TaqMan® Gene Expression Assays for Validating Hits From Fluorescent Microarrays

The most accurate, precise, and cost-effective method for validating microarray results.

Introduction

Gene expression microarrays are used by nearly every major target discovery laboratory to screen for sets of genes involved in various biological processes, whether they are working with human samples or with model organisms such as mouse, rat, and *Drosophila*. Genes identified as being up-regulated or down-regulated on a DNA microarray become "hits" for further study in the biology of interest. Further experimentation, often involving a larger sample set, is required to validate that up- or down-regulation of the identified genes is a direct consequence of the experimental conditions (i.e., a disease state) or if the regulation of that particular gene is instead related to some other cellular processes (i.e., not directly linked to the disease state). Thus, candidate targets are carried through to the next phase of the process, and further studies are usually done using other technologies that are more focused on just that particular candidate gene set of interest.

Screening for Targets Using DNA Microarrays

Microarrays became available in the mid-1990s, providing a huge leap forward in the number of genes that could be analyzed, in parallel, in a given sample. Their tremendous capability for global, parallel gene expression monitoring makes them an ideal starting point for target discovery, enabling a quick assessment of what differentiates the transcriptional profiles of different tissues or cell types.[1,2]

Microarrays are based on the same principles as Northern blots, differential hybridization, and other hybridization-based techniques.[3] With microarrays, DNA fragments are spotted or printed to a specific address (location) on a solid surface such as a glass slide, or nylon membrane. When fluorescently-labeled cDNA (from an mRNA template) hybridizes to the DNA probe on the solid surface, the location of the DNA probe on the microarray surface emits a fluorescent signal. The intensity of the signal on the fluorescent array is relatively proportionate to the amount of fluorescently-labeled cDNA bound to the probe. Thousands or even tens of thousands of oligonucleotides or PCR products can be placed in a precise grid on one microarray, allowing the simultaneous determination of the presence or absence of a very large number of potential gene transcripts in one sample relative to a control sample.[4] Commercial, high-density oligonucleotide microarrays typically contain 1,000s to 10,000s of targets. The lengths of the probes on the arrays vary from manufacturer to manufacturer. Common probe lengths are 25 nucleotides, 45 nucleotides, or 60 nucleotides. The length of the probe, as well as the sequence selected for each probe, greatly affects the probe performance (i.e., its ability to uniquely detect a specific transcript). While longer probes (longer sequences) enable a greater degree of differentiation between transcript sequences and typically result in higher signal, they also have the potential to stringently hybridize to targets containing more mismatches than shorter probes, thus potentially resulting in a higher incidence of false positive signals. Furthermore, sophisticated informatics tools are needed to select the appropriate probe sequence to enable detection of single genes. Because of these limitations, array manufacturers typically array multiple spots of varying probe sequence for each gene, some as many as 20 spots. The numerous data points per gene, along with the large number of genes per chip (some more than 20,000), make data analysis with statistical software a critical step in the evaluation of array results.

Microarray experiments generally require 200 µL of a cDNA sample, representing at least 10^6 cells. This heavy sample requirement leads most microarray users to "preamplify" their RNA. The sample is generally amplified after the reverse transcription (RNA to cDNA) step, and the cDNA is converted to labeled cRNA or aRNA (amplified RNA). Often, the labeling step is incorporated with the preamplification step. In a typical experiment, two RNA samples, each labeled with different fluorophores, are hybridized simultaneously to a microarray.

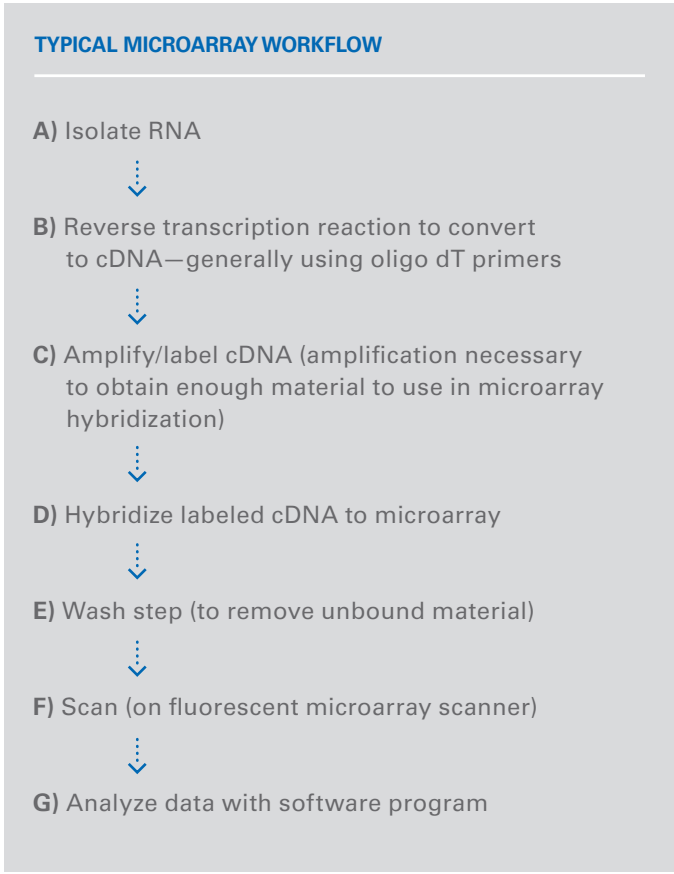

Figure 1. This outline highlights the many steps involved in a typical microarray experiment.

This competitive hybridization allows the direct comparison of the relative gene expression in the two RNA samples within each array[5] (Figure 1).

Validation of Microarray Results

Researchers validate hits from microarrays for two reasons: 1) to verify that the observed changes are reproducible in a larger number of samples, and 2) to verify that array results are not the result of problems inherent to the array technology. Although microarrays are an excellent tool for initial target discovery, there is a broadly recognized variability in microarray results from lab to lab, user to user, platform to platform, and every combination thereof. Additionally, the relatively limited dynamic range of fluorescent microarrays (10^2-10^3) places limits on the technology with respect to sensitivity and specificity. It is thus essential to use independent means to verify that the genes of interest are truly differentially expressed, and to what extent.[1, 6, 7, 8, 9]

Microarray Variability

A number of factors, including array-to-array variability, statistical analysis of microarray data, inherent false positive and false negative rates arising from laboratory processes, and general limits of fluorescent microarray technology often lead to results that fail validation by downstream techniques.

Array Variability

Replica experiments have shown that there can be a large degree of variability, or high coefficient of variation (CV), between identically designed but physically different fluorescent arrays (i.e., two arrays with the same probe patterns hybridized with the same sample). The solid surfaces on which arrays are built are one source of potential problems. For example, the surfaces themselves can emit fluorescence that interferes with the accurate reading of the fluorescent probes.

Another issue, which most frequently arises with *in situ* arrays (where probes are synthesized on the array [*in situ*] using an adaptation of technology developed by the electronics industry for etching integrated circuits), is related to the coupling efficiency during the oligonucleotide synthesis. For example, low coupling efficiency can lead to a spot containing a high percentage of incomplete oligonucleotides, increasing the false positive rate.

Data Analysis

The large number of genes analyzed on a microarray statistically leads to random fluctuations, producing anomalous results.[2] The large quantity of data generated from a single experiment requires statistical analysis to determine which transcripts are producing positive signals. However, different software applications, or even the same software application programmed with different parameters, will produce different sets of hits from a single microarray experiment. Finally, because there are so many genes being assessed on one microarray, it is typical for multiple members of gene families to be represented on the array. Cross-hybridization of probes to these highly-homologous, but biologically different genes can produce false positive results for some of these genes.

General Laboratory Processes

General laboratory processes contribute to the need to validate results observed with fluorescent microarrays. Artifacts can be introduced in both the RNA extraction steps, preamplification of RNA, and the fluorescent labeling reactions. Hybridization of fluorescently-labeled cDNA to microarrays is, like many experiments, a technique-dependent process, meaning that different researchers may carry out the experiments differently, leading to varying results.[10] Simply put, good science always requires validation of results.

Microarray Technology Limitations

It is important to understand the limits of any technology and interpret results within those limits. Otherwise, inaccurate conclusions can be drawn. When researchers use fluorescent microarrays, they are making a tradeoff between the number of targets evaluated in parallel and the ability to detect a signal from a low abundance transcript(s).

Microarrays are an excellent tool for screening gene expression levels across many transcripts in a given sample. However, the relatively narrow dynamic range ($<10^3$) limits the

ability to accurately quantify gene expression levels from a fluorescent microarray.

> "As a scientist, you are going to have to replicate your results; and it is always better if you do it with a different approach."
>
> – Raymond Samaha, Senior Manager
> Genomic Applications, Applied Biosystems

Additionally, the limited sensitivity of microarrays translates into an inability to detect low expressors when samples contain less than 10^6 cells.[2, 11] Depending on the level of sensitivity that researchers need for their particular experiments, and the level of accuracy they need in determining the differences in gene expression, they may want to use a secondary, more sensitive method to validate that small differences are real, and, in fact, are actually small.

Most high-density oligonucleotide arrays use probes that are biased towards the 3′-end of mRNA transcripts. The reason for this is that many of the labeling methodologies used for signal detection on microarrays incorporate a reverse transcription step that uses an oligo dT primer that binds to the poly-A tail of mRNAs for the first-strand cDNA synthesis. Reverse transcriptase is an enzyme with poor processivity, and tends to produce cDNA of only 1,000–1,500 bases in length. This 3′ bias can limit the number of alternatively spliced transcripts that a high-density oligonucleotide hybridization array can discriminate.

TaqMan® Reagents for Real-Time PCR Technology Advantages

Quantitative, real-time PCR was commercialized in 1995. This method capitalizes on the fact that there is a quantitative relationship between the amount of starting target sample and the amount of PCR product at any given PCR cycle number. The Applied Biosystems real-time quantitative PCR solution, using our Real-Time PCR Systems and 5′ nuclease chemistry with TaqMan® probes is considered the gold standard technology for quantitative gene expression. This chemistry exploits the exonuclease activity of AmpliTaq Gold® DNA polymerase by using a cleavable fluorescent probe in combination with forward and reverse PCR primers. The 5′ nuclease assay is sensitive (ability to detect 1 copy in 10–100 cells), reproducible (reliably low CVs), and has a very large dynamic range (10^6). Because of these characteristics, quantitative real-time PCR (qPCR) with TaqMan® probes is an ideal technology for validating hits from microarrays.

5′ Nuclease Chemistry (Using TaqMan® Probes)

TaqMan® probes contain a reporter dye (6-FAM™ dye) linked to the 5′ end of the probe, and a non-fluorescent quencher (NFQ) at the 3′ end of the probe. When the probe is intact, the proximity of the reporter dye to the quencher results in suppression of the reporter fluorescence, primarily due to Förster-type energy transfer (Förster, 1948:Lakowicz, 1983). During PCR, the TaqMan® probe anneals specifically to a complementary sequence between the forward and reverse primer sites.

Only probes that are hybridized to the complementary target are cleaved by the 5′ exonuclease activity of AmpliTaq Gold® DNA polymerase. Cleavage separates the reporter dye from the NFQ, resulting in an increase in reporter dye fluorescence at each PCR cycle. This increase in fluorescent signal occurs only if the target sequence is complementary to the probe and is amplified during PCR (Figure 2).

The amount of fluorescence produced from the TaqMan® probe is measured at each amplification cycle, providing a look at the "real-time" changes in the amplification product as the PCR process unfolds.[12] Identification of the PCR cycle when the exponential growth phase is first detectable (C_T) provides extremely accurate quantitation of gene expression in the starting samples. In fact, real-time RT-PCR with TaqMan® probes is acknowledged to be the most sensitive, reliable method of quantifying gene expression.[1, 6, 8, 9]

Historically, a microarray experiment would generate more hits than could easily be validated by real-time PCR. The

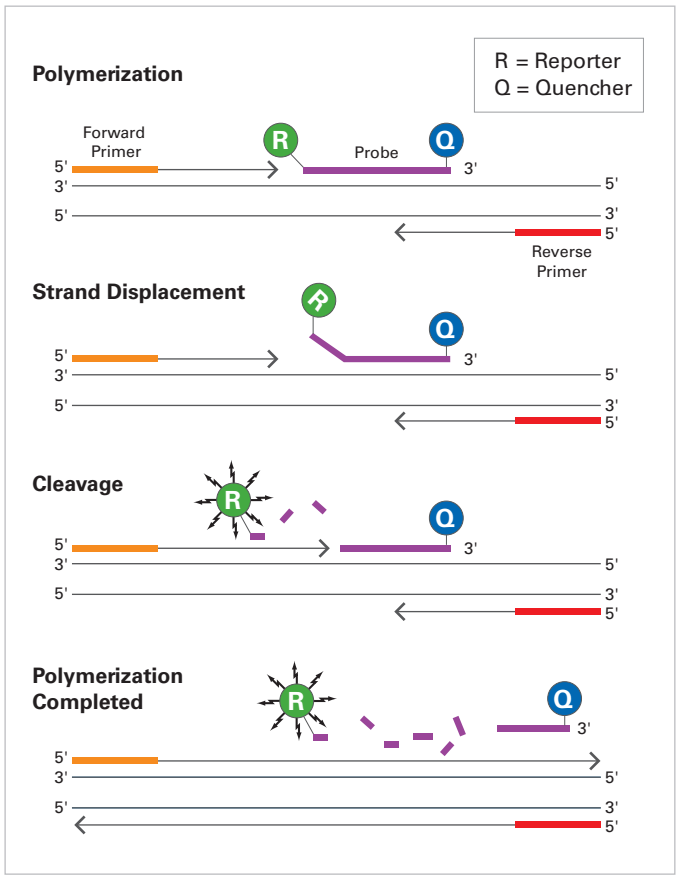

Figure 2. This figure details the 5′ nuclease reaction showing the mechanism of probe cleavage resulting in a fluorescent signal.

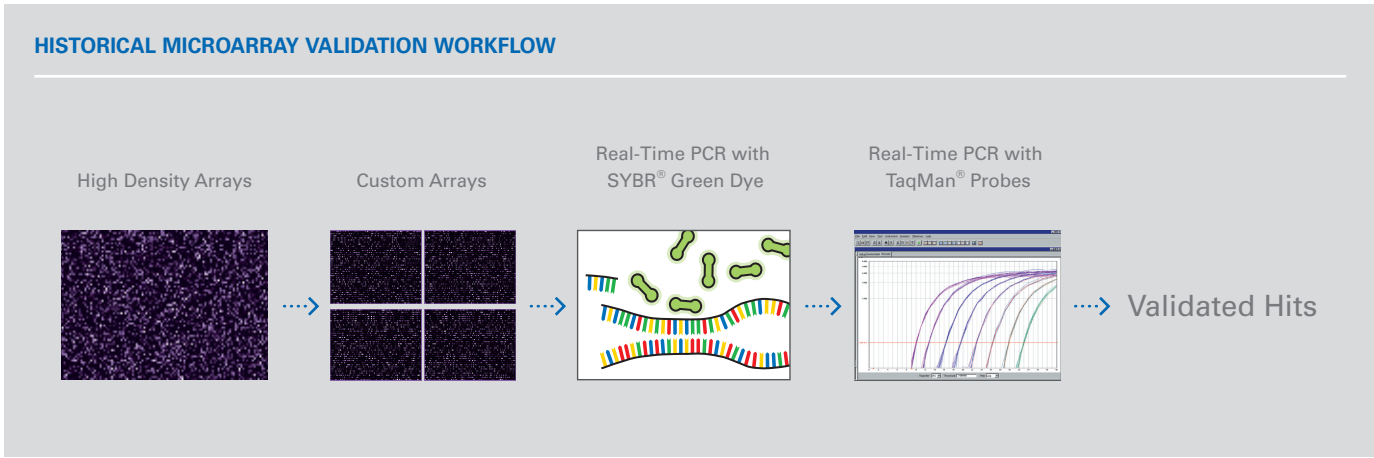

Figure 3. In the past, low accuracy, low sensitivity methods were used to screen the high number of "hits" from a microarray experiment before determining absolute quantitation of expression levels.

limiting factors in using real-time PCR with TaqMan® probes previously had been the labor involved in designing the probe and primer sequences for each target of interest, as well as the cost of the fluorescently-labeled probe. In order to reduce the number of targets to a more manageable number, semi-quantitative protocols were often implemented for relative gene expression measurement before proceeding to an absolute gene expression measurement of the few targets of interest. Absolute quantitation was usually done with real-time PCR[2, 11] (Figure 3).

TaqMan® Gene Expression Assays for Array Validation

TaqMan® Gene Expression Assays are a comprehensive set of over 680,000 pre-designed TaqMan® probe and primer sets (based on 5′ nuclease chemistry) for human, mouse, rat, *Arabidopsis*, *Droshophila*, and *C. elegans* genes. At less than the cost of a probe synthesis, these assays are cost-effective and eliminate the labor barriers historically associated with TaqMan® probe-based 5′ nuclease chemistry. The removal of these barriers allows researchers to realize the benefits of gold standard gene expression results much earlier in their target screening workflow, eliminating the need for less specific, more laborious methods and greatly reducing the time to results (Figure 4). Now, researchers can perform their validation within a few days of ordering. Given the breadth of assays available, "it suddenly makes studies possible that were once too impractical," states a university researcher.

TaqMan® Gene Expression Assays are designed differently than microarrays and have no positional bias within a transcript. Each probe and primer set is designed via the Applied Biosystems Genome Aided Assay Design Pipeline. This sophisticated bioinformatics pipeline leverages information from both public and private genome databases to design gene-specifc TaqMan® probe and primer sets for quantitative gene expression. The comprehensive sequence analysis system masks SNPs, repeats, and areas of sequence discrepancy between the public and the private databases to identify suitable locations along mRNA transcripts for assay design. Where possible, assays span exon junctions, eliminating the possibility of detecting genomic DNA, which may be present in your cDNA sample. Once assays have been designed at each suitable transcript location, extensive *in silico* quality control (QC) is carried out to ensure that the assay chosen for manufacturing is gene specific. The *in silico* QC process includes BLASTing each assay design against transcript and genome databases. This process insures that the assays do not detect highly homologous genes, in addition to the target gene.

The assays are optimized for two-step RT-PCR reactions using random oligonucleotide primers for first strand cDNA synthesis. This random-priming protocol eliminates 3′ bias of the generated cDNA and is therefore more representative of the mRNA used as starting material for the sample being analyzed.

All TaqMan® Gene Expression Assays are designed to run under universal thermal cycling conditions, and are formulated into a single 20X tube (250 nM probe, and 900 nM each primer). This format requires fewer set-up and pipetting steps to assemble your reactions. Simply mix the TaqMan® Assay with your cDNA sample and TaqMan® Universal PCR Master Mix. All assays are optimized for use on all Applied Biosystems Real-Time PCR Systems.

Assay tubes are shipped in a 1D-barcoded 96-position rack designed to accommodate standard liquid handling robotics and fit seamlessly into automated, high-throughput laboratory processes. Each order of assays includes a compact disc with an assay information file (AIF) containing not only the assay IDs and location on the plates, but also all of the content associated with each assay (transcripts detected, associated gene symbol, protein classification, etc.). The AIF can be easily uploaded into LIMS systems or into Applied Biosystems Sequence Detection System software.

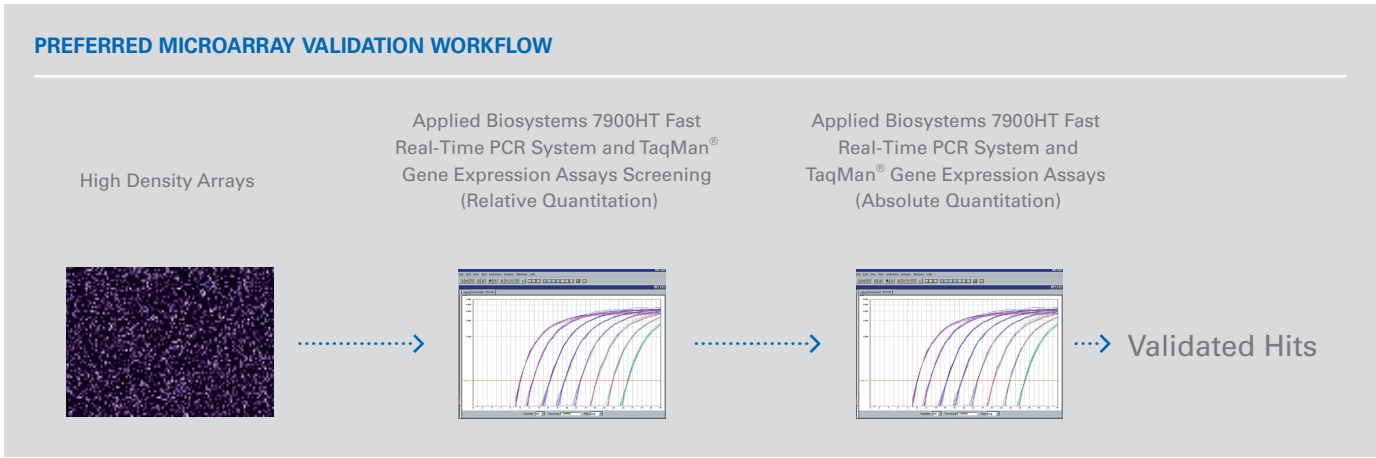

Figure 4. Using TaqMan® Gene Expression Assays in Relative Quantification mode on the 7900HT system enables the screening of a larger number of samples, using highly accurate and specific chemistry.

Applied Biosystems TaqMan Gene Expression Assays are an attractive method for validation of microarray hits because probes and primers are designed to discriminate between highly-homologous gene family members. Additionally, the online assay catalog provides several options for locating the assay of interest. Notably, the "Batch Search" mode can be used to upload tab-delimited files of search terms (for instance, the GenBank accession numbers associated with microarray probe hits), to find all of the assays associated with the terms in the file. Because many of these assays are inventoried (off-the-shelf), customers can receive them within a few days, greatly speeding up the validation of microarray experiments.

TaqMan Gene Expression Assay Formats

TaqMan Gene Expression Assays are available in a single-tube format for human, mouse, rat, and a number of other model organisms. Custom TaqMan® Gene Expression Assays are also available, designed specifically for your target sequence for any organism for which Inventoried or Made-to-Order TaqMan Gene Expression Assays have yet to be made. Simply submit a target sequence and receive a custom designed 5' nuclease assay in a single-tube format that is QC-verified and ready to run under the same conditions as our TaqMan Gene Expression Assay products. For gene expression applications, Custom TaqMan Gene Expression Assay save you time and effort, enabling you to focus on obtaining results, not designing and testing assays.

TaqMan Low Density Array

When a microarray experiment generate tens to hundreds of interesting "hits", running individual real-time PCR reactions becomes a labor-intensive task that is prone to numerous errors. Additionally, validating all of the hits with just a few different samples in triplicate quickly becomes impractical, as it requires the use of automated robotics and multiple 96- or 384-well plates. For more throughput, TaqMan Gene Expression Assays are available pre-loaded onto the TaqMan Low Density Array—a 384-well microfluidic card that enables 1 to 8 samples to be run in parallel across 12 to 384 targets without the need for liquid handling robotics. Choose from over 47,000 Inventoried TaqMan Gene Expression Assays to customize your TaqMan Array, and run on the Applied Biosystems 7900HT Fast Real-Time PCR System. Gold standard TaqMan Assay design and chemistry make TaqMan Low Density Arrays the ideal solution for increased sample throughput, and improved results over microarrays.

For information on all TaqMan Gene Expression Assays products, visit **www.allgenes.com**

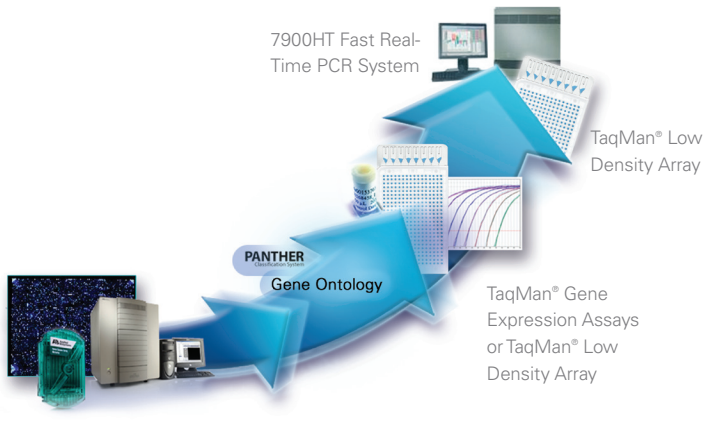

Figure 5. TaqMan® Low Density Arrays provide a reproducible and sensitive method for validating microarrays.

References

1. Reidhaar-Olson, J.F. and Hammer, J. 2001, April. "The impact of genomics tools on target discovery." *Current Drug Discovery*, 20–24.
2. Gu, C.C., et al. 2002. "Role of gene expression microarray analysis in finding complex disease genes." *Genetic Epidemiology*, 23:37–56.
3. Southern, E.M. 1975. "Detection of specific sequences among DNA fragments separated by gel electrophoresis." *Journal of Molecular Biology*, 98:503.
4. Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. 1995, October 20. "Quantitative monitoring of gene expression patterns with a complementary DNA microarray." *Science*, 270(5235):368–369, 371.
5. Yuen, T., et al. 2002. "Accuracy and calibration of commercial oligonucleotide and custom cDNA arrays." *Nucleic Acids Research*, 30(10):e48.
6. Simon, R., Radmacher, M.D., and Dobbin, K. 2002. "Design of studies using microarrays." *Genetic Epidemiology*, 23:21–36.
7. Huang, Y., et al. 2001, December 18. "Gene expression in papillary thyroid carcinoma reveals highly consistent profiles." *Proceedings of the National Academy of Sciences*, 98(26):15044–15049.
8. Rajeevan, M.S., et al. 2001. "Use of real-time quantitative PCR to validate the results of cDNA array and differential display PCR technologies." *Methods*, 25:443–451.
9. Haslett, J.N., et al. 2002. "Gene expression comparison of biopsies from Duchenne muscular dystrophy (DMD) and normal skeletal muscle." *Proceedings of the National Academy of Sciences*, USA, 99:15000–15005.
10. Petricoin, E.F., III, et al. 2002. "Medical applications of microarray technologies: a regulatory science perspective." *Nature Genetics*, 32(supp.):474–479.
11. Nature Genetics. 1998. "Getting hip to the chip." *Nature Genetics* [editorial], 18(3):195. http://www.nature.com/ng/microarray/0398.html, cited June 21, 2004.
12. Cunningham, B. 2001. "Assessing differential gene expression." *The Scientist*, 15(23):27.

Printed in the USA. 08/2006 Publication 127WP01-02

Applied Biosystems

Headquarters
850 Lincoln Centre Drive | Foster City, CA 94404 USA
Phone 650.638.5800 | Toll Free 800.345.5224
www.appliedbiosystems.com

International Sales
For our office locations please call the division headquarters or refer to our Web site at
www.appliedbiosystems.com/about/offices.cfm