Gene Expression Profiling and Validation Using Agilent SurePrint G3 Gene Expression Arrays

Application Note

Authors

Bahram Arezi, Nilanjan Guha and Anne Bergstrom Lucas
Agilent Technologies Inc.
Santa Clara, CA USA

Abstract

Gene expression analysis is currently carried out using three methods that provide complementary data: microarrays, RNA-Seq, and RT-qPCR. Microarrays provide comprehensive whole-genome coverage in single experiments, with rapid data acquisition and results at a low price per sample. RNA-Seq provides the most comprehensive data across all RNA transcripts present in a sample. However, RNA-Seq also requires the largest sample amounts, the most time, and the highest investment. RT-qPCR provides the fastest and the most cost effective solution for reliable expression analysis, but it is most suitable for experiments involving a small number of genes across a large number of samples.

In this study, we compare Agilent SurePrint G3 gene expression microarray data with the results achieved using RT-qPCR and RNA-Seq. The data demonstrates excellent correlation between G3 microarray data and these two complementary techniques. A key factor in achieving strong correlation between microarrays and RT-qPCR data is the design of the primers and probes to the same target sequence on the microarray. This study also demonstrates the coverage of Agilent's gene expression workflow from discovery through biological validation.

Agilent Technologies

Introduction

Agilent Technologies offers a complete solution for gene expression analysis, from microarray profiling to reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) for follow-up biological validation across many samples (Figure 1). SurePrint G3 Gene Expression microarrays provide updated content and higher throughput for whole-genome analysis with eight samples per slide, enabling the evaluation of a broad spectrum of coding and non-coding RNAs in a single experiment.

RT-qPCR is a commonly used tool for corroborating microarray gene expression results and further interrogating targeted genes of interest. However, microarray and RT-qPCR results sometimes disagree due to inefficient primer design or the varying efficiencies of reverse transcriptase. RT-qPCR analysis using Agilent's Brilliant III qPCR reagents, IDT PrimeTime assays, and Agilent's Mx3005P QPCR System provides reliable expression analysis as demonstrated here. Primers and probes can be synthesized to fit the same target sequence from the microarray. By designing the 5′ nuclease dual-labeled probes to the G3 Gene Expression array probe sequence, one can ensure that the RT-qPCR assay is interrogating the same transcript as the microarray. This specific design criterion, combined with sensitive and reproducible microarray and RT-qPCR systems from Agilent, overcomes some of the common issues seen when correlating gene expression data across platforms.

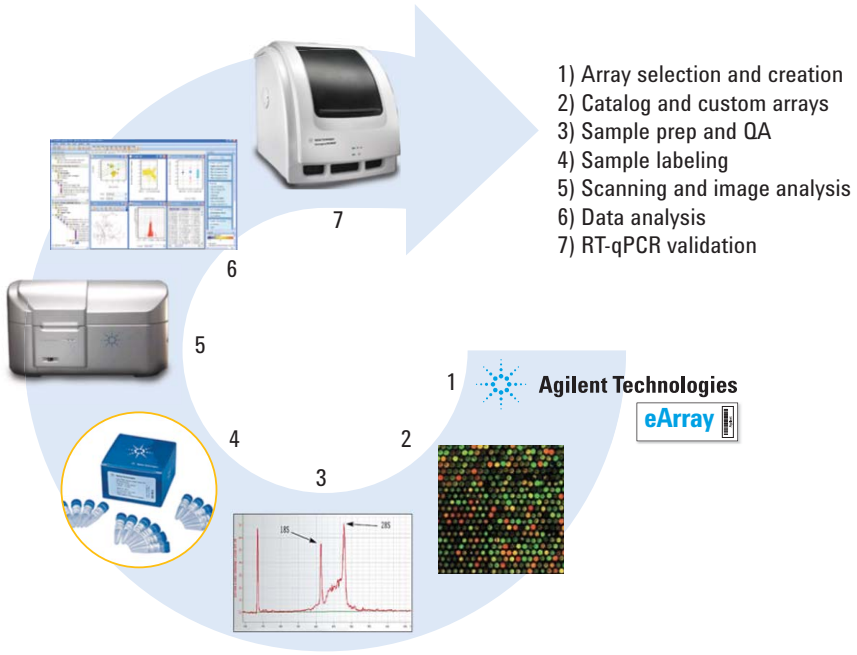


Figure 1. Agilent's complete gene expression analysis workflow offers solutions from microarrays through RT-qPCR validation.

Applying next-generation sequencing to RNA samples (RNA-Seq) to detect and measure expression levels has recently become a common tool for discovery in gene expression studies. The method provides high-throughput quantitative measurement for transcriptome profiling and effectively corroborates microarray results.

Here, we assessed the reproducibility of gene expression predictions for technical replicates using Agilent SurePrint G3 8x60K microarrays and the Agilent Low Input Quick Amp (LIQA) Labeling Kit. Similarly, the correlation of RT-qPCR data from technical replicates was also evaluated. Both studies demonstrate that these approaches are highly reliable for the measurement of gene expression values. Furthermore, a comparison of results between microarrays and RT-qPCR and between microarrays and RNA-Seq shows a close correlation, indicating that Agilent's microarrays and RT-qPCR solution give accurate and reproducible results.

Experimental

Samples

The RNA samples used for microarray, RT-qPCR, and RNA-Seq analysis were the Agilent Universal Human Reference RNA (MAQC A, 740000) and Ambion Brain Reference RNA (MAQC B, AM6050). The quality of the RNA samples were confirmed with the Agilent 2100 Bioanalyzer (G2940CA) using an Agilent RNA 6000 Nano LabChip Kit (5067-1511).

Microarray Analysis

Twenty-five ng of total RNA was amplified and labeled using the Agilent Low Input Quick Amp Labeling Kit, One-Color (5190-2305), and labeled RNA was hybridized to Agilent SurePrint G3 Human Gene Expression 8x60K Microarrays (G4851A). Agilent Feature Extraction Image Analysis Software (Version 10.7.3) was used to extract data from raw microarray image files. Data visualization and analysis was performed using GeneSpring GX (Version 11.0) software.

RT-qPCR Analysis

For RT-qPCR, assay primers and probes (PrimeTime Mini qPCR Assay) were ordered from Integrated DNA Technologies (IDT). IDT's Oligo Analyzer program was used to design the primers and probe sets. The probes were labeled with 6-FAM at the 5'-end and Iowa Black FQ at the 3'-end. The Agilent AffinityScript QPCR cDNA Synthesis Kit (600559) and Brilliant III Ultra-Fast QPCR Master Mix (600880) were used for quantitative RT-PCR. All amplifications were carried out on the Agilent Mx3005P QPCR system (401449).

cDNAs were synthesized using Agilent AffinityScript QPCR cDNA Synthesis Kits. qPCR assays were performed using Brilliant III Ultra-Fast QPCR Master Mix using the synthesized cDNA. Duplicate reactions were run on an Agilent Mx3005P using the **Comparative Quantitation** experiment within the MxPro Software. The differences in gene expression between the two sources of RNA were analyzed and reported as the normalized fold change of target to control. Microarray and RT-qPCR data were analyzed using Agilent MxPro software and exported for correlation analysis in GeneSpring.

RNA-Seq Analysis

The RNA samples were prepared using Illumina mRNA Sequencing Kit (RS-100-0801) with 1 µg of total RNA input. The resulting cDNA was validated using the Agilent 2100 Bioanalyzer DNA-1000 Kit (5067-1504) to ensure that the final product was approximately 200 bp. The validated cDNA library underwent cluster amplification on single read flow cells using the Illumina Single Read Cluster Generation Kit v2 (GD-203-2001-1), followed by sequencing with the Illumina 36 Cycle Sequencing Kit v3 (FC-104-3002) on the Illumina Genome Analyzer IIx. The single end sequencing was performed with short 36 bp reads. The raw sequencing data was aligned using the GERALD software and processed with the Illumina Consensus Assessment of Sequence and Variation (CASAVA) software to convert raw image data into intensity scores, base calls, and quality score alignments.

Microarray and RT-qPCR Comparison

The Agilent MxPro text output file was modified with column headers **Sample**, **Detector**, **Task**, and **Ct Avg** for uploading into the Agilent GeneSpring GX program. Both the microarray and the RT-qPCR data were baseline transformed to MAQC A. The RT-qPCR entity list was translated into Agilent Single Color technology (Gene Spring Technology 27220) using normalized signal values. An entity list was created for the microarray data in which only the genes studied by RT-qPCR were included. In cases where multiple probes on the microarray were available for a single gene, which occurs when there are multiple splice variants at the 3′ end, only the probe that matched the RT-qPCR primer design was used for comparison. Correlation plots were made using **Plot List Associated Values** in GeneSpring with the translated entity list.

RNA-Seq Comparison to Microarray

Microarray data was extracted from the raw image files using Agilent Feature Extraction software (version 10.5.1) and the text output files were imported into Agilent GeneSpring GX software (version 11.0). In GeneSpring GX, the microarray data was 75th percentile normalized and baseline transformed to the MAQC A samples to create $\log_2$ MAQC B/MAQC A ratios. The biological probes were filtered in GeneSpring to remove non-detected probes. The filtered microarray data, along with the sequencing text output files and RT-qPCR text files, were imported into Microsoft Access. A Microsoft Access query was used to generate the sequencing $\log_2$ MAQC B/MAQC A ratios and was used to compare the sequencing and microarray data. The resulting query was visualized in Spotfire DecisionSite (version 19.1.0.89).

Results and Discussion

Gene expression studies rely on the sensitive detection of genes as well as the accurate representation of gene expression levels. Microarrays offer a simple, fast, and affordable approach to genome wide expression analysis. SurePrint G3 Gene Expression Microarrays are found to be highly reliable, with reproducibility demonstrated by a correlation of R > 0.99 for signals across technical replicates in one-color assays (Figure 2). This high reproducibility spanned a dynamic range greater than 5 orders of magnitude, providing dependable measurement of both low and high expressors.

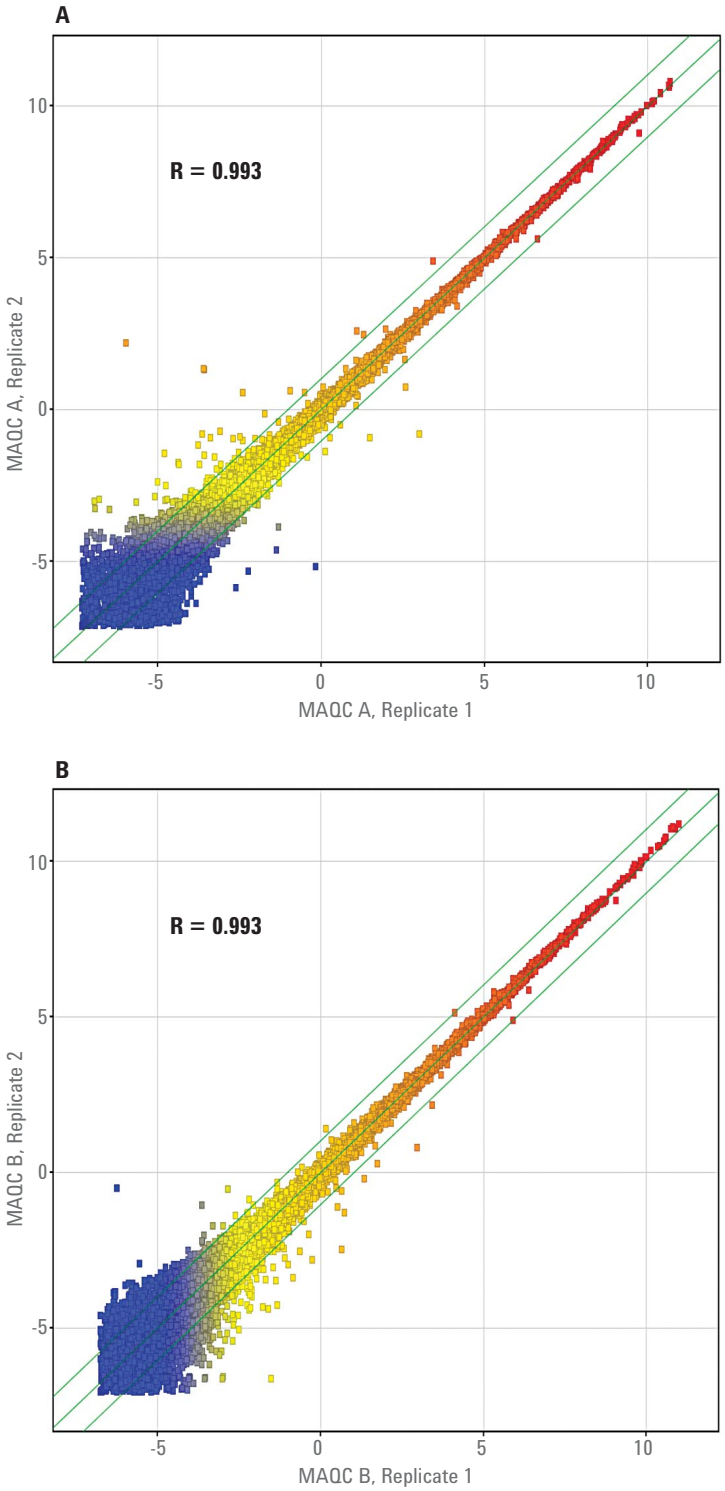


Figure 2. Correlation of microarrays using scatter plots to show the reproducibility of signal intensities between two technical replicates from MAQC A (A) and MAQC B (B).

The comparison of the ratios of the MAQC B/MAQC A sample between the RT-qPCR predictions, with levels determined by microarray analysis, indicates a very high concordance between both techniques (Figure 3) with a correlation of R > 0.99 between the two orthogonal assays.

Ratios calculated from expression levels based on RNA-Seq data were compared to microarray expression predictions and found to be highly concordant as well, with a correlation of R > 0.90 (Figure 4).

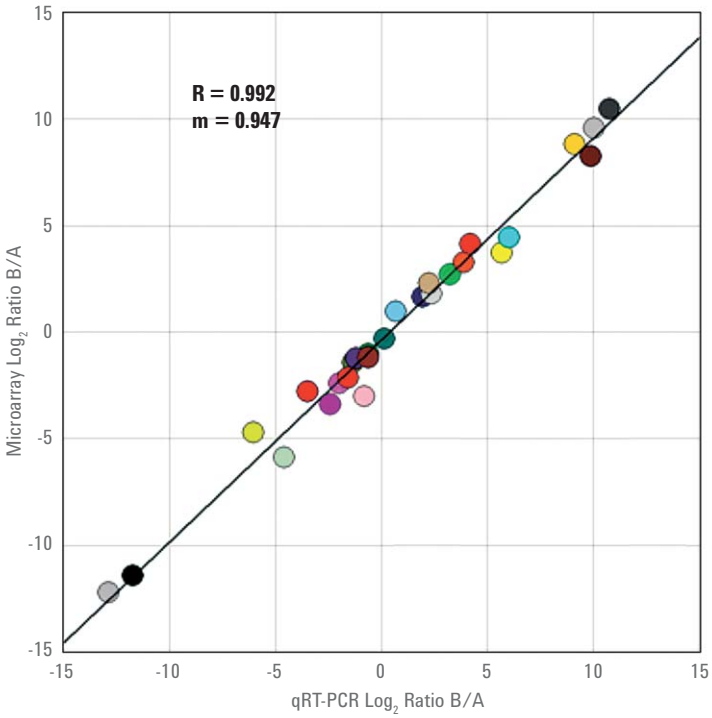


Figure 3. Correlation plot of Microarray and RT-qPCR log ratios demonstrates that Agilent microarrays provide accurate gene expression values. Both the microarray and RT-qPCR data were baseline transformed to MAQC A and log_2 ratios were plotted. The microarray entity list contained only the genes studied by RT-qPCR.

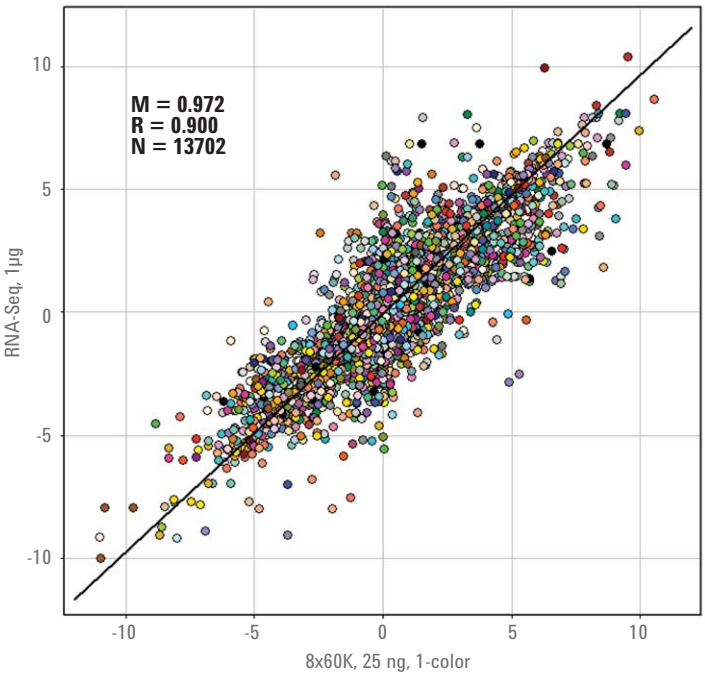
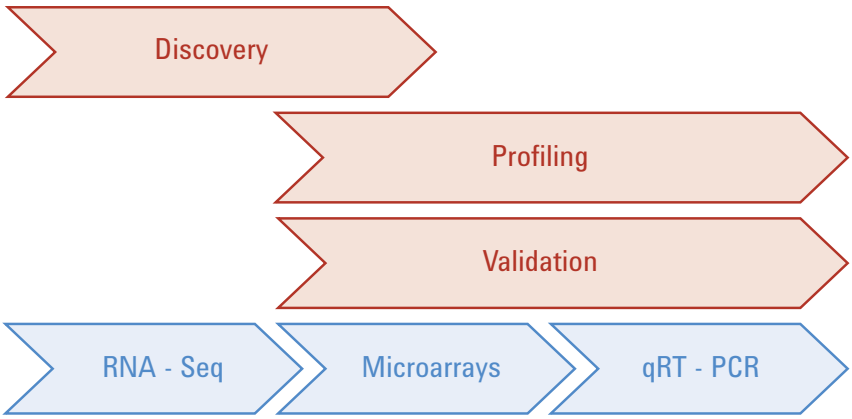


Figure 4. Comparison of SurePrint G3 microarrays and RNA-Seq data. The log_2 ratios of microarray and RNA-Seq data, baseline transformed to MAQC A, were plotted.

The correlation of Agilent SurePrint G3 gene expression microarray data with RT-qPCR and RNA-Seq results confirms that each provides an appropriate and complementary method for gene expression analysis (Figure 5). While RNA-Seq data provides the most comprehensive results across all RNA transcripts present in a sample, it also requires the highest RNA input amount, the longest time to collect data, the most complex data analysis, and the highest cost. Therefore, RNA-Seq is well-suited to discovery projects, but less optimal for profiling and high-throughput studies. In contrast, microarrays provide comprehensive whole-genome coverage in single experiments, with rapid results and low cost per sample. This makes microarrays highly suitable for profiling of hundreds or thousands of targets, or for validating results identified in RNA-Seq experiments. The possibility of creating custom gene expression microarrays using the Agilent SurePrint G3 gene expression platform opens a window for validating novel information attained through a discovery project using RNA-Seq. Alternatively, when only a handful of genes require profiling or validation, RT-qPCR provides the fastest and the most cost-effective solution for reliable expression analysis.

Conclusions

Agilent SurePrint G3 Gene Expression Microarray data tightly correlates with RT-qPCR data generated on the Agilent Mx3005P QPCR System using Brilliant III Ultra-Fast QPCR reagents. A key factor in achieving strong correlation of data is designing primers and probes to the same target sequence from the microarray. Data from both platforms was easily imported and analyzed in GeneSpring GX to further streamline the workflow. We also demonstrate the consistency of the microarray data with RNA-Seq results. Overall, this study establishes the reliability of Agilent's gene expression product offerings in giving accurate and reproducible results with complementary technologies.



Starting Material	1 µg	10 ng	5 ng
Sample Prep Total Time	11.5 h	6.5 h	1 h
Sample Prep Hands-on TIme	3.75 h	1.5 h	30 min
Sample to Data Total Time	~ 1 week	1.5 d	2 h
Analysis	Challenging	Simple	Simple
Price per sample	$$$	$$	$

Figure 5. Workflow comparing three different gene expression analysis platforms.

www.agilent.com/genomics/microarrays

Published in the USA, March 30, 2012
5990-9953EN

Agilent Technologies