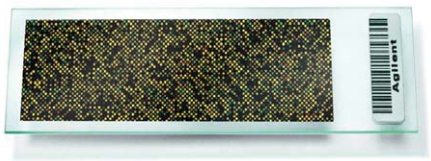
Whole Transcript Profiling Using SurePrint G3 Exon Microarrays in a Complete Workflow

Application Note

Authors

Becky Mullinax
Jennifer Venneri
Anne Bergstrom Lucas
Vinayak Kulkarni
Michael Janis
Jayati Ghosh
Jing Gao
Stephanie Fulmer-Smentek

Agilent Technologies, Inc.
La Jolla and Santa Clara, CA USA



Abstract

Whole transcript expression profiling is an important tool for understanding biological mechanisms and development, classifying tissue and tumor types, and identifying indicators for diagnosis and prognosis of disease. This type of profiling measures the expression of alternatively spliced exons from a given RNA transcript, which leads to the translation of a variety of proteins from a single RNA transcript. Each of these proteins has a potentially different, and even opposing, function in the cell. Profiling the expression level of individual exons thus provides a more complete representation of protein diversity than traditional microarrays, which measure expression at the gene level.

To address the need for whole transcript expression profiling from low quantities of total RNA, we have developed an exon expression workflow that includes custom and catalog human, mouse, and rat SurePrint G3 Exon Microarrays, a whole transcript cRNA amplification and labeling method using the Low Input Quick Amp WT (LIQA WT) Labeling Kit, and integrated GeneSpring GX data analysis software. Gene- and exon-level data analysis detected differential gene expression and alternative transcript splicing of both low- and high-abundance transcripts with a dynamic range of 4 to 5 orders of magnitude. Comparisons to both qRT-PCR and RNA-Seq showed high correlations. Using the exon microarrays, LIQA WT Labeling Kit, and integrated analysis software significantly decreased the amount of input RNA needed, reduced experiment time and cost, and increased throughput when compared to qRT-PCR and RNA-Seq methods.

Methods

Microarray design

The probes on the human, mouse, and rat exon microarrays were designed using public database content from RefSeq, Ensembl, and others (Table 1). Each catalog format includes probes to detect the exons in the coding regions of the transcripts. The expanded density of the 2x400K format allows the inclusion of probes to detect exons as short as 35 nucleotides (nt) in length and probes corresponding to the 5′ and 3′ untranslated regions (UTRs) (Figure 1). The 4x180K catalog microarrays focus on the highest quality database content, targeting a lower number of genes for a more economical and focused profiling option.

Agilent offers catalog exon human, mouse, and rat microarrays in the 4x180K and 2x400K formats, allowing researchers to select the most appropriate approach for their research. In addition, custom microarrays can be created using eArray, Agilent's free online design tool.

RNA and spike-in control preparation

The human exon microarrays were validated using the same total human RNA samples that were used in the Microarray Quality Control study (MAQC Consortium, 2006) — MAQC A (Agilent Universal Human Reference RNA, 740000) and MAQC B (Ambion Brain reference RNA, AM6050).

Table 1. Agilent Exon Microarray Catalog Content

Species	Array Format	Genes Targeted	Exon Probes	Databases Used for Design
Human	4x180K	20,411	174,458	RefseqBuild 36.3, RSNM only
	2x400K	27,696	233,164	Refseq Build 36.3 Ensembl Release 52 Unigene Build 216 (Apr. 2009) GenBank mRNA (Apr. 2009)
Mouse	4x180K	23,215	165,984	RefseqBuild 37, RSNM only
	2x400K	33,795	235,714	Refseq Build 37 Ensembl Release 55 Unigene Build 176 (Apr. 2009) GenBank mRNA (Apr. 2009) RIKEN 3
Rat	4x180K	20,483	160,141	RefseqBuild 36.2, RSNM only
	2x400K	26,276	214,270	Refseq Build 36.2 Ensembl Release 55 Unigene Build 177 (Oct. 2008) GenBank mRNA (Jan. 2009)

Table 1. *Agilent Exon Microarray Catalog content. Both catalog microarray formats target exons in protein-coding regions. The 2x400K catalog microarrays also include probes targeting exons as small as 35 nucleotides and probes targeting all 5′ and 3′ UTRs.*

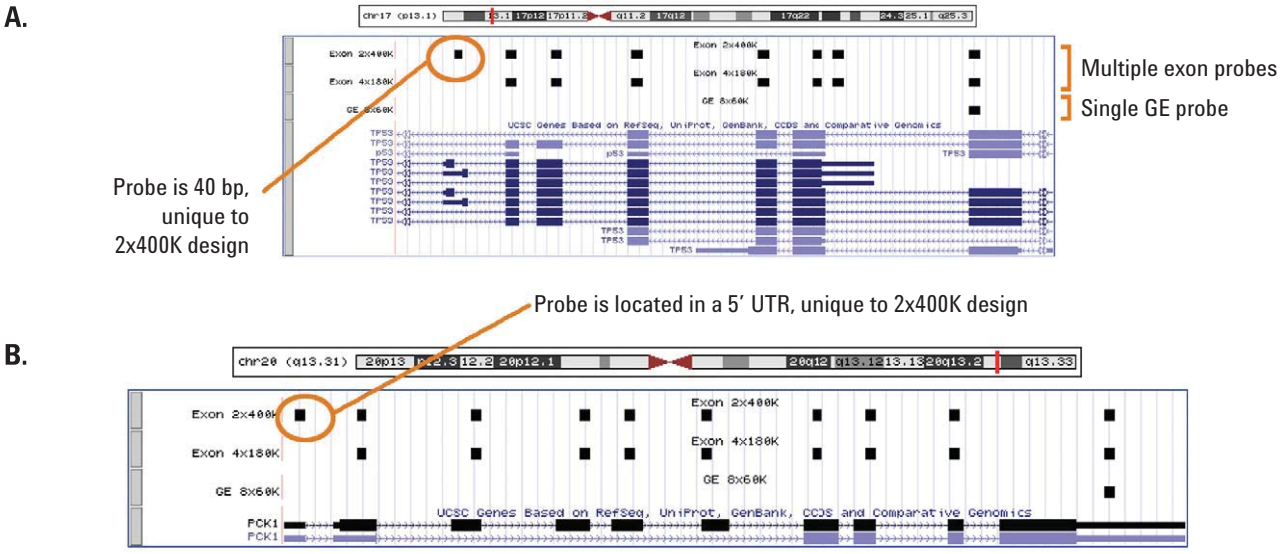


Figure 1. *UCSC Genome Browser representations of TP53 (A) and PGK1 (B) genes illustrating the number of exons in each gene and their representation on the Agilent SurePrint G3 Exon 2x400K and Exon 4x180K microarrays as compared to the single probe on the Agilent Gene Expression 8x60K microarray .*

Mouse brain and Agilent Universal Mouse Reference RNA (UMRR, 740100) and rat brain and thymus RNA were used to validate the mouse and rat exon microarrays, respectively. The quality of the RNA was assessed with the Agilent 2100 Bioanalyzer (Agilent, G2943C) using the RNA 6000 Nano Kit (Agilent, 5067-1511). Spike-in controls from the One-Color RNA Spike-In Kit (Agilent, 5188-5282) and Two-Color RNA Spike-In Kit (Agilent, 5188-5279) were added to the total RNA in one- and two-color microarray assays.

RNA labeling and hybridization

The Low Input Quick Amp WT (LIQA WT) Labeling Kit for one-color and two-color microarray assays uses a modified linear amplification procedure to generate high yields of Cy-labeled cRNA from low RNA inputs using a mixture of oligo dT- and random nucleotide-based T7 promoter primers.

All RNA labeling and hybridizations were performed as described in the Agilent One-Color Microarray-Based Exon Analysis – LIQA WT Labeling Protocol (G4140-90042, Version 1.0, November 2010) and the Agilent Two-Color Microarray-Based Exon Analysis – LIQA WT Labeling Protocol (G4140-90052, Version 1.0, November 2010). cRNA yields and specific activities were measured using a NanoDrop ND1000 spectrophotometer (Thermo Scientific), and 50 ng input total RNA was used in all experiments except in the experiment where the amount of RNA input was varied between 25 and 200 ng per assay.

Microarray hybridization and scanning

Cyanine-labeled cRNA was hybridized to the Agilent SurePrint G3 Catalog Exon microarrays. Fragmentation mixes were prepared for one-color and two-color assays and hybridizations were carried out as described in the user manuals.

Fluorescent signals were detected with the Agilent High Resolution C Scanner (G2656CA) using the scanning protocols specific for the microarray assay (one-color or two-color) and raw microarray image files were created.

Agilent Feature Extraction Image Analysis Software (Version 10.7.3) was used for data extraction from raw microarray image files. The software also generated QC reports using the protocols specific for the microarray assay (one-color or two-color) as well as data files for analysis with Agilent GeneSpring GX (Version 11.5).

Microarray data analysis

In these experiments, the exon microarray data were analyzed for both gene- and exon-level expression using the Agilent exon gene expression workflow in GeneSpring GX. The workflow includes exon splicing analysis using t-tests or multivariate splicing ANOVA and filtering for transcripts based on a splicing index.

For data visualization and analysis with GeneSpring GX, detection calls were made employing the default flag settings using the advanced analysis workflow provided for the Agilent exon one- or two-color experiment type. Signal intensities for each probe were normalized to the 75th percentile without baseline transformation. The software was used to visualize QC metrics and correlations of technical replicates. Exported data files, including probe annotation, were then used to compare microarray, qRT-PCR, and RNA-Seq platforms.

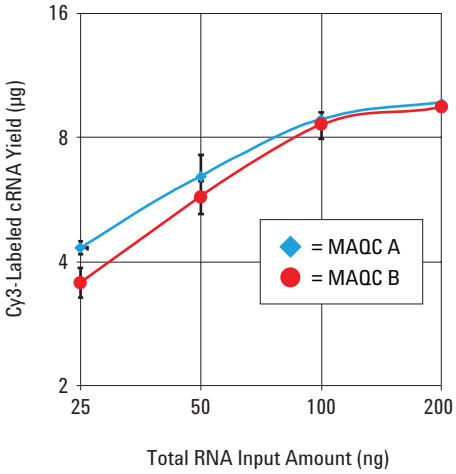
Gene-level analysis was performed to identify differentially expressed genes by evaluating the expression data for probes that detect exons for a given gene. Probe-level analysis was performed, from which splicing indexes were calculated to identify differentially expressed exons with associated p-values to provide confidence levels. Normalized probe-level expression data for a given gene was visualized to identify differentially spliced gene transcripts.

qRT-PCR and RNA-Seq data

The Brilliant III Ultra-Fast QRT-PCR Master Mix (Agilent, 600884) was used with custom designed primers and probes to detect 24 genes of varying abundances in MAQC A and MAQC B RNA. For comparison with TaqMan assays, data generated with catalog primers and probes to 774 genes were analyzed (MAQC Consortium, 2006). The $\log_2$ ratios of MAQC B/MAQC A qRT-PCR data were calculated to make comparisons between platforms.

RNA-Seq data was generated using the Illumina mRNA-Seq 8-Sample Prep Kit (RS-100-0801), Sequencing Kits v4 (FC-104-4001), and Genome Analyzer IIe, followed by analysis of the sequence output using Genome Studio (Illumina), which aligned the sequences to a reference genome (hg18) and calculated the normalized counts at the exon level. Log_2 ratios of MAQC B/MAQC A were calculated for each exon identified by RNA-Seq to allow for comparisons to be made between platforms.

A.



B.

		cRNA Hybridization Requirements	
Array Format	Recommended RNA Input	1-color	2-color
1x1M	100 ng	5.00 µg	2.500 µg
2x400K	50 ng	3.75 µg	1.875 µg
4x180K	50 ng	1.65 µg	825 ng
8x60K	50 ng	600 ng	300 ng

Figure 2. *Cy3-labeled cRNA yields for varying total RNA input amounts (A) were used to define the recommended RNA input in order to meet the cRNA hybridization requirements (B) for the different microarray formats for one- and two-color assays. The blue diamonds are MAQC A sample and the red circles are MAQC B sample labeled with the LIQA WT Labeling Kit.*

Platform comparisons

Scatter plots were generated to compare the $\log_2$ signal intensities of MAQC B/MAQC A for each platform and then orthogonal fit statistics were calculated. To make comparisons at the exon-level, the chromosome and genomic coordinates detected by the given exon microarray, qRT-PCR, and RNA-Seq assay were identified. Probe IDs from the exon microarray annotation were assigned to each data point and used as a common identifier when making the platform comparisons.

Results

Labeled cRNA yields and specific activities

To assess a range of total RNA input levels for use with the LIQA WT Labeling Kit and SurePrint G3 Exon Microarrays, the amount of input RNA of the MAQC samples was titrated from 25 to 200 ng. The reactions were performed in triplicate, resulting in average cRNA yields (Figure 2A), which were used to define the required and recommended amounts of input RNA for each microarray format (8x60K, 4x180K, 2x400K, and 1x1M) and for each assay type (one-color and two-color) (Figure 2B).

Assay sensitivity and dynamic range

Assay sensitivity and dynamic range were measured using the Agilent One-Color Spike-In controls that range in concentrations from approximately 1 in 12.5 billion to 1 in 12,500 transcripts in each assay. The exon microarray assay maintains a high sensitivity to low-abundance transcripts as low as 1 in 125 million and a dynamic range of 4 orders of magnitude (logs) (Figure 3A). In addition, dynamic ranges of slightly less than 4 to greater than 5 orders of magnitude were observed with

biological RNA when using 25 and 50 ng input RNA (Figure 3B). The wide dynamic range and high assay sensitivity demonstrate the ability to accurately detect exon expression at varying abundances.

Assay reliability

Scatter plot correlations of technical replicate one-color or two-color assays were visualized and correlation coefficients were calculated to assess assay reliability. GeneSpring GX was used to generate scatter plot signal correlations for one-color assays with MAQC A using biological probes that were detected in all technical replicates (Figure 4A). Orthogonal fit correlations were >0.99 for the comparisons of the 75th percentile-normalized $\log_2$ signals for the technical replicate microarrays in the one-color assay.

GeneSpring GX was used to generate scatter plot $\log_2$ ratio correlations for technical replicate assays in two-color assays with dye swaps of MAQC B/MAQC A using biological probes that were detected in all replicates (Figure 4B). Orthogonal fit correlations were >0.99 for the comparisons of the $\log_2$ MAQC B/MAQC A ratios for the technical replicate microarrays in the two-color assay. The high correlations demonstrated in the one-color and two-color assays illustrate the assay provides reliable and dependable results.

Assay reproducibility

Reproducibility and repeatability of microarray assays are measured by the variation in signals between technical replicates. For this purpose, the 75th percentile-normalized green processed signals (gProcessedSignals) of probes that were detected as well above background on all technical replicates were compared to calculate the coefficient of variation (CV). The results for the human, mouse, and rat microarrays in both the catalog and custom microarray formats are illustrated in a box-whisker plot (Figure 5). As can be seen on the secondary Y-axis in Figure 5, the number of probes detected varies with the number of unique features on each microarray format. In most microarray formats, the number of probes detected correlates with the number of microarray features. However, the number of probes detected is similar on the 2x400K and 1x1M human microarray formats because the same probes were used on each microarray with the features of the 2x400K microarray replicated 2.5 times on the 1x1M microarray. Overall, the median % CVs for all of the samples shown are <12%, demonstrating that use of the LIQA WT Labeling Kit and SurePrint G3 Exon Microarrays results in highly reproducible and repeatable data between microarrays and slides. These results are independent of the microarray format and species, indicating the robustness of the microarray assay, including the probe design, microarray manufacturing, labeled cRNA preparation, hybridization, microarray scanning, and data analysis methods.

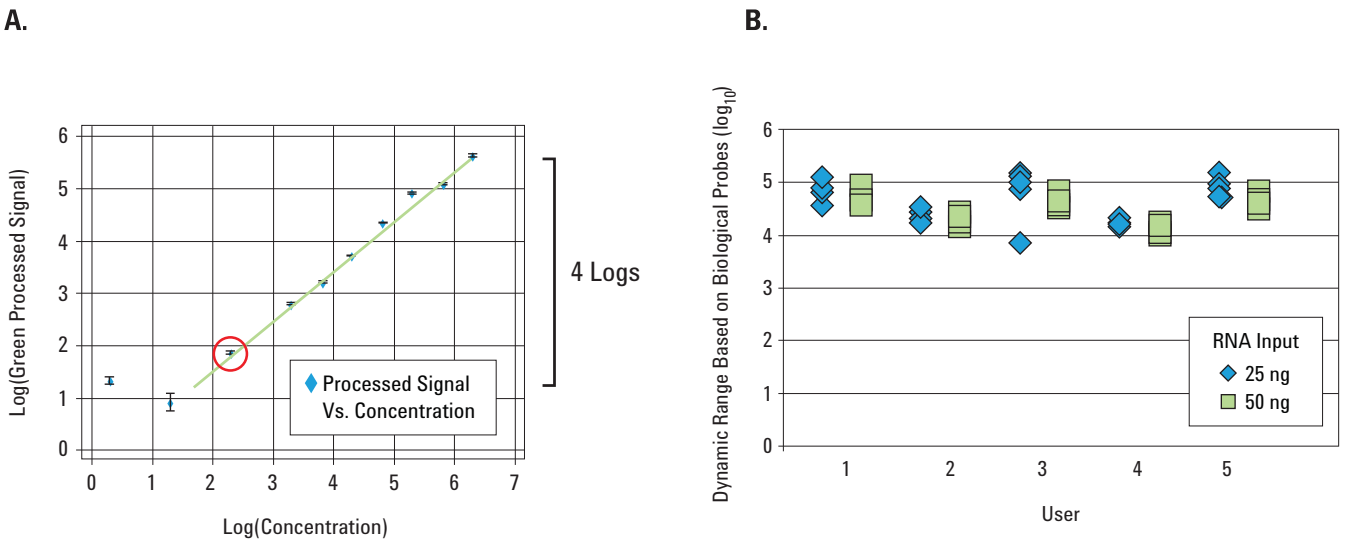


Figure 3. *Concentration-response plots based on the signal intensities of Agilent's One-Color Spike-Ins (A) and the dynamic range of biological probes (B) demonstrate assay sensitivity and dynamic range. The LIQA WT Labeling Kit generates labeled cRNA for an exon microarray assay that has the ability to detect rare transcripts as low as 1 in 125 million (circled in red) based on the One-Color Spike-Ins with over 4 logs of dynamic range (A). The dynamic range of log_{10} signals of the biological probes in assays performed by five different users ranged from slightly below 4 to above 5 logs with 25 ng (blue diamonds) and 50 ng (green squares) input RNA (B).*

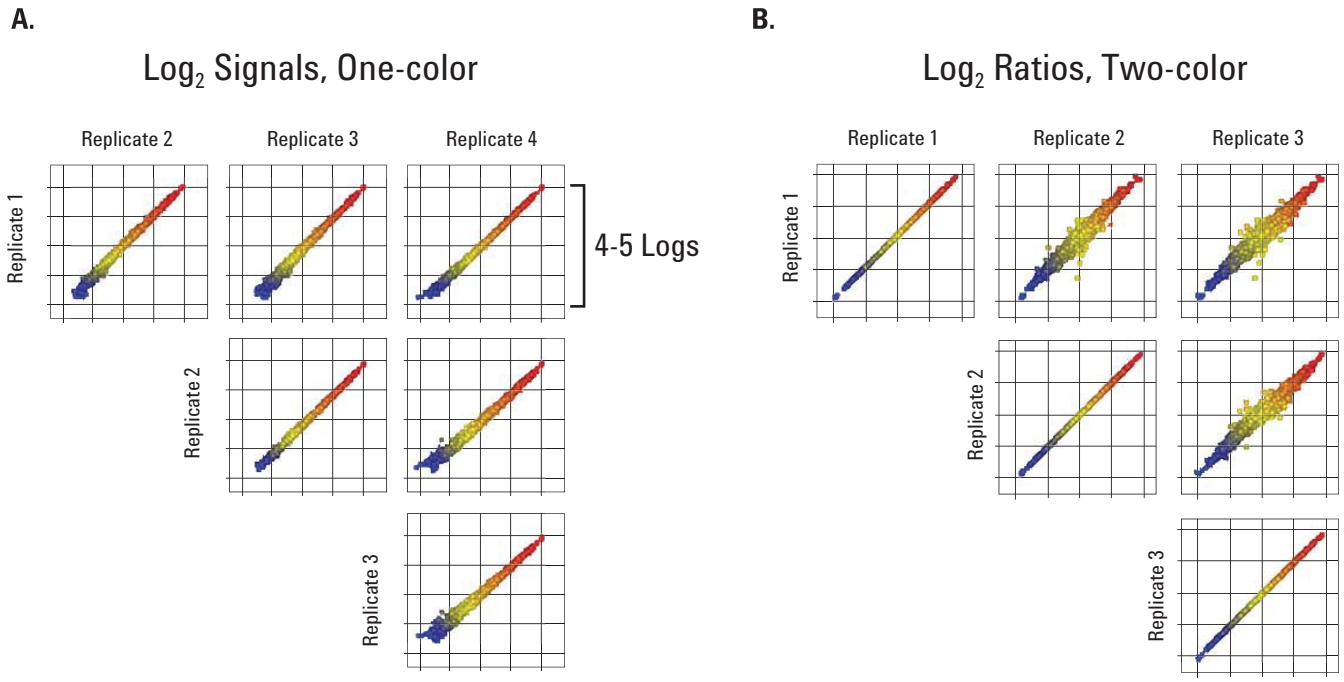


Figure 4. *High correlation between replicates is demonstrated in scatter plots of log_2 signals of replicates in one-color assays of MAQC A Cy3-labeled cRNA (A) and log_2 ratios in two-color assays of MAQC B/MAQC A (B). Log_2 signals in one-color assays demonstrate a 4-5 logs dynamic range based on biological probes (A). Orthogonal fit correlations are >0.99 for the both assays (A and B).*

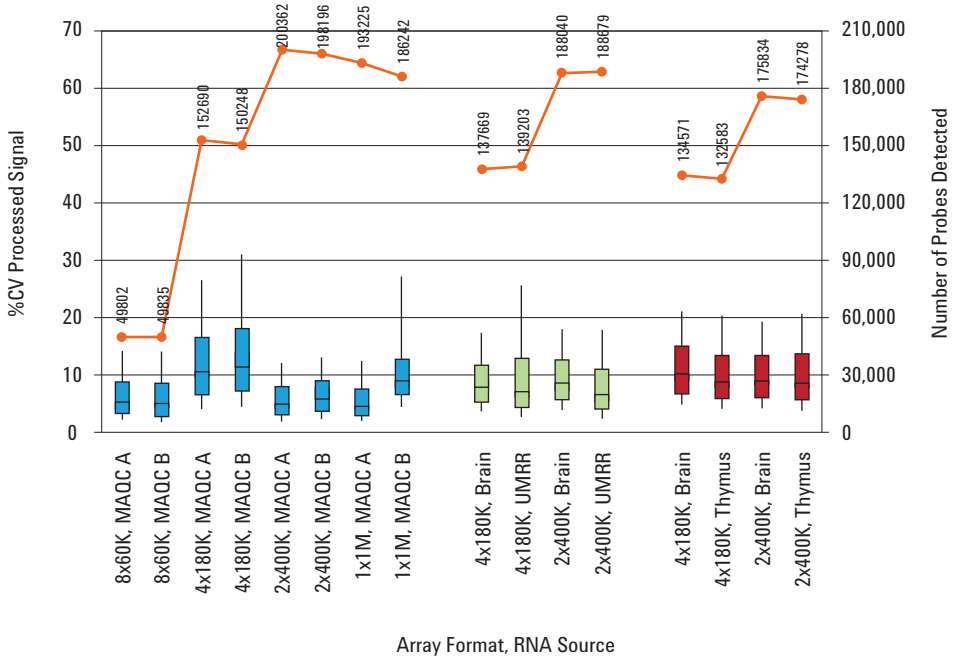


Figure 5. *Median inter-array % CVs are below 12% for all microarray formats. The median % CV calculated from the gProcessedSignal is represented by the horizontal line in the middle of each box and is given on the primary Y-axis. The top and bottom edges of the box are the inter-quartile range (25th percentile to 75th percentile) of the % CVs. The ends of the whiskers above and below the boxes indicate the 90th percentile and the 10th percentile of the % CVs. The % CV calculation was based on the number of probes given on the secondary Y-axis and is indicated by the orange lines and numbers above each box-whisker plot. Microarray formats are colored as follows: human (blue), mouse (green), and rat (red).*

Identification of differentially spliced transcripts

The ability to quickly and easily identify differentially spliced transcripts based on differences in exon expression in the MAQC A and MAQC B RNA was demonstrated using the GeneSpring GX software. A good example of visualizing differential splicing is the CLASP2 gene, where seven exons at the 5′ end of the gene were expressed at significantly higher levels in MAQC A relative to MAQC B (Figure 6). The $-\log_{10}$(corrected p-value) provided the confidence level for each comparison and the splicing index provided a convenient visual tool to quickly identify the relative expression levels of each exon. Representative gene transcripts provided structural information useful in evaluating the observed differences in exon expression. This visualization was generated using an intuitive workflow and provided all of the elements needed to identify differentially expressed exons with high confidence.

Comparison of microarray data to orthogonal qRT-PCR and RNA-Seq data

Gene expression microarrays, qRT-PCR, and RNA-Seq are tools for the analysis of differential gene expression. In addition, exon microarrays and RNA-Seq are useful in whole transcriptome analysis, which includes the detection of alternatively spliced exons. With the wide availability of these tools, it is useful to know the correlation between data generated with each tool. To this end, data generated with the microarray platform were compared to Brilliant III and TaqMan qRT-PCR data and RNA-Seq

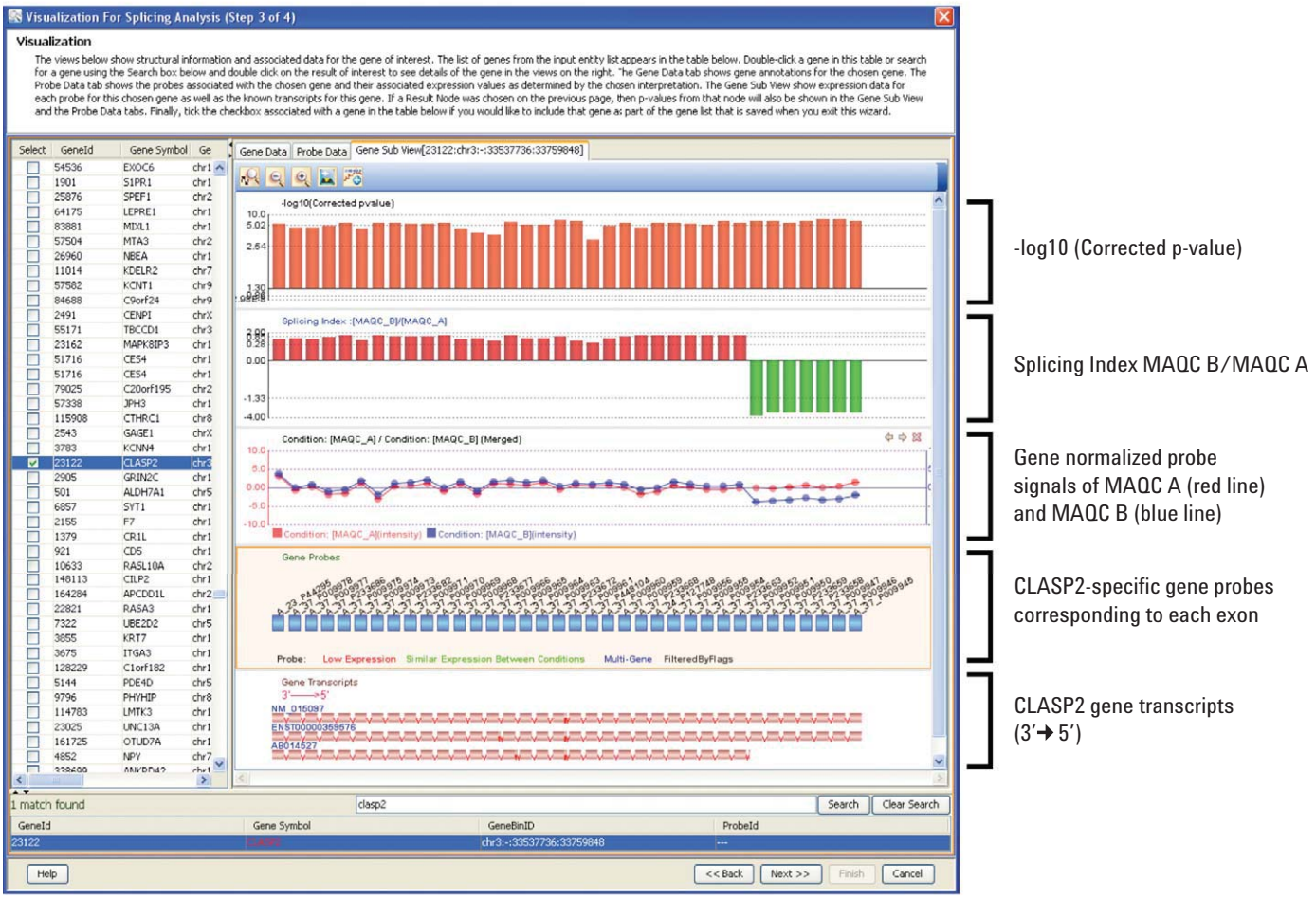


Figure 6. *Differentially expressed exons show that CLASP2 is clearly identified and visualized using GeneSpring GX software. This view includes the -log$_{10}$ (corrected p-value), splicing index (MAQC B/MAQC A), gene normalized probe data for MAQC A and B, gene-specific probes, and gene transcript structure for CLASP2. Visualizations of additional differentially spliced transcripts are selected using the panel on the left or through a built-in search function.*

data (Figure 7). Two parameters were assessed for each platform comparison: the slope (m), which indicates the accuracy of the comparisons at all levels of abundance, and the correlation coefficient (R), which indicates how well the data from the two platforms correspond. The number of data points compared is given by N. These plots indicate the highest correlation between the log$_2$ MAQC B/MAQC A ratios of the microarray and the Brilliant III qRT-PCR platform (R>0.98). While the correlation with RNA-Seq was lower (R>0.86) than with either of the qRT-PCR platforms, we anticipate that the correlation will increase as improvements are made to this relatively new technology. The comparison of the microarray and TaqMan qRT-PCR reagents generated the slope closest to the ideal value of 1.0 (m=0.968). These results demonstrate the log$_2$ signals are accurately represented for low- and high-abundance exons when using these assays. The high correspondence between all three platforms indicates the usefulness of all platforms in measuring differential gene expression. In addition, the high correspondence between the whole transcriptome platforms, exon microarrays and RNA-Seq, indicate the value of the exon microarray assays when compared to RNA-Seq based on the low cost and high throughput of exon microarrays with similar results.

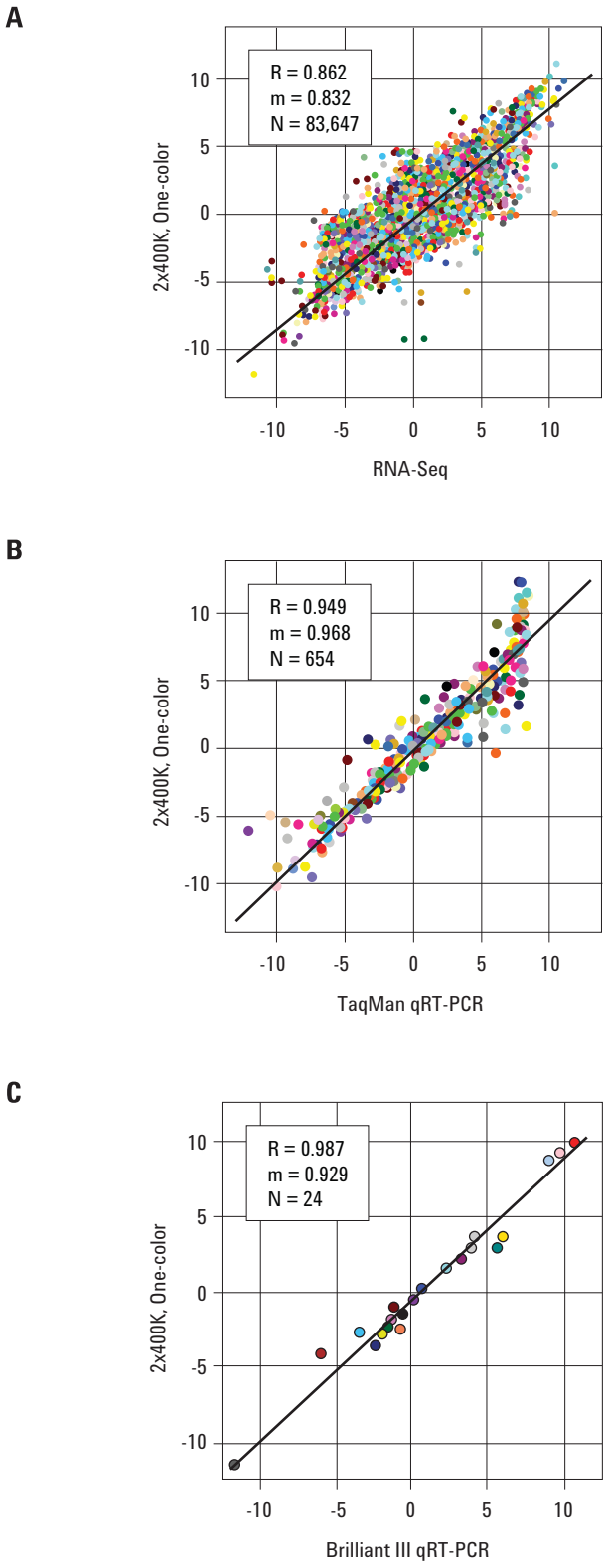


Figure 7. *Scatter plots comparing MAQC B/MAQC A log_2 data generated with the human SurePrint G3 2x400K Exon Microarray and LIQA WT Labeling Kit using 50 ng input RNA in a one-color assay are compared to RNA-Seq (A), TaqMan qRT-PCR (B), and Brilliant III qRT-PCR (C). The scatter plots compare the log_2 ratios of the MAQC B/MAQC A of the microarray assay on the Y-axes to the other platforms on the X-axes. The orthogonal fit (R), slope (m), and number of probes compared (N) are given for each platform comparison.*

Conclusion

Agilent has addressed the need for quick and easy whole transcript expression profiling by developing an exon expression workflow that includes custom and catalog human, mouse, and rat SurePrint G3 Exon Microarrays, a whole transcript cRNA labeling method, and Agilent's GeneSpring GX data analysis software. GeneSpring provides powerful, accessible statistical tools for fast visualization and analysis of a variety of multi-omic data.

Probes were designed to detect differential gene expression and alternative transcript splicing in custom and catalog SurePrint G3 Exon Microarray assays. The probe design was based on high-quality transcript information in public databases to provide users with the most useful and current exon microarray content.

Gene- and exon-level expression experiments were performed using a low amount of total input RNA with the Low Input Quick Amp WT (LIQA WT) Labeling Kit and SurePrint G3 Exon Microarrays. Efficient labeling resulted in high amounts of labeled cRNA with high specific activities from low amounts of total input RNA in less than one day.

We were able to detect One-Color Spike-In control RNA present at 1 in 125 million transcripts and biological transcripts with a dynamic range of 4 to 5 orders of magnitude reliably and reproducibly, providing quality data for low- and high-abundance genes. High correlations were also demonstrated in platform comparisons to both qRT-PCR ($R>0.98$) and RNA-Seq ($R>0.86$).

The Agilent complete workflow provides several advantages over the use of competitive exon microarrays by significantly reducing the amount of input RNA and providing integrated analysis software. In addition, the Agilent complete workflow can be performed in significantly less time and at lower cost with higher throughput when compared to RNA-Seq.

References

Shi, L., *et al*. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol.* 2006. 24(9):1151–61.

Agilent SurePrint G3 Exon Microarrays

Description	Slides/Kit	Part Number
SurePrint G3 Human Exon 4x180K Kit	3	G4832A
SurePrint G3 Mouse Exon 4x180K Kit 3	3	G4833A
SurePrint G3 Rat Exon 4x180K Kit	3	G4834A
SurePrint G3 Human Exon 2x400K Kit	3	G4848A
SurePrint G3 Mouse Exon 2x400K Kit	3	G4849A
SurePrint G3 Rat Exon 2x400K Kit	3	G4850A
SurePrint G3 Custom Exon 8x60K	1	G4863A
SurePrint G3 Custom Exon 4x180K	1	G4864A
SurePrint G3 Custom Exon 2x400K	1	G4865A
SurePrint G3 Custom Exon 1x1M	1	G4866A

Required Processing Components

Description	Part Number
Low Input Quick Amp WT Labeling Kit, no dye	5190-2942
Low Input Quick Amp WT Labeling Kit, One-Color	5190-2943
Low Input Quick Amp WT Labeling Kit, Two-Color	5190-2944
Low Input Quick Amp WT Labeling Kit, Cy5	5190-3386
Two-Color RNA Spike-In Kit	5188-5279
One-Color RNA Spike-In Kit	5188-5282
Gene Expression Hybridization Kit	5188-5242
Gene Expression Wash Buffer Pack	5188-5327
Stabilization and Drying Solution	5185-5979

Description	Academic Part Number	Commercial Part Number
GeneSpring GX Standalone 1 year license	G3784AA	G3778AA
GeneSpring GX Concurrent 1 year license	G3783AA	G3777AA
GeneSpring GX Standalone 2 year license	G3782AA	G3776AA
GeneSpirng GX Concurrent 2 year license	G3781AA	G3775AA
GeneSpring GX Standalone 3 year license	G3780AA	G3774AA
GeneSpring GX Concurrent 3 year license	G3779AA	G3773AA
GeneSpring Workgroup Server 1 year license	G1754AA	G1771AA

www.agilent.com/genomics/exon

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

Published in USA, January 4, 2016
Publication Number 5990-8058EN
PR7000-0091

Agilent Technologies