8-1-2005

Analysis of Affymetrix GeneChip Data Using Amplified RNA

Leslie Cope
Department of Oncology, The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, cope@jhu.edu

Scott M. Hartman
The Procter & Gamble Company, Miami Valley Innovation Center, Cincinnati, OH

Hinrich W.H. Gohlmann
Functional Genomics, Johnson & Johnson Pharmaceutical Research & Development, Belgium

Jay P. Tiesman
The Proctor & Gamble Company, Miami Valley Innovation Center, Cincinnati, OH

Rafael A. Irizarry
Johns Hopkins Bloomberg School of Public Health, rafa@jhu.edu

Analysis of Affymetrix GeneChip Data Using Amplified RNA

Leslie Cope[1*], Scott M. Hartman[2], Hinrich W.H. Göhlmann[3], Jay P. Tiesman[2], and Rafael A. Irizarry[4*]

[1]Department of Oncology, The Sidney Kimmel Comprehensive Cancer Center Johns Hopkins University, 550 North Broadway, suite 1103, Baltimore, MD, 21205. cope@jhu.edu

[2]The Procter & Gamble Company, Miami Valley Innovation Center, P.O. Box 538707, Cincinnati, OH 45253-8707

[3]Functional Genomics, Johnson & Johnson Pharmaceutical Research & Development, Turnhoutseweg 30, 2340 Beerse, Belgium

[4]Department of Biostatistics, Bloomberg School of Public Health, 615 N. Wolfe E3627, Baltimore, MD 21205. rafa@jhu.edu

[*]To whom correspondance should be addressed

Abstract

The standard method of target synthesis for hybridization to Affymetrix GeneChip® expression microarrays requires a relatively large amount of input total RNA (1-15 micrograms). 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 amplification technique used is to perform two successive rounds of T7-based in vitro transcription. However, the use of random primers required to re-generate cDNA from the first round transcription reaction results in shortened copies of the cDNA, and ultimately the cRNA, transcripts 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 Small Sample Target Labeling Protocol V 2 to that of data obtained using the standard 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

Several recent studies have investigated the reliability of gene expression measures obtained using amplified RNA in both cDNA arrays and Affymetrix 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 non-amplified 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 labeling 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. The position of each probe within an amplified transcript influences the intensity value of the probe and those that are more 3'-ward 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 distributional differences between data generated under the two protocols, and to compare probeset summaries with respect to performance on data obtained from both protocols. In the remaining sections we briefly describe the data, and the data analyses used and then present results. In the discussion we make specific recommendations about the processing and use of data from both protocols.

Materials and Methods

Human total RNA from breast, cervix, and testes (Clontech) was mixed to create two composite samples. The first is ninety percent breast and ten percent testes, the second ninety percent cervix and ten percent testes. This mixture strategy permits us to assume that forty-five Y-linked genes in the testes sample should be identically expressed in the two mixtures. However, since these genes are present only in the testicular tissue, which comprises only 10% of each sample, we do not expect to see a typical range of expression values. Nonetheless, these genes offer particular insight into the differences between the two protocols. For all other chromosomes, we expect many genes to be differentially expressed in the two mixtures.

Human total RNA (Clontech) 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 serve to illustrate differential gene regulation for many probe sets. The 10% testis RNA provides a background set of forty-five Y-linked genes that should not show differential regulation between the two samples and offer particular insight into the differences between the two labeling protocols.

From both RNA mixture samples six technical replicates were labeled each by the Affymetrix standard method (10 _g each) and by the Affymetrix Small Sample Target Labeling Protocol Version 2 method (50 ng each) [10]. The resultant targets were hybridized to twenty-four Affymetrix HGU133A GeneChips® according to standard procedures [11].

In the Affymetrix GeneChip® platform, substantial data processing is required after image analysis to obtain expression level measurements. In this paper, we compared results obtained using three different processing algorithms: Affymetrix's MAS 5.0 [11], RMA [12] and a new protocol called sRMA (for small-sample RMA) 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 $\log_2(Y_{ijk})=\theta_{ik}+\phi_{jk}+\varepsilon_{ikj}$, $i=1,...,I$, $j=1,...,J$, $k=1,...,K$. 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 [13], 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 [14]. 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 (see Figure 1).

Specifically, for the j^{th} probe of the i^{th} probeset, we define α_{ij} to be the relative 5'/3' position of the probe within its transcript, a number between 0 and 1. We compute the position dependent effect by regressing log-scale intensity values $\log_2(y_{ijk})$ on the α_{ij}

values. To compute the position dependent variance we estimated probe-specific standard deviations σ_{ij} using the replicate GeneChips® and regressed these on the α_{ij} values. The α dependent standard deviations and effects were used to compute the coefficient of variation which then defined the weights. All computations were done using the R language and packages from the Bioconductor Project [15,16].

Results and Discussion

A plot of the raw probe level data for each GeneChip® does not show any clear, systematic difference in intensity distribution between the two protocols (supplemental Figure 1). However, a simple clustering algorithm can separate the arrays from the two protocols almost perfectly (supplemental Figure 2). These differences can be explained by looking at the effect of α on the probe intensities. 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 Small Sample 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 to probes from the 3' end (Figure 1a).

To better understand the post-processing effects of these differences, we looked at log-fold- changes for the forty-five genes specific to the Y-chromosome in various two-chip comparisons. These genes should not show differential expression; thus their log-fold-chages should be 0. The variation among these fold changes is higher after the second round of amplification (Figure 2). 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.

Although the second round of amplification clearly affects data, the adverse effect on downstream results may be minor. To see this, let us consider
the most common application of microarray technology: identifying differentially expressed genes. We ask the question of how well results from a study using the small-

sample protocol agree with results obtained using the standard protocol. We compared lists of various sizes using CAT plots [17]. That is, we identify the *n* most differentially expressed 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, the moderated t-test [18] (Figure 3a) and also the average log-fold-change (Figure 3b). Using RMA we obtained good agreement between the lists. Some improvement is possible with sRMA, though the advantage over standard RMA is small.

To assess the statistical significances of the differences between the different expression measures we used a resampling scheme. Specifically, we randomly chose sets of 3 arrays from each sample type and looked at agreement between results from the two protocols. The 90% confidence intervals for the agreements in lists of size 100 obtained using moderated t-tests obtained from the permutation procedure were (51%, 61%), (64%, 81%), and (63%, 81%) for MAS 5.0, RMA, and sRMA respectively. For fold-change the intervals were (48%, 57%), (80%, 91%), and (81%, 91%). Results for several other list sizes were similar, in each case confidence intervals RMA and sRMA were nearly identical throughout.

For comparison, we repeatedly split the 6 standard protocol replicates into two random groups of 3 and tested agreement there. The results (not shown) were comparable to those described above and, in fact, the 90% confidence intervals calculated for the standard/standard comparison in every case substantially overlapped those calculated for the standard/small-sample comparison.

These results are very encouraging. There is a good deal of agreement between results obtained utilizing the two protocols, especially when robust methods are used to summarize probe level data. Weighting the contribution of each probe by position slightly reduces 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.

Based on the results of our study we give the following specific recommendations:

- When possible use either standard or small sample protocol, although it is reasonable to combine them.
- Comparative measures like differential expression will be more similar than absolute measures like expression level.
- When small sample data is used to verify an exploratory finding obtained with a less pure sample, the differences are probably less important.

It appears that the small-sample protocol leads to reproducible estimates of expression, although these may differ somewhat from measures of expression obtained under the standard protocol. It is probably better to use small sample data alone rather than mixing it with data prepared under the standard protocol. However, this separation may not always be feasible. 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. In comparing results from these two, it is reassuring to know that results are as reproducible as possible.

References (in no specific order)

1. King, C., Guo, N., Frampton, G. M., Gerry, N. P., Lenburg, M. E., and Rosenberg, C. L. Reliability and Reproducibility of Gene Expression Measurements Using Amplified RNA from Laser-Microdissected Primary Breast Tissue with Oligonucleotide Arrays. *J Mol Diagn* **7**(1), 57–64 (2005).
2. Luzzi, V., Mahadevappa, M., Raja, R., Warrington, J. A., and Watson, M. A. Accurate and Reproducible Gene Expression Profiles from Laser Capture Microdissection, Transcript Amplification, and High Density Oligonucleotide Microarray Analysis. *J Mol Diagn* **5**(1), 9–14 (2003).
3. Ohyama, H., Zhang, X., Kohno, Y., Alevizos, I., Posner, M.,Wong, D., and R., T. Laser capture microdissection-generated target sample for high-density

oligonucleotide array hybridization. *Biotechniques* **29**(3), 530–6 (2000).

4. McClintick, J., Jerome, R., Nicholson, C., Crabb, D., and Edenberg, H. Reproducibility of oligonucleotide arrays using small samples. *BMC Genomics* **4**(1) (2003).
5. Thelen, P., Burfeind, P., Grzmil, M., Voigt, S., Ringert, R., and Hemmerlein, B. 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**(5), 1085–1092 (2004).
6. Gomes, L., Silva, R., Stolf, B., Cristo, E., Hirata, R., Soares, F., Reis, L., Neves, E., and Carvalho, A. Comparative analysis of amplified and nonamplified rna for hybridization in cDNA microarray. *Anal Biochem* **321**(2), 244–251 (2003).
7. Jenson, S. D., Robetorye, R. S., Bohling, S. D., Schumacher, J. A., Morgan, J. W., Lim, M. S., and Elenitoba-Johnson, K. S. J. Validation of cDNA microarray gene expression data obtained from linearly amplified RNA. *Mol Pathol* **56**(6), 307–312 (2003).
8. Klur, S., Toy, K., Williams, M. P., and Certa, U. Evaluation of procedures for amplification of small-size samples for hybridization on microarrays. *Genomics* **83**(3), 508–517,March (2004).
9. Singh R, Maganti RJ, Jabba SV, Wang M, Deng G, Heath JD, Kurn N, and Wangemann P. Microarray-based comparison of three amplification methods for nanogram amounts of total RNA. Am J Physiol Cell Physiol. 288(5), 1179-89 (2005).
10. www.affymetrix.com/support/technical/other/twocycle_presentation.pdf
11. www.affymetrix.com/support/technical/expression_manual.affx
12. Irizarry, R. A. et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–64 (2003)Hoaglin D.C., Mosteller F., and Tukey, J. W. Understanding Robust and Exploratory Data Analysis, John Wiley and Sons (2000)B. M. Bolstad, F. Collin, J. Brettschneider, K., Simpson, L. Cope, R. Irizarry, T. P. Speed
15. Quality Assessment of Affymetrix GeneChip Data. In: Bioinformatics and Computational Biology Solutions using R and Bioconductor, R. Gentleman, V.

Carey, S. Dudoit, R. Irizarry, W. Huber (eds.), Springer, New York. (Due August 2005)

16. Gentleman, R. C. et al. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
17. Irizarry R.A., Warren D., Spencer F., Kim I.F., Biswal S., Frank B.C., Gabrielson E., Garcia J. G. N., et al. Multiple-laboratory comparison of microarray platforms *Nature Methods* 2, 345 - 350 (2005)
18. Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology 3, No. 1, Article 3

Figure Legends

1. Effect of amplification on probe-level data. A) Smooth lines were fitted to a scatter plot of the log intensities of each probe against the relative 5'/3' (or should this be 3'/5'? position in the transcript. Each line represents a different chip. The different colors/line-types represent the different protocols. B) We computed probe-specific standard devtiations of log intensities using the replicated chips. The lines were obtined by smoothing a scatter-plot of these standard deviations against the relative genomic position.
2. Box-plots of the log-fold-changes for the genes on the Y chromosome. The experiment was designed so that these genes are not expected to change; thus all log-fold-changes should be 0. A box-plot is shown for each of the pre-processing algorithm we considered.
3. CAT plots showing agreement in differential expression. The colors represent the different pre-processing algorithms. A) Lists created using moderated t-tests. B) Lists created using fold-change.

Supplemental Figure Legends

1. Box-plots of the probe-level log intensities for all chips hybridized in the study.
2. We used hierarchical clustering with correlation on the log-intensity of all probes as a similarity metric. The heatmap was organized using the results of this clustering algorithm. Although the all probes were used for the clustering procedure only 1000 probes are shown in the heatmap. These were the 1000 genes with most variance across chips and were clustered for the heatmap.

Figure 1

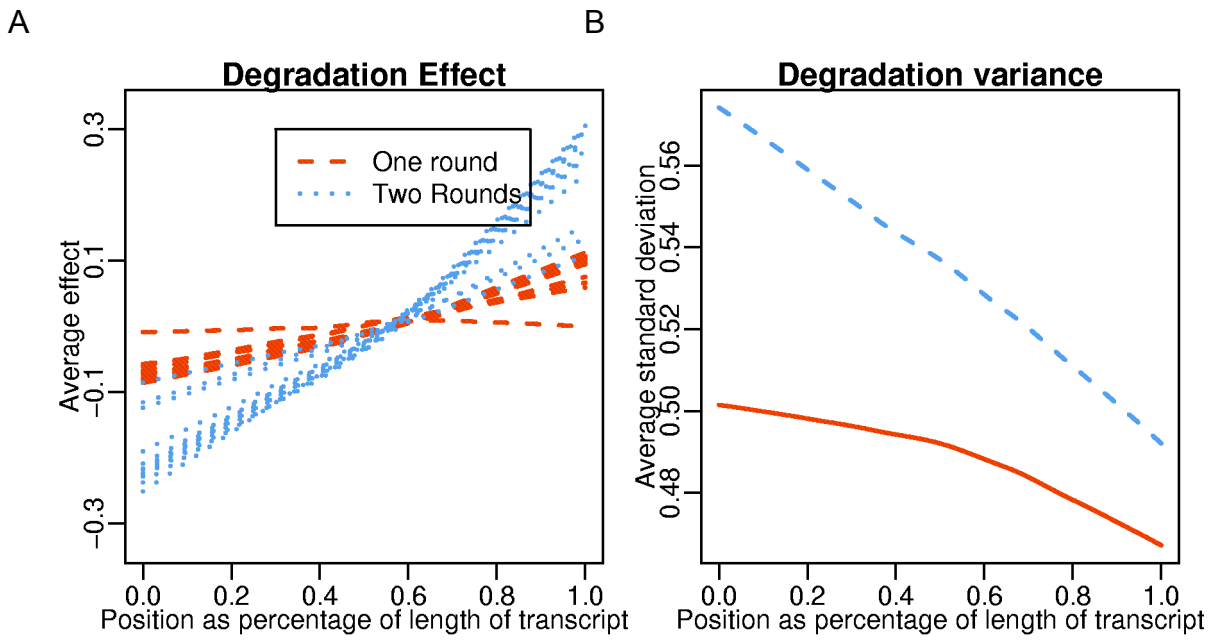


Figure 2

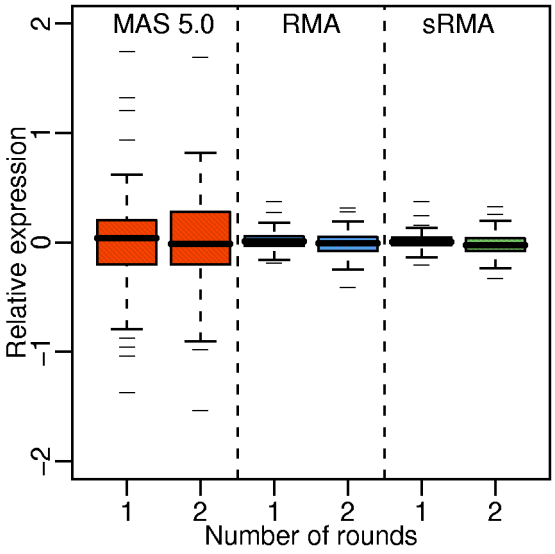
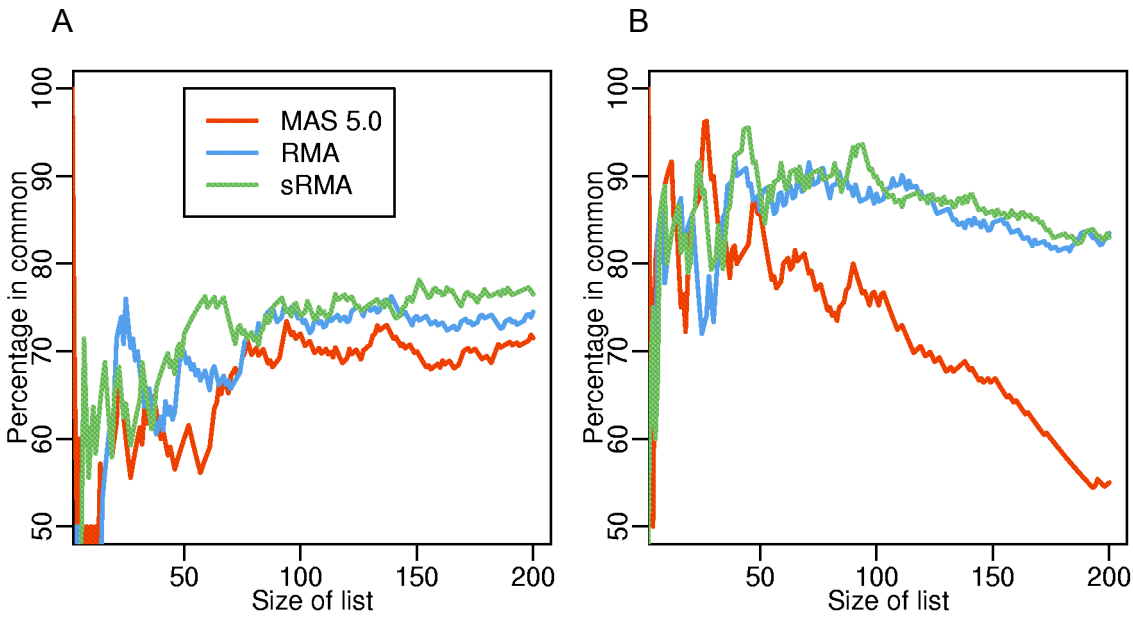
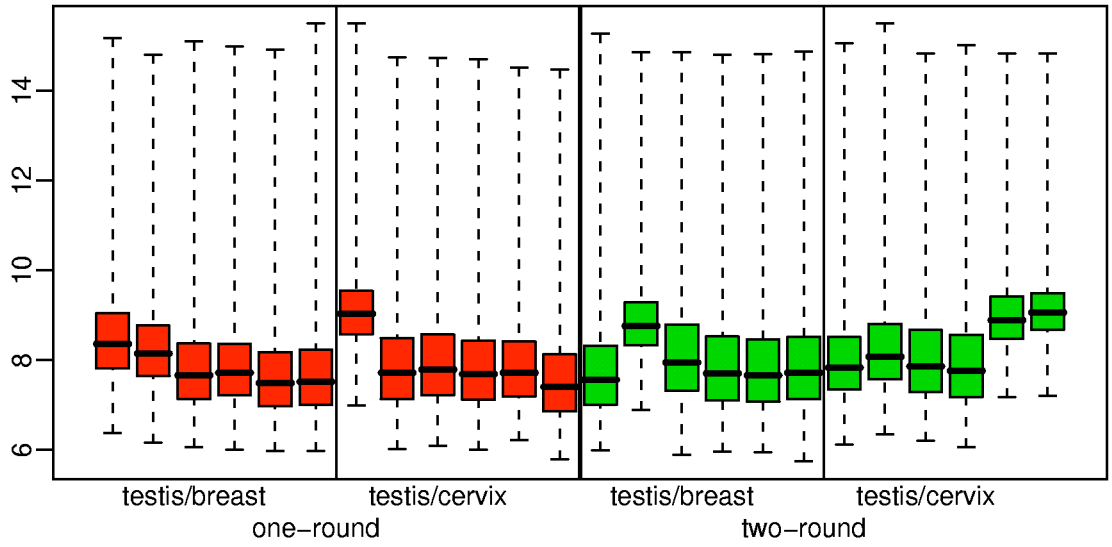


Figure 3



Supplemental Figure 1



Supplemental Figure 2