

Validating Microarray Data Using RT² Real-Time™ PCR Products

Introduction:

Real-time PCR monitors the amount of amplicon as the reaction occurs. Usually, the amount of product is directly related to the fluorescence of a reporter dye. Because it detects the amount of product as the reaction progresses, real-time PCR provides a wide linear dynamic range, demonstrates high sensitivity, and is very quantitative. The initial amount of template DNA is inversely proportional to a parameter measured for each reaction, the threshold cycle (C_t). While not requiring post-reaction processing (such as characterization by agarose gel electrophoresis), real-time PCR does require dedicated and expensive equipment. The method is expensive to set up initially, but it becomes rather cost-effective once used routinely particularly when employing SYBR Green as the fluorescent reporter.

SYBR Green-based detection is the least expensive and easiest method available for real-time PCR. Other methods (such as TaqMan) require an expensive third primer labeled with a dye and a quencher. Most real-time systems detect and accommodate SYBR Green making the method very flexible; however, some instrumentation may also require the simple addition of a reference dye to normalize the system's optics. SYBR Green specifically binds double-stranded DNA by intercalating between base pairs, and fluoresces only when bound to DNA. Detection of the fluorescent signal occurs during the PCR cycle at the end of either the annealing or the extension step when the greatest amount of double-stranded DNA product is present. However, SYBR Green detects any double-stranded DNA non-specifically. Therefore, the reaction must contain a combination of primers and master mix that only generates a single gene-specific amplicon without producing any non-specific secondary products.

RT² Real-Time™ PCR Primer Sets and Master Mixes from SuperArray

RT² PCR Primer Sets and any one of our PCR master mixes combine to generate a complete assay optimized for SYBR Green-based detection on any real-time instrument. An experimentally verified computer algorithm designs each set, and a quality control assay guarantees that they yield a single band of the predicted size by agarose gel electrophoresis. Because primers self-designed by researchers fail nearly 50 percent of the time, the RT² Primer Sets save time and effort in primer design. RT² Primer Sets are available for any gene in the human, mouse or rat genome and are available in a 24-reaction scale, to verify a few samples and many genes, and in a 200-reaction scale, to verify many samples and a few genes.

RT² Real-Time™ PCR Master Mixes and Support Protocols are instrument and application specific providing the primer sets with enough flexibility to be used with the instrument already in your lab or with the one that you plan to purchase. Universal master mixes are available without reference dye for instruments that do not need it (such as the Cepheid SmartCycler) or for you to add the appropriate reference dye (such as the ROX reference dye for the ABI instrumentation). Other master mixes are also available with the appropriate reference dye already included for a specific instrument (such as fluorescein specifically for the BioRad iCycler). Each PCR master mix features ready-to-use convenience. These 2X solution formulations in a 200-reaction scale already contain SYBR Green and allow for low and high-throughput studies with any number of samples and genes.

Protocol for Microarray Data Verification:

This demonstration and tutorial attempt to validate a result obtained by a microarray analysis showing that the expression of the human TNFAIP3 gene increases in HeLa cells upon treatment with TNF-alpha (Figure 1). Therefore, this experiment represents simplest possible example of one control sample, one experimental sample, and one gene. More complicated analyses require a correspondingly more complicated experimental setup and protocol. Also for demonstration purposes, the protocol will involve products from SuperArray including a complete reagent kit for

reverse transcription, the ReactionReady™ First Strand cDNA Synthesis Kit ([C-01](#)), as well as the RT² PCR Primer Sets and Master Mixes.

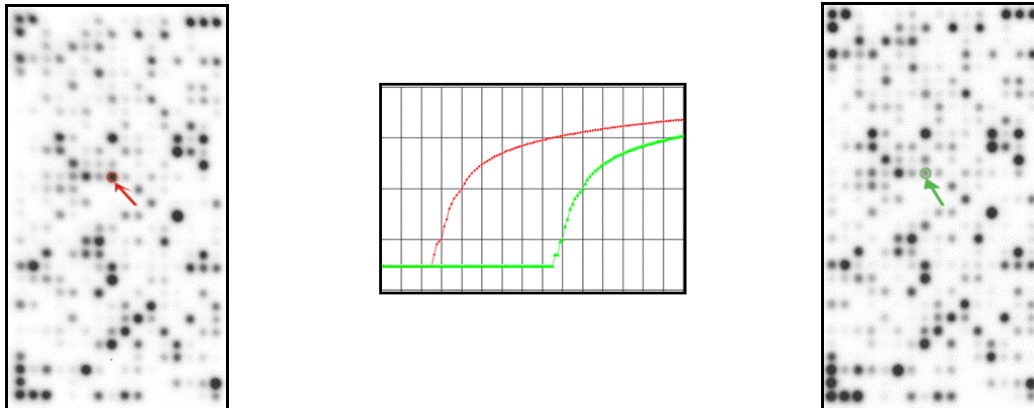


Figure 1: Induction of TNFAIP3 gene expression in HeLa cells upon TNF-alpha treatment. Total RNA was isolated from HeLa cells that were either treated with TNF-alpha (20 ng/ml, 30 min) or left untreated. RNA (3 µg) was characterized with an Oligo GEArray, and the results are displayed in the left and right panels, respectively. The results demonstrated that among many other genes the expression of TNFAIP3 increased in treated cells relative to untreated cells. The same RNA was also used for real-time PCR verification of the up-regulation (middle panel).

Reverse Transcription:

Reverse transcription enzymatically converts RNA into first strand cDNA, the substrate or template for the polymerase chain reaction. Each RNA sample requires one standard reverse transcription reaction: the control RNA (mock-treated), the experimental RNA (TNF-alpha treated), and a source of reference RNA, for example, XpressRef™ Universal Reference Total RNA from SuperArray ([GA-004](#)). The reference RNA sample will serve as a source of gene expression for either the standard (or calibration) curve or the positive control. Each reaction generates enough of each cDNA template for at least 20 polymerase chain reactions. Follow the protocol below using the ReactionReady™ First Strand cDNA Synthesis Kit from SuperArray.

1. Prepare Annealing Mixtures:

For each RNA sample, combine the following in a sterile PCR tube:

Total RNA	1.0 to 5.0 µg
Buffer P (Random Primers)	1 µl
RNase-free H ₂ O	Adjust the final volume to 10 µl

Mix well. Incubate at 70 °C for 3 minutes, then at 37 °C for 10 minutes. This step melts any RNA secondary structure and allows the random primers to anneal to the RNA.

2. Prepare the RT Cocktail:

Combine the following in a sterile PCR tube:

5X RT Buffer (BC)	16 µl
RNase-free H ₂ O	16 µl
RNase Inhibitor (RI)	4 µl
Reverse Transcriptase (RE)	4 µl
Final Volume	40 µl

3. Perform the RT Reaction:

Pre-warm the RT Cocktail at 37 °C for 1 minute. Add 10 µl of the RT Cocktail to each Annealing Mixture. Incubate at 37 °C for 60 minutes to generate first strand cDNA. To prevent them from interfering with the next step, heat-inactivate the reverse transcriptase and degrade the RNA at 95 °C for 5 minutes. Store the reaction on ice while planning the reactions below.

Standard Curve:

Relative gene expression profiling does not require absolute quantification, and fold-changes in gene expression are unit-less numbers. The absolute amount of message in the original RNA sample does not need to be known. The RNA source that you choose simply needs to express the gene of interest. *In vitro* transcripts or full-length cDNA clones are not necessary.

Instead, choose your template based on your microarray results and the relative availability of the RNA. Specifically, use your control template (from RNA from control-treated cells) for down-regulated genes, or use your experimental RNA (from RNA from TNF-alpha treated cells) for up-regulated genes. However, you may be verifying several up- and down-regulated genes at the same time. Or, you may be verifying the regulation of genes in several samples in which gene expression may increase or decrease depending on the sample or the gene. In these more complicated cases, generate a mixture of equal volumes of each template for the standard curve. This method still provides a calibration curve that faithfully represents each gene. Also use this method if the amount of RNA or cDNA template material available from your control and/or experiment samples is precious or limiting. Alternatively, use template generated from reference RNA to conserve control and/or experimental RNA or if you are still not sure which template or what combination of templates is the best.

Carefully plan the setup of your standard curve. Prepare duplicate sets of five (5) 10-fold serial dilutions of one of the cDNA templates. More replicates are not necessary, because the serial dilution themselves serve to control for systematic variation. However, very accurate pipeting is absolutely critical for generating accurate real-time PCR results. Use a calibrated P10 or even a P2 pipettor.

Controls:

Also, plan to include two different control reactions. Set up one positive control using the cDNA template generated from the reference RNA. If you are already using this template to setup the standard curve, there is no need for this extra reaction. Alternatively and if available, you may also use a full-length cDNA clone of the gene of interest. The positive control reaction should generate product and give a reasonable threshold cycle value. (See below.) Also set up one negative control (also known as the water control) that lacks any cDNA template entirely. This reaction provides an indication of the relative amount of DNA contamination or rare non-specific secondary products, in other words, the background amount of template in your reagents. Any other reactions generating threshold cycles greater than the negative control should be ignored because the amount of template lies below the limit of detection of the assay.

Some investigators also perform “minus-RT” controls to test for genomic DNA contamination in the original RNA sample. These controls require another mock reverse transcription reaction that lacks the enzyme for each RNA sample. Only genomic DNA in the sample would generate a PCR product, because no cDNA derived from mRNA would be present. In most cases, genomic DNA contamination is at a low enough level (yielding ΔC_t values greater than 6, see below) that it does not interfere with relative gene expression profiling. However, this control can be performed once to test your RNA isolation technique.

Unknowns:

Finally, plan the set up of your unknown control and experimental samples. Prepare replicate sets of serial dilutions of the control and experimental templates, for example, triplicates of two different 10-fold dilutions. For most purposes, three is a sufficient number of replicates. The preparation of these serial dilutions also helps insure that the resulting reactions will fall on the standard curve.

Set up replicates of all of these reactions (standard curve, controls, and samples) for all of the genes of interest and for an appropriate housekeeping gene. The relative expression of this housekeeping gene will be used to normalize the expression of the genes of interest to control for sample-to-sample systematic variation. If applicable, use the same housekeeping gene that was used for your microarray data analysis. In summary, the total number of reactions required for each gene in this experimental set up is:

10 (5-point standard curve in duplicate)
 2 (positive and negative controls)
 6 (two control template amounts in triplicate)
 + 6 (two treatment template amounts in triplicate)
 = 24 reactions

Polymerase Chain Reactions:

Each of the 24 reactions for each gene contains:

2X PCR Cocktail	12.5 μ l
Completed RT Reaction (template cDNA), or dilution thereof	1.0 μ l
RT ² Real-Time™ PCR Primer Set	1.0 μ l
10X reference dye stock (if needed)	2.5 μ l
<hr/>	
Adjust final volume with ddH ₂ O	25 μ l

Very accurate pipeting is the most critical factor in generating accurate real-time PCR results. Use a P10 or even a P2. Alternatively, generate a scaled-up pre-mix containing PCR cocktail, dyes, and primers for the same gene. Aliquot into individual PCR tubes using a repeating or perhaps even a digital, electronic pipettor, and then add the appropriate serial dilution of the template.

Place tubes in real-time thermal cycler. Enter the following program:

95 °C, 15 min; 40 cycles of (95 °C, 30 sec; 55 °C, 30 sec; and 72 °C, 30 sec); 72 °C, 5 min

Turn the OPTICS ON at end of annealing step. Proceed with the melting (dissociation) curve described below and then perform the data analysis with your instrument's software to determine the threshold cycle for each reaction.

Gene Expression Profiling with Real-Time PCR: Generating Standard Curve

Reactions containing a greater amount of the initial template substrate generate a detectable amount of product (a detectable signal) earlier than reactions containing a smaller amount of template. Appearance of detectable fluorescence at an earlier cycle number indicates a greater amount of initial template substrate. The detectable amount of fluorescence, a signal significantly greater than background, is known as the threshold. The cycle during which a reaction emits that threshold level of fluorescence is known as the threshold cycle, abbreviated C_t. (See Figure 2A.) Assuming the reverse transcription faithfully represents the initial RNA sample, the initial amount of template substrate measures the relative gene expression level. Therefore, gene expression is inversely proportional to the reaction's threshold cycle.

To generate the standard (or calibration) curve, plot the threshold cycle for the standard curve reactions against the fold dilution of the template cDNA on a semi-logarithmic (base 10) plot. Determine the fit to a straight line including the slope and the correlation factor (R²). (See Figure

2B.) Check the observed threshold cycle of the negative (water only) control if the instrument was able to detect product in that reaction at all. Make sure that all C_t values used in the analysis are less than that control. Be sure that all experimental C_t values lie in the linear dynamic range of the assay (on the standard or calibration curve).

Using the standard curve, determine the relative level of expression of the genes of interest and the housekeeping gene for all experimental samples. (See Figure 2B.) Normalize the expression level of the gene of interest by dividing by the relative expression level for the housekeeping gene for the same sample. Finally to calculate the fold-change in gene expression, divide the normalized number for the experimental sample by the normalized number for the control sample as seen in Figure 2C.

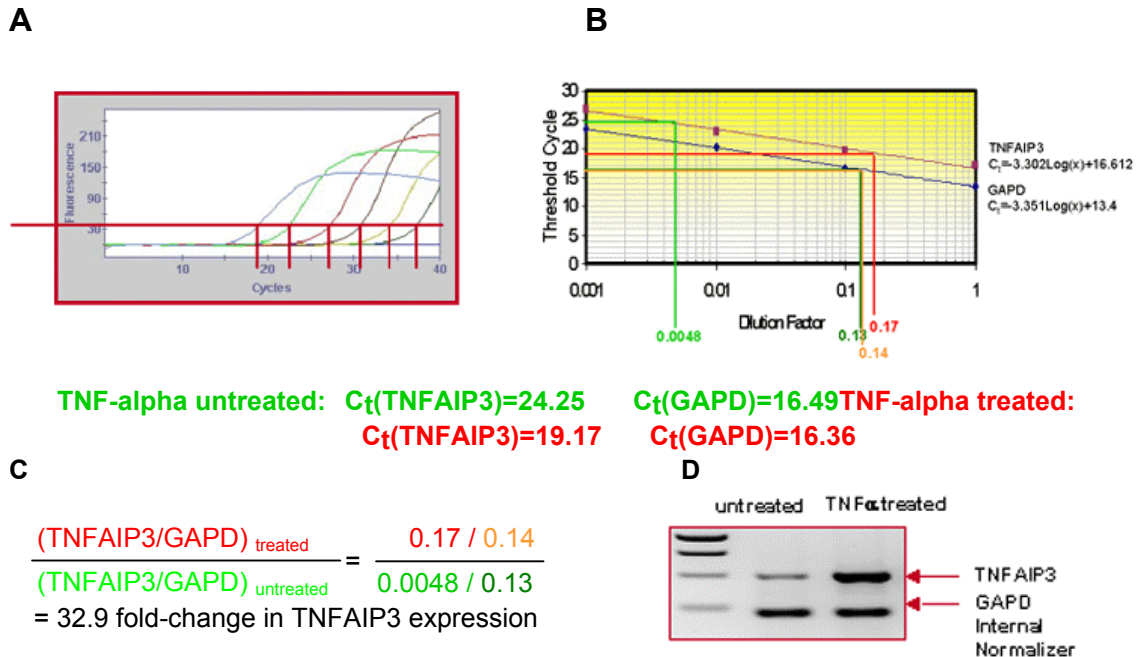


Figure 2: The method for determining relative gene expression profiles by real-time PCR involves the generation and use of a standard or calibration curve. In Panel A, the threshold cycle for each reaction occurs when its fluorescent reading increases significantly above the background. In Panel B, the threshold cycle for each standard curve reaction, for both the gene of interest and the housekeeping gene, is plotted against its corresponding dilution factor. The relative amount of gene expression in the experimental samples is determined using their threshold cycles and the standard curve. Panel C calculates the fold-change in TNFAIP3 gene expression in TNF-alpha treated cells relative to control-treated cells. Panel D displays the conventional or end-point RT-PCR verification results for the same experiment.

Note in Figure 2 that all of the unknown samples lie on the standard curve and that the change in TNFAIP3 gene expression is also clearly visible on that curve. The real-time PCR result at least confirms the direction of the change in gene expression observed by the microarray analysis. However, the magnitude of that fold-change in gene expression may be larger in the real-time experiment, because microarrays tend to suppress those changes due to their narrower dynamic ranges relative to real-time PCR. The change in gene expression is also very clearly visible by conventional (end-point) PCR on an agarose gel (Figure 2D); however, the magnitude here is also suppressed to a level similar to the microarray result.

Checking Specificity:

As mentioned, SYBR Green-based detection for real-time PCR only works if only one gene-specific amplicon is generated during the reaction. Unlike TaqMan-based assays, SYBR Green detection also uniquely allows you to check the specificity of the PCR using melting (also known

as dissociation) curves. After the 40 reaction cycles, perform a temperature ramp using the following program:

95 °C, 1 min; 65 °C, 2 min (OPTICS OFF); 65 °C to 95 °C at 2 °C / sec (OPTICS ON)

At low temperature, the PCR DNA product is double stranded, and it binds SYBR Green, which fluoresces. With increasing temperature, the DNA product melts or dissociates becoming single stranded, releasing SYBR Green and decreasing the fluorescent signal. Most real-time instruments usually plot melting curves as a first derivative. The inflection point in the melting curve then becomes a peak. (See Figure 3A.) Single peaks indicate a single product, which can be verified upon characterization of the product by agarose gel electrophoresis, as in Figure 3B. Multiple peaks usually indicate multiple products. These other products can have many sources including primer dimers, and genomic DNA contamination, but more frequently, they arise from unreported (un-annotated) splice variants of the gene of interest. (See Figure 3C.)

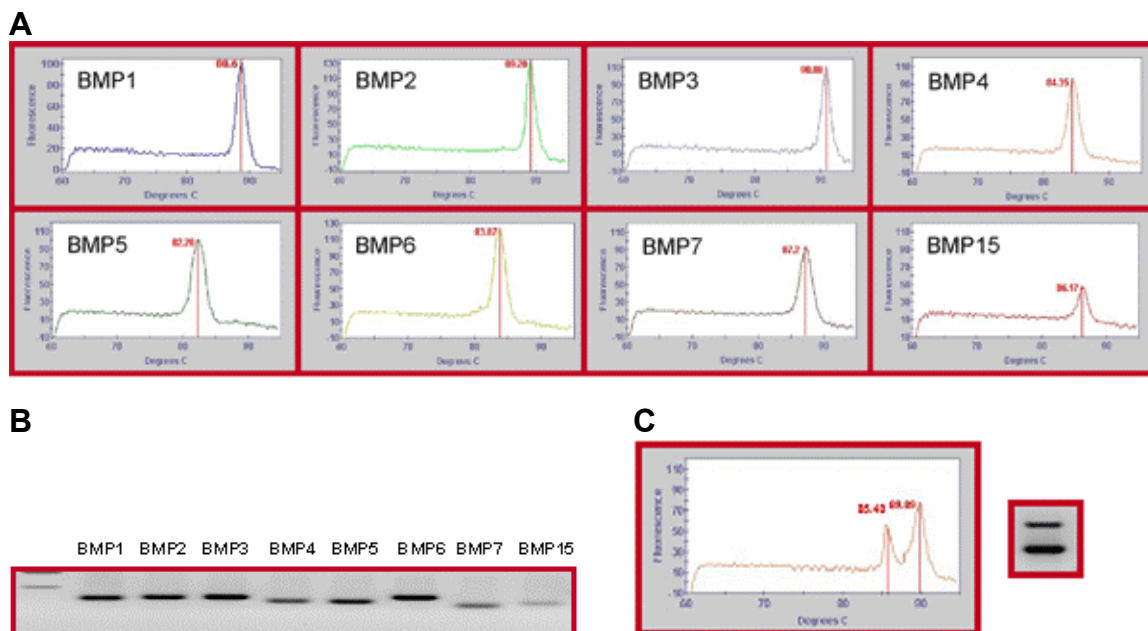


Figure 3: The SYBR Green based detection method uniquely permits a check of the reactions' specificity. Panel A displays the first-derivative melting curve for reactions each characterizing the expression of a different member of the BMP gene family. In each case, the curve contains only one peak indicating that the reaction generates only one product. Panel B confirms the melting curve results by demonstrating that these same reactions also generate only one band of the predicted size by agarose gel electrophoresis. Panel C includes an example of a primer set that detected a previously unknown splice variant or alternative transcript of a gene by generating a second amplicon product. As a result, two peaks appear in the first derivative melting curve and two bands appear by agarose gel electrophoresis.

Shortcuts for High-Throughput Applications: $\Delta\Delta C_t$ Methods

For high-throughput real-time PCR applications particularly when the expression of a few genes across multiple samples is being analyzed, mathematical shortcuts can be applied to replace the repeated determination of standard or calibration curves. However, these methods only work with low error rates if the repeated determination of the calibration curves on the same genes of interest and the housekeeping genes demonstrate reproducibility and high or at least similar amplification efficiencies. Use this method only if the replication efficiencies for your gene of interest and the housekeeping gene are the same or similar.

Simply perform replicate reactions for each gene of interest and for the housekeeping gene for each sample under conditions known to yield C_t values that lie on the standard or calibration curve. To compare gene expression between biological samples, first calculate the $\Delta\Delta C_t$ in the

following fashion. (See also Figure 4.) A ΔC_t value is calculated for each sample as the difference between the C_t values for the gene of interest and the housekeeping gene in each sample. The $\Delta\Delta C_t$ value is the difference between the ΔC_t values of an experimental sample and the control sample. The fold-change in gene expression is equal to $2^{-\Delta\Delta C_t}$ if the PCR replication efficiency for all genes is 100 percent. That is, the amount of the amplicon product perfectly doubles with each cycle. The PCR amplification efficiency can be determined from the slope of the calibration curve. A slope (m) equal to -3.3 indicates 100 percent efficiency. If the PCR efficiency is less than 1 (that is, if the slope of the calibration curves is greater than -3.3), then the fold-change in gene expression is equal to $10^{\Delta\Delta C_t/m}$, where m is the average slope of the calibration curves for the gene of interest and the housekeeping gene.

$$\begin{aligned} \text{Standard curves: TNFAIP3: } C_t &= -3.302\text{Lg}(x) + 16.612 \\ \text{GAPD: } C_t &= -3.351\text{Lg}(x) + 13.4 \\ \text{Average slope } m &= \frac{1}{2} (-3.302 - 3.351) = -3.327 \\ \text{TNF}\alpha \text{ untreated: } C_t(\text{TNFAIP3}) &= 24.25 \quad C_t(\text{GAPD}) = 16.49 \\ \text{TNF}\alpha \text{ treated: } C_t(\text{TNFAIP3}) &= 19.17 \quad C_t(\text{GAPD}) = 16.36 \\ \Delta C_t (\text{treated}) &= C_t(\text{TNFAIP3}) - C_t(\text{GAPD}) = 19.17 - 16.36 = 2.81 \\ \Delta C_t (\text{untreated}) &= C_t(\text{TNFAIP3}) - C_t(\text{GAPD}) = 24.25 - 16.49 = 7.76 \\ \Delta\Delta C_t &= \Delta C_t (\text{treated}) - \Delta C_t (\text{untreated}) = 2.81 - 7.76 = -4.95 \\ \text{The fold-change in TNFAIP3 expression} &= 2^{-\Delta\Delta C_t} = 2^{4.95} = 30.9 \\ \text{Or the fold-change in TNFAIP3 expression} &= 10^{\Delta\Delta C_t/m} = 10^{-4.95/-3.327} = 30.8 \\ \text{Error} &= |(30.9 - 32.9)/32.9| = 6\% \end{aligned}$$

Figure 4: The fold-change in TNFAIP3 gene expression between TNF-alpha and control treated cells is calculated using the data from Figure 2 and the mathematical shortcuts described in the text.

Summary:

For publication purposes, microarray results require verification and validation by an alternative and complementary gene expression profiling method. Real-time PCR is the most rigorous and commonly used technology for this purpose, and SYBR Green is the easiest and least expensive real-time PCR detection method. SuperArray Bioscience's RT² Real-Time™ Gene Expression Assays, by including validated primers and instrument-specific master mixes, are flexible enough for the equipment and systems in your laboratory as well as for your research needs. When using these assays to validate your microarray data, be sure to plan your experiment carefully. Include standard curves, positive and negative controls, and replicates of different unknown template amounts. Perform reactions for all genes of interest and a suitable housekeeping gene. With this setup, the determination of relative gene expression using real-time PCR for microarray data verification becomes very simple. SYBR Green also has the unique ability to allow for a specificity check and is applicable to high-throughput analyses.