

RNA profiling for biomarker discovery: Practical considerations for limiting sample sizes

Danny J. Kelly and Sujoy Ghosh*

Genetics Research, GlaxoSmithKline R&D, Research Triangle Park, NC 27709, USA

Abstract. We have compared microarray data generated on AffymetrixTM chips from standard (8 micrograms) or low (100 nanograms) amounts of total RNA. We evaluated the gene signals and gene fold-change estimates obtained from the two methods and validated a subset of the results by real time, polymerase chain reaction assays. The correlation of low RNA derived gene signals to gene signals obtained from standard RNA was poor for less to moderately abundant genes. Genes with high abundance showed better correlation in signals between the two methods. The signal correlation between the low RNA and standard RNA methods was improved by including a reference sample in the microarray analysis. In contrast, the fold-change estimates for genes were better correlated between the two methods regardless of the magnitude of gene signals. A reference sample based method is suggested for studies that would end up comparing gene signal data from a combination of low and standard RNA templates; no such referencing appears to be necessary when comparing fold-changes of gene expression between standard and low template reactions.

Keywords: Data comparability, submicrogram RNA, low template amplification, reference sample, gene expression

1. Introduction

DNA-based microarrays [1] enable us to quantify the transcript levels of tens of thousands of genes simultaneously and is one of the most powerful approaches towards understanding genome function. Microarrays are rapidly emerging as a frontier technology for the identification of potential biomarkers by application to biological materials most relevant to the phenotypes under investigation. These include biopsy materials from fine needle aspirates (FNA), cell sub-populations, or enriched isolates from laser capture microdissection (LCM). However, materials obtained from LCM or FNA samples very rarely provide enough RNA that is required for standard microarray reactions (5–100 micrograms, depending on the technology). Newer methods, involving synthetic amplification of the starting

RNA, are therefore necessary to make such samples amenable to microarray analysis [2–7].

Consequently, an important question that needs to be addressed is whether or not the data generated from low amounts of RNA are comparable to data generated from a standard reaction. Two prior publications [4,5] have evaluated the reproducibility of fold-change estimates between low template and standard reactions. One study employed Affymetrix oligonucleotide arrays [5] and the other study employed cDNA based arrays and an exponential amplification strategy [4]. However, none of these studies addressed the issue of comparability of gene signals (as opposed to gene fold-changes) between low template and standard reactions. In practice, however, we are often interested in determining the expression pattern of a gene or genes across a wide variety of experiments (resident in public and proprietary databases) that may contain a mixture of low template and standard microarray hybridizations. The present report is aimed towards evaluating the comparability of gene signals as well as gene fold-change estimates. We

*Corresponding author: Dr. Sujoy Ghosh, GlaxoSmithKline, 5 Moore Drive, RTP, NC 27709, USA. E-mail: sg45653@gsk.com.

present a practical approach for making gene signals from low-template and standard reactions more comparable. We also provide evidence that fold-change estimates between low template and standard reactions are comparable across genes of varying signals and also comparable to results obtained on an independent gene expression platform (real time, quantitative polymerase chain reaction, reference 10).

2. Materials and methods

2.1. Sample preparation and processing

Experiments were carried out with total human RNA samples from heart, kidney, lung, spleen and skeletal muscle (Ambion). The standard microarray reactions (SR) utilized 8 μ g of total RNA whereas the low-template amplification reactions (LTA) employed 100 ng of total RNA. SR reactions per tissue were single reactions and the LTA were duplicates. The protocols for synthesis of cDNA and cRNA from SR or LTA were performed according to recommendations from Affymetrix [7–9]. For LTA reactions, we followed the Affymetrix protocol modified for small sample labeling [9]. The quality and integrity of RNA samples were assessed on a Bioanalyzer 2100 instrument. For both SR and LTA samples, 20 μ g of the final cRNA product was fragmented and 15 μ g of the fragmented cRNA was hybridized to Affymetrix human U95Av2 chips. The hybridized chips were scanned on a confocal GeneArray scanner (Hewlett Packard, Santa Clara, CA).

2.2. Real time quantitative PCR

Gene specific primers for real time quantitative PCR were designed in Primer3 (http://www-genome.wi.mit.edu/genome_software/other/primer3.html) using gene sequences around Affymetrix probesets as input (from www.affymetrix.com). Primers were obtained from IDT. Reagents for the PCR reaction and SYBR Green I dye were obtained from Molecular Probes (Eugene, OR).

Real time PCR was conducted in the ABI PRISMTM 7700 Sequence Detection System (Applied Biosystems, Foster City, CA). Each reaction contained 20 mM Tris-HCl (pH 8.3), 100 mM KCl, 7 mM MgCl₂, 0.4 mM each dNTP (dATP, dCTP, dGTP, TTP), 0.05 unit/ml Taq DNA Polymerase, SYBR Green I dye, 300 nM each of gene-specific primers, and cDNA from

heart and kidney samples at 237 ng per reaction. The PCR was initiated with a 10-minutes denaturation step at 95°C. Initial denaturation was followed by 40 cycles at 95°C for 15 seconds, 1-minute annealing at 60°C. Cycling was followed by a 4°C hold. Analyses of data were accomplished using the ABI PRISM 7700 Sequence Detection Software.

2.3. Data analysis

Gene expression data from LTA or SR were generated by the Microarray Analysis Suite 5.0 software (MAS 5.0) provided by Affymetrix. Signal intensities were obtained by scaling all hybridized chips to a target intensity of 150. Data consisting of signal intensities, call values (Present, Marginal or Absent, provided in MAS5.0) and detection p-values were exported from MAS 5.0 into two external applications, Microsoft Excel and JMP Statistical software (SAS Institute, Cary, NC) for downstream analysis.

3. Results

3.1. Comparison of gene expression values

In the following sections, results from human heart samples are primarily used as the basis for analysis. Very similar results were obtained with the other RNA samples from liver, lung, kidney and muscle. Typically the LTA reactions resulted in chips with higher background signals although the differences were not statistically significant. In the LTA, we also observed preferential hybridization to probes representing the 3'-ends of control genes (GAPDH and β -actin) compared to probes in the 5'-ends of these genes, indicating a progressive shortening of cRNA product with each round of linear amplification (data not shown). This is possibly caused by the limited processivity of the oligo-dT primed reverse transcription reaction which could result in incomplete cDNAs that are biased towards the 3'-ends of genes.

Comparison of the gene signals from the LTA sample to the SR displayed good overall correlation ($r = 0.88$, Fig. 1). Similar results have also been reported in the literature [5] and has been used to infer the reproducibility of oligonucleotide arrays using small samples. However, no further amplification beyond the standard Affymetrix protocol was employed in [5] and thus the results obtained are not strictly comparable with the results discussed in our study which employed

Table 1

Correlation of gene signals and fold-changes between low-template and standard microarray reactions on human heart samples. Gene signal ranges for each bin (based on signals from standard reaction heart sample) are shown in row A. The correlation between the gene signals obtained from the two methods are depicted as a function of different data treatment strategies (described in the text). The last row contains data on the correlation of fold-changes in gene expression (heart vs kidney samples) between low-template and standard reactions. Row descriptors are as follows: Untrans-Pearson, Pearson product moment correlation on untransformed data; Untrans-Spearman, rank-based Spearman's rho values on untransformed data; Log2 trans, Pearson correlation on logarithmically transformed data (to base 2); Square root trans, correlation on square root transformed gene signals; Absent removed, correlation on filtered data excluding genes called 'absent' in both LTA and SR; Absent & 2-fold removed, correlation on filtered data excluding genes called 'absent' in LTA and SR plus genes showing a 2-fold or greater change in signal between low template duplicates; Bin regression, correlation of gene signals adjusted by regression coefficient of each bin; Reference set, correlation between samples adjusted by use of a 'reference set'; LogRatio, correlation of log ratios (heart vs. kidney) between samples processed either as LTA or SR

Signal bin signal range	Bin1 0.2–9	Bin2 9–20	Bin3 20–35	Bin4 35–58	Bin5 58–93	Bin6 93–144	Bin7 144–222	Bin8 222–382	Bin9 382–787	Bin10 787–13870	TextID A
Untrans-Pearson	0.11	0.12	0.19	0.18	0.15	0.18	0.20	0.25	0.32	0.80	B
Untrans-Spearman	0.31	0.17	0.21	0.17	0.17	0.17	0.19	0.26	0.33	0.67	C
Log2 trans.	0.31	0.16	0.20	0.16	0.16	0.15	0.18	0.25	0.31	0.68	D
Square root trans.	0.25	0.15	0.21	0.17	0.17	0.17	0.20	0.26	0.33	0.77	E
Absent removed	0.35	0.33	0.34	0.23	0.18	0.18	0.27	0.22	0.27	0.77	F
Absent & 2-fold removed	0.28	0.26	0.26	0.25	0.12	0.25	0.25	0.11	0.18	0.74	G
Bin regression	0.15	0.12	0.18	0.13	0.14	0.18	0.21	0.22	0.30	0.79	H
Reference set	0.30	0.15	0.21	0.23	0.31	0.29	0.34	0.52	0.68	0.97	I
LogRatio (Heart/Kidney)	0.42	0.44	0.39	0.56	0.61	0.75	0.78	0.86	0.91	0.95	J

two rounds of target amplification. We then investigated whether the correlation of expression between LTA and SR was dependent on the magnitude of the gene signal. We first ranked the genes from the standard reaction (heart samples) based on their signals and then partitioned the ranked genes into 10 groups containing approximately equal number of genes. The first group (hereafter referred to as Bin 1) consists of genes with very low expression intensities (range 0.2–8.7) whereas the final bin (Bin 10) is enriched for highly expressed genes (range > 780). Intermediate bins contain genes with intermediate intensities (Table 1, row A). We then compared the correlation of expression intensities between the LTA and SR separately for each bin. Correlation was measured either on the expression intensities of the genes in each bin (Pearson product-moment correlation) or on the signal-based ranks of the genes in each bin (Spearman's rho). The results are shown in Table 1 (row B,C). From Table 1 we observed that the correlation between the two methods was in fact quite low ($r = 0.12 - 0.3$) for the low and medium intensity bins (Bins 1–9), and improved ($r = 0.78$) only for the highest intensity bin (Bin 10). This result indicates that for low to medium intensity genes, the gene expression signals obtained from the standard or the low template methods can differ considerably for a given gene. It is only with the high intensity genes (signal >780) that the two methods produce comparable values. When looking at the entire dataset, the correlation among the high-signal genes overshadow the poor correlation in the low and medium signal ranges. The lack of corre-

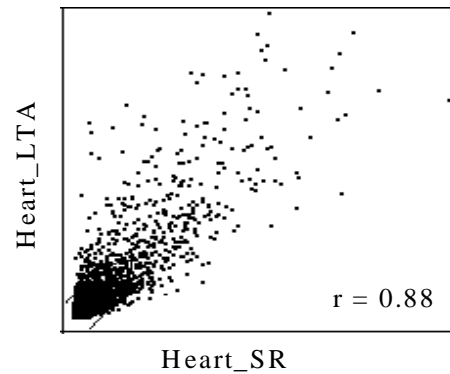


Fig. 1. Correlation analysis of gene signals from human heart samples processed through a low-template amplification reaction (LTA) or standard reaction (SR). Correlation is computed based on Pearson product-moments.

lation in the low and medium signal ranges might arise from two sources: (a), the higher instrumental noise (during scanning and recording of low-intensity pixels from the hybridized chip) for lowly expressed genes and (b), greater variability in the amplification reactions for the low to medium-expressed genes compared to highly expressed genes. Based on our analysis, we concluded that it would be erroneous to consider all gene signals between LTA and SR to be comparable.

We next evaluated several data processing methods to determine if the LTA gene signals could be made more similar to the SR gene signals. We explored three different data processing approaches: (a) data transformation, (b) data filtering and (c) data extrapolation.

4. Data transformation

The scaled expression data from the SR and LTA was transformed by either taking the log of the intensities (base 2) or their square roots. The log and square root transformations are often applied for variance stabilization and we reasoned that such a transformation might benefit the high variation observed between LTA and SR data. The transformed data were then plotted by bins to compare the SR and LTA (Table 1, row D,E). Neither transform improved the bin-wise correlation between the two methods when compared to the untransformed data.

5. Data filtering

Data filtering consisted of excluding genes whose expression signals cannot be measured with high confidence, before comparing LTA to SR data. We reasoned that such genes may be adding unnecessary noise in the analysis and removing them could improve the correlation. We filtered genes in two separate ways. In the first method, we eliminated all genes that were not assigned a call of “Present” in both the LTA and standard reaction by the Affymetrix software (MAS5.0). In the second method, we increased the stringency by additionally excluding genes that showed a greater than 2-fold difference in expression intensity between the replicates of the LTA reaction. After filtering, the remaining genes were subjected to the same binning strategy (with fewer genes per bin) and compared for expression between SR and LTA (Table 1, row F,G). However, we did not see any significant improvement between the SR and LTA correlation as a function of data filtering, suggesting that the lack of good correlation is not driven entirely by the so called ‘absent’ or noisy genes.

6. Data extrapolation

The data extrapolation strategy involved processing an unrelated sample set (hereafter called the ‘reference set’) through both SR and LTA reactions. The ‘reference set’ was created by averaging the gene signals from four tissues (kidney, lung, liver and muscle) separately for LTA or SR data obtained from these tissues. A gene-by-gene comparison was made between the SR and LTA for the reference set to obtain an ‘equalizing factor’ for every gene that would make its LTA and SR values identical in the reference set. The ‘equalizing

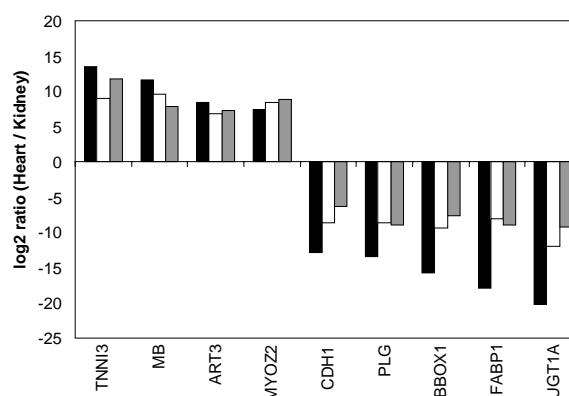


Fig. 2. Validation of results from LTA reaction by quantitative PCR. Nine genes, differentially expressed between heart and kidney samples in the LTA reaction, were selected for quantitative PCR. A comparison of fold-change results obtained with the microarray and quantitative PCR assays are shown (quantitative PCR, black bar; fold-change from SR, white bar; fold-change from LTA, hatched bar). Gene identifiers are as follows: TNNI3, troponin, cardiac; MB, myoglobin; ART3, ADP-ribosyltransferase 3; MYOZ2, myozenin 2; CDH1, cadherin 1; PLG, plasminogen; BBOX1, gamma-butyrobetaine hydroxylase 1; FABP1, fatty acid binding protein 1; UGT1A, UDP-glycosyltransferase 1.

factor’ for any gene was determined by dividing the gene’s average signal in the SR by its average signal in the LTA (for the reference sample). Any gene in the experimental LTA sample (heart tissue) was multiplied by its ‘equalizing factor’ (obtained from the reference set) to obtain an estimated expression value for the gene in a standard reaction. The estimated values were then compared to the experimentally obtained values for the standard reaction after binning the genes as described above (Table 1, row I). For the very low intensity bins (Bins 1–3), no improvement in the correlation is observed compared to other data processing methods. However, for Bin 4 and higher bins, the agreement between LTA and standard reaction is significantly improved compared to the bin-wise correlation observed in the other methods. Thus, data processing through the ‘reference set’ approach appears to generate more comparable data between the LTA and standard reaction compared to other methods (although the agreement is not perfect). We also tried replacing the gene-specific ‘equalizing factor’ with a bin-specific normalization factor represented by the slope of a regression line through the LTA and standard reaction data per bin (Table 1, row H). However, this approach did not improve the bin-wise correlation over the untransformed data.

7. Comparison of fold-change

We next compared fold-change estimates of gene expression between two samples processed either through LTA or the standard reaction. This is an issue of practical importance since in many instances it will not be feasible to compare gene signals between similar samples processed through the LTA and standard reactions or the question of interest will be around gene expression changes (fold-changes) and not gene signals. Where it is only possible to obtain data on fold-changes in gene expression from LTA samples, it is important to determine the validity of such fold-change estimates. We addressed this issue by comparing fold-change estimates of gene expression between two samples (heart and kidney) that are processed through both LTA and standard reaction. Genes were further binned by intensities as described above. We then compared the bin-wise correlation of fold-changes (expressed as log ratios to the base 2) between LTA and standard reaction samples. The observed results are in Table 1 (row J). In contrast to the results with gene signals, we observed significant correlation between the fold-change estimates between LTA and standard reaction samples with the exception of the very low intensity bins (Bins 1–3). The majority of the genes in the low intensity bins are also associated with an ‘absent’ call (from Affymetrix software) and are therefore, likely to be removed from the data analysis anyway. This result implies that despite significant discrepancies between LTA and SR signals, the reported fold-change values for genes are quite comparable. Similar reports on the comparability of fold-change estimates between small samples and standard reactions exist in the literature [4]; however the small-sample amplification protocol and the microarray platform used in the report are different from the one discussed here.

8. Validation of results by quantitative PCR

We next determined the agreement between the fold-change estimates obtained for genes in the LTA procedure (heart vs. kidney samples) to results obtained on an independent RNA-measurement platform (real-time PCR). This step is necessary for determining the validity of the LTA results. A total of 9 genes were selected for validation and included genes that are upregulated or downregulated in heart compared to kidney. As shown in Fig. 2, we observed considerable agreement between the LTA and RT-PCR methods in the directions

of fold-change observed, further validating the results from the LTA reactions. It is to be noted that for some of the 9 genes, the fold-change estimates obtained by RT-PCR is greater in magnitude than fold-changes reported from the microarray studies. This is attributable to the greater sensitivity of the RT-PCR reaction which is based on an exponential amplification of signal, compared to signal detection on the Affymetrix chips which rely on a linear amplification of the sample RNA. These differences are further exacerbated for low abundance transcripts.

9. Conclusion

We focus on a key issue of data comparability when applying microarrays to limiting biological samples for the identification of biomarkers. We used correlation analysis to evaluate the agreement between the low template and standard reactions either on gene signals (within one sample) or fold-change of gene expression (between two samples). Our findings indicate that the correlation between gene signals is highly dependent on the magnitude of gene expression; genes with low or medium expression do not correlate well between the low template and standard methods. Such discrepancies can arise from multiple experimental factors, including the RNA amplification steps and the ability of the scanners to reliably detect low expression levels. Common data processing techniques, such as log transforms or data filtering failed to improve the correlation between the two methods significantly. However, including a non-experimental reference sample in the experimental design did significantly improved the level of correlation between the low template and standard reactions. When gene fold-changes are compared between two samples, the two methods provide results that are in good agreement with each other (an unfavorable bias towards low expressors, although not entirely removed, is much less) and does not necessitate further treatment of the data. Finally, the fold-change estimates obtained from an LTA study can be reproduced on an independent RNA-measurement platform.

Based on our findings we propose the following. If there is a need to compare gene signals from low template samples to gene signals from standard microarray reactions, the method of data extrapolation through the use of a reference sample is advised. The reference sample can be constructed in-silico. In our experience, more reliable estimates of gene expression are obtained if the reference samples are physiologically not too dis-

tinct from the experimental samples (e.g. a cell-line based reference sample is not optimal for an experiment involving tissues). However, if the only comparisons of interest are fold-changes of gene expression, then the results obtained from the low template samples can be accepted as comparable to standard reaction results without further manipulation of the data (excepting for the very low expressed genes). In both cases, better agreement is to be expected for genes with higher levels of expression.

Acknowledgments

We thank Ms. Elizabeth Keiper, Dr. Chris Clayton and Dr. Hongtao Fan for critical review and suggestions towards the manuscript. We thank Mr. Laurie Scott, Mr. K.B. Tan, and Ms. Yifeng Shen for their help with the small sample amplification protocols. We also thank Dr. Karin Au and Dr. Mary Brawner for their continued encouragement and support of this work.

References

- [1] P.T. Spellman, G. Sherlock, M.Q. Zhang, V.R. Iyer, K. Anders, M.B. Eisen, P.O. Brown, D. Botstein and B. Futcher, Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization, *Mol Biol Cell* **9** (1998), 3273–3297.
- [2] M. Mahadevappa and J.A. Warrington, A high-density probe array sample preparation method using 10- to 100-fold fewer cells, *Nat Biotechnol* **17** (1999), 1134–1136.
- [3] H. Ohyama, X. Zhang, Y. Kohno, I. Alevizos, M. Posner, D.T. Wong and R. Todd, Laser capture microdissection-generated target sample for high-density oligonucleotide array hybridization, *Biotechniques* **29** (2000), 530–536.
- [4] N.N. Iscove, M. Barbara, M. Gu, M. Gibson, C. Modi and N. Winegarden, Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA, *Nat Biotechnol* **20** (2002), 940–943.
- [5] J.N. McClintick, R.E. Jerome, C.R. Nicholson, D.W. Crabb and H.J. Edenberg, Reproducibility of oligonucleotide arrays using small samples, *BMC Genomics* **4** (2003), 4–49.
- [6] R.J. Lipshutz, S.P. Fodor, T.R. Gingeras and D.J. Lockhart, High density synthetic oligonucleotide arrays, *Nat Genet Suppl* **1** (1999), 20–24.
- [7] D.J. Lockhart, H. Dong, M.C. Byrne, M.T. Follettie, M.V. Gallo, M.S. Chee, M. Mittmann, C. Wang, M. Kobayashi, H. Horton and E.L. Brown, Expression monitoring by hybridization to high-density oligonucleotide arrays, *Nat Biotechnol* **14** (1996), 1675–1680.
- [8] Affymetrix. GeneChip? Expression Analysis Technical Manual. Santa Clara, CA, 2001.
- [9] Affymetrix. GeneChip Eukaryotic Small Sample Target Labeling Assay Version II Technical Note. Santa Clara, CA, 2003.
- [10] K.B. Mullis and F.A. Faloona, Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction, *Methods Enzymol* **155** (1987), 335–350.

