Analysis of Affymetrix GeneChip® data using amplified RNA

Leslie Cope¹, Scott M. Hartman², Hinrich W.H. Göhlmann³, Jay P. Tiesman², and Rafael A. Irizarry⁴

¹The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD, ²The Procter & Gamble Company, Cincinnati, OH, USA, ³Johnson & Johnson Pharmaceutical Research & Development, Beerse, Belgium, and ⁴Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

BioTechniques 40:165-170 (February 2006)
doi 10.2144/000112057

When small biological samples are collected by microdissection or other methods, amplification techniques are required to provide sufficient target for hybridization to expression arrays. One such technique is to perform two successive rounds of T7-based in vitro transcription. However, the use of random primers, required to regenerate cDNA from the first round of transcription, results in shortened copies of cDNA from which the 5' end is missing. In this paper we describe an experiment designed to compare the quality of data obtained from labeling small RNA samples using the Affymetrix Two-Cycle Eukaryotic Target Labeling procedure to that of data obtained using the One-Cycle Eukaryotic Target Labeling protocol. We utilized different preprocessing algorithms to compare the data generated using both labeling methods and present a new algorithm that improves upon existing ones in this setting.

INTRODUCTION

The standard method of target synthesis for hybridization to Affymetrix GeneChip® expression microarrays requires a relatively large amount of input total RNA (1–15 µg). When small biological samples are collected by microdissection or other methods, amplification techniques are required to provide sufficient target for hybridization to expression arrays. One such technique is to perform two successive rounds of T7-based in vitro transcription. However, the use of random primers required to regenerate cDNA from the first round of transcription results in shortened copies of cDNA from which the 5' end is missing.

Several recent studies have investigated the reliability of gene expression measures obtained using twice-amplified RNA in both cDNA arrays and GeneChips (1–9). In most cases the investigators conclude that amplified RNA produces quality microarray data. They find that the expression levels of amplified samples are highly correlated to one another and have reduced, but significant, correlation with nonamplified samples. Some of these studies

have compared methods for sample preparation and amplification, to develop optimal laboratory protocols (8,9). To date, however, little effort has been made to optimize data processing procedures for microarray studies using the alternative labeling strategies required for small amounts of RNA.

For many microarray platforms, there is probably little that can be done beyond identifying and marking bad players among the probes. Affymetrix GeneChips, where multiple probes are used to measure expression for each transcript, are a possible exception. Given the loss of the 5' end of cDNA transcripts, the position of each probe within an amplified transcript should influence the hybridization of the probe, and those that are closer to the 3' terminus are expected to be more reliable than probes taken from the 5' end of a transcript.

Our goals in this study are to investigate the effect of probe position on absolute and relative expression measurements and to evaluate the performance of a new probe set summary, small-sample RMA (sRMA), designed to minimize the random primer effect in two-round labeling protocols.

MATERIALS AND METHODS

Human total RNA (Clontech, Mountain View, CA, USA) was obtained from both male (testes) and female tissues (breast and cervix) and mixed to make two separate samples (90% breast/10% testis and 90% cervix/10% testis) that should exhibit differential gene regulation for many probe sets. The 10% testis RNA provides a background set of 45 Y-linked genes that should not be differentially regulated between the two samples.

Six technical replicates (10 µg each) of each RNA mixture were labeled according to the Affymetrix One-Cycle Eukaryotic Target Labeling method, and six (50 ng each) were labeled according to the Affymetrix Two-Cycle Eukaryotic Target Labeling procedure (www.affymetrix.com/support/technical/expression_manual.affx; Affymetrix, Santa Clara, CA, USA). Briefly, the one-cycle method consists of an oligo dT/T7 promoter-mediated reverse transcription of total RNA, followed by a T7-based in vitro transcription reaction incorporating biotin rNTPs. The two-cycle labeling protocol consists of two successive rounds of T7-based in vitro transcription incorporating biotin rNTPs in the second round reaction. The conversion of first round cRNA back to cDNA for the second round, via a random primer-mediated reverse transcription reaction, is believed to be the major source of increased 3' bias in targets generated using this labeling protocol. The resultant targets from both labeling methods were hybridized to 24 Affymetrix HGU133A GeneChips according to the manufacturer's instructions.

In the GeneChip platform, substantial data processing is required after image analysis to obtain expression level measurements. In this study, we compared four different processing algorithms: Affymetrix MAS 5.0, Li and Wong's Model Based Expression Index (10), RMA (11), and a new protocol, called sRMA (available at www.biostat.jhsph.edu/~ririzar/Software/srma.R), introduced for the first time here.

The small-sample version of RMA uses the same background adjustment and cross-chip normalization procedure as the standard RMA algorithm and, in keeping with the RMA philosophy, uses a robust linear model to summarize probe level expression values. Specifically, we model background adjusted and normalized probe intensities as $\log_2(Y_{ijk}) = \theta_{ik} + \phi_{jk} + \epsilon_{ijk}$, $i = 1, \dots, I$, $j = 1, \dots, J$, $k = 1, \dots, K$. Here, k represents array, i represents probeset, and j represents probe, θ represents a quantity proportional (in the log scale) to the amount of RNA, ϕ represents a probe-specific effect, and ϵ , representing measurement error, is assumed to have a probe-specific variance σ^2_{ij} . The standard RMA algorithm uses median polish (12), an ad hoc robust procedure, to estimate θ , but code has recently been made available to fit the model above using formal, robust statistical procedures (13). The new implementation accommodates user-defined weights for each probe, and sRMA takes advantage of this by weighting the contribution of each probe according to its relative 5'/3' position in the transcript using the inverse of the position-specific coefficient of variation.

Specifically, for the j th probe of the i th probe set, we define α_{ij} to be the relative 5'/3' position of the probe

within its transcript, a number between 0 and 1. Thus, if a probe is located at the 7000th base, counting from the 5' end of a transcript that includes a total of 10,000 nucleotides, α will be 0.7. The location of each 25-mer probe is defined according to the position of the middle base. We compute the position dependent effect $\mu(\alpha_{ij})$ by regressing log-scale intensity values $\log_2(y_{ijk})$ on the α_{ij} values. To compute the position dependent variance $\sigma^2(\alpha_{ij})$, we estimated probe-specific standard deviations σ_{ij} using the replicate GeneChips and regressed these on the α_{ij} values. The position-dependent standard deviations and effects were used to compute the coefficient of variation, which is then normalized so that $\sum_j \sigma(\alpha_{ij})/\mu(\alpha_{ij}) = 1$ to define the weights.

Expression values for Li and Wong's Model Based Expression Index were calculated using the dChip software. All other computations were done using the R language (14) and packages from the Bioconductor Project (15).

RESULTS AND DISCUSSION

A box plot of the raw probe level data for each GeneChip does not show any clear, systematic difference in intensity distribution between the two protocols

(see Supplementary Figure S1 available online at www.BioTechniques.com). Likewise, MA plots comparing the two protocols, while showing that probe intensities are substantially different across protocols, reveal no apparent patterns (Supplementary Figure S2). However, a simple clustering algorithm can separate the arrays from the two protocols perfectly (Supplementary Figure S3). The labeling method certainly influences expression patterns, but it is difficult to characterize the effects, and it would take a very large-scale experiment to reliably identify individual bad probes. We also looked at log-fold-changes for the 45 genes specific to the Y chromosome in various two-chip comparisons. These genes should not show differential expression; thus their log-fold-changes should be 0. The variation among these fold-changes is higher after the second round of amplification (Supplementary Figure S4). However, with RMA and sRMA, the increase is not substantial, and even the most extreme log-fold-changes are quite close to 0 as expected.

Additional effects of a second round of amplification on expression values can be clearly seen when the probe position α is taken into account. Specifically, if we regress log intensity values against α , we obtain very different slopes for the amplified and standard protocols (Figure 1A). As is expected, the additional round of amplification in the two-cycle protocol labeled targets results in lower expression levels (Figure 1A) and greater variation (Figure 1B) for probes at the 5' end of a transcript as compared with probes from the 3' end. These observations motivated the sRMA algorithm described above, which is designed to limit the influence that probes from the 5' end of the transcript can have on the final expression estimates.

So far we have looked at absolute expression levels. However, gene expression arrays are typically used to assess relative gene expression rather than to quantify RNA abundance. Technological effects are likely to cancel out when expression values are compared and so may not present practical problems. Accordingly, for

Figure 1. Probe position specific effects. (A) Loess was used to model log intensity as a function of probe position within the transcript. Transcripts were standardized to have a length of 1, and the x-axis depicts probe position on this scale with 0 denoting the 5' terminus. Expression values were centered before fitting, so the y-axis indicates deviation of mean probe intensity. Each line represents a different chip, and the different colors/line-types represent the different protocols. (B) Loess was used to model the variance of probe intensity as a function of probe position within the transcript. As in panel A, position is a value between 0 and 1. Replicate chips were used to estimate variance for each probe, so separate lines were not fit for individual chips.

Table 1. Agreement Between Results from the One-Cycle and Two-Cycle Protocols

Agreement Between Protocols								
	<i>t</i> -test				Log-Fold-Change			
	<i>n</i> = 50		<i>n</i> = 100		<i>n</i> = 50		<i>n</i> = 100	
	Median (%)	90% CI	Median (%)	90% CI	Median (%)	90% CI	Median (%)	90% CI
MAS 5.0	58	(50,64)	56	(51,61)	72	(64,76)	52	(48,57)
dChip	64	(58,72)	68	(62,73)	62	(56,68)	53	(45,58)
RMA	64	(56,72)	68	(61,73)	86	(84,90)	85	(79,88)
sRMA	70	(60,80)	74	(63,81)	88	(86,92)	87	(81,91)
Agreement Within Standard One-Cycle Protocol								
MAS 5.0	62	(58,68)	62	(58,65)	80	(74,84)	53	(50,56)
dChip	66	(58,72)	71	(67,75)	70	(66,76)	56	(51,59)
RMA	64	(58,72)	65	(60,70)	90	(90,94)	88	(84,91)
sRMA	68	(60,72)	69	(64,84)	92	(88,96)	90	(86,93)

Percentage of genes in common is shown for lists of *n* = 50 and *n* = 100 genes. A list was created for each protocol using both the moderated *t*-test and log-fold-change to rank genes. CI, confidence interval; sRMA, small-sample RMA.

Figure 2. CAT plots of cross-protocol agreement. (A) Using the moderated *t*-statistic as the measure of differential expression, all genes were ranked according to amount of differential regulation. This was done separately for each processing algorithm and within each labeling protocol. This plot shows the amount of overlap observed between protocols when lists containing the top *n* genes are intersected. The percentage agreement is calculated as (no. of genes common to 2 lists)/(size of lists) × 100%. The four preprocessing algorithms are represented in different colors. (B) This plot is identical to that in panel A, but log-fold-change is used as the measure of differential expression instead of the moderated *t*-statistic. sRMA, small-sample RMA.

our main analysis, let us consider the most common application of microarray technology: identifying differentially expressed genes. To examine how well results from a study using the two-cycle protocol agree with results obtained using the one-cycle protocol, we compared lists of various sizes using CAT plots (16). That is, we identify the *n* most differentially expressed genes under each of the two competing protocols, and plot the percentage of common genes as a function of *n*. We

used two different measures of differential expression to construct the lists, a moderated *t*-test (17) (Figure 2A) and also the average log-fold-change (Figure 2B). When the moderated *t*-test was used, RMA achieved about 65%–70% agreement for lists between 25 and 100 genes, typical of gene lists reported in studies. When fold-change was used, agreement rose to 85%–90%. MAS 5.0 had very high rates of agreement when fold-change was used as the measure of differential expression and

as long as lists were small, but performance fell off rapidly as the number of genes under consideration increased. With the moderated *t*-test, MAS 5.0 agreement was consistently 5–10 percentage points lower than that of RMA. The new expression summary sRMA out performs RMA on average, but the advantage is slight. The results for dChip were similar to those of RMA when using the moderated *t*-test and similar to MAS 5.0 when using fold-change.

We are not able to directly evaluate the significance of the differences in performance seen in the CAT plots. To get a better sense of how the methods compare, we used a resampling scheme.

We randomly chose sets of three arrays from each sample type and looked at the agreement between results from the two protocols. Results are shown in the top half of Table 1. In each case, confidence intervals for RMA and sRMA were similar, and MAS 5.0 and dChip did not perform as well. We suspect that sample size is an important consideration here, because in this case each group has only three samples as opposed to the six samples that were used in the CAT plots. Recall that the weights used in the sRMA are estimated from the data. The extremely small sample sizes used in this analysis may destabilize the weight estimates enough to eliminate any advantage the approach offers.

For comparison, we repeatedly split the six standard protocol replicates into two random groups of three and tested agreement within protocol as well. The results are very interesting, because agreement is in some cases worse within the one-cycle labeling protocol than it is between protocols. This may be due to the fact that one of the arrays prepared under the one-cycle protocol is slightly different than the rest, even after normalization. Because all one-cycle arrays are used each time agreement is assessed within a protocol, the anomalous chip appears in one of the two groups and could reduce the level of agreement. For most

of the comparisons used, agreement within the protocol is slightly better than is observed between protocols, but the confidence intervals overlap substantially.

These results are very encouraging. The level of agreement across protocols is similar to that within protocol for each expression summary and each measure of differential expression. This suggests that the two-cycle labeling protocol does not degrade the quality of the hybridization significantly. Analytic methods that work well for the standard protocol will also work well for small samples. This stands to reason: Affymetrix probes are preferentially selected from the 3' end of each transcript in order to minimize the effects of both natural RNA degradation, and the labeling protocol. Although the probe position effects shown in Figure 1 are distinct and statistically significant, they are small compared with overall expression levels. And these effects may be partially cancelled out when relative measures of expression are considered. Weighting the contribution of each probe by position seems to slightly reduce the bias introduced under the small sample protocol and improves agreement between gene lists obtained under the two processing methods. There is also slight improvement in variance, since the more variable probes are less influential. Another reason why RMA performs nearly as well as sRMA may be that the median polish algorithm used in standard RMA naturally limits the influence of the 5' most probes on the chip. Those probes tend to have low expression values, and so are less likely to influence the median. In light of this study, we can conclude that the improvement is slight, not because sRMA fails to address the problems that arise when analyzing small sample expression data, but because these problems are minor ones and the data does not require special treatment.

It appears that the two-cycle protocol leads to reproducible estimates of expression, although these may differ somewhat from measures of expression obtained under the one-cycle protocol. It is probably better to use two-cycle data alone rather than mixing it with

data prepared under the standard one-cycle protocol as a matter of normal caution. However, this separation may not always be feasible and does not appear to be a very important consideration. In particular, we can foresee that an exploratory study using a larger and less pure tissue sample might be followed by a confirmation study using RNA obtained from a purer set of cells. Every indication from this study is that data quality and reproducibility is high.

ACKNOWLEDGMENTS

Many thanks to James MacDonald at University of Michigan for providing dChip expression measures.

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES

1. King, C., N. Guo, G.M. Frampton, N.P. Gerry, M.E. Lenburg, and C.L. Rosenberg. 2005. Reliability and reproducibility of gene expression measurements using amplified RNA from laser-microdissected primary breast tissue with oligonucleotide arrays. *J. Mol. Diagn.* 7:57-64.
2. Luzzi, V., M. Mahadevappa, R. Raja, J.A. Warrington, and M.A. Watson. 2005. Accurate and reproducible gene expression profiles from laser capture microdissection, transcript amplification, and high density oligonucleotide microarray analysis. *J. Mol. Diagn.* 5:9-14.
3. Ohyama, H., X. Zhang, Y. Kohno, I. Alevizos, M. Posner, D.T. Wong, and R. Todd. 2000. Laser capture microdissection-generated target sample for high-density oligonucleotide array hybridization. *BioTechniques* 35:530-536.
4. McClintick, J., R. Jerome, C. Nicholson, D. Crabb, and H. Edenberg. 2003. Reproducibility of oligonucleotide arrays using small samples. *BMC Genomics* 4:4.
5. Thelen, P., P. Burfeind, M. Grzmil, S. Voigt, R. Ringert, and B. Hemmerlein. 2004. cDNA microarray analysis with amplified RNA after isolation of intact cellular RNA from neoplastic and non-neoplastic prostate tissue separated by laser microdissections. *Int. J. Oncol.* 24:1085-1092.
6. Gomes, L., R. Silva, B. Stolf, E. Cristo, R. Hirata, F. Soares, L. Reis, E. Neves, and A. Carvalho. 2003. Comparative analysis of amplified and nonamplified RNA for hybridization in cDNA microarray. *Anal. Biochem.* 321:244-251.
7. Jensen, S.D., R.S. Robetorye, S.D. Bohling, J.A. Schumacher, J.W. Morgan, M.S. Lim, and K.S.J. Elenitoba-Johnson. 2003. Validation of cDNA microarray gene expression data obtained from linearly amplified RNA. *Mol. Pathol.* 56:307-312.
8. Klur, S., K. Toy, M.P. Williams, and U. Certa. 2004. Evaluation of procedures for amplification of small-size samples for hybridization on microarrays. *Genomics* 83:508-517.
9. Singh, R., R.J. Maganti, S.V. Jabba, M. Wang, G. Deng, J.D. Heath, N. Kurn, and P. Wangemann. 2005. Microarray-based comparison of three amplification methods for nanogram amounts of total RNA. *Am. J. Physiol. Cell Physiol.* 288:1179-1189.
10. Li, C. and W.H. Wong. 2001. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl. Acad. Sci. USA* 98:31-36.
11. Irizarry, R.A., B. Hobbs, F. Collin, Y.D. Beazer-Barclay, K.J. Antonellis, U. Scherf, and T.P. Speed. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4:249-264.
12. Hoaglin D.C., F. Mosteller, and J.W. Tukey. 2000. Understanding Robust and Exploratory Data Analysis. John Wiley & Sons, New York.
13. Bolstad, B.M., F. Collin, J. Brettschneider, K. Simpson, L. Cope, R.A. Irizarry, and T.P. Speed. 2005. Quality assessment of Affymetrix, p. 33-49. In R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber (Eds.), *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. Springer, New York.
14. R Development Core Team. 2004. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.
15. Gentleman, R.C., V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5: R80.
16. Irizarry R.A., D. Warren, F. Spencer, I.F. Kim, S. Biswal, B.C. Frank, E. Gabrielson, J.G.N. Garcia, et al. 2005. Multiple-laboratory comparison of microarray platforms. *Nat. Methods* 2:345-350.
17. Smyth, G.K. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3:3.

Received 29 July 2005; accepted 20 September 2005.

Address correspondence to Leslie Cope, Department of Oncology, The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, 550 North Broadway, Suite 1103, Baltimore, MD 21205, USA. e-mail: cope@jhu.edu; or to Rafael A. Irizarry, Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, 615 N. Wolfe E3627, Baltimore, MD 21205, USA. e-mail: rafai@jhu.edu

To purchase reprints

of this article, contact

Reprints@BioTechniques.com