⟨1128⟩ NUCLEIC ACID-BASED TECHNIQUES—MICROARRAY

INTRODUCTION

Microarrays are microscopic spots of DNA (measured in micrometers) arranged in an ordered manner (columns and rows) on a planar surface so that each DNA spot can be uniquely identified to facilitate an accurate analysis of the data. The DNA spots, also called array elements, are specific DNA molecules of known or unknown sequences and can be of similar or different nucleotide lengths. Samples of these mixtures are placed in fixed locations on the microarray.

Unlike conventional probes, which are a specific DNA or RNA sequence labeled with radioactive, fluorescent, or chemiluminescent tags (see *Nucleic Acid-Based Techniques—General* ⟨1125⟩*, Glossary*), the array elements are referred to as probes when the sequence information of the array elements is known, despite not being labeled. In this context, the target refers to labeled nucleic acids in solutions that are hybridized to the array elements or probes. The purpose of a microarray experiment is to identify the sequence of these labeled nucleic acids and/or determine their content. Compendial applications at this time are limited but may increase with wider use of microarrays in diagnostics and in drug discovery, development, registration, and control applications. When used for compendial purposes, standard assay development and validation approaches with availability of suitable reference materials are likely to apply.

Microarrays can range from hundreds to thousands of array elements (low density), tens to hundreds of thousands of array elements (high density), to millions of array elements (very high density). In addition to the use of planar surfaces for microarrays, the array elements can also be immobilized on individual support particles, such as beads. In these cases the array elements are identified by the particles themselves rather than specific locations on an array. The advantages of using microscopic spots on the array include high density, fast hybridization kinetics, and low sample volumes. Microarrays greatly speed up the acquisition of data, and in some cases increase the predictive power of results, by comparison with conventional nucleic acid-based assays. This is achieved by miniaturization, multiplexing, and parallel execution of nucleic acid-based tests that traditionally are performed in tubes, plates, or capillaries as described in general chapter ⟨1125⟩ (see also *Nucleic Acid-Based Techniques—Extraction, Detection, and Sequencing* ⟨1126⟩, *Nucleic Acid-Based Techniques—Amplification* ⟨1127⟩, *Nucleic Acid-Based Techniques—Genotyping* ⟨1129⟩, and *Nucleic Acid-Based Techniques—Approaches for Detecting Trace Nucleic Acids (Residual DNA Testing)* ⟨1130⟩).

The principle of microarray analysis is the specific binding of the target DNA molecules to the probes or array elements. The ordered array of rows and columns of spots allows highly automated detection and analysis. DNA microarrays are manufactured, processed, detected, and analyzed in a number of different ways and have many applications. With the aid of computers, laboratory automation, and high-resolution detection devices, microarrays produce large amounts of data and are the analytical tool of choice to unravel the molecular complexity of DNA or expressed RNA.

The basic principles of nucleic acid amplification technologies (NAT) and definitions of the various techniques are described in chapter ⟨1127⟩. The present chapter covers the general field of microarrays, but detailed treatment of various application-specific microarrays, including data analysis and validation, are excluded from this chapter at this time. The following sections address the major applications of microarrays, sample processing, labeling, workflow, detection, and analysis of data. Several of these sections, for example, sample preparation and labeling, overlap with chapters ⟨1126⟩ and ⟨1127⟩, and cross references are made accordingly. Finally, regulatory aspects of microarrays will be discussed.

GENERAL PRINCIPLES OF MICROARRAY EXPERIMENTS

Types and Applications

Microarrays are most widely used in three types of analysis: gene expression, microarray-based comparative genomic hybridization (or array comparative genome hybridization, aCGH), and single nucleotide polymorphism (SNP). In brief, gene expression microarrays generally measure messenger RNA in a cell; aCGH analyzes DNA copy number variations, chromosomal additions, and deletions in genomic DNA; and SNP microarrays are used in genotyping to analyze single nucleotide polymorphisms (see ⟨1129⟩). Within each type of microarray, various platforms, both manual and with various levels of automation, are available. *Table 1* summarizes the three major types and most common applications, as well as the target for each application, the probe, and the complementary nucleic acid techniques (see ⟨1126⟩, ⟨1127⟩, and ⟨1129⟩).

Table 1. Major Types and Applications of Microarrays

Types	Application	Target	Probe	Complementary Technology
Gene Expression	Gene Expression	mRNA	Oligonucleotide/cDNA	qRT/PCR, Northern Blotting
aCGH	aCGH CNV	DNA	Oligonucleotide/ cDNA/Pac, Yac, Bac	Cytogenetic chromosome analysis
SNP	SNP Genotyping	DNA	Oligonucleotides	Sequencing
	SNP	Amplicons	Oligonucleotide	Sequencing
	SNP	Oligonucleotide	Amplicon	Sequencing

GENE EXPRESSION MICROARRAYS

Gene expression microarrays are used to measure the relative level at which a certain gene is expressed. They are a powerful tool for target gene discovery, molecular tumor characterization, diagnosis, classification, treatment, and monitoring of diseases. Underlying molecular subgroups that are active in diseases have been identified by observing distinct and recurring gene expression subsets found within diseased tissues. Gene expression microarrays are also used to measure changes in gene expression over a given period of time, e.g., within various stages of a cell cycle or by identification of gene mutation(s) that lead to cancerous growth. Another application for gene expression microarrays is the development of new drugs, e.g. by measuring the down-regulation of a gene associated with a particular disease to monitor the effectiveness of a new drug. When the expression levels from a set of genes are measured, the term gene expression signature (biomarker or classifier) is used. Other examples of biomarkers or classifiers are drug activity classifiers that are used to diagnose the mechanism of action of a drug or toxicity classifiers that are used to diagnose and develop dosage parameters for a patient.

aCGH MICROARRAYS

In contrast to gene expression microarrays, aCGH microarrays target segments of DNA rather than individual genes (this is, similar to chromosomal banding and traditional comparative genomic hybridization). In an aCGH microarray, the array elements, which are large pieces of genomic DNA or specially designed oligonucleotides, are used to identify a known chromosomal location or changes. The primary advantage of aCGH is the ability to detect DNA copy changes at multiple loci in a single assay and to do so at a much greater resolution compared to traditional CGH. Depending on their design, aCGH microarrays provide distinct advantages over conventional cytogenetic analysis such as karyotyping and fluorescence in situ hybridization (FISH) because they have the potential to detect the majority of microscopic and submicroscopic chromosomal abnormalities. Compared to aCGH, these conventional cytogenetic techniques have low throughput, are labor-intensive, and often require specially trained staff to perform tests in a consistent manner. aCGH microarrays are also useful for the detection of cancer by monitoring the loci of oncogenes and tumor suppressor genes.

SNP MICROARRAYS

SNP microarrays identify the presence of known sequence polymorphisms by analysis of the pattern of hybridization to a series of probes that are specifically complementary either to wild-type or mutant sequences. If the SNP or set of SNPs associated with a particular disease are known, SNP microarrays can be used to identify a disease in an individual. SNP microarrays provide an efficient and inexpensive tool for simultaneously studying multiple genetic variations in multiple samples.

Design of Microarrays

The following sections discuss the design of the three types of microarrays described above and the suitability of the materials used for the microarray probes for each of the three types.

GENE EXPRESSION MICROARRAY

These microarrays are the most common type of microarray in use today. The array elements consist of either cDNA derived from mRNA of known genes but of unknown sequence, or oligonucleotides for which detailed sequence information is available. Oligonucleotides are preferred array elements because of the affordable cost of synthesis and the large amount of sequence information now available for specific genes or gene fragments. These can be arrayed in specific patterns to enable accurate analysis of related gene sequences and gene families in a single hybridization assay. The following general principles apply to oligonucleotide design for gene expression microarrays:

1. Oligonucleotides should be 25–70 mers.
2. Oligonucleotides should include appropriate controls (i.e., oligonucleotides corresponding to sequences from a different organism).
3. All oligonucleotides should map to within 1000 nucleotides of the 3′ end of cDNAs and should correspond to the coding strand.
4. Sequence repeats, stretches of polyA, G, C, and T and extremes of T_ms should be avoided.
5. Oligonucleotides should be compared to sequences in existing databases to avoid cross-reactivity (less than 70% sequence identity with nontarget sequences is preferable).

In addition to oligonucleotides, PCR amplicons and double-stranded DNA (dsDNA) are also used as probes. However, the PCR amplicons require purification to remove enzymes, salts, nucleotides, and other contaminants from the amplification process that could interfere with the binding of the probes and could also inhibit hybridization. In addition, the preparation of dsDNA probes for spotting is labor intensive and expensive. Moreover, dsDNA probes can have repetitive sequences that compromise hybridization specificity. When sequence information is unavailable, dsDNA remain the probes of choice because unknown dsDNA probes can still be used to study gene expression.

aCGH MICROARRAYS

These microarrays traditionally use bacterial artificial chromosomes (BACs) of 100–200 kilo-base pairs per DNA segment as the array elements. However, the large-scale DNA isolations or PCR amplifications of such large-insert clones are elaborate and time consuming. As is the case in expression profiling applications, aCGH microarrays have transitioned from dsDNA targets to oligonucleotide targets. Oligonucleotide libraries or ready-made microarrays can now be purchased, saving considerable time and effort.

Depending on the application, SNP microarrays can use both amplicons and oligonucleotides as probes. In one of the most common formats to detect mutations in a gene sequence, the probe is that of a single gene in which the sequence differs by a single nucleotide polymorphism from the sequence of the other probes for that gene in the same microarray. For the discrimination of only one mismatch, short oligonucleotide probes (15–30 bp) maximize the destabilization caused by mispairing and are therefore used for the detection of SNPs.

Manufacturing of Microarrays

Microarray elements are deposited onto a solid support, the most widely used of which is glass. Microarray manufacturing can be divided into two main categories, direct synthesis of the probes on the microarray (in situ) or synthesis of the probes before spotting on the microarray (ex situ). In situ synthesis is generally used for higher density microarrays but is limited to nucleotides of approximately 25–100 bases. With increasing nucleotide length, the likelihood of truncated products increases because of the limited stability of building oligonucleotides in situ. In contrast, ex situ microarray manufacturing can put any premade material into a microarray format, including oligonucleotides, PCR products (amplicons), complementary DNA (cDNAs), and BACs.

The main techniques for in situ synthesis are photolithography, maskless lithography, and ink jetting. Microarrays are generally manufactured commercially, although for a small number of low-density microarrays, the end user can manufacture the microarrays using a low-throughput microarray manufacturing robotic instrument (a personal microarrayer). However, only maskless lithography and ink jetting are available for end user manufacturing. In photolithography, a glass substrate containing a photomask, which is chemically prepared so that particular nucleotides bind to specific positions, is used to synthesize the oligonucleotides on the substrate. The masks predetermine which of the nucleotides are activated when flooded with one of the four types of nucleotides. The process is repeated until the required number of bases is synthesized. The manufacture of these microarrays uses computer algorithms and multiple spots to cover the gene of interest. Maskless lithography uses a digital micromirror device that uses a solid-state array of miniature aluminum mirrors to create virtual masks that replace the physical photomasks. A computer controls the desired pattern of UV light via individual mirrors. Each digital micromirror in turn controls the pattern of UV light projected onto the glass in the reaction chamber, which is coupled to a DNA synthesizer. The UV light selectively cleaves a UV-labile protecting group at the precise location where the next nucleotide will be coupled. The patterns are coordinated with the DNA synthesis chemistry in a parallel, combinatorial manner so that hundreds of thousands of unique oligonucleotides can be synthesized in a single microarray. Ink jetting is accomplished by building up the nucleotides, base-by-base, in repetitive print layers using standard phosphoramidite chemistry. Inkjet heads similar to those used in commercial inkjet printers are connected to bottles that contain the four different phosphoramidite nucleotides that make up the building blocks of in situ nucleic acid synthesis. The advantages of inkjetting and maskless lithography are flexibility in design and the ability to make small batches of arrays quickly.

The two main types of ex situ manufacturing techniques are microspotting pins (contact printing) and piezoelectric printing (noncontact). The technology excels at printing multiple probes many times over numerous surfaces with one small-volume loading of probe. Spot size and delivery volume are controlled by the size of the end of the tip, and many tip sizes are available. A piezoelectric printing mechanism uses a small dielectric crystal in contact with a glass capillary that holds the sample fluid. Application of the voltage results in ejection of fluid from the tip, resulting in drop volumes from hundreds of picoliters to several microliters.

General Experimental Considerations

Regardless of the type and application, all microarray experiments have a similar workflow: amplification step, labeling, hybridization, and wash steps, followed by scanning, quantitation, and reporting. The experimental design determines the type of microarray used, number of spots required, and the specific sets of nucleic acids on the microarray. The experimental design also influences the platform used, such as the number of spots, surface type, nucleic acid type, throughput, resolution, and number of colors that can be detected in a single assay. Platforms can be open (support is available from multiple vendors) or closed (support from a single vendor). In general, experimental designs that require a high density of spots and quantitation are more difficult and expensive to implement than qualitative assays.

Microarray Sample Considerations

Sample extraction, isolation, and preparation should be carefully chosen in order not to alter the ability of the resulting target to hybridize to the microarray. In general, sample preparation issues are the same for microarrays as for other laboratory techniques such as qPCR (quantitative PCR) and sequencing (described in chapters ⟨1126⟩ and ⟨1127⟩). RNA, cDNA, genomic DNA, and PCR products are some of the sample types analyzed with microarrays. In some genotyping applications, specific alleles are used both as array elements and targets.

As with any nucleic acid technique, the quality of the nucleic acid is critical for the microarray experiment. The nucleic acid should be pure, intact, and accurately quantitated before use ⟨1126⟩. In particular, the presence of contaminating DNA in total RNA samples may cause problems in microarray analysis because some labeling methods label both RNA and DNA with equal efficiency. For some applications in which even trace contaminants with either RNA or DNA may interfere, pretreatment with DNase or RNase may be necessary. For example, contaminating, labeled DNA can hybridize with microarray targets leading to high-level hybridization signals that are not derived from RNA transcripts, thus resulting in an inaccurate estimation of the target RNA concentration because both nucleic acid species are quantitated at the same wavelength.

A major consideration in any microarray experiment is the availability of adequate amounts of sample nucleic acid for analysis. For example, sample from laser-capture microdissection, needle tissue biopsies, or other small clinical samples do not yield

sufficient RNA (for expression microarrays) or DNA (for aCGH microarrays) and must be amplified before analysis. It is critical that the amplification procedures for amplification of mRNA be so designed that the final mixture of amplicons accurately reflect the distribution of mRNA species in the sample. Uniform amplification of genomic DNA for aCGH microarrays can be achieved by the use of multiple displacement amplification (MDA), which overcomes the nonuniform amplification of genomic DNA that occurs in PCR-based amplification methods that use degenerate oligonucleotide primed PCR (DOP-PCR). For SNP arrays where specific alleles are the target of interest, nonuniform amplification is not an issue, and samples can be amplified (and labeled) by PCR, multiplex PCR, and WGA (see ⟨1127⟩).

Microarray Labeling

The targets for a microarray are a population of nucleic acids that are extracted from a sample and are appropriately labeled. Many methods can be used for labeling targets (see ⟨1127⟩), but fluorescent labeling is the most widely used because it offers high sensitivity and a superior dynamic range. An added advantage is the ability to detect two or more signals in a single experiment. The method of labeling depends on the microarray type. The two methods used to fluorescently label targets for gene expression microarrays, direct and indirect labeling, have been described in ⟨1127⟩. In general, the second method (indirect labeling), in which the label is added via a linker, requires less starting material and is less expensive. Published reports have shown that this method yields results similar to those obtained from directly labeled samples. In microarray aCGH, a patient's DNA and reference DNA (300–1000 ng) are typically fluorescently labeled with red and green fluorescent dyes, respectively, often using a random priming protocol. Random prime labeling uses a high concentration of Klenow enzyme whereby genomic DNA is digested with restriction enzymes and hybridized with random primers. The primers are extended by the 5′–3′ polymerase activity of Klenow, resulting in a strand displacement activity with the direct incorporation of labeled nucleotides. SNP microarrays using oligonucleotides as array elements are labeled using fluorescently labeled nucleotides in both single and multiplexed PCR reactions, followed by a purification step to remove unincorporated dyes. Where amplicons are used as array element, labeled oligonucleotide probes are synthesized using phosphoramidite chemistry.

Hybridization and Wash

Hybridization should be carried out under conditions that minimize annealing of noncomplementary fragments. The wash steps following a hybridization reaction are optimized to provide the highest possible specificity, signal-to-noise ratio, and reproducibility (see ⟨1126⟩). Before hybridization, double-stranded probes and targets should be denatured, and nonspecific sites should be blocked. Microarray surface chemistries are designed to capture all nucleic acids with high efficiency, so the free-binding groups on the surface must be blocked or inactivated to prevent nonspecific binding of labeled material that could compromise the signal-to-noise ratio. Surfaces are blocked and washed with various aqueous-based buffers that typically include salts, detergents, and blocking agents such as low molecular weight, hydrolyzed proteins. The purpose of the posthybridization washes is to remove all unattached and nonspecifically bound label from the surface and probes. In general, both automated and manual washes are done in saline sodium citrate/sodium dodecyl sulfate (SSC/SDS) buffers of various concentrations and at different elevated temperatures depending on the stringency required. After the final wash step, microarrays using fluorescent targets are dried immediately by centrifugation or in a nitrogen stream. Hybridized microarrays must be stored in the dark and should be scanned as soon as possible. Some fluorescent dyes used in microarray analysis are subject to degradation by environmental ozone, and in these cases ozone levels in the experimental environment must be less than 5 parts per billion. Specialized ozone-free hoods are made to protect microarray dyes.

Microarray Detection

Regardless of the microarray type, each spot on a microarray represents a unique probe sequence to which a single, labeled target is bound, and this specific binding allows detection and quantitation of the target. This is achieved by the emission of light (photons) at a particular wavelength by the fluorescently labeled duplexes when the microarray is exposed to light of specific wavelength from an excitation source. The emitted fluorescent light is converted to electrical energy by a detector. The detector is either a photomultiplier tube (PMT) or a charge-coupled device (CCD) with specially designed optical paths that collect the raw data from microarrays (scanning). The detector filters and optical paths are designed to detect specific fluorescent dyes at sufficient resolution while eliminating crosstalk when two or more dyes are used on a single microarray. The resulting signal is proportional to the number of photons emitted by the microarray. These signals are used to create a digitized image showing the presence and quantitation of specific targets.

Samples can be scanned from a single wavelength channel or can be sequentially scanned from two channels. For instance, for a single-channel microarray platform, a sample is typically labeled with a fluorophore that emits a signal in the red channel. For a dual-channel microarray format, a second sample can be labeled with a dye that emits in the green channel. Dual labeling is used in some experimental designs, such as expression microarrays, to measure the overexpression of a gene associated with a disease state. In such experiments, cDNAs derived from the mRNA of normal and diseased tissues are differentially labeled, mixed, and tested on the same slide in a competitive hybridization reaction. The resulting ratios of the two colors reflect the relative abundance of the labeled material within each sample. Similarly, calculating the fluorescent ratios from each target on an aCGH microarray allows the mapping of gains and losses for a chromosome of interest.

Microarray Image Processing

Most microarray scanners detect and acquire one, two, or more colors (via one, two, or more channels). The optical path of the system minimizes overlap between the spectra (crosstalk) and allows acquisition of two spectrally separate images. In many cases, the images are represented as a red and a green image. When two colors are used, the ratio of the two fluorescence images eliminates artifacts caused by regional bias and irregular spot size. When one color is used, the fluorescence signals from

two or more microarrays are normalized and can be compared with each other. Diameters of spots printed on the arrays range from 10 µm to just under 1000 µm, and the resolution of scanners ranges from 1 to 50 µm. Thus, depending on scanner resolution, variable amounts of pixel data can be collected per scan over an entire microarray.

MICROARRAY IMAGE ANALYSIS

The analysis of scanned images usually involves three tasks: spot finding or gridding, image segmentation, and spot quantification.

Spots are initially assigned specific coordinates, and the process of spot finding or gridding can range from manual to fully automatic, depending on the image-processing software used. This takes into account the individual size and shape of each spot and adjusts for uneven rows and columns that may be produced by the printing process.

The process of segmentation partitions the entire image to foreground or background pixels and relies on the spatial and intensity properties of each pixel. There are four main types of signal segmentation that have been used for spotted arrays. The simplest method is spatial segmentation which places two circles (inner and outer circles) of fixed but different sizes over each spot to demarcate probe signal from the immediate background signal. On the one hand, because of the irregularity of spot sizes on some microarrays, the actual area inner circle may be larger than the diameter of a spot and thus will contain background pixels. On the other hand, artifacts and signal can be found in the area between the inner and outer circles and contribute to the background signal. The second method, intensity-based segmentation, distinguishes signal pixels from background pixels based on the spot intensities within a target region. In this case, a certain percentage of pixels within the top-ranked intensities may be classified as signal pixels. The advantages of this method are simplicity and speed, but the drawback is the inability to distinguish between artifacts and signal and the tendency to detect low signals that are close to background. The third method is a statistical approach known as Mann-Whitney segmentation that combines information from spatial and intensity-based analysis. Here, background pixels located outside the inner circle set are used to determine a threshold intensity level for a signal within the inner circle. The limitation of this method is that a large amount of spot irregularities and artifacts can reduce its accuracy. The fourth method, the trimmed measurement segmentation method, also combines spatial and intensity information and measures signal distributions inside and outside the inner circle. The method trims the upper and lower extremes of each distribution to allow removal of signal from artifacts and incorrectly located background or foreground signal pixels.

The main assumption of spot quantification is that the total fluorescent intensity from a spot is proportional to the expression level of the labeled transcript. This is highly dependent on a number of factors, including target preparation, hybridization conditions, and signal detection within the linear dynamic range. If the amount of probe deposited during the microarray manufacturing procedure varies from spot to spot and from array to array, thus resulting in different sized spots, the sum or total signal intensity can be variable and inaccurate. To correct for this variation, microarrays should be spotted via homogenous surface chemistry that has a fixed binding capacity. This ensures the same amount of probe at each spot location. Alternatively, spots can be quantified by taking the mean, median, or mode of intensities of all signal pixels determined to be foreground signal. The more robust methods that protect against outlier signals are the trimmed mean (where a certain percentage of top and bottom signals are trimmed before calculation of the mean) and median signal intensities. When two different fluorophores are used, the intensity ratio can be used to correct for variable probe amounts and can be calculated from mean, median, and mode intensities from each channel.

MICROARRAY DATA ANALYSIS

Particularly dense formats of microarrays that contain tens of thousands to millions of probes per chip or slide generate a large volume of raw data per array, which requires the use of specialized data-analysis software. Microarray software programs are designed to extract primary data, normalize the data to remove the influence of experimental variation, and link probes to relevant gene and sequence-derived targets. Software programs are also available to apply statistical methods, analyze, visually display, and manage data in order to extract biologically meaningful information. The major parts of data analysis are normalization, background correction, and ratio calculation.

Normalization systematically adjusts microarray raw data in an effort to reduce the variability brought about by differences in the manufacture and processing of the microarrays and by technical variables so that true biological differences between samples can be detected. The wide range of normalization methods precludes a detailed discussion of the topic in this chapter, and currently there are no standards for normalization. Commonly used algorithms are selected based on the microarray type, the number of fluorophores used, and the samples being studied. Some methods are built into the manufacturer's software, but others are available from commercial sources or open-source software providers.

Background correction eliminates low levels of noise in microarrays stemming from both the inherent noise of the detection instruments and from the surface chemistry used in manufacturing. Several contaminants acquired from microarray processing can cause high levels of background that must be corrected before data analysis.

In two-color microarrays, the ratio of signal intensities of array elements of two co-hybridized samples is used as a relative measure of gene expression. In single-channel systems, the ratio can be calculated between signals taken from two different samples (one sample is a reference sample) hybridized on individual microarrays. Thus, the resulting data from microarrays does not represent an absolute quantification but rather a relative level of RNA or DNA against a reference sample or control.

Quality Control and Quality Assurance

As with any diagnostic assay, quality control and quality assurance are critically important. Microarrays must demonstrate robustness and reproducibility. The general quality control and assurance steps outlined in chapter ⟨1127⟩ for nucleic acids and NAT also apply to microarrays. Unlike other diagnostic tests, no reference reagents are available at present for quality control

of microarrays, and regulatory guidance is emerging. FDA has issued a draft guideline titled "In Vitro Diagnostic Multivariate Index Assays." This guidance addresses the definition and regulatory status of a class of in vitro diagnostic devices referred to as in vitro diagnostic multivariate index assays (IVDMIAs), and microarrays fall into this category. The guidance addresses premarket pathways and postmarket requirements with respect to IVDMIAs.

Several unintended sources of variability that are specific to microarrays can extensively affect signal intensities and the accurate derivation of a true signal that accurately reflects the labeled transcript. A major source of variability is spot quality. Measurements of spot quality at the processing stage permit removal of spots with poor or questionable quality. Other sources of variability are artifacts, for example, regional shifts (rise or fall) in an array's overall signal that can be visualized within single chips or in-composite data derived from multiple chips. These changes can be distinguished from actual variability because they are nonrandom, and patterns can be detected by visualizing signals over the entire area of the chip. When dyes of different spectral properties are used to label two different samples in a competitive hybridization reaction using a single array, differences may arise because of labeling bias rather than gene expression level. For instance, the green channel may appear consistently brighter than the red channel despite the fact that there are no real differences in expression. Hybridization with reverse dyes can ensure detection and elimination of dye bias effects. As with any quantitative assay for RNA, the integrity of the sample affects its measurement, and sample quality is an important determinant for accuracy. For instance, because labeling is directed from the 3′ to the 5′ end but RNA degrades from the 5′ end, degraded RNA leads to high 3′/5′ ratios, resulting in nonuniform labeling across the entire transcript. Finally, variability can be introduced during the processing of microarrays, which is a relatively complex procedure that involves multiple steps such as labeling, hybridization, washing, and staining (technical variables). Such variability can mask true differences in the samples tested.

When used as a diagnostic test, the microarray should demonstrate robustness, reproducibility, a high degree of correlation to the original format, and reliable prognosis prediction. The microarray ideally should contain at least 2–3 replicate spots for each reporter gene to ensure intra-assay reproducibility. With a two-channel microarray, a reference sample pool can be hybridized in the complimentary fluorescent channel so that data can be expressed as log ratios, which reduces the need for extensive normalization. Interassay reproducibility of test results and stability over time can be tracked by using a number of reference samples that, when labeled and hybridized, represent a spectrum of predictive endpoints (for instance, high risk, borderline risk, low risk) and that should fall within a predetermined range of results. Failure of these controls should result in rejection of results of samples in the same assay run. If the assay is performed at many sites, site-to-site reproducibility is imperative and must be assessed. The reproducibility of the assay with regard to tissue extraction also must be determined, and the quality of tissue specimens or RNA should be specified clearly (for instance, percentage of tumor cells within a specimen).

In conclusion, microarray experiments should be carefully designed and conducted in order to minimize variability and to yield data that accurately relate to the samples analyzed. In addition, biomarkers of interest should be analyzed and verified using an alternative platform such as qRT-PCR that should be shown to be reproducibly detected in the same and different samples. The development of reference standards, especially when microarrays are used as diagnostic tests, is the next step to ensuring the quality and validation of microarray results. With the shift from custom-built to commercial microarrays, issues with reproducibility, standardization, and quality control have been largely addressed by the stringent quality controls used in commercial manufacturing.