TECHNICAL REPORT #2

Tissue Mixing as a Differential Expression Model for Ovation® RNA Amplification System V2

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 start with less 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 as little as 5 ng total RNA.

NuGEN's amplification products have been employed successfully for amplification of minute amounts of total RNA in order to generate sensitive and accurate differential gene expression results. NuGEN's Ovation® RNA Amplification System V2 (Part No. 3100), allows amplification of as little as 5 ng of total RNA, yielding approximately 6 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, formerly FL-Ovation® cDNA Biotin Module V2) and analyzed on standard Affymetrix GeneChip® arrays. 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 alter in expression in predictable amounts. We chose a tissue mixing model 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 known dilutions employed. Such tissue mixing approaches have been often used to compare results generated from different labs and different gene expressions 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, we used a 10% placenta, 90% spleen total RNA mixture allowing us to understand 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 does address the ability of a gene expression approach to detect transcripts with low signal values (in the 100% parent tissue) further diluted 1:10. In this design, we demonstrate that high quality data can be collected from samples yielding as little as 5 ng of total RNA using the Ovation RNA Amplification System V2. 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 were 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). In order to assess the quality of amplified cDNA, an Agilent Bioanalyzer was used to obtain traces for the amplified cDNA with an RNA 6000 Nano LabChip® (Agilent Cat. #5065-4476), and the Eukaryotic Total RNA Nano program (Nano assay in the Expert 2100 software) were used, according to the manufacturer's instructions.

Triplicate amplifications of the mixed RNA samples were performed using 5 ng of total RNA (15 amplifications total) according to the standard Ovation RNA Amplification System V2 user guide protocol. 3.75 µg of the resultant, amplified and purified cDNA was fragmented and labeled according to the standard Encore Biotin Module protocol.

Each of these samples was then hybridized to Affymetrix GeneChip® HG-U133A 2.0 Arrays (Cat. #900469) at a final target concentration of 17 ng/µL and treated according to manufacturer's recommendations and

NuGEN®
A Tecan Group company

the Encore Biotin Module user guide. Microarray data were probe normalized using RMA in all cases except those stated. All analyses were performed in MAS 5.0 (Affymetrix) or Bioconductor.

Real-time qPCR assays were designed within 1.5 kb of the 3′ end of the transcripts for 18 placenta-specific transcripts using the Universal ProbeLibrary™ and primer design software. Primers were ordered from Integrated DNA Technologies (Coralville, IA). Assays were screened for good efficiency as close to 100% as possible with a slope of 1 + 0.1. **Table 4** contains the list of genes and accession numbers; further primer and probe information are provided upon request.

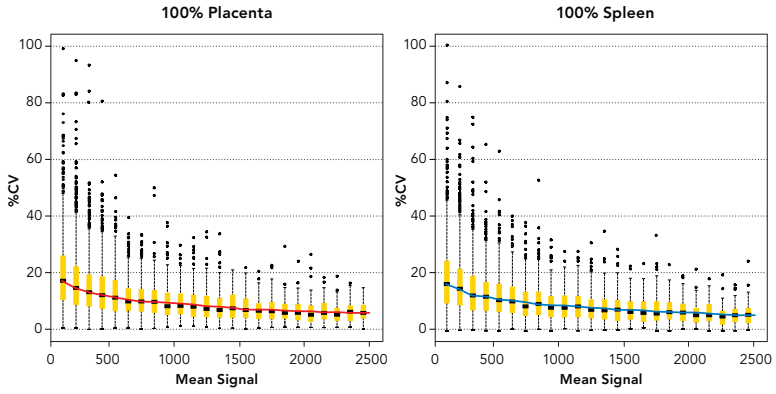
The cDNA from the triplicate amplifications used for the arrays was diluted 1/20 in 1X TE (prior to fragmentation and labeling) for qPCR analysis using TaqMan® Fast Universal PCR master mix with 5 nM each of the forward and reverse primers and 100 nM ProbeLibrary probe or with ABI's Assays-on-Demand™ primer and probe mix following vendor's instructions. The assays were analyzed using the ABI 7500 with the Fast block installation using the default settings.

The housekeeping gene RPL35 was used as a normalizer for the qPCR to calculate ΔCt values (ΔCt = Ct test gene – Ct RPL35). The fold changes between cDNA samples were calculated from the ΔΔCt values ($\Delta\Delta Ct = \Delta Ct_{(X\% \text{ placenta})} - \Delta Ct_{(Y\% \text{ placenta})}$, fold change = $22_{(-\text{ delta delta Ct})}$).

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

Tissue Dilution	Mean of Scale Factor	SD of Scale Factor	Mean of %P	SD of %P
100% Spleen	0.77	0.005	72.0%	0.4%
10% Placenta	0.78	0.010	73.8%	0.3%
50% Placenta	0.78	0.010	74.1%	0.4%
90% Placenta	0.78	0.004	72.2%	0.1%
100% Placenta	0.78	0.010	70.9%	0.4%

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 interquartile range. Only signal values below 2500 are shown on these plots as CVs for signals above 2500 continue to steadily decrease (data not shown).

Results and Discussions

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**.

These data are 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 GeneChip HG-U133A 2.0 Arrays, the analysis was performed using MAS 5.0 with target intensity (TGT) of 250. With only 5 ng of input total RNA per reaction, the amplification resulted in 71–74% 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 were 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**.

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 12% for rare transcripts (signal ~500) and falls steadily as probe set signals increase with a mean CV of ~7%, 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**, we calculated a representative change in gene expression that would be detectable in this data set at a signal level of 500. In order to perform this analysis, we assumed a standard deviation of 100. 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**.

TABLE 2. **Ability to detect changes in gene expression abundance using 'average' probe set with signal of 500.**

Number of Replicate Chips per Group	Detectable Difference of Means (from 500)	Percent change Detectable as Different (from 500)
3	136	27%
4	104	21%
5	88	18%
6	77	15%

The above analysis indicates that in the case of an 'average' transcript with a signal of 500, if triplicates are run then 27% (136/500 x 100%) changes in signal are detectable. When 4, 5 and 6 replicates are run, the corresponding detectable changes are 21%, 18% and 15%, respectively. It bears repeating that this data was performed with triplicate amplifications from a common RNA sample, with 5 ng 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 associated with replicate RNA source.

TABLE 3. **Limit of detection of tissue-specific probe sets. (% of Placenta-specific probe sets detected at each dilution)**

Signal level	100–500	500–2000	>2000	Total
Dilution	n=174	n=84	n=38	n=296
10%	42%	98%	100%	n=193
50%	91%	100%	100%	n=87
90%	97%	100%	100%	n=10

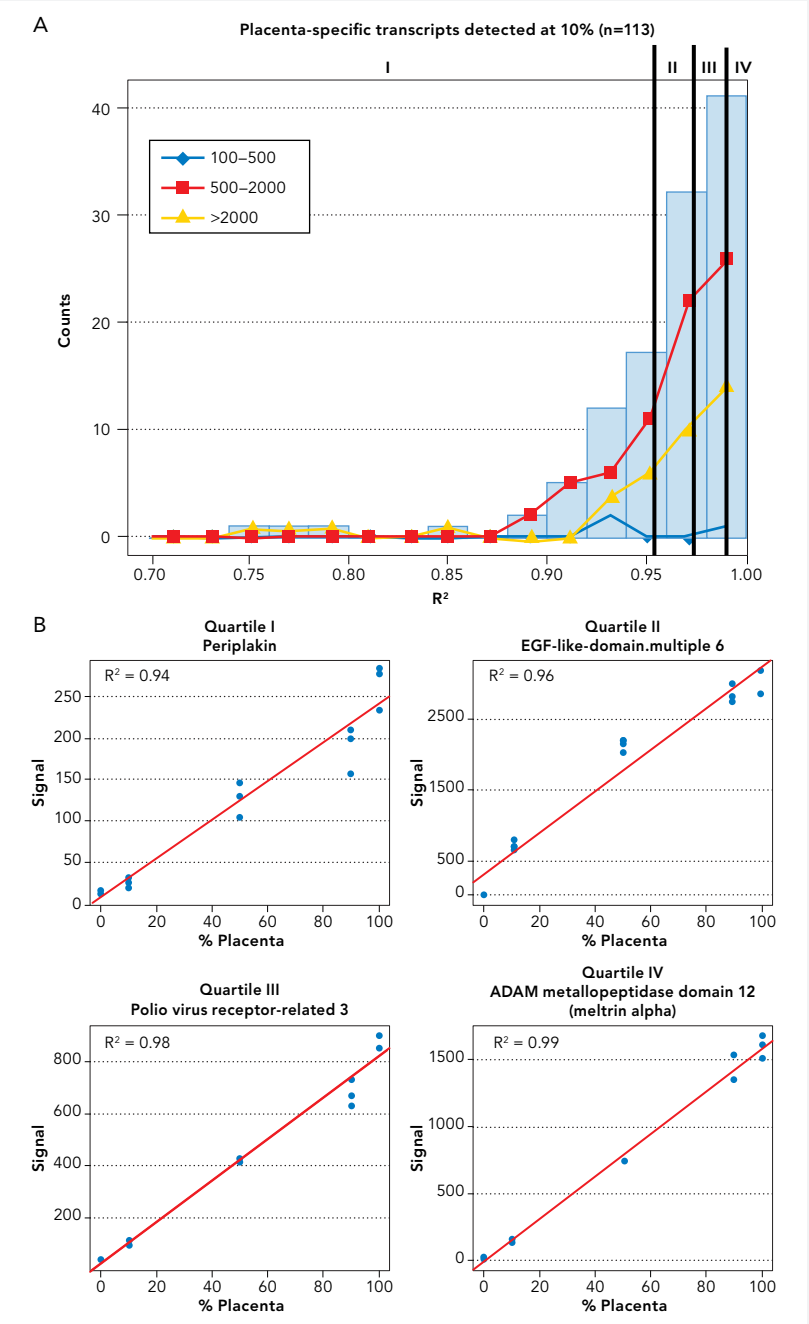
Given that the tissue specific probe sets were identified only using the 100% placenta and 100% spleen samples, the ability to detect rare transcripts can be explored by assessing the detection of these transcripts in the total RNA mixtures. **Table 3** shows the results of this analysis. The limit of detection for a transcript is the lowest tissue dilution at which a probe set is determined as statistically 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 5 ng input RNA. For example, 42% (73 of 174) of the rare probe sets were detectable in the 10% placenta sample. This indicates 73 of the 174 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, we performed an analysis of signal linearity. The signals of the 193 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 193 transcripts was calculated, as were the slopes 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 (red 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. We identified 4 transcripts, 1 from each quartile of the R^2 frequency distribution, and plotted them in **Figure 2**, panel B. The overall analysis indicates that the number of transcripts with R^2 values greater than 0.95 is 113, with the majority of transcripts falling above 0.96. The low abundance transcripts (those with 100% placenta signals between 100 and 500 n=73) again show generally very high R^2 values, and the majority of these rare transcripts have linearities above 0.94 (37 of 73). The transcripts of intermediate signals in the 100% placenta samples (500–2000) show even better linearity with the majority demonstrating linearity between 0.98 and 1.0 (43 of 82). 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 propor-

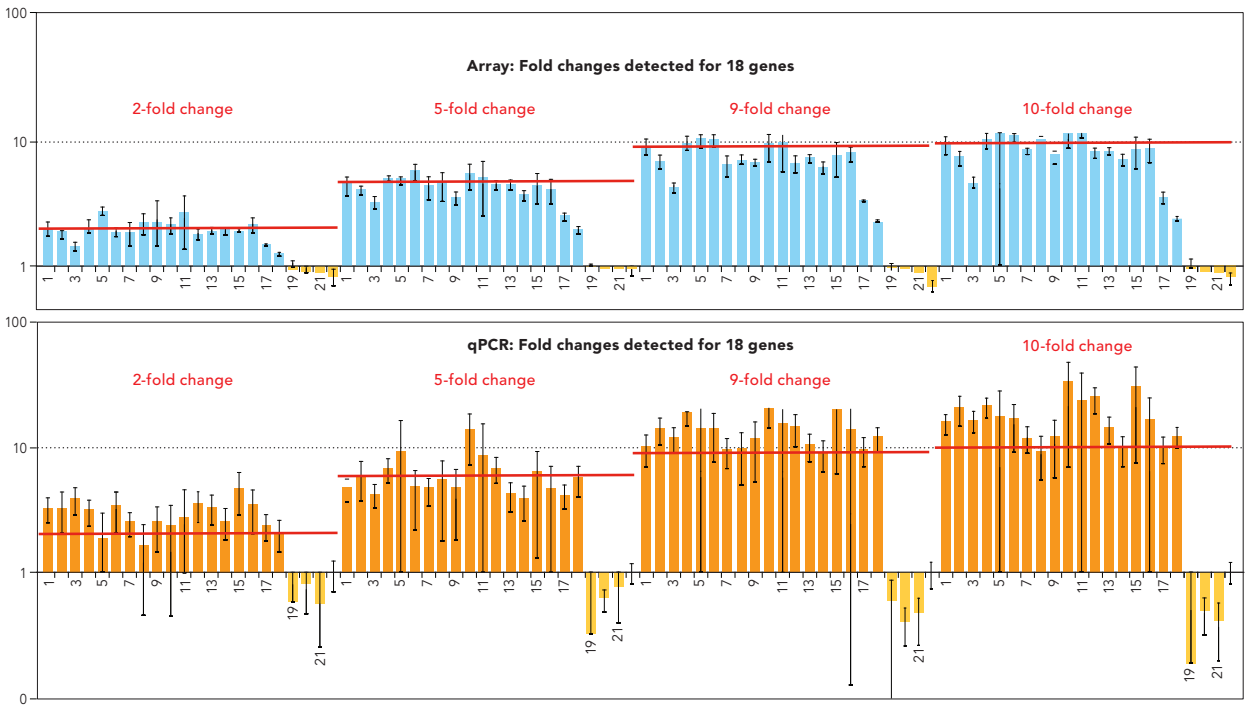
FIGURE 2. **Individual transcript linearity assessment.**



Panel A: Frequency distribution of R2 for all 107 placental specific transcripts assessed (red histograms). Lines indicate frequency distributions for the 107 transcripts categorized by their mean signal level in 100% placenta replicates. Dashed 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.

FIGURE 3. **Both array and qPCR data show fold changes consistent with the expected levels.**



The fold change for a panel of 18 placenta-specific genes is plotted in a log scale for both arrays (blue histograms) and validating qPCR assays (orange histograms). The error bars represent the confidence interval for one standard deviation. The observed fold changes agree with the expected fold change for the mixtures' comparisons (shown as a red bar): 2-fold change between 100% and 50%, 5-fold change between 50% and 10%, 9-fold change between 90% and 10%, and 10-fold change between 100% and 10%. The right-most 4 genes in the graph are housekeeping genes, and the very right-most gene, RPL35-1 was used as a normalization gene.

tional to the amount of transcript in the sample. This is true even for rare transcripts despite starting with very small amounts of RNA (5 ng).

To investigate the accuracy of fold change values resulting from the amplification and detection of placenta specific transcripts using this model, the fold change values of a subset of placenta specific transcripts were calculated from the microarray analysis and validated using qPCR. 18 placenta-specific genes were chosen from the array results covering a range of signal levels to test using real-time qPCR. Four additional genes were chosen as housekeeping genes. RMA-normalized signals from the triplicate arrays were log2 transformed to calculate the log2 ratios of differential gene expression. The fold changes were calculated as 2 log2 ratio and are shown in **Figure 3**. Both the microarray and qPCR fold change results strongly agree with the expected fold change values based on the known total RNA mixtures.

Conclusions

This study was designed to assess the capability of using minute amounts of total RNA (such as those typically available from some clinical tissue sources and experimental models) for the purposes of differential gene expression analysis. A tissue-mixing model system was employed in order to generate a large pool of transcripts that alter in abundance in known ratios. Using NuGEN's new Ovation RNA Amplification System V2, sufficient quantities of cDNA were generated from 5 ng of input total RNA to fragment and label the amplified product and perform microarray analysis on Affymetrix GeneChip HG-U133A 2.0 Arrays. The data demonstrate good sensitivity in detecting rare placental transcripts even at the 10% placenta/90% spleen total RNA mixture, excellent reproducibility with an overall average of less than 10% CVs across triplicate amplifications, and very good linearity of tissue specific transcript signal relative to known total RNA dilution. These data clearly demonstrate that high quality gene expression results can be achieved even with 5 ng total RNA.

Originally published in December 2006.

TABLE 4. **Placenta-specific genes tested with qPCR.**

Transcript	Accession #
KRT19-1	NM_002276
HSD3B1-3	NM_000862
CGA-1	NM_000735
ALPP-1	NM_001632
HSD17B2-2	NM_002153
CYP2J2	NM_000775
EPS8L1-4	NM_017729
IGFBP1-3	NM_001013029
FN1-2	NM_054034
PSG11-2	M94890
LRP2-2	NM_004525
ISL1-1	NM_002202
GCM1-2	NM_003643
FLJ13546-1	AK023608
PP3227-1	AF193053
TAC3-2	NM_013251
GPC3-2	NM_004484
ADAM12-F3	NM_021641
PAGE4-1	NM_007003
M4-2	AF061832.1
DAD1-1	NM_001344.1
RPL35-1	NM_007209.1
SEPT2-1	NM_004404.1

NuGEN®
A Tecan Group company

Tecan Genomics, Inc.

Headquarters USA
900 Chesapeake Drive
Redwood City, CA 94063 USA
Toll Free Tel: 888.654.6544
Toll Free Fax: 888.296.6544
cservice-gn@tecan.com
techserv-gn@tecan.com

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

Worldwide
For our international distributors contact information, visit our website
www.nugen.com