate

calculation of the extent of change from a true control value gains as much importance as the impact of variability on statistical significance. Dye-bias correction of each of the single individual microarrays of a dye-swap hybridization pair yields mean ratio values for each transcript that differ very little from means of the two microarrays of the dye-swap pair, and therefore this dye-bias correction approach using split-control hybridization data maintains high accuracy. Dye-bias correction further reduces the standard deviations of the gene expression ratios. This increase in precision may allow more genes to be identified from technical replicate measures as statistically significant, depending on the technical measurement filtering criteria of an individual experimenter.

For any given set of dual-labeled cDNA microarray hybridizations performed on the same day, we saw a dye bias that varied in intensity across genes. When dye bias was noted in an experiment, certain transcript targets tended to be affected more consistently than others. The magnitude of the dye bias was variable across eight pairs of split-control hybridizations on eight separate dates, but the strongest correlation between split-control hybridization sets was generally seen when the experiments were performed on the same day. On some days there was evidence of a greater dye-bias effect than on other days. Because the dye bias can be measured, it can be mathematically mitigated by performing a pair of split-control hybridization arrays on the same day, then using

these data to correct the corresponding concurrently processed experimental treated/control arrays. Furthermore, experimental dye-swap hybridization data derived from treated and control cell culture samples can be highly positively correlated with the magnitude and direction of the same genes of the corresponding (Cy3/Cy5) split-control hybridization pair when the treated sample is labeled with Cy3. Conversely, dye-swap hybridizations can be highly negatively correlated with the magnitude and direction of the control hybridization when the treated samples are labeled with Cy5. After dye-bias correction, these strong correlations are mitigated over the entire gene set, indicating the successful removal of a strong experimental artifact. Because the contribution of dye bias to the signal intensity ratios in some experiments will be high for certain transcripts, the impact of the mathematical correction on accurate study-data interpretation could be great. On days when the contribution of dye bias to signal intensity ratios is low, less impact on overall data interpretation would be expected. If data were averaged without first always correcting for the variable strength of dye-bias signal, there would be greater measurement variability among certain transcripts, and therefore greater difficulty for these transcripts to achieve a level of statistical significance needed to drive support for biological consideration.

Given the numerous variables impacting microarray performance, irreproducibility of data generation across laboratories may be



**Figure 5.** Dye-bias correction maintains expression ratio accuracy and reduces technical measurement variance. (A) Data from a representative pair of dye-swap hybridizations are shown. Log<sub>2</sub> treated/control ratios are determined for each gene (four spots per microarray) and then either corrected for dye bias (red and green) using the corresponding ratios from the concurrently processed split-control microarray pair or not corrected (blue) before calculating mean  $\pm$  SD of the treated/control expression ratio for each gene. The genes are ranked in descending order by the mean treated/control log<sub>2</sub> ratios calculated from the eight dye-swap hybridization pair values (blue). The yellow dashed line indicates the 1.5-fold thresholds. The inset graph represents only the true standard deviations of the eight measurements for each dye-swap hybridization pair transcript (blue) or of the four measurements for each dye-bias corrected treated/control experimental microarray (red and green). The log<sub>2</sub> means for these same eight measurements are all set to zero for presentation purposes only. (B) The mean log<sub>2</sub> ratios from the microarray for which the treated sample is labeled with Cy3 and the control with Cy5 are plotted against the mean log<sub>2</sub> ratios of the dye-swap pair of microarrays. Data for the single microarray are either first dye-bias corrected (blue) or not corrected (red).



highly likely using the same microarray platforms if experimenters each develop their own set of labeling and hybridization conditions. Adherence to standardized labeling, hybridization, and data generation protocols that have been optimized by commercial microarray platform providers will be important. It is unclear whether small changes in such protocols could further minimize or even worsen dye bias with the specific cDNA microarray used in these experiments. It is also unclear whether the same magnitude and variability of dye bias shown here can be generalized and expected across all two-color directly labeled or indirectly labeled cDNA or oligonucleotide microarrays.

For a simple experimental design requiring a control and three concentrations of test agent, all harvested at two time points, eight samples of RNA will be generated for microarray analyses (Figure 1). The conventional dye-swap approach to dye-bias correction would stipulate the use of two arrays at each dose and time point to derive one set of values constituting a single biological replicate value for each transcript. When repeated

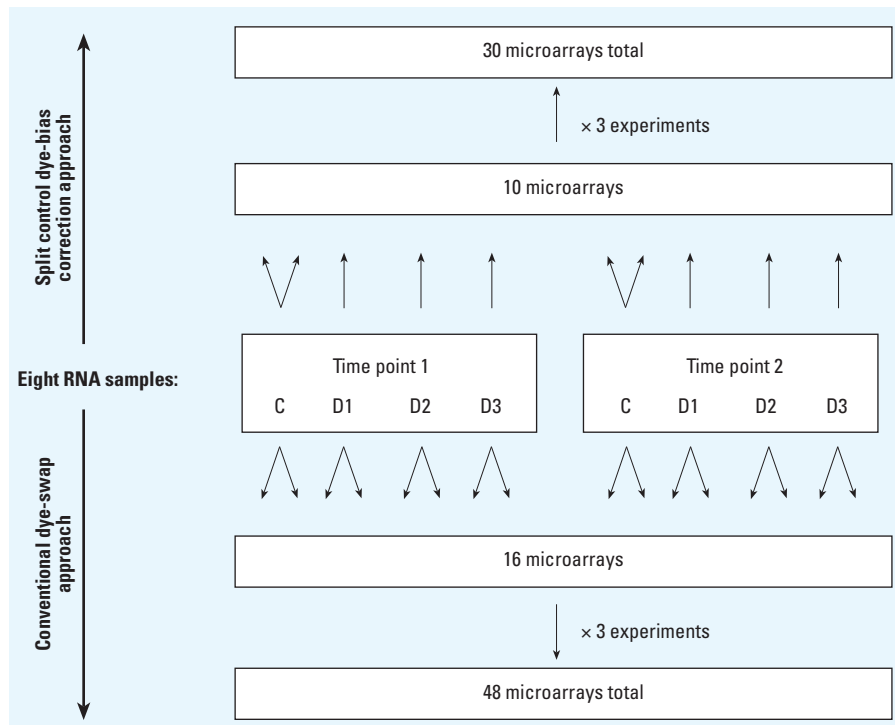
two additional times, a total of 48 microarrays would be needed to complete the study with three biological replicates (Figure 6). For the split-control hybridization dye-bias correction approach we have described, two microarrays would be needed for each of the split controls at both time points, but for each of the treated samples at the three doses and the two times, only one microarray hybridization is employed. This design for deriving microarray gene expression results from the same eight samples would require only 10 concurrently processed microarrays per biological replicate and therefore 30 rather than 48 microarrays to complete the study with three biological replicates (Figure 6).

In summary, our approach to mathematical dye-bias correction of dual-labeled cDNA microarray hybridization experiments using concurrent split-control hybridization data has little effect on mean gene expression ratio values compared with a more conventional replicate dye-swap approach to mitigate dye bias and establish accuracy but greatly improves the standard deviations of technical replicate

measurements. The dye-swap replicate approach requires an expenditure of nearly 40% more microarrays and associated reagents. The magnitude of dye bias is variable across experiments but reproducible within any given set of dual-labeled cDNA microarray experiments processed concurrently. By running a pair of split-control hybridizations, dye bias can be mitigated, resulting in an accurate set of gene expression microarray data at reduced cost.

## REFERENCES

- Akerman GS, Rosenzweig BA, Domon OE, McGarrity LJ, Blankenship LR, Tsai CA, et al. In press. Gene expression profiles and genetic damage in benzo[a]pyrene diol epoxide-exposed TK6 cells. *Mutat Res*.
- Cleveland WS, Grosse E, Shyu WM. 1992. Local regression models. In: *Statistical Models in S*. (Chambers JM, Hastie TJ, eds). Pacific Grove, CA:Wadsworth & Brooks/Cole. 309–376.
- Dobbin K, Shih JH, Simon R. 2003. Statistical design of reverse dye microarrays. *Bioinformatics* 19:803–810.
- Goryachev AB, Macgregor PF, Edwards AM. 2001. Unfolding of microarray data. *J Comput Biol* 8:443–461.
- Ideker T, Thorsson V, Siegel AF, Hood LE. 2000. Testing for differentially-expressed genes by maximum-likelihood analysis of microarray data. *J Comput Biol* 7:805–817.
- Kerr MK, Martin M, Churchill GA. 2001. Analysis of variance for gene expression microarray data. *J Comput Biol* 7:819–837.
- Qiagen, Inc. 1999. *RNeasy Midi/Maxi Handbook*. Valencia, CA:Qiagen, Inc.
- Quackenbush J. 2002. Microarray data normalization and transformation. *Nat Genet* 32(suppl):496–501.
- Tseng GC, Oh M, Rohlin L, Liao J, Wong WH. 2001. Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res* 29:2549–2557.
- Wang X, Ghosh S, Guo S. 2001. Quantitative quality control in microarray image processing and data acquisition. *Nucleic Acids Res* 29:E75.
- Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, Bushel P, et al. 2001. Assessing gene significance from cDNA microarray expression data via mixed models. *J Comput Biol* 8:625–637.
- Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, et al. 2002. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 30:E15.
- Yang YH, Dudoit S, Luu P, Speed TP. 2001. Normalization for cDNA microarray data. In: *Microarrays: Optical Technologies and Informatics* (Bittner ML, Chen Y, Dorsel AN, Dougherty ER, eds). Vol 4266. Proceedings of SPIE.
- Zhou Y, Gwadyr FG, Reinhold WC, Miller LD, Smith LH, Scherf U, et al. 2002. Transcriptional regulation of mitotic genes by camptothecin-induced DNA damage: microarray analysis of dose- and time-dependent effects. *Cancer Res* 62:1688–1695.



**Figure 6.** Reduction in microarray resources using the split-control dye-bias correction approach. By adopting the split-control dye-bias correction approach as opposed to the conventional dye swap approach, the total number of microarrays and associated reagents can be reduced from 48 arrays to 30. Only one microarray is needed from each of the corresponding treated samples (D1, D2, D3) as long as replicate microarrays are performed on each of the respective control samples (C), which we refer to as the split-control hybridization, to accurately measure and calculate transcript-dependent dye-bias correction factors.