

Effect of local background intensities in the normalization of cDNA microarray data with a skewed expression profiles

Jin Hyuk Kim^{1,3}, Dong Mi Shin², and Yong Sung Lee²

¹Department of Physiology, College of Medicine, Hanyang University, 17, Haengdang-dong, Sungdong-ku, Seoul 133-791, Korea

²Department of Biochemistry, College of Medicine, Hanyang University, 17, Haengdang-dong, Sungdong-ku, Seoul 133-791, Korea

³Corresponding author: Tel, +82-2-2290-0614; Fax, +82-2-2281-3603; E-mail, jhkim1@hanyang.ac.kr

Accepted 5 July 2002

Abstract

Normalization of the data of cDNA microarray is an obligatory step during microarray experiments due to the relatively frequent non-specific errors. Generally, normalization of microarray data is based on the null hypothesis and variance model. In the Yang's model (Yang *et al.*, 2001), at least two types of noises are included. The one is additive noise and the other is multiplicative noise. Usually, background is considered as one of additive noise to the signal and the variation between the signal pixels is the representative multiplicative noise. In this study, the relation between the signal (spot intensity minus background intensity) and background was observed and the influence of background on normalization as a representative additive factor was investigated. Although the relation has not been considered as a factor affecting the normalization, it could improve the accuracy of microarray data when the normalization was carried out considering signal/background ratio. The background dependent normalization decreased the number of genes whose expression levels were changed significantly and it could make their distribution more consistent through the whole range of signal intensities. In this study, printing pin dependent normalization was also carried out regarding the printing pin as a representative multiplicative noise. It improved the distribution of spots in the Cy3-Cy5 scatter plot, but its effect was slight. These studies suggest that there are some influences of the signals on the local backgrounds and they must be considered for the normalization of cDNA microarray data.

Keywords: cDNA microarray, normalization, background, printing pin, LOWESS

Introduction

Experiments using cDNA microarray allow the monitoring of expression levels for thousands of genes simultaneously. As more and more genes are being identified, it has become extremely important to understand the function of these genes and their pathways. Great efforts have been invested in elucidating the relationship between the phenotypes and genotypes of biological systems. To achieve it, it is essential to upgrade the high-throughput technologies and improve the accuracy of analysis tools.

Brown and coworkers developed the cDNA microarray protocols that are widely used (Pat Brown Laboratory Protocol, 2002). The basic strategy for microarray study is to isolate RNA from two sources, a control and an experimental sample. The extracted RNAs from two samples are converted cDNAs and labeled with fluorescent probes, Cy3 and Cy5 respectively. cDNAs labeled with fluorophores are hybridized to the microarray and washed. After those procedures, microarray is scanned at both 532 nm (Cy3) and 635 nm (Cy5) for the labeled cDNA that has hybridized to the microarray. Therefore, two unsigned 16 bit TIFF or BMP image files are generated from the scanner. Various measures for each spot are collected with the image analysis and the ratio of experimental to control expression for subsequent analysis should be calculated from those measures. One of the most critical processes in the microarray experiments is accurate assessing of the expression ratios between the experimental and control samples in dual color experiments because they are used for subsequent cluster analyses and path finding procedures. Therefore, those analyses depend on the accuracy of expression ratio calculated by image analysis and data processing.

The steps, which affect the accuracy of the expression ratio in the microarray image analysis, are (1) accurate discrimination of the spots from the background, and (2) normalization of signals. Several methods have been developed for isolation of spots from the background. There appeared to be little difference between them when applied to the low-density microarray, however edge or shape detection method is appears to be more prominent when applied to the high-density images (Kim

et al., 2001). Regarding the normalization process, which is essential for adjusting the distorted expression ratio distribution, it can be divided into several ways including use of housekeeping genes or spiked controls and global normalization. Although all of them have some advantages and disadvantages, global normalization has been used widely. Global normalization is based on the assumption that the summation of expression ratios between experimental and control group must converge on zero (null hypothesis). Generally, it is correct if the biological properties of them are similar and number of genes whose expressions increase or decrease significantly is not large. Another assumption of global normalization is that the signals (spot intensity minus local background intensity) must be independent of local background intensity. However, it is necessary to consider the effect of local background for the normalization if there is a relationship between the intensities of signals and backgrounds. In this study, the influences of background intensities and printing pin on the normalization are investigated.

Materials and Methods

cDNA microarray experiment

cDNAs for microarray fabrication were amplified by PCR with 5' amino-linked primers in 100 μ l volumes using the Perkin-Elmer Cycler 9600 (Perkin-Elmer Applied Biosystems, Foster, CA, USA). The PCR products were purified, and spotted onto silylated slides (CEL Associates, Houston, TA, USA) with a Pixsys 5500 arrayer (Cartesian Technologies). Printed slides were dried overnight, washed twice in 0.2% SDS for 2 min, washed twice in distilled water for 2 min, and boiled in water for 2 min. Then, the slides were treated with sodium borohydride solution (1.0 g NaBH₄ in 300 ml phosphate buffered saline and 150 ml ethanol) for 30 min, rinsed three times in 0.2% SDS for 1 min, rinsed once in distilled water, boiled in water for 2 s, and dried in air.

Total RNA was extracted using the Trizol reagent (Gibco BRL Life Technologies, Grand Island, NY, USA) and concentrated using Microcon 30 (Amicon). Fluorescence-labeled cDNA probes were made from total RNA by oligo (dT)-primed polymerization. The Cy5- and Cy3-labeled probes were mixed. Hybridization was performed and the slides were washed with 2 x SSC, 0.2% SDS at room temperature for 5 min. The slides were scanned separately to get the Cy3 and Cy5 images with GMS 418 Array Scanner (Affymetrix, Santa Clara, CA, USA).

Relation between signal and background

Data output from a dual color microarray experiment

contains the average fluorescent intensities for the spot and local background for both the control and experimental samples. On each microarray, let

$$s_{ij} \text{ and } b_{ij} \quad (1)$$

represent the spot and local background fluorescent intensities for sample (color) i , ($i = 1, 2$), at spot j ($j = 1, 2, \dots, n$), where n is the number of spots in the microarray (Yang *et al.*, 2001). The amount of fluorescent probe available for hybridization to spot j is

$$a_i d_j \theta_{ij} \quad (2)$$

where a_i , d_j , and θ_{ij} represent total amount of mRNA used for labeling, amount of cDNA in the j th spot of microarray, and the fraction of mRNA for gene j among the total mRNA for all expressed genes (Yang *et al.*, 2001). Spot intensity s_{ij} is the function of true signal $a_i d_j \theta_{ij}$, background b_{ij} and noise.

$$s_{ij} = f(a_i d_j \theta_{ij}, b_{ij}, \text{noise}) \quad (3)$$

If the noise and background intensities are multiplicative and additive, respectively, equation (3) is

$$s_{ij} = a_i d_j \theta_{ij} \exp(\varepsilon_{ij}) + b_{ij} \quad (4)$$

where ε_{ij} is multiplicative variation (Yang *et al.*, 2001).

Within-print-tip-group normalization

Usually, a subgrid is printed with the same pin. The microarrayer used in this study prints the cDNA spots with 16 pins and all subgrids are arranged in 4 by 4. Especially, a large number of pins is used to print the high-density microarray but several factors such as a little difference in the length or in the opening of the tips, and deformation after many hours of printing affect the amount of cDNA on the j th spot (d_j). If a part of spots are printed with the k th pin, equation (4) must be modified into

$$s_{ijk} = a_i d_{jk} \theta_{ij} \exp(\varepsilon_{ijk}) + b_{ij} \quad (5)$$

The true expression ratio (r_j) between the experimental and control samples and its log transformation (λ_j) are defined as (Yang *et al.*, 2001)

$$r_j = \theta_{2j} / \theta_{1j}, \quad \lambda_j = \ln r_j \quad (6)$$

Therefore, the observed log expression ratio (Yang *et al.*, 2001) (y_j) is

$$y_j = \ln \frac{t_{2jk}}{t_{1jk}} = \lambda_j + \rho + (\varepsilon_{2jk} - \varepsilon_{1jk}) \quad (7)$$

where ρ is $\ln(a_2/a_1)$ and t_{ijk} is true signal ($s_{ijk} - b_{ijk}$). Equation (7) means that it is necessary to normalize the expression ratio for two components. The first component is the inequality of total mRNA between experimental and control sample and the second one is due to noise and properties of the pin. In this study, the

influence of the pin on the Cy3-Cy5 scatter plot was investigated.

Background dependent normalization

In equation (4), it can be assumed that local background (b_{ij}) is almost similar to the background in the spot (b_{ij}') (Yang *et al.*, 2001). The accurate form of equation (4) is

$$s_{ij} = a_i d_j \theta_{ij} \exp(\varepsilon_{ij}) + b_{ij}' \quad (8)$$

If the assumption ($b_{ij}' = b_{ij}$) cannot realize, equation (4) cannot replace the equation (8). It is necessary to show that the influence of signal (t_{ij}) does not affect intensity of local background. If there is any relationship between the spot and local background intensity, it must be composed of the true local background and influence of the spot. Therefore, the process that eliminates the influence of spot on the local background is needed. In this study, the relationship between the local background and signal was examined to investigate the influence of it on the signal. Signals were normalized and combined to minimize the influence of spot on the local background within each categorized ranges of $\log t_{ij} / \log b_{ij}$.

Results

In order to get the intensity distribution of microarray background pixels, three rectangular areas that does not include the spot, were randomly extracted from the Cy3 and Cy5 microarray images, and the intensity distribution histogram was fitted to the log-normal distribution (Table 1 and Figure 1). The distribution of fluorescence intensities of the extracted backgrounds was different by the location and the fluorescence dyes, and showed

different values for various parameters in fitted log-normal curves. It suggests that analysis of microarray images must be carried out in consideration of the local backgrounds and some errors might occur when differences of mean or variance between Cy3 and Cy5 backgrounds are not considered during normalization.

It is assumed that background intensities are additive to the signal as in equation (4). If it is true, local background cannot be affected by the signal and the slope of linear regression between $\log b_{ij}$ and $\log(s_{ij} - b_{ij})$ must approach to zero. In this study, relationship between signals and background intensities was investigated (Figure 2). The slopes of Cy3 and Cy5 are 0.12 and 0.27 and their correlation coefficients are 0.11 and 0.31 (p -values are less than 0.001), respectively. Especially, slope of Cy5 is greater than that of Cy3 and might suggest the assumption that background is additive may cause a serious problem in normalization.

To minimize the effect of local background and find more reliable normalization process, printing pin dependent normalization was carried out in this study. Spots were divided into 16 groups by the printing pin and their distribution was expressed in Cy3-C5 scatter plot. The

Table 1. Comparison between Cy3- and Cy5- background distribution for three different areas

	Cy3			Cy5		
	a	b	x_0	a	b	x_0
Area 1	1180.81	0.35	20.00	1915.59	0.57	6.81
Area 2	1407.82	0.38	15.35	3377.42	1.01	1.66
Area 3	1355.87	0.35	17.00	3949.12	1.00	1.39

The distribution of Cy3 and Cy5 background intensity is fitted to log-normal distribution $y = a \exp[-0.5(\ln(x/x_0)/b)^2]$. a and b are coefficients and x_0 represents mean.

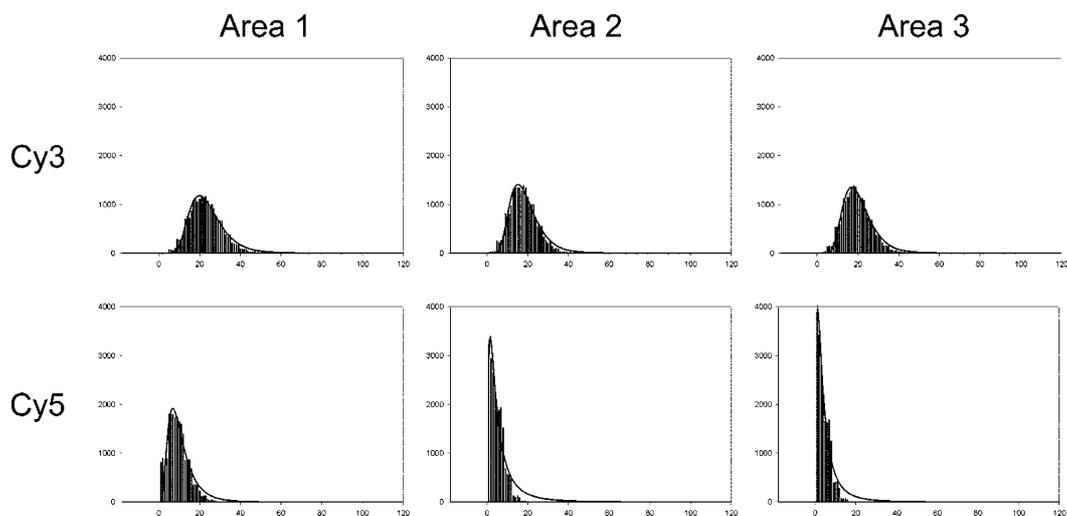


Figure 1. The distribution of background intensities for three different areas. Random three background areas were extracted and the intensities of pixels were measured. The distribution of Cy3 and Cy5 background intensity is different from area to area, showing the distribution of Cy5 background intensity is much more inconsistent when compared to that of Cy3. Y axis: frequency of pixels, X axis: intensity value

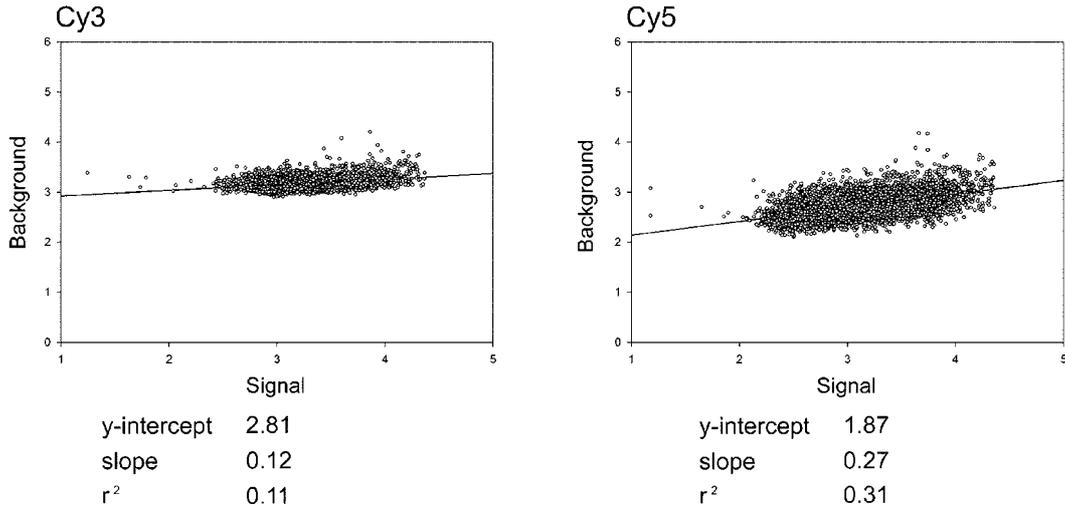


Figure 2. The correlation between signal intensity and background intensity. Local background intensities of whole spots were measured and linear regression was carried out for Cy3 and Cy5. Signal and background intensity are positively correlated in both Cy3 and Cy5. The correlation is more prominent in Cy5 than in Cy3, with a larger variation in background intensities. Y axis: log (background intensity), X axis: log (signal)

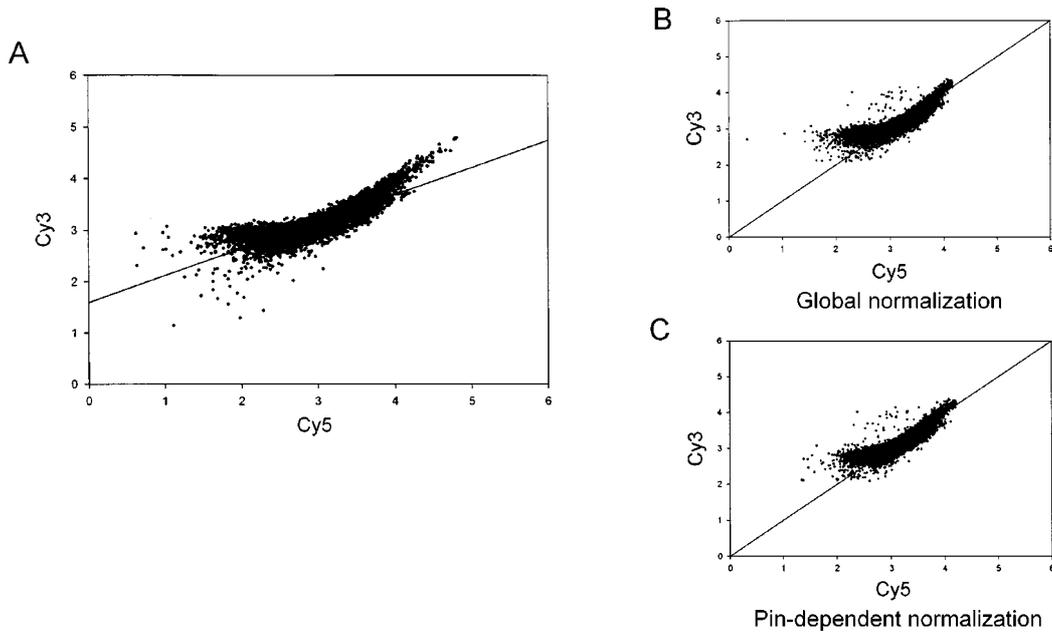


Figure 3. Comparison between global and pin-dependent normalization. Scatter plot of Cy3 versus Cy5 before normalization (A), global normalization (B), and pin dependent normalization (C) are shown. The number of up regulated genes ($Cy5/Cy3 > 2.0$) was 763 and that of down regulated genes ($Cy5/Cy3 < 0.5$) was 279 in global normalization, while those were 403 and 220, respectively, in pin-dependent normalization.

distribution of spots in each group, which were printed by the same pin, was relatively similar between groups according to the Cy3-Cy5 scatter plot, but inconsistent values were observed in the slope and Y-intercept between groups. For example, in the case of pin number 5, points were inclined toward the low intensity range for both Cy3 and Cy5 (Figures 3 and 4). If the distributions of spots in Cy3-Cy5 scatter plot are even, the normalization according to printing pin can be similar to that of global normalization. Because printing pin effect of ϵ_{2jk}

ϵ_{1jk} in equation (7) is relatively small (Table 2), normalization according to printing pin has a little effect on the expression ratio (Figure 3). However, normalization according to printing pin made a shift of the spots, which were located in the low intensity region of Cy3-Cy5 scatter plot, a little to the neutral line ($\log Cy5 = \log Cy3$) (Figure 3).

In Figure 2 and Table 1, the logarithmic local backgrounds intensities (b_{ij}) showed a linear relation to the logarithmic signal intensities (t_{ij}) and their slopes were

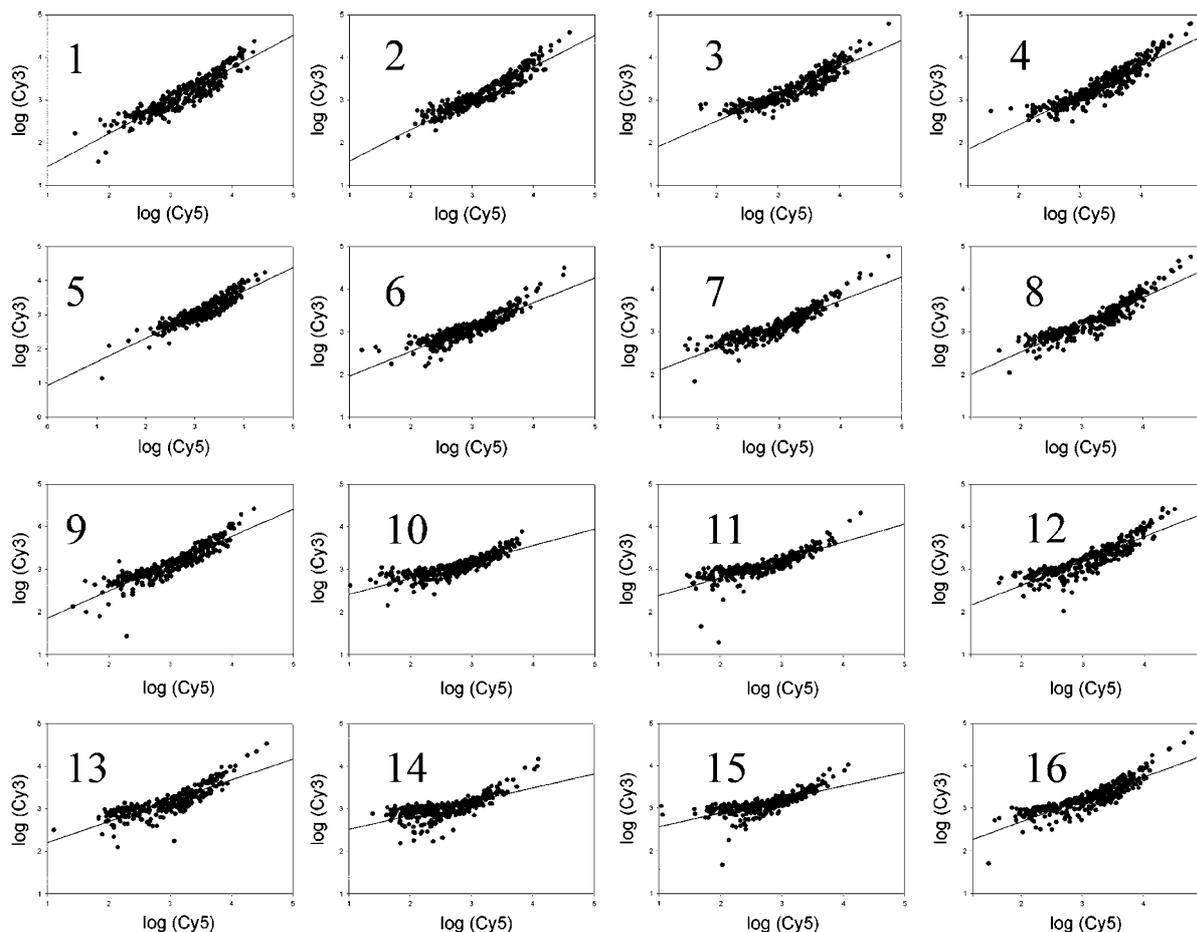


Figure 4. Scatter plots of log Cy3 versus log Cy5 for each printing pin. Scatter plots of Cy3 versus Cy5 for each printing pin were expressed after linear regression.

Table 2. Printing pin-dependent normalization with linear regression

	SLOPE	Y-INTERCEPT	r^2	Mean of $(\varepsilon_{2jk} - \varepsilon_{1jk})$
pin #1	0.7693	0.6757	0.8617	0.000954
pin #2	0.7372	0.8324	0.8813	-0.011600
pin #3	0.6278	1.2509	0.8225	-0.017620
pin #4	0.6960	1.0252	0.8273	-0.021590
pin #5	0.6916	0.9275	0.8300	-0.008770
pin #6	0.5742	1.3915	0.8041	0.002602
pin #7	0.5523	1.5223	0.8118	0.014775
pin #8	0.6437	1.2349	0.8428	0.031476
pin #9	0.6378	1.2197	0.8163	0.049362
pin #10	0.3821	2.0345	0.7199	0.059628
pin #11	0.4191	1.9607	0.6798	0.068868
pin #12	0.5689	1.4886	0.7895	0.098219
pin #13	0.4893	1.7175	0.7179	0.117347
pin #14	0.3241	2.1898	0.4773	0.104766
pin #15	0.3187	2.2599	0.5428	0.053239
pin #16	0.5247	1.6392	0.7648	-0.001120

According to the printing pins, slope, y-intercept, and r^2 of Cy3 versus Cy5 scatter plots showed different values. $\varepsilon_{2jk} - \varepsilon_{1jk}$ is the logarithm of difference between the pin dependent normalization and global normalization.

not zero. It means that local backgrounds affect the signals through unknown way and the difference of correlation coefficients between Cy3 and Cy5 can disperse and bend the distribution of points on Cy3-Cy5 scatter plot. Correlation coefficient of Cy5 between log background and log signal was observed to be larger than that of Cy3 in this experiment. Therefore, the points in the scatter plot should be more dispersed along the Cy5 axis at the low intensity region (Figure 3A), and difference of slopes also made a skewed pattern in Cy3-Cy5 scatter plot. To compensate such a phenomenon, all the spots were categorized into five groups according to the value of $\log t_{Cy5j} / \log b_{Cy5j}$ (<0.5, 0.5-1.0, 1.0-1.5, 1.5-2.0, and >2.0) and normalizations were carried out separately for each group (Figure 4). Before normalization, the slope of regression line in the Cy3-Cy5 scatter plot was decreased gradually when $\log t_{Cy5j} / \log b_{Cy5j}$ value was reduced. Spots within each group were normalized with simple regression and all spots were combined. After normalization with background dependent method, all the spots on Cy3-Cy5 scatter plot were gathered near the neutral line and the dispersion of spots has dis-

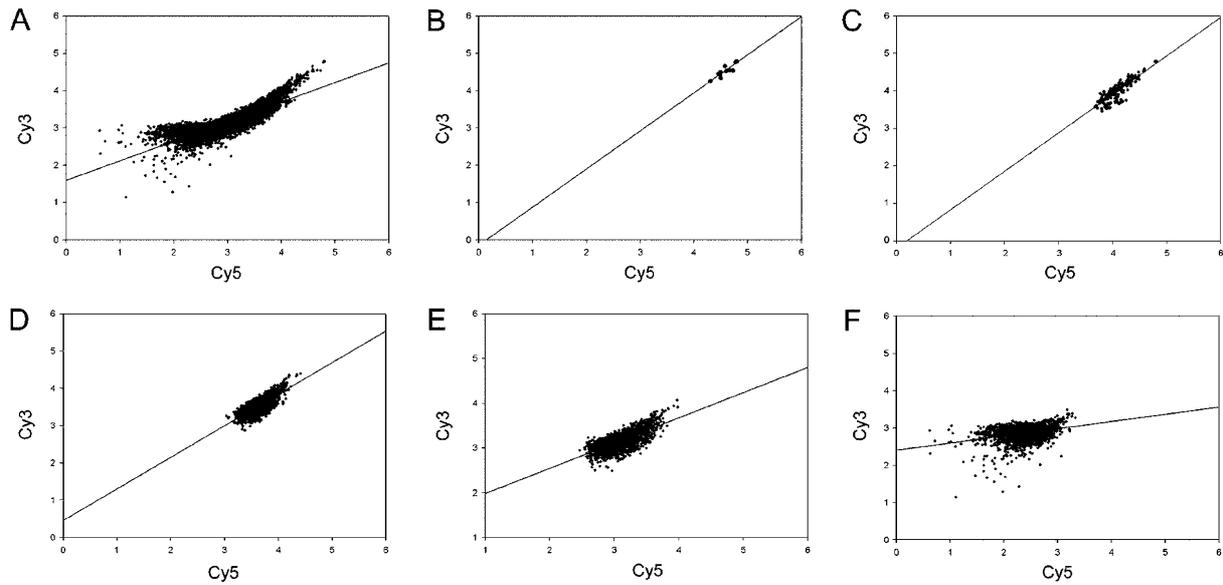


Figure 5. Scatter plots for groups categorized by the range of $\log(\text{Cy5 signal intensity})/\log(\text{Cy5 background intensity})$ value. Scatter plots of Cy3 versus Cy5 for whole spots (A) and spots categorized into five groups (B-F) according to the ratio of $\log(\text{Cy5 signal intensity})/\log(\text{Cy5 background intensity})$. The ratios in group B-F are >2.0 , $1.5\text{--}2.0$, $1.0\text{--}1.5$, $0.5\text{--}1.0$ and <0.5 , respectively.

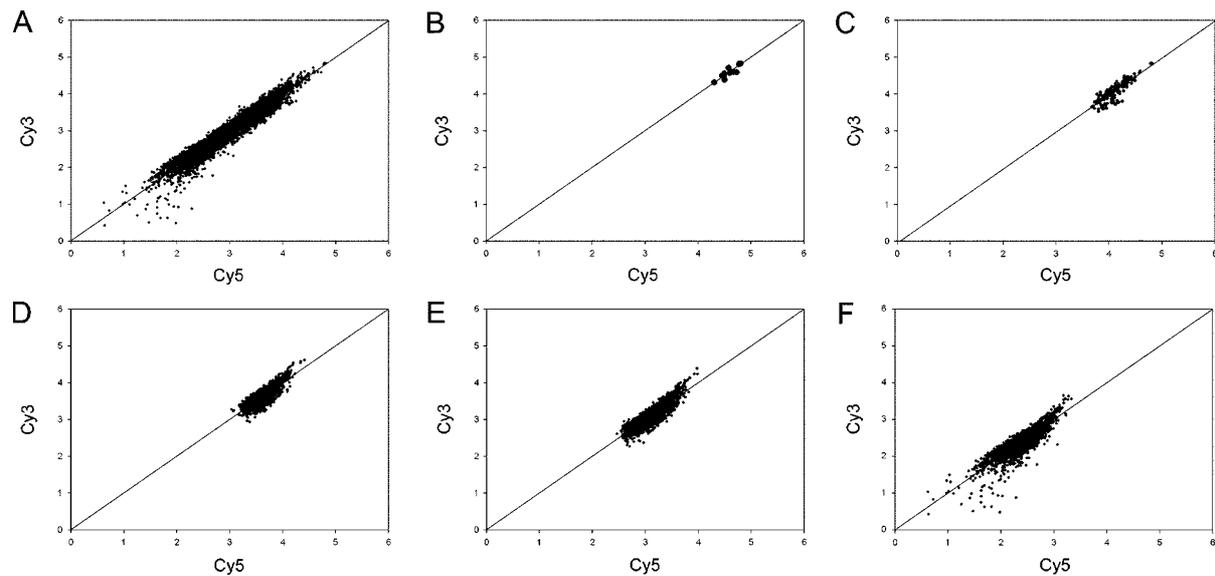


Figure 6. Scatter plots for groups after normalization. After normalization of each group (B-F) with simple linear regression, all the spots are combined and shown in figure A.

appeared, especially at low intensity region. Also, the skewed expression pattern, which remained even after global normalization, was changed to a linear pattern (Figure 6).

Comparison of the efficacy of background dependent normalization to global normalization as indicated in the summary of changes in log expression ratios is shown in Table 3 (Figure 7). Normalization with LOWESS smoothing (Yang *et al.*, 2002), showed the lowest $\Sigma\log(\text{Cy5}/\text{Cy3})$ value among three methods. However, number of spots whose expression levels were changed

Table 3. Comparison between Cy5 background dependent normalization and other conventional normalization methods

	Background normalization using ImaGene	Normalization with LOWESS smoothing	Normalization with LOWESS smoothing
$\text{Cy5}/\text{Cy3} > 2.0$	213	375	523
$\text{Cy5}/\text{Cy3} < 0.5$	38	1318	203
$\Sigma\log(\text{Cy5}/\text{Cy3})$	5.539	-653.397	0.019

significantly ($r_j > 2$ or $r_j < 0.5$) was decreased by the order of normalization with ImaGene[®], normalization

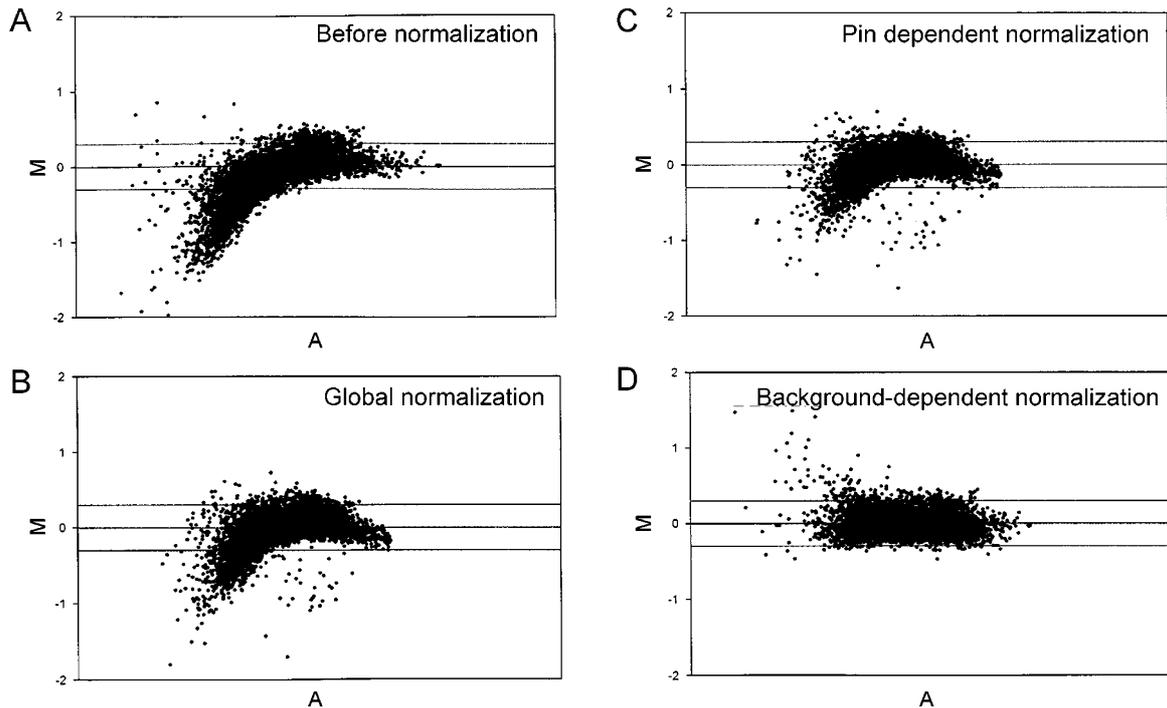


Figure 7. Comparison between conventional and Cy5 background-dependent normalization. MA plots before normalization (A), after global normalization (B), printing pin-dependent normalization (C), and Cy5 background-dependent normalization (D) are shown with lines indicating 2 fold increase or decrease in expression levels. The M is $\log(\text{Cy5 signal}) - \log(\text{Cy3 signal})$ and A is $[\log(\text{Cy5 signal}) + \log(\text{Cy3 signal})]/2$.

with LOWESS smoothing, and background dependent normalization. It could be explained that LOWESS smoothing succeeded in making the Cy3-Cy5 scatter plot symmetrical to the neutral line, but failed to decrease the perturbation by noise at low intensity region of the Cy3-Cy5 scatter plot. Regarding the distribution of spots, which were increased or decreased significantly in expression levels, background dependent normalization showed even distribution through the whole range of signal intensities. Comparing to the background dependent normalization, normalization with ImaGene[®] showed a much skewed pattern with a large number of decreased genes at low signal intensity region and a large

number of increased genes at relatively high signal intensity region (Figure 8).

Discussion

In this study, the influences of local background (b_{ij}) on the signal intensities were investigated and new normalization method was designed for microarray experiments. Generally, it is accepted that the background is additive to the spot intensity (s_{ij}) of microarray and another noise (ϵ_{ij}) that is due to the probe preparation and hybridization efficiency is multiplicative to the spot

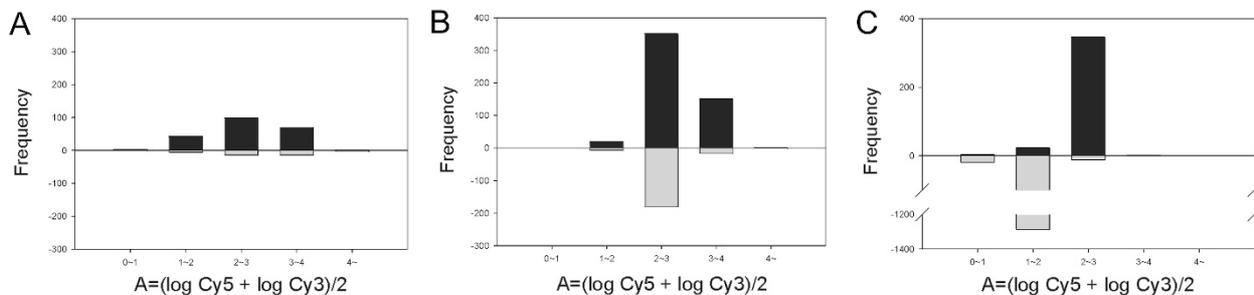


Figure 8. Comparison between Cy5 background dependent normalization and other conventional normalization methods. The number of up and down regulated gene is dramatically decreased in Cy5 background dependent normalization method (A) when compared to global normalization method using LOWESS smoothing (B) or ImaGene (C) software (Biodiscovery, USA). The number of genes up-regulated (■ $\text{Cy5/Cy3} > 2.0$), or down-regulated (■ $\text{Cy5/Cy3} < 0.5$), are expressed as a function of Cy5 signal/Cy5 background ratio. Y axis: frequency, X axis: $[\log(\text{Cy5 signal}) + \log(\text{Cy3 signal})]/2$

intensity (equation 4). Several investigators have developed such models. Chen *et al.* (1997) first proposed the assumption of multiplicative noise, and the likelihood function was based on the null hypothesis that all genes have the same expression. Therefore, the application of their model is limited to known house keeping genes. Analysis of variance model that have the same basis as equation (7) was proposed Kerr *et al.* (2000). But they omitted background intensities in their model. In 2001 Yang *et al.* designed the similar model but they considered the additive and multiplicative noise in their model (equation 4). In that model, backgrounds were regarded as an additive factor. Therefore, spot intensity is composed of multiplicative factors such as total mRNA amount (a), cDNA amount in every spot (d_j), the fraction of mRNA for a gene among the total mRNA (θ_j), and multiplicative noise (ε_{ij}) as well as additive factor such as background (b_{ij}) in the model. If the background is additive, observed signal ($t_{ij} = s_{ij} - b_{ij}$) is independent of background intensities and the simple regression of background to signal intensities must show the slope to converge into zero at log scatter plot of background-signal intensities and their correlation coefficients does not differ significantly from each other. However, their distribution in scatter plot did not make a slope of regression zero and their correlation coefficients were quite different from each other (Figure 2). The reason why the backgrounds affect the signals in microarray is not known but likely explanation may be either background was contaminated with the Cy3 and/or Cy5 fluorescent dye or chemicals coating the surface of microarray, or base material (glass, nylon and etc) was more or less sensitive to the laser that have a specific wave length. In the case of base material, nonspecific fluorescence must be uniform over the whole range of the microarray so that it can be additive. Therefore, it can be excluded. Chemicals coating the surface maybe have similar properties so that it may be out of bounds of possibility. The last one is the possibility of contaminating local background by Cy3 and Cy5 fluorescent dye. If it is possible, the opportunity of contamination can be increased for the local backgrounds around the bright spots, which have high intensity. However, bright backgrounds contaminated with fluorescent dye from vague spots can change the expression ratio very little. Then, Cy3 image that has relatively bright background can show smaller slope than Cy5 image.

Yang *et al.* (2002) proposed printing pin dependent normalization because there may be systematic differences between the printing pins such as slight differences in the length or in the opening of the pin tips, and deformation after many hours of printing. Such factors can affect the shape of the spot, quantity and distribution of cDNA in the spots so that another multiplicative noise must be considered. In this study, normalization by printing pin makes the spots gathered around

the neutral line in Cy3-Cy5 scatter plot. Especially, the effect of printing pin dependent normalization was increased in the low intensity range of both Cy3 and Cy5. When the logarithmic transformation is taken for the expression ratio, most of $\Sigma \lambda_j$ are zero as in Kerr *et al.* (2000). Therefore, it is meaningful to count the number of genes whose expression levels are increased or decreased significantly. If the control and experimental samples are biologically similar, normalization must decrease the number of them and evenly distribute their location in Cy3-Cy5 scatter plot. Printing pin dependent normalization is more excellent than the normalization with whole spot in this aspect.

Because it is apparent that the background intensity is the function of signal, it is reasonable to categorize the relation between local background intensities and signals. In this study, $\log t_{Cy5j} / \log b_{Cy5j}$ was categorized into five graded ranges and the spots included in each categorized range were normalized and they were combined. The reason, why Cy5 was chosen, was that the slope and correlation coefficient of log Cy5 signal to background scatter plot was larger and smaller than those of Cy3, respectively. The result of background dependent normalization makes the dispersed spots' compact around the neutral line in the low intensity region of Cy3-Cy5 scatter plot and significantly increased or decreased spots could be decreased very much. And also it could improve the symmetry of spots distribution in the Cy3-Cy5 scatter plot. The improvement of symmetry following the normalization with LOWESS smoothing was a little bit better than the background dependent normalization but spots, whose expression level changed significantly, tended to crowd in the specific range (especially, low intensity range). It means that the normalization with LOWESS smoothing can straighten the distribution of spots but fail to improve the dispersion of them.

In summary, these results indicate that the local backgrounds must be considered when the normalization of data is performed.

Acknowledgements

This study was supported by the IMT2000 contribution from Ministry of Health and Welfare, and Ministry of Information and Communication (01-PJ11-PG9-01BT-00A-0045).

References

- Chen Y, Dougherty ER, Bitter ML. Ratio-based decisions and quantitative analysis of cDNA microarray images. *J Chem Optics* 1997;2:364-74
- Kerr MK, Martin M, Churchill GA. Analysis of variance for

gene expression microarray data. *J Comput Biol* 2000;7:819-37

Kim JH, Kim HY, Lee YS. A novel method using edge detection for signal extraction from cDNA microarray image analysis. *Exp Mol Med.* 2001;33(2):83-88.

Pat Brown Laboratory Protocol[Online]. Department of Biochemistry, Stanford University, 2002(<http://cmgm.stanford.edu/pbrown/protocols/>)

Yang MC, Ruan QG, Yang JJ, Eckenrode S, Wu S, McIndoe RA, She JX. A statistical method for flagging weak spots improves normalization and ratio estimates in microarrays. *Physiol Genomics.* 2001;7(1):45-53

Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* 2002;30(4):e15