Technical Note

Affymetrix®
Revolutionize life

Performance of Affymetrix® Human, Mouse, and Rat Gene 1.1 ST Array Strips for the Gene Atlas® System

Introduction

Affymetrix® Human, Mouse, and Rat Gene 1.1 ST Array Strips, in conjunction with the GeneAtlas® System, enable whole-transcript expression analysis of these well-annotated genomes in an easy-to-use format.

Gene 1.1 ST Array Strips contain four arrays that are processed in parallel using the GeneAtlas System, which can process up to two strips per day or eight strips per week. The system includes intuitive GeneAtlas® Instrument Control Software and one-year licenses* for additional statistical and pathway analysis packages to simplify data analysis. In addition, using Affymetrix® Expression Console™ Software enables you to evaluate data at the transcript or exon level, allowing alternative splicing detection.

This technical note demonstrates that whole-transcript expression arrays on the GeneAtlas System achieve high levels of reproducibility, sensitivity, and specificity.

Overview

Target preparation and array strip processing

Fragmented, labeled, single-stranded DNA target was prepared from commercially available total RNA using the Ambion® WT Expression Kit and the Affymetrix® GeneChip® Terminal Labeling and Controls Kit according to the *GeneAtlas WT Expression Kit User Manual*. Pooled target was hybridized to Human, Mouse, or Rat Gene 1.1 ST Array Strips and the strips were processed on the GeneAtlas System according to the *GeneAtlas System User's Guide*. CEL files were sketch-quantile normalized in Expression Console Software using the Robust Multi-Chip Analysis (RMA) algorithm for probe set signal summarization.

Analysis

Prior to calculating the coefficient of variation (CV), the RMA signal was linearized and the CV was calculated for each probe set. The median CV was calculated from all probe sets for a given sample and the Pearson product moment correlation coefficient (R) was calculated from comparisons of median RMA signal.

Results

Reproducibility

A number of commercially available human, mouse, and rat total RNAs were used as starting material to test reproducibility. Four human tissues with three biological replicates per tissue were used, as well as total RNA from the HeLa cell line. Three tissues with two biological replicates per tissue were used for both mouse and rat. A pooled hybridization sample was used to assess reproducibility within and between array strips. Table 1 shows the median CV for Human, Mouse, and Rat Gene 1.1 ST Array Strips. We observed a median CV equal to or less than 8.3 percent for all tissues and array types.

Table 1: Median CV for human, mouse, and rat tissues. Pooled, labeled target from the samples shown was hybridized in quadruplicate to Human, Mouse, or Rat Gene 1.1 ST Array Strips. The RMA signal was converted from log to linear scale and the CV for each probe set was calculated from four replicates (N = 4). The median CV was calculated from all probe sets. Each sample represents a different biological replicate of the same tissue type.

	Sample 1	Sample 2	Sample 3
Human Gene 1.1 ST Array Strip			
Brain	6.1%	5.6%	8.3%
Heart	5.8%	5.8%	7.5%
Prostate	7.0%	5.6%	8.2%
Testes	4.6%	5.3%	8.0%
HeLa	6.5%	N/A	N/A
Mouse Gene 1.1 ST Array Strip			
Brain	6.4%	6.6%	N/A
Heart	6.2%	6.7%	N/A
Liver	6.7%	6.5%	N/A
Rat Gene 1.1 ST Array Strip			
Brain	6.2%	6.2%	N/A
Heart	8.0%	6.7%	N/A
Liver	5.6%	6.4%	N/A

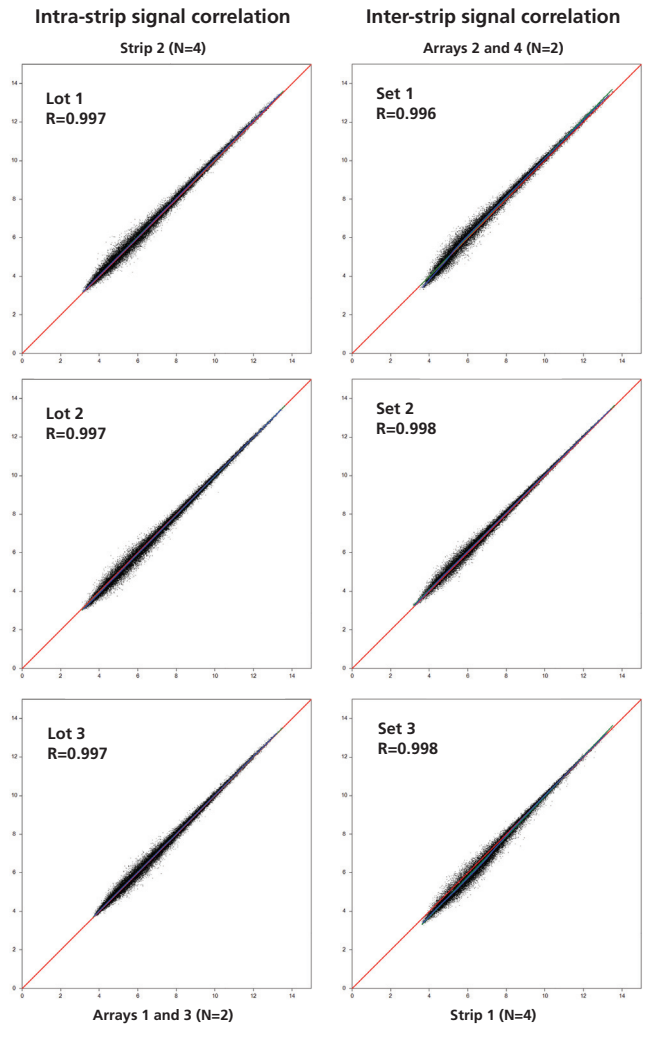
*One-year licenses for academic customers only

To further explore the reproducibility, we evaluated the correlation of median RMA signal for arrays within and between strips (Figure 1). We observed high correlation for both comparisons (R > 0.990).

Sensitivity and specificity

To assess the sensitivity and specificity of Gene 1.1 ST Array Strips on the GeneAtlas® System, we employed a mini-Latin square experimental design. A set of 61 in vitro transcribed (IVT), unlabeled transcripts was added to total HeLa RNA. The transcripts were selected based on their absence of expression in the HeLa cell line under normal growth conditions. The spiked transcripts were assembled into four groups such that each transcript was present at one of four relative abundance levels: 0, 1:200,000, 1:100,000, or 1:50,000 (Table 2). The 1:200,000 relative abundance represents a concentration of 0.75 pM, which is equivalent to approximately 1.75 copies per cell.

Figure 1: Signal scatter plots. Pooled sample prepared from total HeLa RNA was used to compare signal correlation within and between strips. The median RMA signal was calculated from two or four replicates per sample. Within-strip comparisons used the median RMA signal of duplicate arrays on a four-array strip (N = 2). For between-strip comparisons, the median of four replicates from one strip was compared to the median of four replicates from a different strip (N = 4). Not all data are shown.



Array performance at the gene and exon level was measured by calculating the *t*-statistic for the discrimination of the same spike at different concentrations. Each concentration was represented by three technical replicates. The *t*-statistic (alpha significance level of 0.95) was calculated for each of the 61 spikes in each of two concentration comparisons (1:200,000 to 0 and 1:100,000 to 1:200,000). The *t*-statistic is a measure of the separation of signal intensities for the two concentrations relative to the variance of the three replicates within each concentration. The more pronounced the separation or the smaller the variance within the replicates, the higher the *t*-statistic value.

Table 2: Mean Pearson correlation coefficient for within- and between-array-strip comparisons. Pooled, labeled target from total HeLa RNA was hybridized in quadruplicate to Human Gene 1.1 ST Array Strips. The median RMA signal was calculated from two replicates (N = 2) for within-strip comparisons or four replicates (N = 4) for between-strip comparisons. The average Pearson correlation coefficient (R) from 12 within-strip and 66 between-strip comparisons is summarized.

Number of comparisons	Mean R	Range R
Within-strip (N = 12)	0.996	0.992–0.997
Between-strip (N = 66)	0.994	0.986–0.998

Table 3: Latin square design. The rows represent spike groups (A–D) and the columns represent samples (pools 1–4). Each sample contained spikes at each of the four abundance levels (0, 1:200,000, 1:100,000, and 1:50,000). Each spiked transcript was represented at all four abundance levels across the four samples.

Spike group	Pool 1	Pool 2	Pool 3	Pool 4
Group A	0	1:50,000	1:100,000	1:200,000
Group B	1:200,000	0	1:50,000	1:100,000
Group C	1:100,000	1:200,000	0	1:50,000
Group D	1:50,000	1:100,000	1:200,000	0

The sensitivity and specificity for discrimination of transcripts at different concentrations can be illustrated by receiver operator characteristic (ROC) curves (Figure 2). The area under the curve (AUC) is a measure of the ability to discriminate between two concentrations, with perfect discrimination resulting in an AUC of 1.0. The percentage of spikes with a significant *t*-statistic was also calculated for the two comparisons above (Table 4). These comparisons represent the ability to detect the presence of spikes at very low concentrations when compared to a sample in which that transcript is not expressed (1:200,000 vs. 0) and the ability to discriminate two-fold changes between transcripts at very low concentrations (1:200,000 vs. 1:100,000).

Conclusion

Human, Mouse, and Rat Gene 1.1 ST Array Strips and the easy-to-use GeneAtlas® System enable whole-transcript expression profiling. The data presented demonstrate a high level of sensitivity, specificity and reproducibility. AUC values observed for gene-level comparisons of spikes at 1:200,000 and 1:100,000 were greater than 0.990, indicating very good sensitivity for a two-fold change at low concentrations. The CV was less than 8.3 percent for multiple tissue types, and signal correlation within and between strips exceeded 0.990.

Figure 2: Gene- and exon-level ROC curves. ROC plots of the 1:200,000 relative abundance compared to 0 (**A**, **C**) or the 1:200,000 relative abundance compared to 1:100,000 (**B**, **D**). Plots **A** and **B** represent gene-level comparisons; plots **C** and **D** represent exon-level comparisons. The x axis represents sensitivity (true positive) and the y axis represents 1 – specificity (false positive). Perfect performance (100 percent true positive, 0 percent false positive) would be shown as a line reaching the upper left corner.

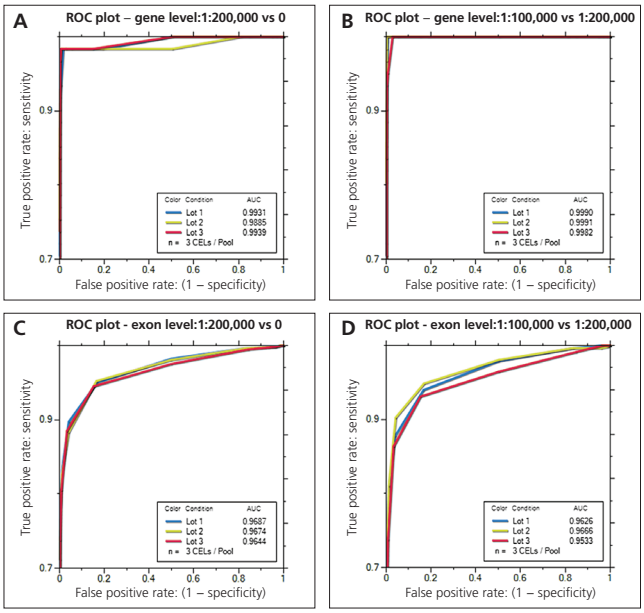


Table 4: Area under curve (AUC) values and percentage of spikes with significant t-statistic. The AUC values and percentage of the 61 spiked transcripts with significant *t*-statistic were calculated for transcripts at relative abundance levels of 1:200,000 compared to 0 and 1:200,000 compared to 1:100,000.

Gene level	1:200,000 to 0 concentration		1:200,000 to 1:100,000 concentration	
	AUC	Percentage of spikes with significant *t*-statistic	AUC	Percentage of spikes with significant *t*-statistic
Lot 1	0.993	98.4%	0.999	100.0%
Lot 2	0.988	98.4%	0.999	100.0%
Lot 3	0.994	98.4%	0.998	100.0%
Exon level	**1:200,000 to 0 concentration**		**1:200,000 to 1:100,000 concentration**	
	AUC	Percentage of spikes with significant *t*-statistic	AUC	Percentage of spikes with significant *t*-statistic
Lot 1	0.969	89.0%	0.963	87.1%
Lot 2	0.967	87.8%	0.967	88.5%
Lot 3	0.964	88.0%	0.953	83.7%

Affymetrix, Inc. Tel: +1-888-362-2447 ■ **Affymetrix UK Ltd.** Tel: +44-(0)-1628-552550 ■ **Affymetrix Japan K.K.** Tel: +81-(0)3-6430-4020
Panomics Products Tel: +1-877-PANOMICS www.panomics.com ■ **USB Products** Tel: +1-800-321-9322 www.usb.affymetrix.com

www.affymetrix.com Please visit our website for international distributor contact information.

For Research Use Only. Not for use in diagnostic procedures.

P/N EXP00906 Rev. 1