TECHNICAL REPORT

Tissue Mixing as a Differential Expression Model for the WT-Ovation® Pico RNA Amplification System

Introduction

Gene expression profiling has become a powerful approach to address a variety of biologically and therapeutically important questions and is currently transitioning into clinical settings. In these studies, the validation of gene expression signatures as accurate predictors of clinical phenotypes has been hindered by the inability to generate high-quality gene expression results due to limitations in the quality or quantity of RNA obtained from these sources. The need to address purified cell fractions will further increase the need for amplification technologies that require less starting material and yet accurately reflect the initial transcript composition of the sample. These studies utilize a model system to demonstrate that gene expression patterns are faithfully reproduced even when starting with picogram amounts of RNA.

NuGEN's amplification products have been used successfully with minute amounts of total RNA in order to generate sensitive and accurate differential gene expression results. NuGEN's WT-Ovation Pico RNA Amplification System employs whole transcriptome (WT) amplification starting from as little as 500 pg of total RNA, yielding 6–10 µg of cDNA in 5 hours. The amplified cDNA can be analyzed using standard qPCR techniques or fragmented and labeled with the Encore Biotin Module (Part No. 4200) and analyzed on standard Affymetrix GeneChip® arrays. Note: the Ovation Pico WTA System V2 (Part No. 3302) has replaced the WT-Ovation Pico RNA Amplification System.

In order to evaluate the performance of this system with an extremely limiting amount of input total RNA, an artificial model was designed to generate a set of samples containing a pool of transcripts which vary in expression in predictable amounts. A tissue mixing model was chosen in which total RNA generated from placenta was mixed in various ratios with total RNA generated from spleen.

This model enables a researcher to assess general reproducibility of the resulting gene expression data, the sensitivity of detecting rare transcripts, and the linearity of the expression values relative to the dilutions used. Such tissue mixing approaches have been often used to compare results generated from different labs and different gene expression platforms (see Microarray Quality Control [MAQC] Consortium study, *Nature Biotechnology* 24(9) Sept 2006).

This approach has the advantage of generating many transcripts that change in expression according to the tissue ratios employed.

The limitation to this approach includes the inability to determine the lower limit of detection of transcripts in terms of absolute copy number or concentration. In order to, in part, address the question of sensitivity in this experiment, a mixture of 10% placenta, 90% spleen total RNA mixture was used to determine the number of placental specific transcripts that can still be detected when highly diluted. While not allowing the determination of sensitivity in terms of copy number, this addresses the ability of a gene expression approach to detect transcripts with low signal values (in the 100% parent tissue) further diluted 1:10. This experiment demonstrates that high quality data can be collected from samples yielding as little as 500 pg of total RNA using the WT-Ovation Pico RNA Amplification System. Reproducibility and sensitivity in detecting presumably rare transcripts, as well as linearity of the expression results are all reviewed here.

Materials and Methods

Placenta and spleen RNA was purchased (Ambion, Cat.# 7950 and Cat.# 7970 respectively). The RNA ratios chosen for this study were 100% placenta, 90% placenta/10% spleen, 50% placenta/50% spleen, 10% placenta/90% spleen. The quantity of total RNA was assessed using the Nanodrop ND-1000 spectrophotometer (Wilmington, DE). To assess quality, an Agilent Bioanalyzer was used to obtain traces for the amplified cDNA with an RNA 6000 Nano LabChip® (Agilent Cat. #5065-4476). The Eukaryotic Total RNA Nano program (Nano assay in the Expert 2100 software) was used for analysis, according to the manufacturer's instructions.

Triplicate amplifications of the mixed RNA samples were performed using 500 pg of total RNA (15 amplifications total) according to the standard WT-Ovation Pico RNA Amplification System user guide protocol (note: the Ovation Pico WTA System V2, Part No. 3302, has replaced the WT-Ovation Pico RNA Amplification System). Five micrograms of the resultant, amplified and purified cDNA was fragmented and labeled according to

NuGEN
imagine more from less®

the standard Encore Biotin Module protocol.

Each of these samples were then hybridized to Affymetrix HG-U133A 2.0 arrays (Cat.# 900469) at a final target concentration of 22.7 ng/µL and treated according to manufacturer's recommendations and the Encore Biotin Module user guide. Microarray data was probe normalized using RMA in all cases except those stated. All analysis was performed in MAS 5.0 (Affymetrix) or Bioconductor.

Results and Discussion

All 15 arrays were assessed for standard quality control metrics (image quality, signal distribution, pair wise scatter plots) with no arrays eliminated due to quality issues. The general MAS 5.0 array metrics of the data set are shown in **Table 1**.

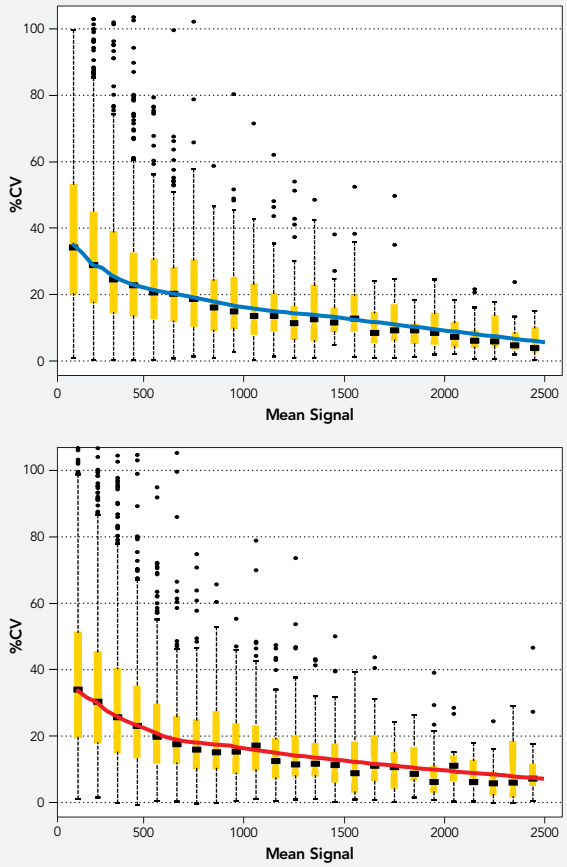
This data is used to assess the general sensitivity as well as the mean and standard deviations of the percent of probe sets called 'present' for each of the sample dilutions (amplifications were done in triplicate). The sensitivity of the system is measured by %P using the 3' biased Affymetrix HG-U133A 2.0 arrays, the analysis was performed using MAS 5.0 with target intensity (TGT) of 250. With only 500 pg of input total RNA per reaction, the WT amplification resulted in 55–59% present calls, demonstrating the ability to detect a large number of transcripts across a wide range of abundance, including rare transcripts.

In order to assess the reproducibility of the RMA normalized signals across multiple arrays, the coefficient of variances was calculated for all probe sets across the triplicate 100% placenta and 100% spleen arrays. These CVs are plotted against mean response and the distribution of the coefficient of variances as a function of mean signal level is shown in **Figure 1**.

TABLE 1. **MAS 5.0 array metrics of triplicate amplifications used in this study.**

Tissue Dilution	Scale Factor (SF)	SD of SF	Mean of %P	SD of %P
100% Spleen	0.65	0.004	59	0.5
10% Placenta	0.64	0.010	59	0.7
50% Placenta	0.66	0.002	59	0.2
90% Placenta	0.66	0.006	56	0.1
100% Placenta	0.68	0.005	55	0.2

FIGURE 1. **Distribution of the coefficient of variances as a function of mean signal level.**



%CV is determined for the 3 replicates at the 100% placenta and 100% spleen tissue levels. The solid line is a loess smoothed fit to the individual transcript CV's and the boxes are calculated by binning the individual transcript CV's by mean signal level at intervals of 100 (101–200, 201–300, etc.). The notches indicate the 95% confidence interval of the bin's median value and the whiskers are drawn at 1.5 times the inter-quartile range. Only signal values below 2500 are shown on these plots as CVs for signals above 2500 continue to steadily decrease (data not shown).

The reproducibility of the signal values generated across triplicate amplifications is directly related to the ability to detect small changes in gene expression between samples. The variances of signal values across triplicate amplifications are around 20% for rare transcripts (signal ~ 500) and fall steadily as probe set signals increase with a mean CV of ~10%, for probe sets set with signals greater than 1000. To demonstrate the ability to detect changes in gene expression given the CV values shown in **Figure 1**, a representative change in gene expression that would be detectable in this data set at a signal level of 500 was calculated. In order to perform this analysis, a standard deviation of 100 was used. The detectable difference in means was determined from the critical t-value at a 95% confidence level. The result of this analysis is shown in **Table 2**.

The above analysis indicates that in the case of an 'average' transcript with a signal of 500, if triplicates are run, then approximately 45% (227/500 x 100%) changes in signal are detectable. When 4, 5 and 6 replicates are run, the corresponding detectable changes are 35%, 29% and 26% respectively. It bears repeating that this data was performed with triplicate amplifications from a common RNA sample, with 500 pg input total RNA into the amplifications. This calculation only includes the technical variability associated with performing replicate amplifications and microarray hybridization and analysis. It does not include the major source of variability which is the RNA source.

Given that the tissue specific probe sets were identified using only the 100% placenta and 100% spleen samples, the ability to detect rare transcripts can be explored by assessing the detection of these transcripts in the total RNA mixtures. **Table 3** shows the results of this analysis. The limit of detection for a transcript is the lowest tissue dilution at which a probe set is determined as statisti-

TABLE 2. **Ability to detect changes in gene expression abundance.**

Number of Replicate Amplifications	Detectable Difference in Mean Signals (Probe Set Signals = 500)	Detectable % Change in Signal Means
3	227	45%
4	173	35%
5	146	29%
6	129	26%

TABLE 3. **Limit of detection of tissue-specific probe.**

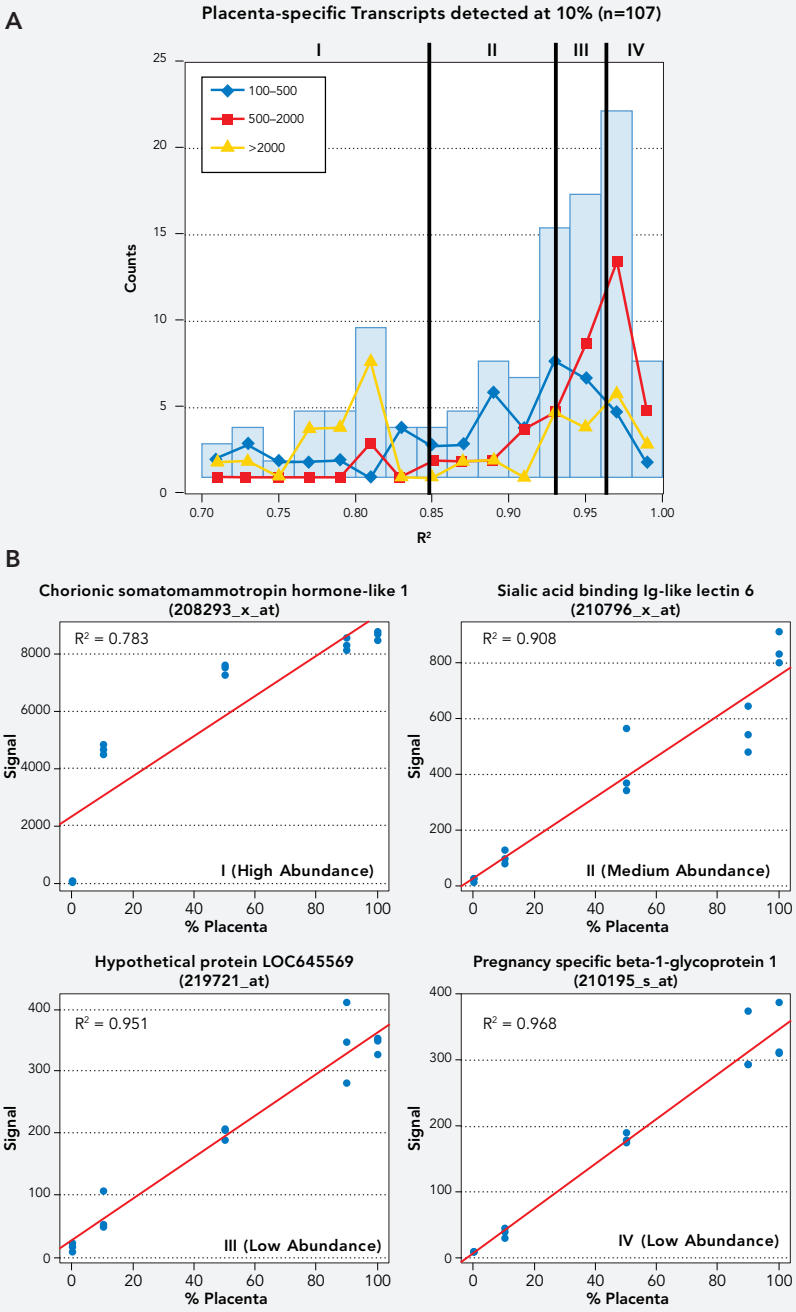
Percent of Placenta-Specific Probe Sets Detected at Each Dilution				
Signal Level	**100–500**	**500–2000**	**>2000**	**Total**
Dilution	n=124	n=47	n=32	n=203
10%	32%	79%	97%	n=107
50%	75%	98%	100%	n=64
90%	89%	100%	100%	n=19

cally above its 0% control group by one-sided t-test at 95% confidence. All subsequent tissue dilutions must also be statistically above the 0% control group. For example, a transcript detected at the 50% dilution level, must be statistically significantly different from its 0% control group at the 50% dilution, 90% dilution, and 100% dilution. Note that all tissue-specific transcripts will be detected at the 100% dilution by definition of 'tissue specific'. This analysis highlights the ability to detect presumably quite rare transcripts even when beginning with only 500 pg input RNA. For example, almost one third (39) of the rare probe sets were detectable in the 10% placenta sample. This indicates 39 of the 124 tissue specific probe sets with mean signal values between 100 and 500 in the 100% placenta sample were detected even when 'diluted' 1:10 in spleen total RNA.

In order to compare signal values of the tissue specific probe sets with the expected signals based on the known total RNA dilutions, an analysis of signal linearity was performed. The signals of the 107 placenta-specific probe sets that were detected at the 10% placenta dilution (and all higher mixtures) were normalized by the mean signal of each probe set at the 100% placenta dilution. The Pearson's correlation coefficient for each of the 107 transcripts was calculated, as was the slope of the best-fit line (data not shown).

A frequency analysis on the R^2 values of all of the placental specific transcripts detected in the 10% placenta dilution. These are shown in **Figure 2**. Panel A shows the frequency distribution of all of the 107 placental specific transcripts assessed (blue histograms) as well as the rare, intermediate and high abundance transcript groups independently (lines). The R^2 values on the x-axis are binned in 0.02 increments. Four transcripts were identified, one from each quartile of the R^2 frequency distribution, and plotted in **Figure 2**, panel

FIGURE 2. **Individual transcript linearity assessment.**



Panel A: Frequency distribution of R^2 for all 107 placental specific transcripts assessed (blue histograms). Lines indicate frequency distributions for the 107 transcripts categorized by their mean signal level in 100% placenta replicates. Vertical lines indicate the quartile divisions of the distribution.

Panel B: Individual linearity plots for representative transcripts from each quartile of the R^2 distribution. Examples were chosen to demonstrate representative behavior for each abundance level by choosing transcripts with R^2 closest to the median R^2 value from each quartile. The abundance level is indicated above each panel, signal values between 100 and 500, 500 and 2000, and above 2000 in the 100% placenta samples for low, intermediate and high abundance groups respectively.

B. The overall analysis indicates that the number of transcripts with R^2 values greater than 0.95 is 40, with the majority of transcripts falling above 0.90 (68 of 107). The low abundance transcripts (those with 100% placenta signals between 100 and 500 n=39) again show generally very high R^2 values, and the majority of these rare transcripts have linearities above 0.90 (22 of 39). The transcripts of intermediate signals in the 100% placenta samples (500–2000) show even better linearity with the majority demonstrating linearity between 0.95 and 1.0 (33 of 37). The more abundant transcripts (signal values above 2000 in 100% placenta samples) perform a bit worse than the rare and intermediate transcript groups, likely due to saturation of the signals for some transcripts (upper left plot **Figure 2**, panel B). Overall, the vast majority of signals show linear signal increases proportional to the amount of transcript in the sample. This is true even for rare transcripts despite starting with as little as 500 pg of RNA.

Conclusions

This study was designed to assess the capability of using minute amounts of total RNA (such as those typically obtained from clinical tissue sources and experimental models) for the purposes of differential gene expression analysis.

A tissue mixing model was employed in order to generate a large pool of transcripts that differ in abundance in known ratios. Suffiicient quantities of cDNA were generated from 500 pg of input total RNA using NuGEN's WT-Ovation Pico RNA Amplification System. The amplified product was fragmented and labeled for microarray analysis on HG U133A 2.0 GeneChip arrays. The data demonstrate good sensitivity in detecting rare placental transcripts even at the 10% placenta/90% spleen total RNA mixture. Excellent reproducibility is also observed, with an overall average of less than 10% CVs across triplicate amplifications, as is very good linearity of tissue-specific transcript signal relative to known total RNA dilution. This data clearly demonstrates that high-quality gene expression results can be achieved with as little as 500 pg of total RNA.

NuGEN imagine more from less®

NuGEN Technologies, Inc.

Headquarters USA

201 Industrial Road, Suite 310
San Carlos, CA 94070 USA
Toll Free Tel: 888.654.6544
Toll Free Fax: 888.296.6544
custserv@nugen.com
techserv@nugen.com

Europe

P.O. Box 109
9350 AC Leek
The Netherlands
Tel: +31-13-5780215
Fax: +31-13-5780216
europe@nugen.com

For our international distributors contact information, visit our website

www.nugen.com

M01036 v2