19

Evaluation of the Reliability of cDNA Microarray Technique

Yao Li,[1§] Yao Luo,[1§] Chengzhi Zhang,[1] Minyan Qiu,[2] Zhiyong Han,[2] Qin Wei,[1] Sanzhen Liu,[2] Yi Xie,[1*] and Yumin Mao[1*]

[1] *State Key Laboratory of Genetic Engineering, Institute of Genetics, School of Life Science, Fudan University Shanghai 200433, P. R. China*

[2] *Shanghai BioStar Genechip Inc., Shanghai, 200092, P. R. China*

[§] *These two authors contribute equally to this paper*

[*] *Corresponding authors. E-mail: ymmao@fudan.edu.cn*

1. Introduction

Microarray is a new technological approach developed in early 1990s, which has found wide applications in studies of gene expression patterns in various tissues[1,2]. The technology enables rapid parallel genetic analysis of tens of thousands of genes in one experiment, and makes possible for the genetic researchers to measure the expression of all genes in an organism simultaneously. cDNA microarray technology was established and applied in the research of hepatoma expression pattern in our laboratory[3]. In the early stage, the information about the genes which expressions changed obviously in a specific condition could be generally obtained by ratio analysis[4]. However, there are many variables that will impact on the quality of the data generated by any microarray experiments[5], therefore it is significant to evaluate the reliability of microarray data. Up to now, a large quantity of useful data has been acquired from these experiments. The problems of how to analyze and deal with the data and how to validate the reliability of results have become the key to utilize the approach more effectively.

Currently, there were only a few papers about comprehensive evaluation of cDNA technology. Incyte Ltd. reported the issue of precision, accuracy and reproducibility of microarray data[6]. In order to study the data more efficiently, we evaluated the reproducibility, reliability and variation at several different aspects, and analyzed the advantages and gave an assessment on the whole.

2. Methods and Material

Microarrays of 4096 or 14112 human cDNAs, were manufactured by BioStar Ltd. All clones were verified by being sequenced. The array included spots of

plant genes and HCV genes as negative control and spots of preparing solution without cDNA as blank control.

2.1. Two Sets of Experiments

The experiments for system evaluation were divided into two sets: self-comparison experiments and differential expression experiments. In the self-comparison experiments, Cy5- and Cy3-labeled cDNA were both prepared from the RNA of the same tissue, while in differential expression experiments, the RNA was from two different tissues to measure the differentiation.

2.2. Preparation of Probes Labeled with Fluorescent Molecules

Donated hepatoma and normal liver tissues were supplied by Changhai Hospital. Two methods for extracting total RNA were used here: Method One[7], as a common method, was used for most total RNA in present article; and Method Two[8] was only used for comparing two methods in 3.1.1. mRNA was purified using Oligotex mRNA Midi Kit (Quagen Company). In the self-comparison experiment, the mRNA (3μ g) from the same tissue or total RNA (50μ g) was labeled with Cy3-dUTP and Cy5-dUTP respectively; While in the differential expression experiments, normal liver tissue was labeled with Cy3-dUTP and hepatoma was labeled with Cy5-dUTP, or vice versa. Labeled cDNA was deposited with ethanol, and then dissolved in hybridizing solution of 20μ L 15×SSC+0.2%SDS.

2.3. Hybridization and Rinse

The methods were as described[3].

2.4. Scanning and Analysis

Microarrays were scanned with a Scanarray 4000 laser induced fluorescence scanner from Packard Biochip Technologies Ltd. and signal intensity for each target element was detected with GenePix 3.01 image software from Axon.

2.5. Data Statistical Study

All data obtained were calibrated on the whole level by Yang's integral correction algorithm[9]. The corrected data were used to calculate the ratio of corresponding signal and determine the cutoff of differential expression by tolerance interval algorithm. CV of each ratio was calculated so as to assess the accuracy and reproducibility of the arrays. According to the cutoff, the distribution of ratio of those differential expression genes was observed and

the possible sources of variation were inspected. All data were screened automatically with the image software. Relevant coefficient *r*, which was usually used to assess the reproducibility of microarray data, was calculated as Pearson relevant coefficient[10]. Therein, *x* and *y* represented the corresponding ratio value of two experiments.

$$r = \frac{\sum_i \mathrm{lrb}\, x_i - \bar{x}\, \mathrm{rrblrb}\, y_i - \bar{y}\, \mathrm{rrb}}{\sqrt{\sum_i \mathrm{lrb}\, x_i - \bar{x}\, \mathrm{rrb}^2 \cdot \sum_i \mathrm{lrb}\, y_i - \bar{y}\, \mathrm{rrb}^2}}$$

The concept of "consistence rate" was put forward here as a new parameter to evaluate the reproducibility of microarray data. The consistence rate was the percent of gene number, which showed differential expression in the same direction in both of two experiments, from the total number of all differentially expressed genes, the formula as follows. N_d was the total number of genes showing differential expression, and N_{id} was the number of elements of differential expression in the same direction in both of two experiments therein.

$$CR = \frac{N_d - 2 \times N_{id}}{N_d}$$

3. Results

3.1. Self-Comparison Experiments

The same normal liver tissue was labeled with Cy3 and Cy5 to perform self-comparison experiments. Theoretically, the ratio of Cy5/Cy3 should be 1 for all elements arrayed on the slide in self-comparison experiment. However, due to some systematic biases, some deviations from the theoretical value were observed of some genes. The cutoff of the ratio of Cy5/Cy3 to screening differential genes was 2, which was recognized all over the world. Thus, in the self-comparison experiments, any gene of which ratio was higher than 2 or lower than 0.5 was regarded as false positive gene. False positive rate (FPR) refers to the percentage of the number of false positive genes from all genes on the array. The values of false positive rate, relevant coefficient (R) and CV of the ratio were used to evaluate the reliability of microarray data.

3.1.1. Impact of Different Kinds and Different Processes of RNA on Hybridization

In order to know the impact of different kinds of RNA on reproducibility of hybridization, we performed the following experiments: (1) Performing self-comparison experiments with total RNA, which included three sets of experiments: A) total RNA from the same extraction method—Method One at

different time was used; C) total RNA from the same extraction method—Method One at the same time was used; (2) Performing self-comparison experiments with mRNA; (3) Performing self-comparison experiments with total RNA and purified mRNA. The results are shown in Table 1.

According to statistic analysis, the results of Table 1 indicated that there was no obvious differentiation (P>0.05) in the first four sets. False positive rate was usually about 1% when cutoff was defined as 2.0, which was similar to the advanced level in the world[11]. It suggested that the approaches and processes of mRNA extraction would not induce any distinct differentiation, but in the self-comparison experiments of mRNA vs. total RNA, false positive rate was more than 10%. The above results were reliable by several reproducible performances. We could conclude that mRNA and total RNA could be both used in experiments of expression pattern, but only the same kind of RNA can be used in one experiment. For instance, if mRNA was labeled with Cy3, then the kind of RNA labeled with Cy5 should also be mRNA.

3.1.2. Impact of Probe Labeling Process on False Positive Rate

We demonstrated the reproducibility of two experiments by another means: x and y axis represented ratio in natural log (Ln) scale of the two of replicate experiments respectively. Thus, it was convenient to compare the identity of them (Figure 19.1). Since the false positive rate was usually less than 1% in common condition, the false positive rate after reproducing twice was very low (1%*1%=0.01%) theoretically. In the experiments, when the cutoff was defined as 1.7, Figure 19.1B (The probes of replicated experiments labeled separately) showed that none of genes appeared false positive in both arrays, which accorded with the theoretical value, while Figure 19.1A (the probes of replicate experiments labeled simultaneously) showed that the most false positive genes only appeared in one experiment (the points in diamond), but 5 genes appeared in both experiments (the points in triangle) in the same direction. Analogous results were obtained in the replicated experiments. (Data were not shown here.) Such false positive was due to the bias in the labeling process. It could be concluded that performing two replicated experiments

TABLE 19.1. The result of self-comparison experiments with RNA (mRNA and total RNA) of the same Tissue RNA but different process methods

	Total RNA/Total RNA				
	A Different extraction methods	B The same extraction method	C Extracting at the same time	mRNA/ mRNA	Total RNA/mRNA (Cy5/Cy3)
FPR (cutoff=1.7)	2.72%	2.83%	2.51%	1.38%	–
FPR (cutoff=2.0)	1.23%	0.97%	0.67%	0.39%	12.40%
correlation coefficient R	0.90	0.97	0.95	0.98	0.76
P value	>0.05				–

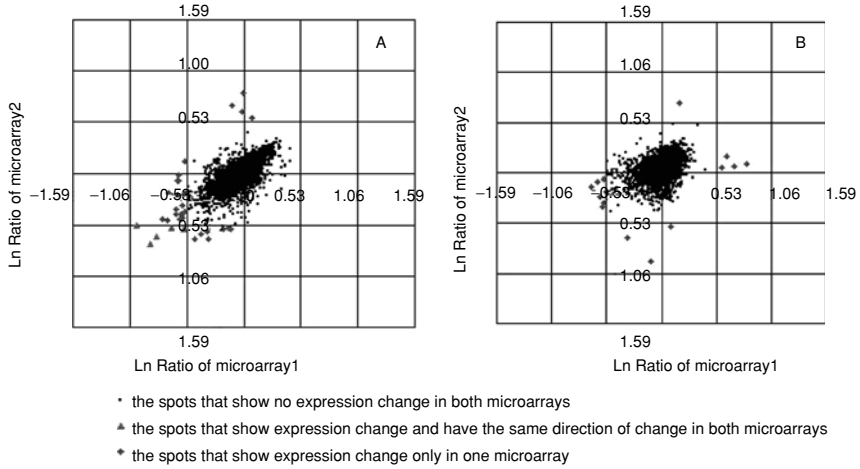

FIGURE 19.1. The identity comparison between self-comparison experiments. A: Replicated experiments with labeling simultaneously; B: replicated experiments with labeling separately.

with labeling separately can avoid false positive, while performing two replicated experiments with labeling simultaneously will remain at a low false positive rate (0.1-0.2%).

3.2. Differential Expression Experiments

Analysis of differential expression experiments was using normal liver and hepatoma tissues, which were labeled by Cy3 and Cy5 respectively. We evaluated the system from the following several aspects.

3.2.1. Impact of Different Concentrations of Target Sequences on Hybridizing Signals on the Multiple-Gene Microarrays

Four genes were chosen to be prepared in 5 concentrations arrayed on a microarray of 14112 genes. The UniGene IDs of the four genes (A-D) were Hs. 181165, Hs. 14376, Hs. 7838, Hs. 148212 respectively. The series of concentrations was 5 ng/μl, 50 ng/μl, 100 ng/μl, 200 ng/μl and 400 ng/μl. Among them, A and B were the genes of high-abundance, which were detected having many copies during sequencing, while C and D were the genes of low-abundance, which were detected having a few copies during sequencing. The results, shown in Figure 19.2 and Figure 19.3, indicated that with the increasing of concentration, the signal intensity enhanced as well, and the intensities of high-abundant genes were obviously stronger than those of low-abundant genes. It also proved that the ratio of these genes were constant in various

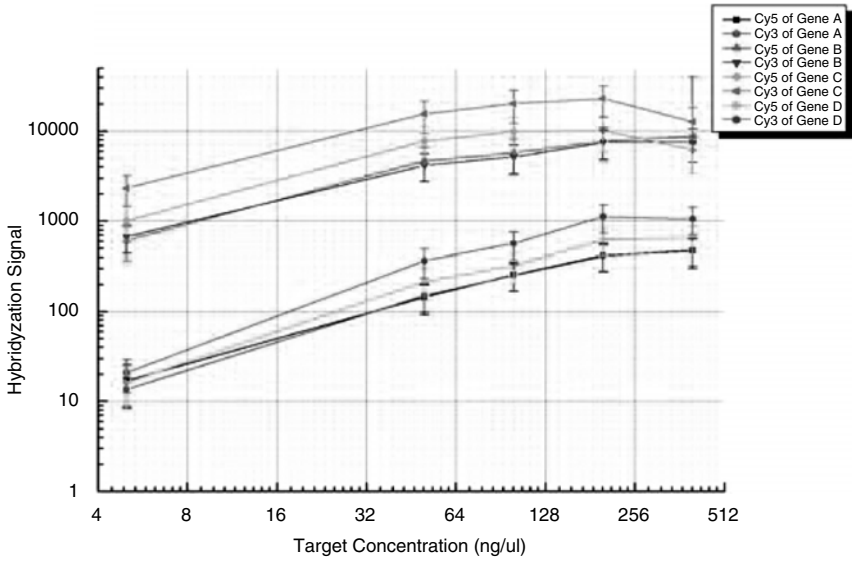

FIGURE 19.2. The value of signal intensity with serial concentrations of target genes (target genes refers to the gene sequences immobilized on the arrays).

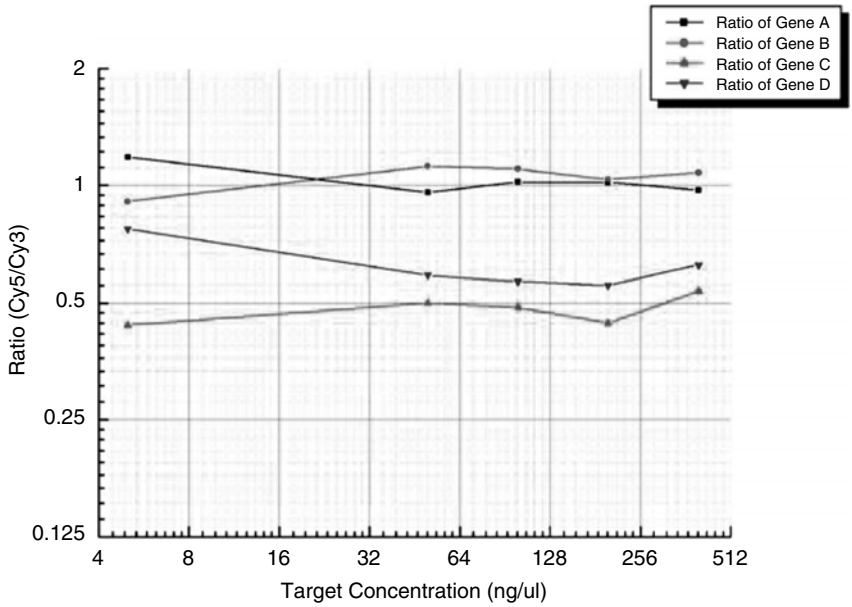

FIGURE 19.3. The ratio of target genes in different concentrations.

concentrations, and the ratio of high-abundant genes were more stable than that of low-abundant genes, which was because deviation increased with intensity decreasing when intensity was lower than 1000.

3.2.2. Impact of Cy5 and Cy3 Reverse Labeling on the Results

In order to investigate the impact of Cy5 and Cy3 labeling on the result, two sets of experiments were performed: in one set, hepatoma RNA was labeled with Cy5, and normal liver RNA was labeled with Cy3; while in the other set, hepatoma RNA was labeled with Cy3, and normal liver RNA was labeled with Cy5. The obtained data were calibrated and then showed in Figure 19.4 and Figure 19.5. Figure 19.4 showed the histogram of corresponding ratio of 50 genes in two sets of experiments. Figure 19.5 showed the scatter plot of corresponding ratio value of all genes in two sets. The relevant coefficient of two sets was –0.909. The histogram and scatter plot demonstrated that labeling with Cy5 or Cy3 has no impact on result. In other words, the ratio of Cy5/Cy3 was not influenced by reverse cross labeling.

3.2.3. Comparison of a Series of Replicated Experiments among Multiple Microarrays

The comparison of microarrays representing different batches was performed with total RNA from normal liver and hepatoma. Meanwhile, the impact of

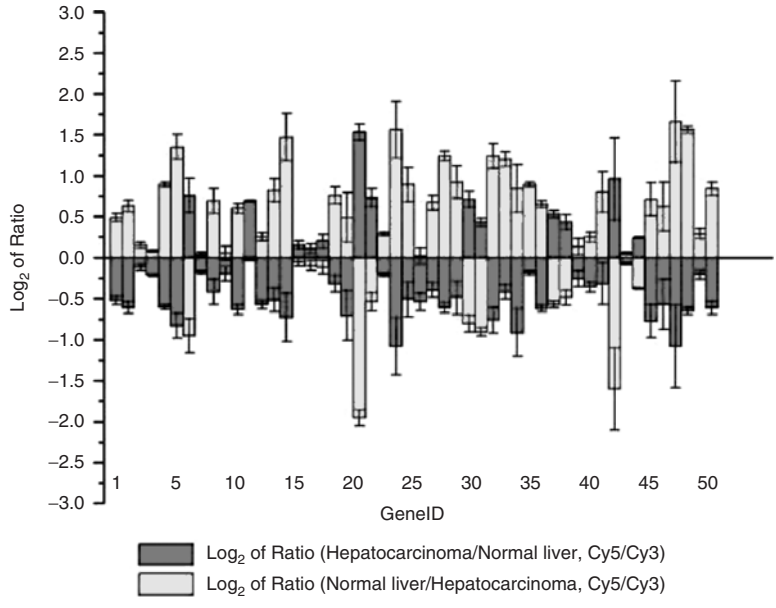

FIGURE 19.4. The histogram of 50 Genes ratio in reverse cross labeling microarrays.

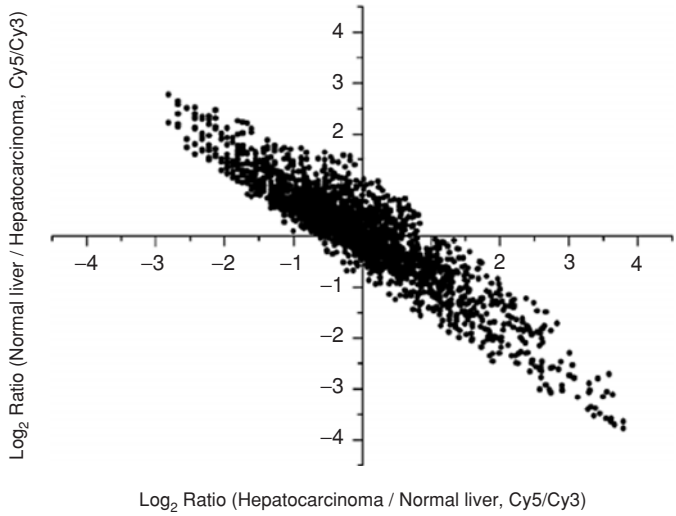

FIGURE 19.5. The scatter plot of ratio in reverse cross labeling microarrays.

fluorescence labeling on the experiments was investigated (Table 19.2). When the cutoff was defined as 0.5 and 2, there were two genes showing contradictory results in the microarray of different batches. Thus, the consistence rate was 99%. There were no more contradictory results in other microarrays and the consistence rate was 100%. But when the cutoff was defined as 0.667 and 1.5, only the same microarray had no genes of contradictory results. Moreover, the microarrays labeling simultaneously had less genes of contradictory results than microarrays labeling separately. The consistence rate commonly ranged from 93.6% to 100%. The consistence rate commonly ranged from 93.6% to 100%. Thus, it was concluded that when the cutoff was defined as 0.5 and 2, microarray had a high reproducibility, and the consistence rate reached 100%.

3.3. Evaluating the Reproducible Microarray Experiments by Consistence Rate Could Reflect the Reproducibility Better

We put forward a concept of "consistence rate" (CR), defined as the percent of gene number, which showed differential expression in the same direction in both of two experiments, from the total number of all differentially

TABLE 19.2. The comparison with different batches and labeling process

	Replicated gene in the same array	Different batches Labeling simultaneous	The same batch Labeling simultaneous	The same batch Labeling separately
Correlation coefficient (R)	0.942	0.910	0.908	0.878
Cutoff	0.5-2.0	0.5-2.0	0.5-2.0	0.5-2.0
CR	100%	99%	100%	100%
Contradictory genes and proportion	0	0.5% (2 genes)	0	0
Cutoff	0.667-1.5	0.667-1.5	0.667-1.5	0.667-1.5
CR	100%	93.6%	99.92%	99.88%
Contradictory genes and proportion	0	3.2% (12 genes)	0.04% (3 genes)	0.06% (6 genes)
Increasing multiple of co-differential genes	1.40	1.83	2.78	2.33

expressed genes. We considered that consistence rate was better than correlation coefficient (R) and coefficient of variation (CV) to reflect the reproducibility of the results. Table 19.2 showed that compared to correlation coefficient, consistence rate reflected the proportion of contradictory genes to differential genes more exactly. In addition, we found that the number of differential genes had a great impact on the value of correlation coefficient. Three sets of replicated experiments has been analyzed here: self-comparison experiments for replicating twice; low-differential expression replicated experiments for replicating twice (A); high-differential expression replicated experiments for replicating twice (B) (Table 19.3). It was indicated that correlation coefficient was related to the number of differential genes. The lower number of differential genes were, the more the correlation coefficient deviated from 1 and the less it can evaluated the correlation quality of two replicated experiments.

Moreover, the linear working range of scanner was from 800-60,000, but some of the signal intensities of high-abundant or low-abundant genes (especially the weak signals) were out of the linear range, which would make serious impacts on variation coefficient and correlation coefficient. However, such shortcomings could be conquered by consistence rate. In addition, cutoff value could also be determined with evaluating consistence rate. For instance, in terms of the replicated gene in the same array in the Table 2, if cutoff value was determined according to the consistent rate 100%, it is considered that the result was reliable when cutoff was 0.667-1.5. Thus, more differential genes could be obtained.

TABLE 19.3. Comparison of correlation coefficient and consistence rate among different replicated experiments

	Percent of differential genes to the total genes		R	Nd	Nid	CR
	Microarray 1	Microarray 2				
Self-comparison replicated experiments	0.68%	0.70%	0.002	0	0	–
Differential Expression replicated experiments A	2.49%	2.11%	0.882	141	0	100%
Differential Expression replicated experiments B	38.64%	36.35%	0.978	3151	0	100%

4. Discussion

We reported our investigation of the precision, accuracy and reliability of microarray data and the sources of variation here.

In term of cDNA microarrays prepared by arraying cDNA on the slides, the sensibility of the microarrays was related to the concentrations of target genes to some extent. We had already studied the sensibility based on the array of a single gene[12]. The research of multiple-gene hybridization showed that the signal intensity changed with concentrations altering, but the ratio was constant. Thus, the change of ratio could be used to represent the differentiation of gene expression.

In the reverse cross labeling experiments, we found no direct impact of labeling with Cy5 or Cy3 on the results. Two sets of experiments showed a rather good pertinence.

The comparison of the reproducibility of self-calibration experiments indicated that false positive rate of the microarrays was controlled below 3%, similar to the other reports (0.5-3%)[11]. It proved that our system was stable and reliable, and the data were reproducible. Moreover, the genes of differential expression screened from the experiments changed in the same direction in the replicated experiments, so we confirmed that these genes are really differentially expressed, but not a false positive signal.

The variation of expression pattern microarray was divided into biological variation and experiment systematic variation. The biological variation was mainly referred to sample variation, which meant that the samples from different persons perhaps had high differentiation. The expression of some special genes may not be the identical even in the cells from the same tissue. Such difference is difficult to calculate. We focused on discussing experiment systematic variation here. The above data suggested that microarray results were

reproducible to some extent, but owing to the existed systematic bias, microarray could only be regarded as a qualitative or half-quantitative approach. We discussed the possible reasons of variation as follows:

1. Linear working range of scanner: Some signal intensities were so weak or so strong that they were out of the linear range, which made the signal values of different arrays fluctuate. Therefore, variation enhanced inevitably.
2. Variation in the process of RNA isolation: The samples of the experiments were from ten different normal liver tissues. The total RNA isolation produced only a little contribution to the variation of the reproducibility (Table 1), but RNA degradation because of improper reservation hasn't been tested.
3. Variation in the process of RNA reverse-transcription: Even to the same gene after normalized, the efficiency of labeling for each time would have a little variation. If we performed labeling separately, we could eliminate the variation by replicating experiments (Figure 19.1).
4. Variation in the process of hybridization: This is the most important source of variation, including the inhomogeneity of solution for hybridization and dilution, different procedure, impurity, background and so on. Thus, the different processes of hybridization brought to different signal intensities.

The sources of variation mentioned in 2., 3. and 4. were always generated randomly. The best method to eliminate the variation was replicating the experiments. Two replicated experiments could avoid most deviations. To biological difference, it was better to replicate more than three times[4]. Comprehensive statistical results could be used to establish a special mathematical model, perform cluster analysis[13] and look for mark genes of diseases and polymorphism.

Another source of variation that could not be ignored was derived from image acquirement and data analysis. This source was related with scanner, analytic software and algorithm[14].

In addition to systematic variation during the experiments, the bias was also owing to the preparation of miraoarray, specially the accuracy of clones. It was reported that accuracy rate of the commercialized clones was about 60%-80%[15]. Thus, every clone should be sequenced to guarantee the reliability of results.

We put forward the "consistence rate" as a new parameter to evaluating the reproducible performance of microarray. Compared to the correlation coefficient and coefficient of variation, consistence rate took advantage in some aspects. Consistence rate, which was put forward based on the microarray as a half-quantitative approach, was used to evaluate the reproducibility by calculating the percent of the genes whose expression changed in the same direction in the replicated experiments from the whole differentially expressed genes. The parameter was not impacted by the number of differentially

expressed genes and very high-abundant and low-abundant genes out of the linear range of scanner. The criterion of cutoff was always disputed. Most researchers adopted 2 as cutoff value, and some used 3 or 1.7. We thought that cutoff value could be determined according to the consistent rate of corresponding experiment depended on detailed purpose. For instance, if in some experiments with a little genes of differential expression, the cutoff value could be redefined by reducing the consistence rate, so as to get more useful information from microarray data.

The results presented in this report demonstrated the performance of the cDNA microarray technology platform, and proved that the platform could provide data of high quality to establish a reliable gene expression database. The usefulness of any data acquired from this platform for scientific researchers depended on a strict method how to test the performance of this technology.

References

1. Schena M., Shalon D., Davis R. W., Brown P. O. 1995, Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science*, 270(20): 467-470.
2. Schena M., Shalon D., Heller R., Chai A., Brown P. O., Davis R. W. 1996, Parallel human genome analysis: Microarray based expression monitoring of 1000 genes. *PNAS*, 93: 10614-10619.
3. Li Y., Qiu M. Y., Wu C. Q., Cao Y. Q., Tang R., Chen X., Shi X. Y., Hu Z. Q., Xie Y., Mao Y. M. 2000, Detection of Differentially Expressed Genes in Hepatocellular Carcinoma Using DNA Microarray. *Acta Genetica Sinica (Chinese)*, 27(12): 1042-1048.
4. Shalon D., Smith S. J., Brown P. O. 1996, A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res*, 6(7): 639-645.
5. Lee M. L., Kuo F. C., Whitmore G. A., Sklar J. 2000, Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci USA*, 97(18): 9834-9839.
6. LifeArray Chip Validation Study [EB/OL]. http://www.incyte.com
7. Chomczynski P., Sacchi N. 1987, Single-Step Method of RNA Isolation by Acid Guanidinium Thiocyanate-Phenol-Chloroform Extraction. *Anal Biochem*, 162: 156-159.
8. Raha S., Merante F., Proteau G., Reed J. K. 1990, Simultaneous isolation of total cellular RNA and DNA from tissue culture cells using phenol and lithium chloride. *Genet Anal Tech Appl*, 7(7): 173-177.
9. Yang P., Otto C. M., Sheehan F. H. 1997, The effect of normalization in reducing variability in regional wall thickening. *J Am Soc Echocardiogr*, 10(3): 197-204.
10. Eisen M. B., Spellman P. T., Brown P. O., Botstein D. 1998, Cluster Analysis and Display of Genome-Wide Expression Patterns. *Proc Natl Acad Sci*, 95: 14863-14868.
11. Yue H., Eastman P. S., Wang B. B., Minor J., Doctolero M. H., Nuttall R. L., Stack R. Becker J. W., Montgomery J. R., Vainer M., Johnston R. 2001, An

evaluation of the performance of cDNA microarrays for detecting changes in global mRNA expression. *Nucleic Acids Res*, 29: e41.
12. Li Y., Li Y. L., Tang R., Xu H., Qiu M. Y., Chen Q., Chen J. X., Fu Z. R., Ying K., Xie Y., Mao Y. M. 2002, Discovery and analysis of hepatocellular carcinoma genes using cDNA microarrays. *J Cancer Res Clin Oncol*, 128: 369-379.
13. Kuruvilla F. G., Park P. J., Schreiber S. L. 2002, Vector algebra in the analysis of genome-wide expression data. *Genome Biol*, 3: RESEARCH0011
14. Hsiao L. L., Jensen R. V., Yoshida T., Clark K. E., Blumenstock J. E., Gullans S. R. 2002, Correcting for signal saturation errors in the analysis of microarray data. *Biotechniques*, 32(2): 330-332, 334, 336.
15. Kothapalli R., Yoder S. J., Mane S., Loughran T. P. Jr. 2002, Microarray results: how accurate are they? *BMC Bioinformatics*, 3(1): 22