

# 2

## Gene Expression Profiling Utilizing Microarray Technology and RT-PCR

Dominick Sinicropi, Maureen Cronin, and Mei-Lan Liu

*Genomic Health, Inc., Redwood City, California USA*

### 2.1. INTRODUCTION

Over the last decade microscale technologies for molecular analysis have become the springboard for a new era in biological investigation. In parallel with nucleic acid sequencing technology improvements that enabled completion of the human genome project years ahead of schedule [1, 2], methods were developed for high throughput analysis of genetic variation and gene expression. These new molecular analytical tools have stimulated a resurgence of non-hypothesis driven biological research and promise to play a key role in the emerging field of personalized medicine [3]. Knowledge gained from the most common type of genetic variation in DNA, single nucleotide polymorphisms (SNPs), has had an enormous impact on the identification of genes involved in disease and is beginning to be of value in tailoring therapeutic regimens for an individual's genetic composition [4]. Application of technologies for gene expression analysis, the subject of this review, has lagged behind analysis of genetic variation, primarily due to the intrinsic complexity of gene expression measurement. However, the number of studies employing gene expression analysis has expanded in the last few years as the available analytical methods mature and become more reliable and affordable.

A variety of methods have been used for gene expression quantification. The first methods to be widely adopted include northern blotting, RNA protection, differential display, serial analysis of gene expression (SAGE), and quantitative competitive reverse transcription-polymerase chain reaction (RT-PCR). Although all of these methods are still used today, real-time RT-PCR and DNA microarrays have achieved prominence in recent years and

are the focus of this review. Both of these methods are also useful for analyzing genetic variation; however, for this review we will concentrate on the use of these technologies for gene expression analysis.

Real-time RT-PCR, though not the first PCR-based method for gene expression analysis, has emerged as the “gold standard” against which other methods are compared. “Real-time”, or “kinetic”, PCR methods are those that measure the accumulation of PCR product with each PCR cycle. The original real-time PCR method measured the fluorescence of ethidium bromide intercalated in the double stranded DNA products of PCR amplification [5]. Subsequently, a variety of alternative methods for real-time quantitation of PCR products (reviewed below) have appeared as commercial products. The primary advantage of real-time quantitation is the relative simplicity of experiments as compared with other PCR-based methods. The ability to obtain a quantitative result in a single reaction is responsible, in large part, for the popularity of real-time PCR as compared with other PCR-based methods that require multiple step reactions for each sample that is analyzed. Another benefit is that commercial systems for real-time RT-PCR are configured for simultaneous analysis of hundreds of samples (or genes) simultaneously. These features, taken together with the high degree of analytical precision that is possible, have made real-time RT-PCR the method of choice for quantitative expression profiling.

Interest in DNA microarray technologies for measuring gene expression has exploded in recent years. Without question, the biggest advantage microarray technology has to offer is the large number of transcripts that can be quantified in a single experiment. DNA microarrays are capable of making tens of thousands of gene expression measurements simultaneously. Major commercial suppliers of DNA microarrays have recently released products in which the entire complement of known expressed human genes (the “transcriptome”; approximately 40,000 expressed sequences) can be measured on a single microarray. The unprecedented ability to monitor the expression of entire genomes has led to biological discoveries that would not have been possible by other methods. Nevertheless, microarray technology has limitations including its relatively high cost and inability to analyze more than one sample per array experiment.

In biological research it is often desirable to explore the expression profiles of two or more conditions (disease vs. normal, treated vs. untreated, etc.) with no hypothesis about which genes may be differentially expressed. Microarray technology is ideally suited for this early “discovery” phase of biological research due to the large number of genes that can be analyzed simultaneously. Microarray studies can reveal changes in expression of a smaller number of genes that are used for subsequent hypothesis generation and testing. Once identified, the smaller gene set can be analyzed by real-time RT-PCR, which is better suited to analyzing multiple samples.

Increasingly, the capabilities of real-time RT-PCR and microarray technologies are beginning to overlap. Several companies that offer high-density microarrays have recently introduced microarray products that are designed for the analysis of relatively small numbers of genes. In addition, several microarray manufacturers are designing products intended for high throughput analysis of multiple samples, such as “arrays of microarrays” in standard microtiter plate format. Likewise, advances in real-time PCR technology have been introduced to enable simultaneous analysis of larger numbers of genes or samples.

In the remainder of this chapter, we will review the most common methods used for real-time RT-PCR followed by a review of alternative microarray technologies. We will

then compare the qualitative and quantitative results obtained by these methods with one another.

## 2.2. REAL-TIME PCR

As PCR methodology matures, new technical developments refining its use continue to emerge. One of these is “real-time” PCR product quantification, made possible by the development of fluorescence-kinetic detection methods. In this technique, PCR product is measured as it accumulates during sequential amplification cycles. The strength of this method is that no post-PCR manipulation or product analysis is required since the quantitative measurement of the reaction product is known at the end of the cycling process. The term “reverse transcriptase-PCR” (RT-PCR) is reserved for quantification of gene expression at the level of mRNA. In real-time RT-PCR, a sample containing the mRNA target of interest is first reverse transcribed to cDNA that is subsequently amplified by real-time PCR.

More recently, a number of detection technologies, allowing both non-specific and specific target detection, have been developed for real-time PCR applications. Non-specific detection systems detect all double-stranded DNA generated during the PCR reaction. Specific detection systems distinguish specific target sequences from primer dimers or non-specific amplification products.

### 2.2.1. Detection Systems

*2.2.1.1. Non-Specific Detection Systems* DNA intercalating dyes, such as ethidium bromide [5] and SYBR Green<sup>®</sup>, have been used to detect PCR products non-specifically [6]. During the polymerization phase of a PCR reaction, these dyes bind to the newly synthesized double-stranded DNA, resulting in increased fluorescence emission that can be detected in real-time as PCR cycling progresses. The level of specificity is limited to that of the reaction primers since dye intercalation detects all double-stranded DNA and is not amplicon sequence specific. When intercalating dyes are used for allelic discrimination or to test PCR specificity, PCR products are assessed by running an end-point analysis of the DNA melting curve after thermal cycling concludes. PCR products of different lengths or different sequence will give distinct dissociation curves as they melt at different temperatures ( $T_m$ ). The temperature at which the double-stranded DNA becomes single-stranded is measured as fluorescence reduction when the intercalator dissociates from the melting double-stranded DNA. A single melting peak is expected in an optimized PCR reaction since it implies a single specific product and the absence of self-priming artifacts. Therefore, a SYBR Green assay is frequently employed to evaluate PCR performance of difference primer pairs, varying cycling conditions or to optimize primer concentrations in PCR reactions. Melting curve analysis on the Applied Biosystems 7900HT Sequence Detection System can typically resolve amplicon  $T_m$  values that differ by one or two degrees. The SYBR Green assay is also commonly used for SNP detection since short distinct sequences, differing by even a single nucleotide, melt at slightly different temperatures. Peak SYBR Green fluorescence measured over a series of PCR cycles can also be plotted to derive a quantitative PCR result similar to the original real-time PCR applications developed with ethidium bromide detection of dsDNA PCR products. However, sensitivity, specificity, accuracy, and precision

of this method do not compare to results that can be achieved using specific detection probes labeled with fluorescent dyes as described in the next section. Nonspecific detection systems work best for short amplicons since  $T_m$  resolution decreases as amplicon length increases.

**2.2.1.2. Specific Detection Systems** Specific PCR detection systems generally rely on using detection probes that are complementary to the target sequence amplicon generated by the forward and reverse PCR primers. The following sections describe some common probe chemistries including hydrolysis probes, Molecular Beacons, Scorpion<sup>TM</sup> probes and FRET hybridization probes.

**Hydrolysis Probes** Hydrolysis probes, often called TaqMan<sup>®</sup> probes, are probably the most widely used detection method for real-time PCR. Typical probe structure is shown in Figure 2.1A. These probes are oligonucleotides with a fluorescent reporter dye conjugated to the 5' end and a quencher dye conjugated to the 3' end that absorbs the light emission of the reporter dye conjugated to the 3' end. The reporter may be any one of a number of fluorescent molecules but most often is FAM<sup>TM</sup> (6-carboxyfluorescein), TET<sup>TM</sup> (tetrachloro-6-carboxyfluorescein), JOE<sup>TM</sup> (2,7-dimethoxy-4,5-dichloro-6-carboxyfluorescein), HEX<sup>TM</sup> (hexachloro-6-carboxyfluorescein), or VIC<sup>®</sup>, a proprietary dye from Applied Biosystems (Foster City, CA). The 3' quencher may be either a fluorescent dye with an emission

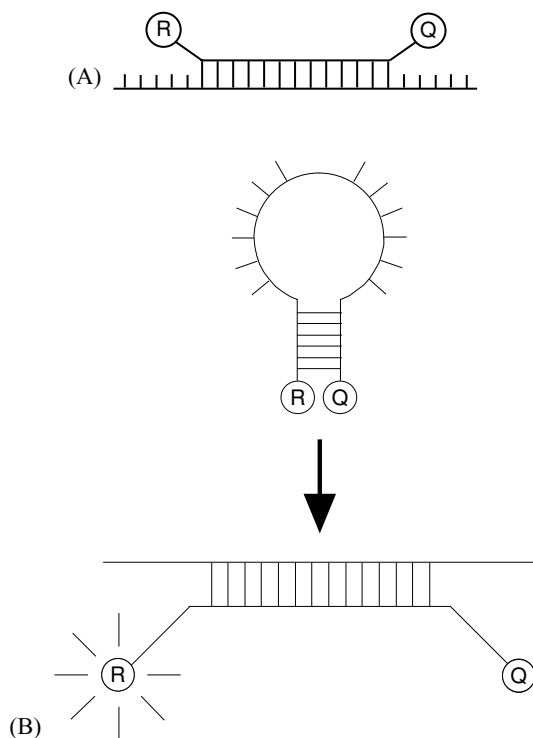


FIGURE 2.1. A. Hydrolysis probe. R, fluorescent reporter dye; Q, quencher dye. B. Molecular beacon shown in nonfluorescent stem-loop structure and as fluorescent hybrid with DNA target.

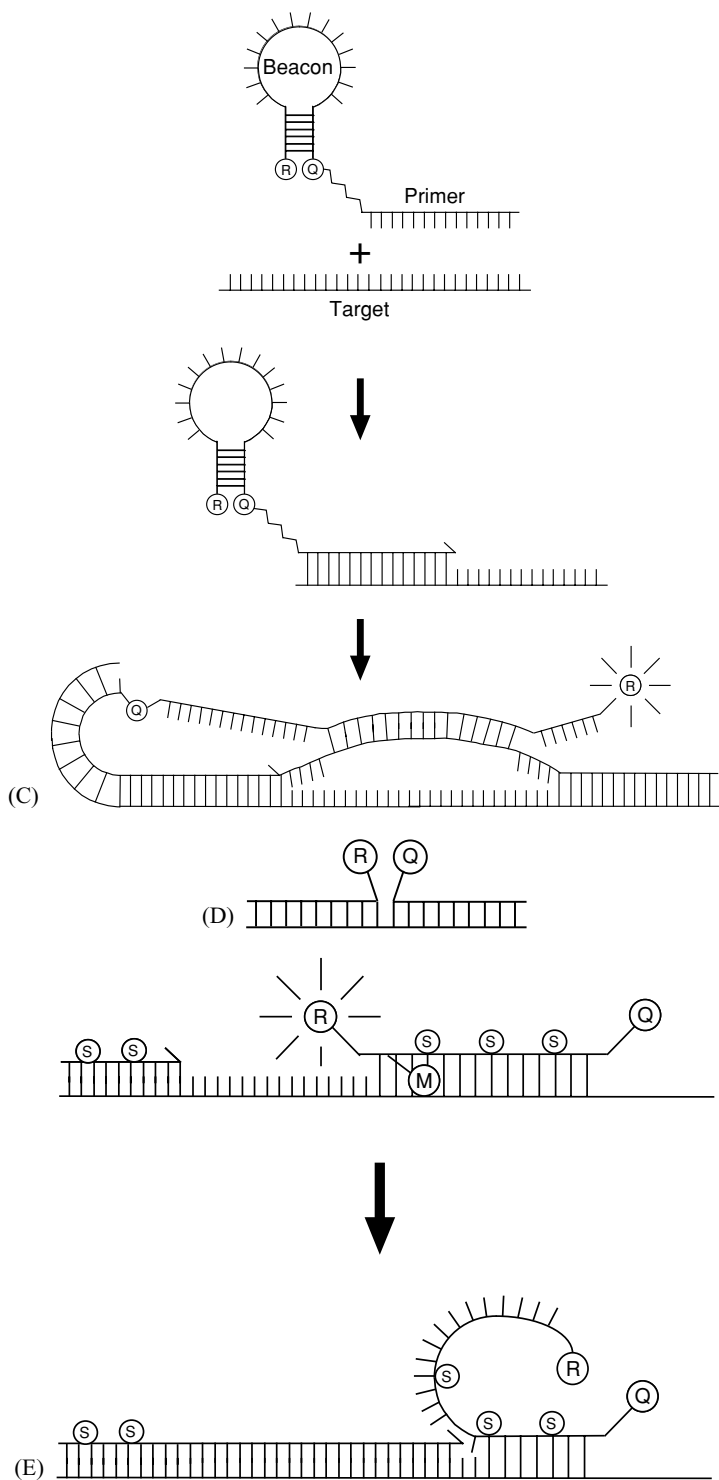


FIGURE 2.1. (Continued) C. Scorpion Primer/Probe; D. Hybridization probes; E. Qiagen QuantiProbe™. M, minor groove binder; S, Superbases™.

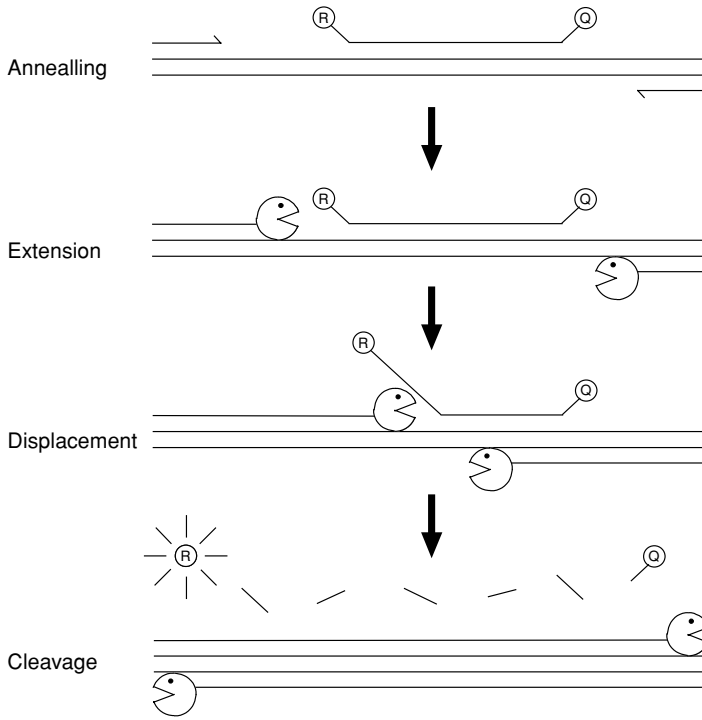


FIGURE 2.2. Mechanism of real-time PCR with an hydrolysis probe.

wavelength well separated from that of FAM and whose excitation wavelength overlaps the FAM emission wavelength, such as TAMRA<sup>TM</sup> (6-carboxytetramethylrhodamine), or a non-fluorescent (“dark”) quencher, such as Black Hole Quencher<sup>TM</sup> (Biosearch Technologies, Inc.) or DABCYL [7]. Dark quenchers absorb the light emission from the reporter dye and release it as energy other than fluorescence. Consequently, the dark quenchers tend to provide improved signal-to-noise ratios since they do not contribute to background signals as the fluorescent quenchers do. These probe constructs rely on a form of fluorescence resonance energy transfer (FRET) [8]. When hydrolysis probes are intact, the report dye and quencher dye are in close proximity. The quencher dye absorbs the fluorescence emitted by the reporter dye, which results in a non-fluorescent probe. During the PCR reaction, the probe anneals specifically to the target template when it is present as shown in Figure 2.2. The 5'-exonuclease activity of Taq DNA polymerase displaces and hydrolyzes the probe while it polymerizes a replicate of the template on which the probe is bound. The cleaved reporter dye becomes separated from its quencher resulting in fluorescence emission. Accumulation of the PCR product is detected by monitoring the increase in fluorescence emission from the cleaved reporter dye at each cycle and is directly proportional to the amount of target present in the sample as illustrated in Figure 2.3. A threshold fluorescence value is established within the exponential amplification stage of the reaction. The PCR cycle number at which the fluorescence emission is equal to the threshold value is termed the threshold cycle ( $C_T$ ). As shown in Figure 2.4, the  $C_T$  value is linearly correlated with the logarithm of the mass of RNA added to the initial reaction mixture. Therefore, TaqMan real-time RT-PCR can be

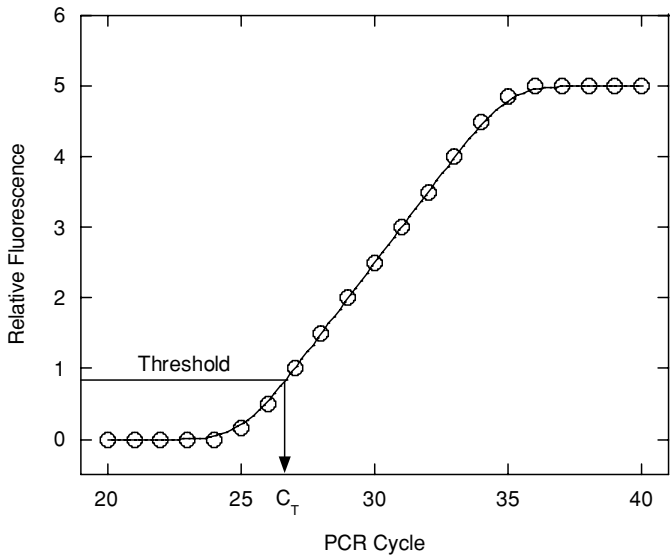


FIGURE 2.3. Idealized amplification plot. Symbols are defined in the text.

used as a quantitative assay for a specific RNA target in a complex background of other sequences. TaqMan hydrolysis probes have been widely used in both research applications, particularly genotyping and gene expression analysis and diagnostic applications, such as HIV, HCV and other RNA viral load teststesting.

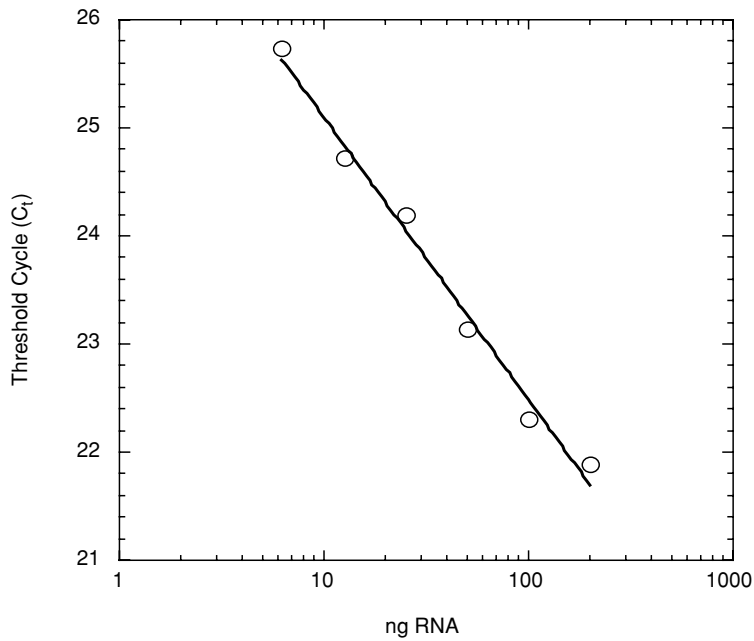


FIGURE 2.4. Calibration curve.

*Molecular Beacons* Molecular beacon probes [9] differ from the hydrolysis probes described above in both their structure and their mode of action. These oligonucleotide probes are designed to form stem-loop structures when they are free in solution as shown in the example in Figure 2.1B. The loop structure contains sequences complementary to the target template while the stem portion consists of a self-complementary stretch of approximately six bases (mainly C's and G's), which holds the probe in a hairpin like configuration. A fluorescent dye is linked to one end of the molecule and a quencher dye to the other end. In solution while the molecular beacon is in a hairpin structure, the stem holds the fluorophore and quencher in close proximity allowing efficient quenching of the fluorophore by FRET. During the PCR reaction, the molecular beacon encounters its complementary sequence in the amplicon and the probe sequence in the loop anneals to the complementary target sequence. A conformational change in the molecule results as the hybridization linearizes the probe, FRET no longer occurs and an increase in fluorescence emission is observed. The probe-target duplex is designed to be more thermodynamically stable than the hairpin structure. Unlike the hydrolysis probes, the increase in fluorescence with molecular beacons is reversible as the probe will dissociate from the target sequence and close back to the hairpin structure if the temperature is increased (for example during the denaturing cycle of PCR). Molecular beacon probes are particularly suitable for detecting point mutations or polymorphisms in target sequences because the probe-target duplex will not form at annealing temperature when a mismatch is present in the duplex.

*Scorpion<sup>TM</sup> Primer/Probes* Scorpions are fluorescent PCR primers with a sequence in the hairpin loop structure at the 5' end of the oligonucleotide that acts as a detection probe [10, 11]. In this real-time PCR detection system, the PCR reaction is carried out using one standard primer and a second Scorpion primer, which serves the dual functions of both a primer and reporter probe. An example of a Scorpion probe structure and function can be found in Figure 2.1C. The fundamental elements of Scorpion probes include (i) a PCR primer in series with, (ii) a PCR "stopper" which prevents PCR read-through of the hairpin loop probe structure, (iii) a specific detection probe sequence complementary to the target held in a hairpin configuration by complementary stem sequences and (iv) a fluorescence detection system consisting of a fluorophore and a quencher held in close proximity by the stem structure. After extension of the Scorpion primer during PCR amplification, the probe sequence located in the loop structure binds to the complementary sequence in the amplicon within the same strand of DNA. This results in increased fluorescence emission because the quencher is no longer in the vicinity of the fluorophore. Scorpions perform better under fast cycling conditions [11] than hydrolysis (TaqMan) probes and molecular beacons. This may result from the unimolecular hybridization mechanism, which is kinetically highly favorable and not dependent on enzyme hydrolysis activity. Scorpion detection technology has been used successfully in allelic discrimination [11] and in SNP genotyping [12].

*2.2.1.3. Hybridization Probes* This FRET-based real-time PCR reaction system requires two primers and two probes. The two hybridization probes are designed to bind to the target template adjacent to each other as shown in Figure 2.1D. One probe is labeled with a donor fluorophore at its 5' end and the second probe with an acceptor fluorophore at its 3' end. In the presence of target sequence, hybridization brings the two probes into close proximity allowing energy transfer via FRET from the donor fluorophore emission



to the acceptor fluorophore, which then emits the detected fluorescent signal. Fluorescence emission increases during sequential PCR cycles in proportion to the quantity of amplicon synthesized during the PCR reaction. This technology has been optimized and validated using the LightCycler<sup>TM</sup> instrument [13]. This detection strategy is particularly suitable for multiplexing since a single donor molecule can be used to excite multiple acceptor fluorophores.

More recently, another type of hybridization probe has been developed based on the MGB Eclipse<sup>TM</sup> (Epoch Biosciences) probe chemistry, namely the QuantiProbe<sup>TM</sup> assays (Qiagen). In this detection system, only one probe is required. QuantiProbe detection reagents are dually labeled oligonucleotides with a fluorophore at the 3' end of the probe and a non-fluorescent quencher at the 5' end (Dark Quencher<sup>TM</sup>, Epoch Biosciences). When these probes are not bound to a target sequence, they form a random structure in solution allowing the fluorophore and quencher to come into close enough proximity to prevent fluorescence. During the annealing step of the PCR reaction, these probes hybridize to the target sequence separating the fluorophore from the quencher and allowing fluorescence emission. An example of this type of probe construct is shown in Figure 2.1E. The amount of fluorescence measured in real time is proportional to the amount of target sequence. During the extension step in PCR, the bound probe is displaced from the target sequence allowing the fluorophore to be quenched during that phase of the reaction. QuantiProbes have both minor groove binder (MGB) and Superbase<sup>TM</sup> nucleotide modifications designed to stabilize their hybridization to target amplicons. The MGB moiety is a protein that associates with DNA by either hydrogen bonding or hydrophobic interactions along the duplex minor groove [14–17]. When the MGB moiety is attached to the 3' or 5' end of a DNA probe, it folds back into the minor groove formed by DNA hybridization and effectively increases the  $T_m$  of the probe. This chemistry allows shorter probes to be used which increases specificity, especially for polymorphism discrimination assays. Modified nucleotides known as Superbases are analogs of the corresponding natural bases [18]. Super A<sup>TM</sup> and Super T<sup>TM</sup> form stronger bonds with their natural complements in the target sequence providing further DNA duplex stabilization and allowing short, discriminating probes to be successfully used in these assay formats.

### 2.2.2. Real-Time RT-PCR Data Analysis

One of two methods is commonly used to quantitatively analyze data obtained by real-time PCR, either the standard curve method or the comparative threshold method.

**2.2.2.1. Standard Curve Method** This method requires using a standard template of known concentration to generate a standard curve of threshold cycle ( $C_T$ ) values relative to input target quantity as was illustrated in Figure 2.4. The quantity of RNA target in test samples is calculated by comparing the threshold cycle value for the sample run under the same conditions as the known standard and assigning the corresponding input copy number from the standard curve. There are several different approaches to making standards including synthesizing a single-stranded oligonucleotide of the amplicon sequence, making *in vitro* transcribed RNA, or purifying plasmid DNA with the target sequence inserted. When absolute quantitation of mRNA expression is required, for example when quantifying viral load or analyzing gene expression, an *in vitro* transcribed RNA standard is preferred to

control for variations that may be introduced by the reverse transcription reaction. This involves constructing a cDNA clone with an RNA polymerase promoter sequence that can serve as a template for *in vitro* transcribing copy RNA. One advantage of including a standard curve in each RT-PCR run is having an internal process control measurement of the PCR efficiency during that individual run.

**2.2.2.2. Comparative Threshold Method** The comparative threshold method for relative RT-PCR quantitation relates the fluorescence signal generated from a test sample to that of an internal control template. Examples where this is particularly useful would include study designs similar to those used in competitive hybridization on microarrays, such as treated and untreated control samples or the time zero control sample in a time-course study that must be compared to post treatment time point samples. Derivation and validation of the arithmetic formula ( $2^{-\Delta\Delta C_T}$ ) used for quantifying the relative change in gene expression using real-time PCR have been described [19]. This method is valid under the assumption that amplification efficiencies of the target gene and reference gene are approximately equal. The  $\Delta C_T = (C_T \text{ of the target gene} - C_T \text{ of the reference gene})$ , and the  $\Delta\Delta C_T = (\Delta C_T \text{ (sample)} - \Delta C_T \text{ (calibrator)})$ . This equation represents the normalized expression of the target gene in the test sample, relative to the normalized expression of the calibrator sample.

The reference gene is used as an internal control for normalization of the amount of RNA input for the RT-PCR reaction and the efficiency of each individual run. The reference can be single or multiple genes; common examples include  $\beta$ -actin, glyceraldehydes-3-phosphate dehydrogenase (GAPDH), or a ribosomal RNA sequence. Selecting a suitable housekeeping gene is crucial to ensure the reliability of the experimental results. It is important to verify that the expression level of the chosen reference gene(s) is not affected by the treatment or tissue type of the experimental design.

### 2.2.3. Qualification of Gene Panels Using Real-Time RT-PCR

Here we describe a study performed in our laboratory to evaluate the analytical performance of the RT-PCR assay for quantifying gene expression in RNA samples isolated from human tumors. In this study, TaqMan technology was employed for quantitative RT-PCR measured on the ABI Prism 7900HT SDS instrument (Applied Biosystems, Inc.). To obtain sufficient RNA for the study, a pooled sample was prepared by combining equal amounts of RNA extracts from 52 breast cancer tumors. The expression levels for sixteen target genes and five reference genes in the pooled RNA sample were assessed. The real-time RT-PCR assay was carried out for fifteen different RNA input levels, ranging from 0 to 8 ng of total RNA per reaction. Assay performance characteristics were measured as described in the next sections.

**2.2.3.1. Amplification Efficiency** As described previously, for reference normalization to be valid and for panels of genes to be comparable with one another for relative expression, the amplification efficiencies of the target gene(s) and reference gene(s) must be very similar. Estimates of amplification efficiencies were obtained for each of sixteen

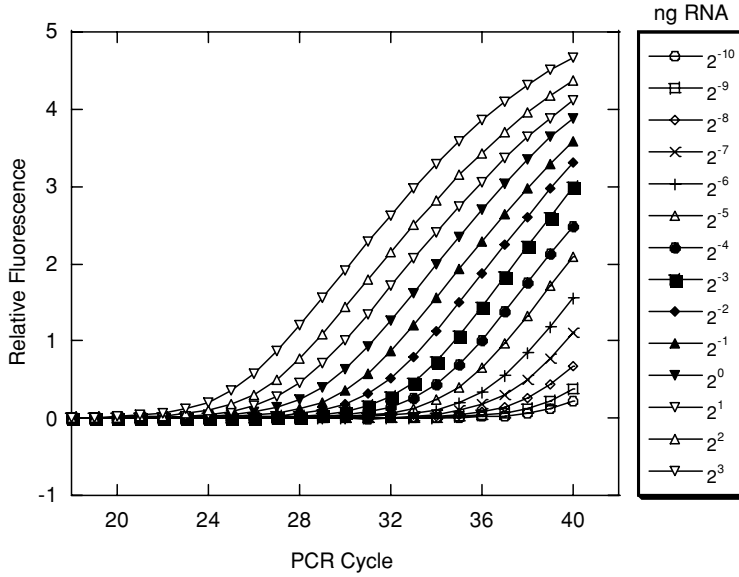


FIGURE 2.5. Amplification plots for serial dilutions of GAPDH target.

target genes and five reference genes according to the formula:

$$\text{Efficiency} = 2^{-1/\text{slope}} - 1$$

where slope is estimated from the regression of  $C_T$  measurements and  $\log_2$  RNA concentration.

The results showed that amplification efficiencies for the sixteen target genes was 96% while the average efficiency among the reference gene set was 88%. This indicates consistent amplification and strong agreement among the target and reference genes.

**2.2.3.2. Analytical Sensitivity and Dynamic Range** The linearity of the  $C_T$  value relative to RNA concentration was evaluated for each individual gene. Data indicated that for both target and reference genes  $C_T$  measurements were proportional to input RNA amounts over at least a 10  $\log_2$  range and generally were consistent over a range of nearly 14  $\log_2$ . A typical set of amplification curves from an input target dilution series is shown in Figure 2.5.

The accuracy and precision of the assay were estimated from calibration curves. Regression of calibration data (Figure 2.4 is an example) was used to calculate predictions of input RNA concentration based upon observed  $C_T$  measurements. Bias for each gene and RNA concentration level was then calculated as the percent difference between the predicted and input RNA concentrations.

$$\%Bias = \frac{\text{Predicted RNA Concentration} - \text{Input RNA Concentration}}{\text{Input RNA Concentration}} \times 100\%$$

Similarly, for each gene and input RNA concentration level, precision was estimated as the coefficient of variation (CV) in predicted RNA concentrations versus input RNA concentration level.

$$CV = \frac{s(\text{Predicted RNA Concentrations})}{\text{Mean}(\text{Predicted RNA Concentrations})} \times 100\%$$

Analytical sensitivity of the assay expressed in terms of the assay limit of detection (LOD) and limit of quantitation (LOQ), were calculated for each individual gene in the panel under standard assay conditions. Results were quite consistent across the full assay panel with the sixteen test genes averaging a LOD  $C_T$  of 40 and a LOQ  $C_T$  of 38.4 while the five reference genes gave an average LOD  $C_T$  of 39.6 and a LOQ  $C_T$  of 38.3.

The dynamic range of expression for each gene was defined by the maximum range of  $C_T$  measurements extending from the LOQ for the specific gene that maintained acceptable amplification efficiency, accuracy and precision. If acceptance criteria are set to have a mean bias within  $\pm 20\%$  and coefficient of variation less than 20%, the test genes in the panel could be reliably measured over a dynamic range of about 0.1 to 8 ng of input RNA while the reference genes had a range of 0.03 to 8 ng of input RNA. The dynamic range would increase for most of these genes by another 2–3  $\log_2$  if acceptance criteria for accuracy and precision were relaxed to mean bias within  $\pm 30\%$  and coefficient of variation less than 40%; a range often considered typical for microarray data.

**2.2.3.3. Reproducibility and Precision** Reproducibility for this RT-PCR assay was evaluated at 2 ng of input RNA to measure how variable the RT and PCR reactions were as different operators assembled and ran them on different instruments over time. Precision was assessed by estimating between-day, between-run (within day), and within-run variability as well as total variability over the test period. Reproducibility was assessed by estimating differences in mean  $C_T$ 's separately for each gene over the duration of the test period. Analyses were performed on both non-normalized and reference normalized  $C_T$  measurements. The combined largest variation ranged from 0.1 to 0.4  $C_T$  (average = 0.28  $C_T$ ) and 0.1 to 0.4  $C_T$  (average = 0.26  $C_T$ ) for the sixteen target genes and five reference genes, respectively. Overall, the differences in non-normalized  $C_T$  between instruments and between operators across the entire gene panel were small. The variability in the reference normalized  $C_T$  measurements were even smaller than the variability in the non-normalized  $C_T$  measurements. These results indicate that real-time RT-PCR analysis can be highly precise and reproducible.

#### 2.2.4. Real-Time RT-PCR Summary

Based on these analyses, it can be concluded that quantitative RT-PCR is a fast, sensitive and accurate assay with a broad quantitative range in total RNA samples. Although RT-PCR is a very powerful and precise technique, achieving performance levels such as those described here requires careful optimization. Factors that have to be taken into consideration for RNA analysis include preventing co-amplification from genomic DNA, evaluating and choosing the optimal primer and probe designs, and identifying appropriate reference gene(s) for normalization. Co-amplifying genomic DNA in RT-PCR assays can be avoided

in two ways. The first is to design primers and probes that generate amplicons spanning more than one exon. In this case, the additional length of an intervening intron separates PCR primers when they prime genomic DNA, which generally results in signal being detected only from the RNA target. Alternatively, a reliable “no RT” control assay may be developed for evaluating the contribution of residual genomic DNA in the RNA sample on RT-PCR performance and indicate when DNase treatment is required.

It is generally useful to evaluate more than one primer and probe set for each intended target to obtain one with optimal performance. Desirable characteristics include high amplification efficiency, absence of primer dimer or non-specific amplification products, high target sensitivity. Finally, if reference normalization is used for data analysis, it is crucial to confirm that the chosen reference gene(s) are stable in the experimental sample set. Once the factors discussed here are verified and all parameters are optimized, real-time RT-PCR is able to provide accurate, reproducible results while being capable of considerable sample throughput. The value of being able to fully characterize and optimize RT-PCR assay performance should not be underestimated in clinical applications.

## 2.3. MICROARRAYS

### 2.3.1. *Technology Platforms*

Microarray technology is based on the principles of nucleic acid hybridization. Since base pairing permits identification of complementary sequence within complex mixtures, single stranded labeled “probes” of known sequence were used to detect the presence of their complements in unknown samples. Adaptation of solution phase nucleic acid hybridization to solid supports was the direct precursor to present microarray technologies [20]. Early macroscopic hybridization methods used flexible, porous filters as the solid support to immobilize detection probes. The advent of solid supports such as glass and the miniaturization of immobilized nucleic acid features marked the arrival of the first microarrays. We will use the terminology established with the publication of a special supplement on microarrays [21] in which the immobilized strand of nucleic acid is referred to as the “probe” and the complementary solution phase strand is termed the “target”.

Microarrays composed of three different types of probes are in common use: cDNA (purified PCR products of cDNA clones), pre-synthesized oligonucleotides, and in situ synthesized oligonucleotides. In the sections that follow, the properties of each of these microarray types are presented separately.

**2.3.1.1. *cDNA Arrays*** cDNA arrays are typically made by “printing” PCR products generated from cloned cDNA libraries, on specially prepared glass microscope slides [22]. This process involves PCR amplification using primers modified to enable covalent coupling of the amplicons to the derivatized substrate surface. These PCR products are generally hundreds to thousands of bases long and are immobilized on individual array features that are typically in the range of 80–200  $\mu\text{m}$  in diameter. The number of features on an array can vary from under one hundred to more than 20,000. Usually, there is a one-to-one correspondence between features on the array and target molecules, i.e., each target sequence is complementary to only a single probe sequence on the array. The immobilized

probes are double-stranded when printed onto the array and are subsequently denatured; therefore, they can be used for hybridization with either sense or antisense targets. The most common strategy for preparing a sample for cDNA array analysis is to reverse transcribe mRNA in the presence of one or more fluorescently labeled deoxynucleotide triphosphates to produce a fluorescently labeled cDNA target.

Detailed methods for microscale printing of nucleic acid probes onto array substrates have been published elsewhere [22–24]. Alternatively, cDNA arrays can be purchased from a few commercial suppliers. cDNA arrays have proven difficult to quality control since PCR products are often heterogeneous, are subject to contamination by other probes and often are only inefficiently coupled to array substrates. Many, but not all, investigators have discontinued use of cDNA arrays in favor of oligonucleotide arrays, which generally have a more carefully and accurately characterized probe composition that perform more uniformly. Among the reasons cDNA arrays continue to be popular, especially with academic investigators is that the technology and materials for making them is easily accessible and affordable. In addition, since the investigator can easily customize cDNA arrays, they offer a degree of flexibility that is not easily achieved with commercial arrays, particularly for investigators studying organisms whose expressed genome is not represented on commercial arrays.

*2.3.1.2. Printed Oligonucleotide Microarrays* Oligonucleotide microarrays are available with probes either “printed” or synthesized “in situ” on the array substrate surface. The latter are synthesized directly on a glass surface using proprietary chemical methods discussed in the next section. Printed oligonucleotide microarrays are made using techniques similar to those used for cDNA microarrays. An important property of printed oligonucleotide arrays is that the nucleic acid probes are pre-synthesized using standard phosphoramidite chemistry including the addition of a reactive group to permit covalent coupling to the array surface. An advantage that synthetic oligonucleotide probes have over PCR products is they can be synthesized with great homogeneity and chemical purity. Consequently, printed oligonucleotide probes are generally more uniform in length and sequence composition as compared even with in-situ synthesized oligonucleotide probes. The size and density of individual features is similar to that of cDNA arrays: up to 40,000 probe features per array that are 80–200  $\mu\text{m}$  diameter. Probe sequences for oligonucleotide arrays are typically designed for optimal target specificity and uniform hybridization performance. Factors contributing to specificity and uniformity include the length of the probe (see below), selection of a sequence that does not have significant homology with other transcripts in the genome, similar probe length, and similar base composition. Therefore it should be sufficient to print a single feature on the array for hybridization to each target.

The length and sequence orientation of printed oligonucleotide arrays are determined by the manufacturer. Practical limits for oligonucleotide length are 25–80 bases. In theory, probes 30–80 nucleotides in length provide greater hybridization specificity than either cDNA [25], or shorter oligonucleotide probes without compromising sensitivity. The probes deposited on commercially available microarrays from Agilent and MWG Biotech are single-stranded oligonucleotides, 50–70 bases in length, in the sense orientation and are suitable for hybridization with antisense cDNA or aRNA (amplified RNA, defined below) targets. Recently, Affymetrix, MWG Biotech, Agilent, and Applied Biosystems have announced the release of whole “human genome” microarrays for simultaneous monitoring of 40,000–61,000 transcripts.

*2.3.1.3. In Situ Synthesized Oligonucleotide Microarrays* Affymetrix, Inc. pioneered the development of microarrays that contain oligonucleotide probes synthesized in situ on the array surface. Alternative methods for in situ synthesis of oligonucleotides were developed subsequently [26–29]; however, Affymetrix microarrays are the predominant platform in this category. Affymetrix GeneChip® microarrays are manufactured by a proprietary light-directed method [30]. The process is similar to the photolithographic method used for mass production of semiconductor chips. Oligonucleotide probes are synthesized in situ by a modification of the phosphoramidite method. The most recent generation of GeneChip microarrays have square features, 11 µm on each side, and the oligonucleotide probes are uniformly 25 bases in length. Twenty-two nucleic acid features are designated as a “probe set” for analysis of each target transcript to be detected. Eleven features in each probe set, called perfect match (or PM) features, are exact complements to the target sequence that hybridize to distinct, albeit sometimes overlapping, sequences of the target nucleic acid. Each PM feature has a paired mismatch (or MM) feature that is identical in sequence except for the middle nucleotide base which is substituted to cause a homonucleotide mismatch (A:A, C:C, G:G or T:T). Any signal generated from the MM feature is used in the computational algorithms as a measure of nonspecific hybridization to its paired PM feature. The recently released “Human Genome U133 Plus 2.0” array contains approximately 1,300,000 features in more than 54,000 probe sets representing 47,400 human transcripts.

In contrast to many cDNA and printed oligonucleotide arrays, specialized Affymetrix scanners are needed for analysis of GeneChip microarrays. The procedures for amplification and labeling RNA are similar to the methods used for printed oligonucleotide arrays but have been optimized for GeneChip arrays and scanners. The biggest difference in sample processing for Affymetrix arrays as compared with most printed oligonucleotide arrays is that only one sample is hybridized to each array; therefore, only one type of fluorophore (ie., one color) is needed. Dual sample hybridization, with differentially labeled samples, one serving as a reference and one as a test sample is a common approach used with printed arrays to compensate for variability in feature probe content and quality. The target amplification and labeling procedures for both types of experimental design are discussed below.

### *2.3.2. Target Amplification and Labeling*

There are a variety of different methods available for amplifying and labeling nucleic acids that are to be hybridized to different types of microarrays. In addition, different microarray platforms recommend a variety of different approaches to experimental design and data analysis. In the sections that follow, we discuss the two general experimental designs that use either one or two fluorophores for labeling nucleic acid targets.

*2.3.2.1. Single Fluorophore Experimental Designs* Affymetrix GeneChip microarrays were designed to be hybridized with an amplified mRNA (aRNA) target labeled with a single fluorophore. Labeled RNA offers two advantages as a hybridization sample. First, single stranded targets are more available for hybridization than denatured, double stranded targets such as PCR products, which tend to self anneal rather than hybridize with array probes. Secondly, RNA:DNA duplexes are more stable than DNA:DNA duplexes. Test samples to be compared with one another are labeled separately and hybridized to different

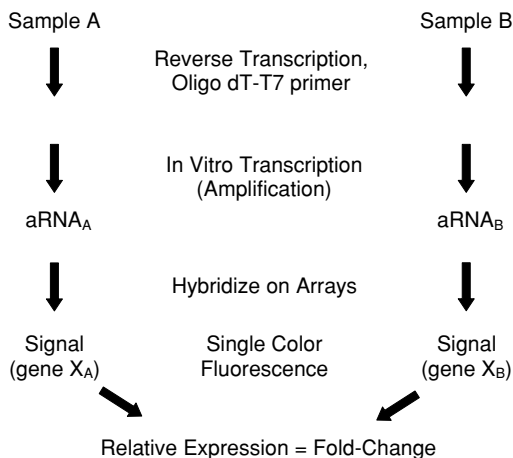


FIGURE 2.6. Single fluorophore experimental design.

microarrays as illustrated in Figure 2.6. The amplification procedure involves reverse transcription of purified RNA followed by in vitro transcription of cDNA using a strategy originally developed by Eberwine and coworkers [31,32]. Briefly, reverse transcription is primed using an oligo-dT primer tailed with a T7 RNA polymerase recognition sequence on the 3' end. A second cDNA strand is then synthesized by the method of Gubler and Hoffman [33]. In vitro transcription of the double stranded cDNA product with T7 RNA polymerase in the presence of biotinylated UTP results in the production of 100–1000 copies of amplified, biotin labeled RNA (aRNA) from each original mRNA target molecule. The biotinylated targets are hybridized to the array followed by detection of RNA: DNA duplexes with the fluorophore phycoerythrin conjugated to streptavidin and biotinylated goat anti-streptavidin antibody.

Affymetrix GeneChip microarrays require the use of a custom scanner that cannot scan other types of microarrays. Several types of controls are used in the GeneChip design to validate the various steps of a microarray experiment. Verification of microarray quality and image orientation is accomplished by hybridization of a biotinylated control oligonucleotide to complementary probes on the microarray. Other biotinylated oligonucleotide controls are added to the hybridization mixture to qualify the performance and sensitivity of each microarray. Synthetic mRNA controls constructed from bacterial genes are added to RNA samples to qualify the amplification and fluorescent labeling steps of the procedure.

The computational algorithms used for processing the raw fluorescence data is beyond the scope of this review but is described in detail elsewhere ([www.affymetrix.com](http://www.affymetrix.com)). In addition to the software provided by Affymetrix, alternative computational algorithms for analyzing data from these arrays have been developed by independent investigators [34–36]. The most recent version of the software provided by Affymetrix computes both qualitative and quantitative metrics of gene expression. The qualitative results classify each transcript as either present, marginal, or absent in an individual sample and will not be considered further here. The quantitative measure of gene expression, termed the signal value, is computed for each transcript by combining the fluorescent intensity data obtained for the 22 probes in each



probe set. Comparative analysis of the probe set data in different samples can also be done to calculate the relative expression or fold-change in expression of that gene. A variety of software packages are available from Affymetrix and other vendors for comparative analysis of more than two samples.

Comparison of samples hybridized on different microarrays requires consistency in the process for manufacturing the arrays as well as the use of the controls mentioned above for GeneChip microarrays to normalize for variation in the immobilized probes, target labeling reactions, and sample hybridization conditions. Similar controls are used with printed cDNA and oligonucleotide microarrays. Many of these sources of variation in printed microarray experiments are controlled by competitive hybridization of two samples that have been labeled with different fluorescent dyes (see below). However, as the methods for production of printed cDNA and oligonucleotide microarrays have improved it is also possible to use them with a single fluorophore experimental design.

*2.3.2.2. Two Fluorophore Experimental Designs* Most experimenters employ dual fluorophore experimental designs for use with cDNA and printed oligonucleotide microarrays. These designs involve competitive hybridization of two samples labeled separately with different fluorophores as illustrated in Figure 2.7. In this example, two different samples were amplified and labeled for co-hybridization to a cDNA microarray although the same strategy applies to printed oligonucleotide microarrays. Each sample is reverse transcribed and amplified independently following the “Eberwine” strategy discussed above for single fluorophore experiments. A variety of methods have been published for fluorescent labeling of nucleic acid targets [37]. In our laboratory, we have made labeled cDNA by reverse

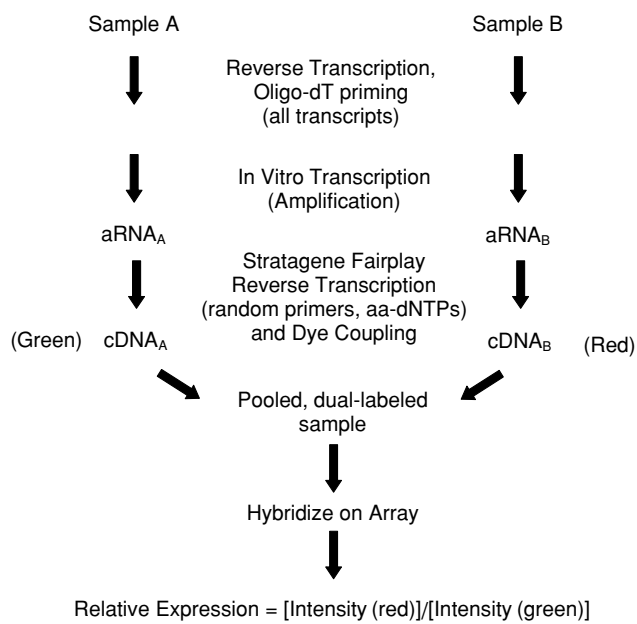


FIGURE 2.7. Two fluorophore experimental design.

transcribing amplified RNA in the presence of aminoallyl-nucleotide triphosphates using a commercially available kit (Stratagene, Inc.; FairPlay™ Microarray Labeling Kit) followed by reaction with N-hydroxysuccinimide esters of fluorescent dyes. Use of aminoallyl NTPs in the labeling reaction eliminates bias that can arise if samples are labeled using different fluorophore conjugated NTP's with different efficiencies of incorporation. In the example shown in Figure 2.2, the fluorophores are designated as green and red, the two colors that are commonly used for labeling microarray targets. The samples are mixed and hybridized to the microarray then the red and green fluorescence intensities are measured separately for each feature on the array. The primary measure of relative gene expression between the two samples is the ratio of fluorescence intensities at each array feature, sometimes referred to as the fold-change or fold-difference. A more useful data transformation is the base 2 logarithm of the intensity ratio, called the signal log ratio, which is symmetrical for increases and decreases of gene expression in one sample versus another [38].

As mentioned above, competitive hybridization of two samples on the same microarray controls for many potential sources of experimental variation. One drawback of dual fluorophore experimental designs; however, is that the results from multiple samples cannot be compared with one another unless they are all hybridized competitively against the same reference sample. To address this need, "universal standards" have been proposed and evaluated [39]. A commercially available mixture of RNA isolated from 10 human cell lines has been widely adopted for this purpose (Stratagene, Inc.).

### 2.3.3. Applications

Numerous studies have been published over the past five years in which microarrays have been used for gene expression profiling. The underlying theme of these studies is that expression data from multiple genes provides much more informative power than can be obtained from a single gene. A formidable challenge is the development of computational and statistical methods to analyze the large datasets generated by such studies. Consequently, the development of algorithms for "mining" data from microarray experiments has become a field of specialization [40]. Despite the continuing evolution of data mining algorithms, gene expression profiling has already revealed biological insights that have not been achieved by other methods. Two recent examples of such studies are discussed below.

In a landmark study, Golub and coworkers [41] used Affymetrix microarrays to profile the expression of 6817 genes in bone marrow samples from 38 patients with acute leukemia. The expression of approximately 1100 genes was found to correlate with the leukemia classification (based on a combination of morphological, histochemical, immunohistochemical, and cytogenetic analyses) as either acute lymphoblastic leukemia (ALL) or acute myeloid leukemia (AML). From this set of 1100 genes, a "class predictor" set of 50 informative genes was defined. The 50 gene class predictor was then tested on an independent group of 34 leukemia samples and was found to be 100% accurate in distinguishing between ALL and AML. Further "class discovery" analysis was able to further subclassify ALL into B-lineage ALL and T-lineage ALL. Although AML and ALL can be distinguished based on a combination of histological and cytogenetic criteria (see above), this was the first study to demonstrate that clinically relevant classifications were possible using only gene expression profiles.

Another study demonstrated that gene expression profiles could be used to predict future clinical outcome. Alizadeh and coworkers [42] designed a specialized cDNA microarray,

named the “lymphochip”, containing 17,856 clones derived from a B-cell library as well as other genes believed to be important in immunology or cancer. Lymphocyte samples were obtained at biopsy from patients with diffuse large B-cell lymphoma (DLBCL) prior to a regimen of standard multi-agent chemotherapy. Hierarchical clustering of the microarray data identified two new subtypes of DLBCL that correlated with long-term (8–10 year) patient survival, similar to the International Prognostic Indicator (an index that takes into account patient age, disease severity, disease location as well as other clinical parameters). Some patients defined as low risk based on the International Prognostic Indicator had gene expression subtypes that predicted substantially worse survival. Thus, gene expression profiling was able to predict poor prognosis in a subgroup that was not predicted by clinical criteria alone. Identification of these high-risk patients by gene expression profiling may, in the future, influence treatment decisions as well as patient selection in clinical trials of new therapeutics.

## 2.4. COMPARISON OF GENE EXPRESSION PROFILING METHODS

Validation of gene expression profiles generated using microarrays presents a challenge due to the high multiplicity of genes that are represented on a single array. It is not feasible to validate the results obtained using microarrays for thousands of genes by comparison with an independent method that measures expression only one gene at a time. Consequently, most investigators have compared their microarray results with data obtained using other technologies for only small numbers of genes.

Generalizations about the comparability of data obtained using different methods for gene expression profiling are complicated due to the use of different technology platforms and experimental design. It is generally accepted that RT-PCR is more sensitive, precise, and able to resolve smaller differences in gene expression than DNA microarrays. However, qualitative and quantitative comparisons of the same samples by different technology platforms have yielded varying results. A recent review concludes that the results of microarray experiments are only partially consistent with RT-PCR or Northern blot analyses in several laboratories [43]. The discussion below cites additional examples of comparisons between technology platforms as well as data generated in our laboratory.

### 2.4.1. *Comparison of cDNA Arrays with Other Gene Expression Profiling Methods*

In a study of the temporal response of fibroblasts to serum growth factors, Iyer and coworkers [44] were among the first to verify expression differences determined using microarrays by an independent method. Expression changes in five genes, determined using custom cDNA microarrays, were qualitatively very similar to expression profiles measured by real time RT-PCR. They found that fold-change values determined by real time RT-PCR were quantitatively larger than fold-change values determined by microarrays for four of the five genes; the fifth was a housekeeping gene whose expression was unchanged throughout the study period. The larger dynamic range of expression profiles generated by RT-PCR as compared with microarrays has been noted in subsequent studies. [45, 46] and may reflect the exponential amplification inherent in the former method as compared with the relatively compressed range of hybridization. In one study, expression changes measured by microarrays were biased towards underestimation of the changes measured by real time

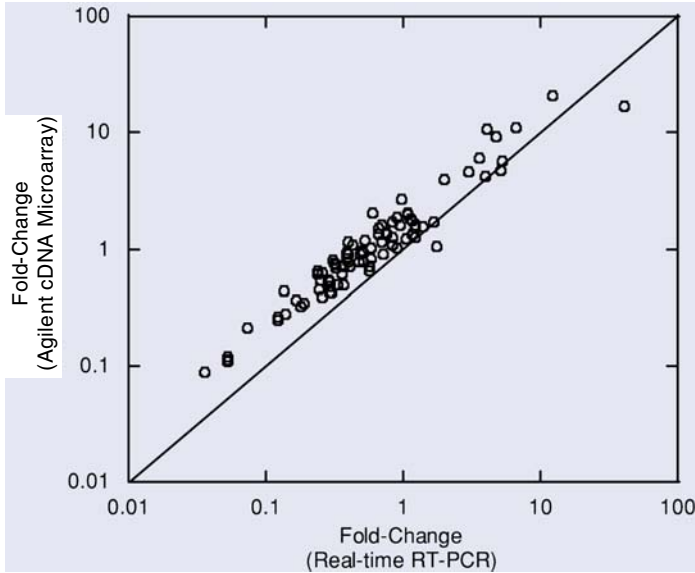


FIGURE 2.8. Comparison of gene expression profiles determined by cDNA microarrays and TaqMan RT-PCR.

RT-PCR, although the degree of bias for 47 individual transcripts was not predictable [46]. The Spearman rank correlation coefficient (our calculation from the supplemental data supplied by Yuen and coworkers) between expression fold-changes for all 47 genes measured by cDNA arrays and real time RT-PCR was 0.725.

Results obtained in our laboratory comparing cDNA microarray and TaqMan expression profiles were in general agreement with the studies cited above. The two samples we analyzed were different mixtures of RNA from various human tissues that we refer to as Human Reference RNA A and B. The relative expression ratio of 91 genes in the samples was determined by TaqMan RT-PCR and was compared with the expression ratio determined by Agilent cDNA microarrays as shown in Figure 2.8. Although the agreement for individual transcripts can vary quantitatively, an excellent correlation was observed between the rank orders of expression profiles determined by these two technology platforms (Spearman's  $R = 0.943$ ). Similar to the results of Yuen and coworkers, we observed a systematic bias in the magnitude of expression ratios measured by microarrays as compared with RT-PCR. Overall, our results provide further support for the qualitative validity of expression differences measured by cDNA microarrays.

#### 2.4.2. Comparison of Oligonucleotide Arrays with Other Gene Expression Profiling Methods

Published studies report varying degrees of correlation between expression profiles determined by oligonucleotide arrays and RT-PCR. Yuen and coworkers reported qualitatively similar expression profiles using Affymetrix GeneChip arrays and real time RT-PCR. However, as with cDNA microarrays, the magnitude of expression differences measured by

oligonucleotide microarrays tended to underestimate the expression differences measured by real time RT-PCR [46]. The Spearman rank correlation coefficient (our calculation from the supplemental data supplied by Yuen and coworkers) between expression fold-changes for all 47 genes measured by Affymetrix arrays and real time RT-PCR was 0.683. As noted above for cDNA arrays, the degree of bias varied for individual genes. In another study, Van den Boom and coworkers [47] studied the expression profiles of different grades of astrocytomas using Affymetrix microarrays and compared their results for 12 genes determined by real-time RT-PCR. Eight pairs of gliomas were profiled by both methods. Correlation coefficients calculated for each of the 12 genes were in the range of 0.48–0.98. In contrast, Baum and coworkers measured the expression profile of 56 genes induced by osmotic shock in yeast [29] using a different in situ synthesized oligonucleotide microarray. They observed a Spearman rank correlation coefficient of 0.972 when their microarray results were compared with the expression profile for the same samples determined by real-time RT-PCR.

In our laboratory, we compared the expression profiles of the Human Reference RNA samples mentioned above as determined using Affymetrix microarrays and TaqMan real-time RT-PCR. Prior to hybridization, the RNA samples were carried through two cycles of amplification before labeling the final aRNA product. The original RNA samples were evaluated by RT-PCR for 180 genes. Comparison of the expression differences determined by the two methods produced a Spearman rank correlation coefficient of 0.911 although disagreement between the methods was observed for several individual genes as seen in Figure 2.9. In addition, the fold-change measured by microarrays was biased towards underestimation of the fold-change measured by real time RT-PCR.

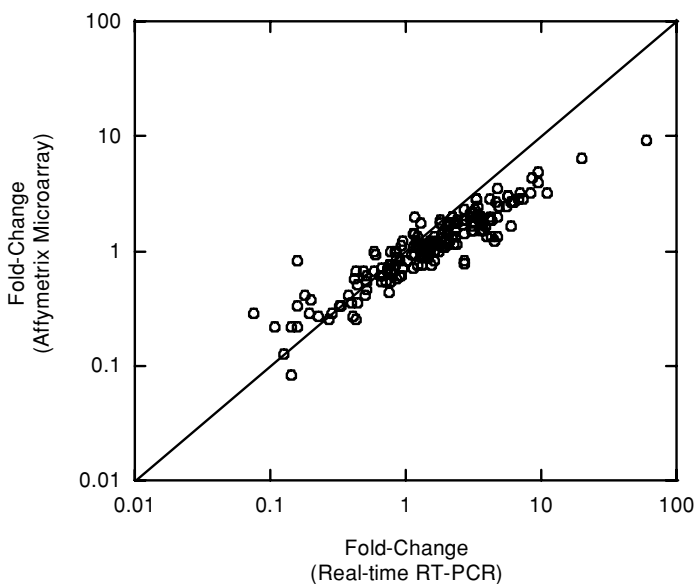


FIGURE 2.9. Comparison of gene expression profiles determined by Affymetrix oligonucleotide microarrays and TaqMan RT-PCR.

### 2.4.3. *Comparison of cDNA and Oligonucleotide Microarray Expression Profiles*

Recently, Lee and coworkers [48] proposed a method for validation of thousands of gene expression levels at a time by comparing results obtained with cDNA and oligonucleotide microarrays that are subject to different artifacts. In their study, several thousand transcripts were profiled in 60 human cancer cell lines (the NCI-60 panel) using both cDNA and Affymetrix GeneChip microarrays. The investigators posit that agreement of results obtained for a single transcript across many samples provides support for the validity of data obtained using both technology platforms. On the other hand, disagreement for an individual transcript does not indicate which (or if either) of the technology platforms generated a valid result. The correlation coefficients calculated for 2,344 Unigene-matched transcripts on the two microarray platforms were broadly distributed between  $-0.5$  and  $1.0$ . A consensus set of transcripts was identified that produced similar expression profiles on both cDNA and Affymetrix GeneChip microarrays. The observation by Lee and coworkers that expression profiles generated by cDNA and oligonucleotide microarrays are discordant for many transcripts supports the view that inaccuracies can arise from cross-hybridization, sequence variability of hybridization efficiency, as well as variability in the design, synthesis, manufacture of probes and target labeling. Other investigators have reached similar conclusions based on comparisons of expression profiles for the same samples generated on different microarray platforms [46, 49]. These studies demonstrate that one cannot compare expression differences identified using different microarray platforms without first cross-validating the methods for the specific genes of interest.

## 2.5. SUMMARY

The advent of technologies for expression profiling of multiple genes has launched a new era of biological research. Real time RT-PCR and DNA microarrays are among the most widely adopted methods employed in this new era. As originally developed, real-time RT-PCR and DNA microarrays were considered complementary technologies. Real-time RT-PCR is ideal for studies involving moderate numbers of genes (up to several hundred) in many biological specimens whereas DNA microarrays are better suited to analysis of many genes (tens of thousands) in fewer biological specimens. Given these characteristics DNA microarrays have more often been applied in the discovery phase of biological research with the aim of identifying the most informative genes. A relatively small number of genes, typically less than 50, can be identified whose differential expression is sufficient for the biological inquiry [50]. Once identified, the expression profile of this smaller set of informative genes can be screened with greater precision, better resolution, and more economically in a larger number of specimens by real-time RT-PCR. Concordance of results obtained from DNA microarray and real-time RT-PCR is critical if the former is used to identify smaller gene sets that will be screened subsequently by the latter technology. Existing data indicate that, although good overall correlations between technology platforms are possible, substantially different results can occur for individual genes. Thus, expression differences identified by microarrays must be verified if they are to be analyzed subsequently by another technology platform.

Microarray and real-time RT-PCR technology continues to evolve and improve. Increasingly, so-called “low density” microarrays targeting a small number of selected genes

are being adopted in formats suitable for analyzing large numbers of specimens. Conversely, improvements in RT-PCR technology and instrumentation have enabled simultaneous analysis of larger gene sets. Thus, we expect that both of these technologies will be used in the future for gene “discovery” as well as for quantitative analysis of gene expression profiles.

## ACKNOWLEDGEMENTS

John Morlan, Ken Hoyt, Debjani Dutta, Jennie Jeong, Ananthu Nguyen and Mylan Pho are thanked for providing microarray and TaqMan data. Chithra Sangli is thanked for assistance with statistical analysis and Joffre Baker is thanked for helpful discussions.

## REFERENCES

- [1] E.S. Lander et al. *Nature*, 409:860–921, 2001.
- [2] J.C. Venter et al. *Science*, 291:1304–1351, 2001.
- [3] L. Mancinelli, M. Cronin, W. Sadee. *AAPS. Pharm Sci.*, 2:E4, 2000.
- [4] A.D. Roses. *Nature*, 405:857–865, 2000.
- [5] R. Higuchi, C. Fockler, G. Dollinger, and R. Watson. *Biotechnology (N.Y.)*, 11:1026–1030, 1993.
- [6] T.B. Morrison, J.J. Weis, and C.T. Wittwer. *Biotechniques*, 24:954–8, 960, 962, 1998.
- [7] S. Nasarabadi, F. Milanovich, J. Richards, and P. Belgrader. *Biotechniques*, 27:1116–1118, 1999.
- [8] J. Ju, C. Ruan, C.W. Fuller, A.N. Glazer, and R.A. Mathies. *Proc. Natl. Acad. Sci. U.S.A.*, 92:4347–4351, 1995.
- [9] S. Tyagi and F.R. Kramer. *Nat. Biotechnol.*, 14:303–308, 1996.
- [10] D. Whitcombe, J. Theaker, S.P. Guy, T. Brown, and S. Little. *Nat. Biotechnol.*, 17:804–807, 1999.
- [11] N. Thelwell, S. Millington, A. Solinas, J. Booth, and T. Brown. *Nucleic Acids Res.*, 28:3752–3761, 2000.
- [12] A. Solinas et al. *Nucleic Acids Res.*, 29:E96, 2001.
- [13] B.E. Caplin, R.P. Rasmussen, P.S. Bernard, and C.T. Wittwer. *Biochemica*, 1:5–8, 1999.
- [14] E.A. Lukhtanov, I.V. Kutyavin, H.B. Gamper, R.B. Meyer, Jr. *Bioconjug. Chem.*, 6:418–426, 1995.
- [15] I. Afonina, I. Kutyavin, E. Lukhtanov, R. B. Meyer, and H. Gamper. *Proc. Natl. Acad. Sci. U.S.A.*, 93:3199–3204, 1996.
- [16] I.V. Kutyavin et al. *Nucleic Acids Res.*, 28:655–661, 2000.
- [17] I.A. Afonina, M.W. Reed, E. Lusby, I.G. Shishkina, and Y.S. Belousov. *Biotechniques*, 32:940–949, 2002.
- [18] E.A. Lukhtanov, M.A. Podyminogin, I.V. Kutyavin, R.B. Meyer, and H.B. Gamper. *Nucleic Acids Res.*, 24:683–687, 1996.
- [19] K.J. Livak and T.D. Schmittgen. *Methods*, 25:402–408, 2001.
- [20] E. Southern, K. Mir, and M. Shchepinov. *Nat. Genet.*, 21:5–9, 1999.
- [21] B. Phimister. *Nat. Genet.*, 21:1, 1999.
- [22] M. Schena, D. Shalon, R.W. Davis, and P.O. Brown. *Science*, 270:467–470, 1995.
- [23] M.B. Eisen and P.O. Brown. *Methods Enzymol.*, 303:179–205, 1999.
- [24] V.G. Cheung et al. *Nat. Genet.*, 21:15–19, 1999.
- [25] M.D. Kane et al. *Nucleic Acids Res.*, 28:4552–4557, 2000.
- [26] M. Beier and J.D. Hoheisel. *Nucleic Acids Res.*, 28:E11, 2000.
- [27] S. Singh-Gasson et al. *Nat. Biotechnol.*, 17:974–978, 1999.
- [28] T.R. Hughes et al. *Nat. Biotechnol.*, 19:342–347, 2001.
- [29] M. Baum et al. *Nucleic Acids Res.*, 31:e151, 2003.
- [30] S.P. Fodor et al. *Science*, 251:767–773, 1991.
- [31] R.N. Van Gelder et al. *Proc. Natl. Acad. Sci. U.S.A.*, 87:1663–1667, 1990.
- [32] J. Eberwine et al. *Proc. Natl. Acad. Sci. U.S.A.*, 89:3010–3014, 1992.
- [33] U. Gubler and B.J. Hoffman. *Gene*, 25:263–269, 1983.
- [34] S. Dudoit, R.C. Gentleman, and J. Quackenbush. *Biotechniques*, Suppl: 45–51, 2003.

- [35] R.A. Irizarry et al. *Nucleic Acids Res.*, 31:e15, 2003.
- [36] M.A. Zapala et al. *Genome Biol.*, 3:SOFTWARE0001, 2002.
- [37] A. Richter et al. *Biotechniques*, 33:620–8, 630, 2002.
- [38] J. Quackenbush, *Nat. Genet.*, 32 Suppl:496–501, 2002.
- [39] M.R. Weil, T. Macatee, and H.R. Garner. *Biotechniques*, 32:1310–1314, 2002.
- [40] G. Stolovitzky. *Curr. Opin. Struct. Biol.*, 13:370–376, 2003.
- [41] T.R. Golub et al. *Science*, 286:531–537, 1999.
- [42] A.A. Alizadeh et al. *Nature*, 403:503–511, 2000.
- [43] P.J. van der Spek, A. Kremer, L. Murry, and M.G. Walker, *Geno. Prot. and Bioinfo.*, 1:9–14, 2003.
- [44] V.R. Iyer et al. *Science*, 283:83–87, 1999.
- [45] M.S. Rajeevan, S.D. Vernon, N. Taysavang, and E.R. Unger, *J. Mol. Diagn.*, 3:26–31, 2001.
- [46] T. Yuen, E. Wurmbach, R.L. Pfeiffer, B.J. Ebersole, and S.C. Sealfon, *Nucleic Acids Res.*, 30:e48, 2002.
- [47] B.J. van den et al. *Am. J. Pathol.*, 163:1033–1043, 2003.
- [48] J.K. Lee et al. *Genome Biol.*, 4: R82, 2003.
- [49] W.P. Kuo, T.K. Jenssen, A.J. Butte, L. Ohno-Machado, and I.S. Kohane, *Bioinformatics.*, 18:405–412, 2002.
- [50] C.H. Chung, P.S. Bernard, and C.M. Perou. *Nat. Genet.*, 32 Suppl:533–540, 2002.



BioMEMS and Biomedical Nanotechnology

Volume II: Micro/Nano Technologies for Genomics and  
Proteomics

Editor-in-chief: Ferrari, M. - Ozkan, M.; Heller, M. (Eds.)

2007, XXIV, 540 p., Hardcover

ISBN: 978-0-387-25564-4