J. Microbiol. Biotechnol. (2005), **15**(6), 1377–1383

JOURNAL OF MICROBIOLOGY AND BIOTECHNOLOGY

Simple Method to Correct Gene-Specific Dye Bias from Partial Dye Swap Information of a DNA Microarray Experiment

KIM, BYUNG SOO[1], SOO-JIN KANG[1], SAET-BYUL LEE[2,3], WON HWANG[2], AND KUN-SOO KIM[2]*

[1]*Department of Applied Statistics, Yonsei University, Seoul 120-749, Korea*
[2]*Department of Life Science and Interdisciplinary Program of Integrated Biotechnology, Sogang University, Seoul 121-742, Korea*
[3]*Current address: Natural Products Division, Korea Institute of Science and Technology, Gangneung Institute, Gangneung, Gangwon-Do 210-340, Korea*

Received: July 20, 2005
Accepted: September 12, 2005

Abstract In a cDNA microarray experiment using Cy3 and Cy5 as labeling agents, particularly for the direct design, cDNAs from some genes incorporate one dye more efficiently than the other, which is referred to as the gene-specific dye bias. Dye-swaps, in which two dyes are switched on replicate arrays, are commonly used to control the gene-specific dye bias. We developed a simple procedure to extract the gene-specific dye bias information from a partial dye swap experiment. We detected gene-specific dye bias by identifying outliers in an X-Y plane, where the X axis represents the average log-ratio from two sets of dye swap pairs and the Y axis exhibits the average log ratio of four forward labeled arrays. We used this information for detecting differentially expressed genes, of which the additionally detected genes were validated by real-time RT-PCR.

Key words: cDNA microarray, differentially expressed genes, direct design, dye-swap, gene-specific dye bias, statistical methods

There are two general experimental designs in the two-color spotted cDNA microarry experiment: direct and indirect. The indirect design, also known as the reference design, uses an aliquot of a reference cDNA generated from mRNA as one of the samples in the array. Each cDNA sample of interest is hybridized together with the reference sample on the same array. Thus, the hybridization intensities of several cDNA samples of interest are indirectly estimated in a relative sense by setting the reference as a standard. The disadvantage of the reference design is that half of the hybridizations are used for the reference samples, which are of no real interest in most cases [19]. Designs without reference have been recently proposed as alternatives, and these are referred to as direct designs.

The direct design measures the hybridization intensities of two samples of interest directly on the same array. A direct design can be further classified as a balanced block design or a loop design depending on, among other things, the allocation of samples to the array and dye labels to the samples [6, 7, 20]. It has been observed that some genes incorporate one dye more efficiently than others [8, 11, 18, 23] and thus tend to appear brighter in one color. This effect in general is referred to as the dye bias. Dobbin *et al.* [5] classified dye bias into four types: (1) dye bias is the same for all genes on an array; (2) dye bias depends on the overall spot intensity; (3) dye bias occurs in a subset of genes, of which the size is consistent for the same gene across samples; (4) dye bias depends on both the gene and the sample. Normalization procedures can elimimates bias of types (1) and (2), but not the bias of type (3). The dye bias of type (4), which is referred to as the gene- and sample-specific dye bias, is difficult to eliminate. Our primary interest in this report was concerned with the bias of type (3), which is referred to as the gene-specific dye bias. The gene-specific dye bias has been reported since the early stage of the cDNA microarray experiment [3, 6–8, 11, 16, 18, 23]. This gene-specific dye bias and systematic effect such as technical artifacts occur simultaneously in a two-color microarray experiment. Normalization can control the systematic effect. Dye-swaps, in which two dyes are switched between each other on replicated arrays, are commonly employed, particularly in the direct design, to control the gene-specific dye bias.

There are three layers of replication in a two-color microarray experiment; i.e., biological units, technical replicates, and duplicate spots. These three layers comprise major sources of variation in a microarray experiment [3]. Technical replicate measurements of each biological unit

*Corresponding author
Phone: 82-2-705-8460; Fax: 82-2-704-3601;
E-mail: kskim@sogang.ac.kr

are generally incorporated at the discretion of the experimenter. The number of technical replicates may depend on, among other things, the amount of sample available, the budget of the experimenter, whether the array has duplicate spots, and whether samples constitute a set of matched pairs. The relative amount of technical to biological replications perhaps depends critically on the cost of the array hybridization relative to the sample.

Matched pair samples often consist of control (C) and treated (T) samples from a batch of cultured cells. It is not clear how one can designate a biological unit in a microarray experiment using cells in culture. We considered a biological unit in this note as a batch of cells independently cultured at a different time point. Forward labeling refers to dye orientation, which places Cy3 at the control sample (C) and Cy5 at the treated sample (T). The backward labeling assigns dyes in the opposite direction.

Let s ($1 \leq s \leq n$) denote the number of biological units from which mRNA samples are extracted and labeled as forward and backward on different arrays. For simplicity, we assume that at most two technical replicates are available for each biological unit. Then, n-s stands for the number of biological units whose samples appear only once on an array. Let $m=n+s$ represent the total number of arrays. Dobbin *et al.* [6] showed that the most efficient design for matched pair samples is the design that runs each sample once in an array and balances the samples with respect to the dye. Design 1 in Table 1 illustrates the most efficient design for n=6, s=2. As the number of biological replicates increases within a balanced block design with s=0, we obtain the greatest statistical efficiency on gene-specific dye bias. When s=n and hence m=2n, each of n biological units is labeled both forward and backward. Thus we have a dye-swap hybridization pair for each individual array. We refer to this experiment as a full scale dye-swap experiment. However, in contrast to the general belief, it is not required in the matched pair samples to have a dye-swap pair for every individual array [6]. This means that, even with s<n, we may achieve the most efficient design and Design 1 of Table 1 is an example. Design 2 is what is often performed in a microarray laboratory without balancing the samples with respect to the dye. We refer to the experiment with s<n as the partial dye-swap experiment. The aim of this report is to show how one can extract the gene-specific dye bias from a partial dye-swap experiment of Design 2 for detecting differentially expressed (DE) genes between control and treated samples. A simple modification of our approach can be applied to Design I.

Rosenzweig *et al.* [18] proposed what is called a split-control hybridization approach as an alternative to the full-scale dye-swap design in a dose-time/response experiment of benzo[*a*]yrene-*trans*-7,8,dihydrodiol-9,10-epoxide (BPDE), an environmentally hazardous chemical. A split-control hybridization, sometimes called a self-self hybridization [16], is defined as follows: The control sample is split and labeled separately with Cy3 and Cy5. The split Cy3- and Cy5-labeled samples are then combined and hybridized to the same microarray. Two such microarrays represent a split-control hybridization pair. Rosenzweig *et al.* [18] reported that a concurrently processed split-control hybridization approach provides information with fewer microarrays on the gene-specific dye bias as accurately as that of a full-scale dye-swap experiment. Thus, they concluded that incorporating split-control hybridization can eliminate the need for the full-scale dye-swap experiment. The basic assumption of this split-control hybridization approach is that only a few genes between the control and the treated samples are differentially expressed; therefore, the gene-specific dye information derived from the control sample can be translated to the treated sample. There are several reports, however, particularly in the cancer microarray, that a substantial number of genes are differentially expressed between tumor and normal tissues [1, 12]. Furthermore, incorporating concurrent split-control hybridization may not be a viable option in a certain design. For example, the concurrent split-control hybridization would not be a practical alternative of a full-scale dye-swap, when we have two comparison groups (e.g., C and T) with an upper limit of two technical replicates for a biological unit under each condition (e.g., C or T).

A number of bacteria associated with eukaryotic hosts employ diffusable compounds as signals to sense their population density, thereby modulating the expression of

Table 1. Design 1 is the most efficient design for matched pair samples for n=6 and s=2. Design 2 is what is often performed in a laboratory without balancing samples with respect to the dye. C and T stand for a control sample and a treated sample, respectively. Each thick square represents a microarray.

	Biological units											
	1		2		3		4		5		6	
	C	T	C	T	C	T	C	T	C	T	C	T
Design 1	Cy3	Cy5	Cy3	Cy5	Cy3	Cy5	Cy3	Cy5	Cy5	Cy3	Cy5	Cy3
	Cy5	Cy3	Cy5	Cy3								
Design 2	Cy3	Cy5	Cy3	Cy5	Cy3	Cy5	Cy3	Cy5	Cy3	Cy5	Cy3	Cy5
	Cy5	Cy3	Cy5	Cy3								

sets of genes involved in physiological responses associated with survival, propagation, and/or virulence [4, 9, 19]. *V. vulnificus* is an opportunistic human pathogen that causes severe wound infections and primary septicemia [21]. Several reports have suggested that *V. vulnificus* possesses a *V. harveyi*-type quorum sensing system rather than the canonical *V. fischeri* system [13, 17]. We recently found that *V. vulnificus* produces cyclic(L-Phe-Pro) as a signal molecule, which affects the expression of numerous genes in the pathogen. In screening genes whose expression is modulated by the compound, we are employing a DNA microarray by comparing expression profiles of cells grown in the presence and absence of exogenous cyclic(L-Phe-Pro) in the medium. In the microarray experiment, we observed a gene-specific dye bias, thereby necessitating the development of a statistical method that facilitates the exploitation of dye-swap information. We performed a cDNA microarray of Design 2 in Table 1 to identify genes modulated by the signal molecule produced by *V. vulnificus*.

We developed a simple idea of deriving gene-specific dye bias information by comparing a summary gene expression ratio based on two sets of dye-swap pairs corresponding to biological units 1 and 2 against a summary gene expression ratio based on four forward labeled arrays associated with biological units 3 to 6 of Design 2 in Table 1. Correction factors were derived for genes with large gene-specific dye bias.

We compared the list of differentially expressed (DE) genes before correcting for the gene-specific dye bias against the list of DE genes after correcting for the gene-specific dye bias. We performed an RT-PCR for 4 DE genes of which 3 were newly detected after correcting for the gene-specific dye bias and confirmed that all 4 genes were differentially expressed. This confirmation indicated that our simple method of correcting for the gene-specific dye bias was effective in detecting DE genes.

MATERIALS AND METHODS

Cell Culture

V. vulnificus MO6-24/O [28] cells were cultured as described previously [14]. Briefly, cells were grown overnight in Luria-Bertani (LB) broth at 28°C. Next day, the cells were inoculated into fresh LB broth and incubated for 2 h. Then, the culture was split into two portions, and one was added with cyclic(Phe-Pro) at a final concentration of 1 mM for induction (the treated group) and the other was not added (the control group). These cultures were further incubated at 28°C for 4 h, and cells were harvested by centrifugation to isolate the total RNA. We used independent batches cultured on 6 different days for each of 6 microarray experiments.

Microarray and Experimental Design

Molecular manipulations of nucleic acids from *V. vulnificus* were carried out as described previously. We prepared a cDNA microarray containing representatives of 162 genes of *V. vulnificus* MO6-24/O generated by PCR. The genes involved were associated with virulence, quorum-sensing, the stringent response, transport, regulation, and signal transduction. Homologues of *V. cholerae* genes encoding similar functions were also included. A DNA fragment spanning approximately 700-bp of the 3' terminal region of each gene was prepared by PCR using specific primers based on the genome sequences of *V. vulnificus* in GenBank (accession number NC_004459 and NC_004460). The PCR products were purified with a Qiagen PCR purification kit (Qiagen, Valencia, CA, U.S.A.) and printed in quadruplicate on Corning GAPS™ II slides (Corning, NY, U.S.A.) using a DNA microarrayer (Affymetrix 427™, Inc., Santa Clara, CA, U.S.A.). As a result, each microarray slide was printed with a total of 648 spots.

Purification of Total RNA and Preparation of Dye-Labeled cDNA

Total RNAs were purified using an RNeasy Mini kit (Qiagen). Synthesis of cDNA and labeling with Cy3 and Cy5 followed the protocols provided by TIGR (http://pfgrc.tigr.org/protocols.shtml). We carried out a total of 8 array experiments (6 independent biological replicates, and 2 dye-swaps). The labeling of 2 dyes in control and treated samples were carried out as indicated in Table 1 (Design 2).

Hybridization and Analysis of Microarray Experiments

The hybridization was carried out following the TIGR protocol, and the results of hybridization were analyzed with an Axon scanner (GenePix® 4000B, Axon, UnionCity, CA, U.S.A.) using the analyzing program GenePix Pro 5.0.

Real-Time Reverse Transcription Polymerase Chain Reaction (RT-PCR)

Two grams of total RNA was reverse transcribed using AMV-reverse transcriptase (Promega, Madison, WI, U.S.A.) to synthesize cDNA. ABI PRISM Primer Express software (PE Applied Biosystems, Foster City, CA, U.S.A.) was used for designing primers for a real time PCR. The primer sets used are listed in Table 2. The relative amount of each mRNA was estimated using the amount of mRNA of the gene *pfkA* encoding 6-phosphofructokinase, which is a house-keeping gene and known to be expressed constitutively [10], as 1.0. Real-time PCR was carried out using a SYBR Green PCR master mix (PE Applied Biosystems) and Applied Biosystems 7500 Real-Time PCR System (PE Applied Biosystems). The PCR was carried out for 40 cycles as follows; hybridization for 2 min at 50°C, activation for 10 min at 95°C, denaturation for 15 sec at 95°C, and annealing/extension for 1 min at 60°C. The results were

Table 2. DNA nucleotide sequences of primers used for real-time RT-PCRs.

Gene ID	Nucleotide sequences of primers used for RT-PCR	
vv10189	Forward	5' CATTCTGCAGGTGAAGGTGA 3'
	Reverse	5' TGAGGAACTCTTTGCCGTCT 3'
vv12604	Forward	5' GGCATGCGTGTAATGGTAAG 3'
	Reverse	5' TCATCGCTTAGCACCACTTC 3'
vv10223	Forward	5' CCAACGCGACAGGTCATATTG 3'
	Reverse	5' CACCAAGCGGATCTGAAATGA 3'
vv10306	Forward	5' TGTGCTGGCCAAACCCGAA 3'
	Reverse	5' TGTTCATGGCAGACGACGTAA 3'
pfkA	Forward	5' CGTGTTCTTGCGTCTCGTAT 3'
	Reverse	5' ACGCTTCATGTTCTCGATTG 3'

analyzed using the comparative cycle threshold (C_t) method. Relative quantification was performed, comparing gene expression between the two cell lines and using *pfkA* for normalization. Relative expression was calculated as $2^{-\Delta\Delta Ct}$, where ΔC_t was calculated by subtracting the average normalization C_t from the average target C_t value. The $\Delta\Delta C_t$ was then obtained by the difference in ΔC_t values obtained for each of the two cell lines.

Statistical Analysis

Data Filtering and Normalization. Each array consists of 648 spots, in which 162 genes were spotted in quadruplicate. We filtered out poor quality genes on an array by excluding spots whose foreground intensities were lower than the background intensities at either channel. The proportion of remaining spots ranged from 42% to 80% throughout 8 arrays. We normalized the log intensity ratio, $M \equiv \log_2(R/G)$, using the within-print tip group, intensity dependent normalization [26]. With only 648 spots on an array, we did not have enough spots to run a single (smoothing parameter) loess normalization. We defined no missing proportion (NMP) of a spot as the proportion of valid observations out of the total number of arrays. For example, if a spot has valid observations for 5 arrays out of 8, its NMP is 0.625. We used 0.75 for the cut point of NMP to delete spots containing missing values for more than 25% of the total number of observations. This filtering procedure yielded 293 spots. We employed the k-nearest neighbor (k=10) method for the imputation of missing values [22]. We averaged values for multiple spots. The numbers of singly spotted, and duplicated, triplicated, and quadruplicated spots were 13, 7, 10, and 59, respectively. These frequencies added up to make up 89 distinct genes.

Now, we have a data set represented by a 89×8 matrix, where 89 implies the number of genes and 8 stands for the number of arrays. Let M values of a gene for 8 arrays be denoted by M_1, M_1^*, M_2, M_2^*, M_3, ..., M_6, where M_1^* and M_2^* represent the dye-swap versions of M_1 and M_2, respectively.

Two sets of dye-swap pairs and four forward labeled microarrys were based on six different batches. We consider that a batch of cells cultured on a different day comprises an independent experimental unit. Therefore, a set of dye-swap pairs that was based on the same batch needs to be averaged to yield a single independent observation. Define $M_1^{DS}=(M_1-M_1^*)/2$ and $M_2^{DS}=(M_2-M_2^*)/2$. We thus have six independent observations, namely, M_1^{DS}, M_2^{DS}, M_3, ..., M_6, for each of 89 genes. The primary purpose of the dye-swap experiment is to balance out the gene-specific dye bias, which may exist for some genes. Since we have two sets of dye-swap pairs and four forward labeled arrays, we could detect gene-specific dye bias by identifying outliers in the scatterplot of $\left(\frac{M_1^{DS}+M_2^{DS}}{2}, \frac{1}{4}\sum_{i=3}^{6} M_i\right)$. For each gene we calculate $D=\frac{M_1^{DS}+M_2^{DS}}{2}-\frac{1}{4}\sum_{i=3}^{6} M_i$. Let D^g denote the D value for the g-th gene, g=1, ..., 89. Let $\bar{D}$ and s_D denote the sample mean and sample standard deviation, respectively, based on D^1, ..., D^{89}. We identify as outlying genes in the scatterplot of $\left(\frac{M_1^{DS}+M_2^{DS}}{2}, \frac{1}{4}\sum_{i=3}^{6} M_i\right)$ those genes whose indices belong to $\{g: |D^g-\bar{D}|>2s_D\}$. For these outlying genes we adjust M values for each of four forward labeled arrays by adding the following correction factor of Eq. (1).

$$\frac{M_1^{DS}+M_2^{DS}}{2}-\frac{1}{4}\sum_{i=3}^{6} M_i \quad (1)$$

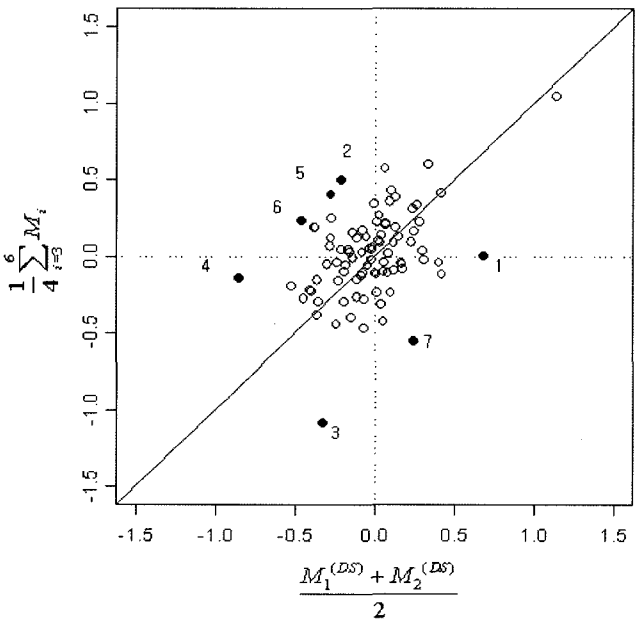

Fig. 1. The scatterplot of dye-swap intensities against non-dye-swap intensities. Each gene corresponds to a spot.
For each gene, we calculated the distance between x and y coordinates, from which we computed the mean and standard deviation. We then identified genes whose distances were deviated by more than two standard deviations from the mean. Seven outlying genes are indicated in the figure using the same ID number as in Table 3.

Table 3. List of genes showing gene-specific dye bias and their correction factors of Equation (1).

ID number[a]	Gene Name	Gene ID[b]	Correction factor
1	HtpG	vv10189	0.6767
2	FlgE	vv10223	-0.7048
3	FOG: EAL domain	vv11132	0.7628
4	Superoxide dismutase	vv11252	-0.7111
5	Type 4 prepilin	vv11623	-0.6793
6	ABC-type spermidine/putrescine transporter	vv12604	-0.6885
7	5'-Nucleotidase/2', 3'-cyclic	vv12771	0.8025

[a]The same ID number is used as in Fig. 1.
[b]GenBank accession number.

Statistical Methods for Detecting DE Genes

We employed the following two statistical procedures for detecting a set of DE genes from the final data set.

(1) Tusher *et al's* SAM procedure [24]
(2) Lönnstedt and Speed's empirical Bayes procedure using B statistic [15], which is referred to as the B procedure.

RESULTS

Figure 1 shows the scatterplot of $\left(\frac{M_1^{DS}+M_2^{DS}}{2}, \frac{1}{4}\sum_{i=3}^{6} M_i\right)$, on which 7 genes with gene-specific dye bias are indicated. For each of these 7 genes, the correction factor of Equation (1) is listed in Table 3. We employed SAM and B procedures on the data set, which was finally corrected for the gene-specific dye bias. The SAM procedure detected 21 DE genes with FDR=0.2622 (26.22%, delta=0.47). The B procedure detected 4 genes whose B statistics were greater than 0. These 4 genes were included in the 21 DE genes detected by SAM. Table 4 shows a subset of DE genes detected by these 2 procedures.

The RT-PCR was performed for 3 newly detected DE genes (highlighted in Table 4). Additionally, a transcriptional regulator gene (vv10306), which has been known to be positively regulated by cFP, was also subjected to RT-PCR analysis as a positive control. For normalizing the RT-PCR experiment results, *pfkA* encoding phosphofructosyl kinase, which is expressed constitutively, was employed as a standard.

DISCUSSION

We have developed a simple procedure of extracting gene-specific dye bias information from a small-scale cDNA microarray experiment of partial dye swap, particularly of Design 2 of Table 1. We showed that the gene-specific dye bias information affected the detection of DE genes. The newly detected DE genes were validated by RT-PCR. The RT-PCR validation suggests that our method of correcting for the gene-specific dye bias is effective in detecting DE genes. This idea was illustrated for Design 1 of Table 1. However, it can easily be modified to Design I of Table 1.

We found that SAM [2, 24] was particularly useful for analyzing this data set, because of its implementation of the fudge factor s_0 in the denominator of the t statistic for modulating it, to prevent a gene having a low intensity ratio with a small variance from being detected as a DE

Table 4. List of a subset of differentially expressed genes detected by SAM and/or B. Highlighted genes are the newly detected after correction for the gene-specific dye bias.

	SAM				B	
	Gene name	Gene ID	Fold change	Score(d)[a]	Gene name	B Stat
Up-regulated	HtpG	vv10189	1.60	3.59	HtpG	1.60
	OmpU	vv11686	2.11	3.26	OmpU	1.67
	Transcriptional regulator	vv10306	1.33	2.46		
Down-regulated	Superoxide dismutase	vv11252	0.55	-3.72	Superoxide dismutase	2.18
	Flagellin FlaC	vv10215	0.80	-2.65		
	ABC-type spermidine/putrescine	vv12604	0.73	-2.60		
	FlgE	vv10223	0.87	-1.34		

[a]Score(d) is an output of the SAM procedure. It is a modulated t statistic in which a small positive constant is added to the standard deviation in the denominator to prevent the possible false-positive occurring in the low intensity region, particularly when the variance is quite small.

Table 5. Relative quantity of transcripts of newly detected DE genes after correcting for the gene-specific dye bias estimated by real-time RT-PCR.

Gene ID[a]	Relative quantity[b]
vv10189	6.345
vv12604	0.605
vv10223	0.595
vv10306[c]	3.235
pfkA[d]	1.000

[a]These IDs are identical to those in Table 4.
[b]Relative expression was calculated as $2^{-\Delta\Delta C_t}$, where ΔC_t was calculated by subtracting the average normalization C_t from the average target C_t value. The $\Delta\Delta C_t$ was then obtained by the difference in ΔC_t values obtained for each of the two cell lines.
[c]This gene is known to be positively regulated by cFP, and therefore, it was included as a positive control.
[d]Gene encoding a phosphofructosyl kinase, which is expressed constitutively, and therefore used as a standard.

gene. It is quite important to maintain a stable variance for non-dye-swap arrays even after being corrected for the gene-specific dye bias. We recommend that at least 4 non-dye-swap arrays be required to maintain the stable variance.

One of the methods of detecting the gene-specific dye bias is to perform the self-self hybridization, where the same RNA sample is hybridized against itself [16, 18]. Our procedure, however, does not require the self-self hybridization. Furthermore, it does not require for detecting the gene-specific dye bias, based upon the normal distribution assumption of the log intensity ratio, which may not sometimes be validated in the real cDNA microarray data.

It has been generally believed that the indirect design avoids the problem of gene-specific dye bias by running all arrays "forward," so that the samples to be compared are always labeled with the same fluorescent dye. This approach assumes that gene-specific dye bias will be "canceled out" when the expression ratios from two or more measurements are compared, as long as the dye orientation is consistent. However, it has been demonstrated even in indirect design that dye orientation can have a significant influence on the measurement and inference of differential gene expression [18]. Dobbin *et al.* [5] reported that the gene-specific dye bias related to the individual samples existed, but they appeared to have minimal impact on the gene expression differences between 7 cell lines considered in their study. However, it remains to be seen whether the gene-specific dye bias is consistent between the control and treated groups.

Acknowledgments

B. S. Kim was supported by a grant from Korea Science and Engineering Foundation (R01-2004-000-10057-0), and K. S. Kim was supported by the 21C Frontier Microbial Genomics and Applications Center Program, Ministry of Science and Technology, Republic of Korea (Grant MG05-0201-4-0).

References

1. Boer, J. M., W. K Huber, H. Sültman, F. Wilmer, A. von Heydebreck, S. Haas, B. Korn, B. Gunawan, B. Vente, L. Füzeshi, M. Vingron, and A. Poustka. 2001. Identification and classification of differentially expressed genes in renal cell carcinoma by expression profiling on a global human 31,500-element cDNA array. *Genome Res.* **11:** 1861–1870.
2. Chu, G., B. Narashimhan, R. Tibshirani, and V. Tusher. 2002. SAM "Significance Analysis of Microarray" Users guide and technical document, http://www-stat.stanford.edu/~tibs/SAM.
3. Churchill, G. A. 2002. Fundamentals of experimental design for cDNA microarrays. *Nat. Genet. Suppl.* **32:** 490–495.
4. De Kievit, T. R. and B. Iglewski. 2000. Bacterial quorum sensing in pathogenic relationships. *Infect. Immun.* **68:** 4839–4849.
5. Dobbin, K., E. S. Kawasaki, D. W. Petersen, and R. M. Simon. 2005. Characterizing dye bias in microarray experiments. *Bioinformatics* **21:** 2430–2437.
6. Dobbin, K., J. H. Shih, and R. Simon. 2003. Statistical analysis of reverse dye microarrays. *Bioinformatics* **19:** 803–810.
7. Dobbin, K. and R. Simon 2002. Comparison of microarray designs for class comparison and class discovery. *Bioinformatics* **18:** 1438–1445.
8. Dombkowski, A. A., B. J. Thibodeau, S. L. Starcevic, and R. F. Novak. 2004. Gene-specific dye bias in micorarray reference designs. *FEBS Lett.* **560:** 120–124.
9. Henke, J. M. and B. L. Bassler. 2004. Bacterial social engagement. *Trends Cell Biol.* **14:** 648–656.
10. Kawase, T., S.-I. Miyoshi, Z. Sultan, and S. Shinoda. 2004. Regulation system for protease production in *Vibrio vulnificus*. *FEMS Microbiol. Lett.* **240:** 55–59.
11. Kerr, M. K. and C. G. Churchill. 2001. Experimental design for gene expression microarrays. *Biostatistics* **2:** 183–201.
12. Kim, B. S., I. Kim, S. Lee, S. Kim, S. Y. Rha, and H. C. Chung. 2005. Statistical methods of translating microarray data into clinically relevant diagnostic information in colorectal cancer. *Bioinformatics* **21:** 517–528.
13. Kim, S. Y., S. E. Lee, Y. R. Kim, C. M. Kim, P. Y. Ryu, H. E Choy, S. S. Chung, and J. H. Rhee. 2003. Regulation of *Vibrio vulnificus* virulence by the LuxS quorum-sensing system. *Mol. Microbiol.* **48:** 1647–1664.
14. Lim, M. S., M. H. Lee, J. H. Lee, H.-M. Ju, N. Y. Park, H. S. Jeong, J. E. Rhee, and S. H. Choi. 2005. Identification and characterization of the *Vibrio vulnificus malPQ* operon. *J. Microbiol. Biotechnol.* **15:** 616–625.
15. Lönnstedt, I. and T. P. Speed. 2002. Replicated microarray data. *Statistica Sinica* **12:** 31–46.
16. Martin-Magniette, M.-L., J. Aubert, E. Cabannes, and J.-J. Daudin. 2005. Evaulation of the gene-specific dye bias

indDNA microarray experiment. *Bioinformatics* **21:** 1995–2000.

17. McDougald, D., S. A. Rice, and S. Kjelleberg. 2001. SmcR-dependant regulation of adaptive phenotypes in *Vibrio vulnificus*. *J. Bacteriol.* **183:** 758–762.
18. Rosenzweig, B. A., P. S. Pine, O. E. Domon, S. M. Morris, J. J. Chen, and F. D. Sistare. 2004. Dye-bias correction in dual-labeled cDNA microarray gene expression measurements. *Environ. Health Perspect.* **112:** 480–487.
19. Salmond, G. P. C., B. W. Bycroft, G. S. A. B. Stewart, and P. Williams. 1995. The bacterial 'enigma': Cracking the code of cell-cell communication. *Mol. Microbiol.* **16:** 615–624.
20. Simon, R. M., E. L. Korn, L. M. McShane, M. D. Radmacher, G. D. Wright, and Y. Zhao. 2003. *Design and Analysis of DNA Micorarray Investigations*, p. 19. Springer, New York, U.S.A.
21. Strom, M. S. and R. N. Paranjpye. 2000. Epidemiology and pathogenesis of *Vibrio vulnificus*. *Microbes Infect.* **2:** 177–188.
22. Troyanskaya, O., M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, R. B. Altman. 2001. Missing value estimation methods for DNA microarrays. *Bioinformatics* **17:** 520–525.
23. Tseng, G. C., M.-K. Oh, L. Rohlin, J. C. Liao, and W. H. Wong. 2001. Issues in cDNA microarray analysis: Quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res.* **29:** 2549–2557.
24. Tusher, V., R. Tibshirani, and G. Chu. 2001. Significance analysis of microarrays applied to transcriptional responses to ionizing radiation. *Proc. Natl. Acad. Sci. USA* **98:** 5116–5121.
25. Wright, A. C., J. G. Morris Jr., D. R. Maneval Jr., K. Richardson, and J. B. Baker. 1985. Cloning of the cytotoxin-hemolysin gene of *Vibrio vulnificus*. *Infect. Immun.* **50:** 922–924.
26. Yang, Y. H., S. Dudoit, P. Luu, D. M. Lin, V. Peng, J. Ngai, and T. P. Speed. 2002. Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30:** e15.