# Agilent's Microarray Platform: How High-Fidelity DNA Synthesis Maximizes the Dynamic Range of Gene Expression Measurements

Author

Emily LeProust, Ph.D.
Agilent Technologies

A key factor controlling the dynamic range of signal intensity in microarray experiments is probe sequence fidelity. Agilent's SurePrint inkjet oligonucleotide synthesis process dramatically increases the sequence fidelity of probe synthesis, resulting in over five orders of magnitude of dynamic range of signal intensity measurements. As a result, the Agilent microarray platform produces accurate relative quantitation of gene expression at a level far exceeding the sensitivity of other array platforms.
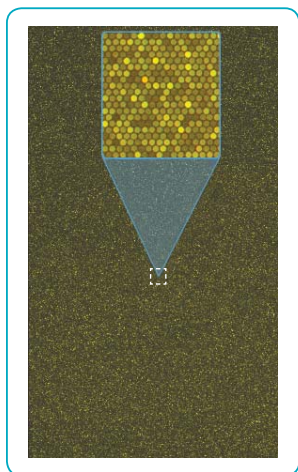
Agilent's million-featured microarray. Inset shows a magnified view of a small region.

Introduction

DNA quality is a crucial factor in obtaining superior biological data from microarray experiments. In differential gene expression studies, where expression levels can vary over an enormous range, only probes synthesized with the highest fidelity can produce the great dynamic range and highest signal-to-noise ratio required for meaningful results.

Agilent Technologies manufactures probes for microarrays using a proprietary DNA synthesis method based upon inkjet printing technology. The method, in which layers of DNA nucleotides are "printed" onto desired microarray feature locations to synthesize probes, circumvents many of the limitations of light-based synthesis methods, and is capable of producing oligonucleotide probes of unprecedented quality and length.

Light-based DNA chemistry is another common method implemented for synthesizing oligonucleotide probes for microarrays. This method has certain inherent limitations, including long reaction times and leakage of light from one microarray feature to its neighbors, both of which introduce unwanted changes into probe sequences. While microarrays with probes synthesized in this way can generate reproducible and significant measurements of gene expression levels, many genes—especially those with lower levels of expression—cannot be resolved above the experimental noise.

Agilent Technologies

This application note describes Agilent's inkjet DNA chemistry, compares it to both conventional and light-based processes, and discusses how probe quality is related to the dynamic range of signal intensity measurements. Experimental results are presented that show Agilent's methodology achieves an extraordinary 99.5% stepwise DNA synthesis yield, and produces probes with very high sequence fidelity. Also described are data that confirm the Agilent microarray platform can produce accurate relative quantitation of gene expression levels over a range of more than five orders of magnitude, far exceeding the capabilities of other systems.

Chemistry of DNA synthesis

A DNA oligonucleotide probe with a specific base sequence must be synthesized serially, one nucleotide at a time. To this end, most probe synthesis processes, including Agilent's, employ repeated cycles of standard phosphoramidite chemistry. This process, shown in **Figure 1**, comprises three steps:

1. **Coupling** – To begin the synthesis, a dimethoxytrityl (DMT)-protected phosphoramidite nucleoside is linked to an available –OH attachment point on the microarray substrate.

This results in the attachment of the first base. The DMT group blocks the incoming nucleoside's 5' –OH, preventing additional nucleosides from coupling to it during that synthesis cycle.

2. **Oxidation** – An oxidizing agent converts the P^{III} phosphite of the added nucleoside into the P^{V} phosphate normally found in DNA nucleotides.

3. **Deblocking** – A deblocking agent removes the DMT group from the added nucleotide, exposing its 5' –OH. This –OH group serves as the attachment point for the next coupling reaction.

This cycle is then repeated as many times as needed, each time using the appropriate phosphoramidite nucleoside.

Constructing a microarray requires many parallel oligonucleotide syntheses at closely spaced locations. To allow different nucleotides to be added to each specific feature on the microarray, at least one step in the synthesis cycle must be spatially controlled. Agilent's inkjet process spatially controls the *coupling* step by printing the desired phosphoramidite nucleosides onto the appropriate feature locations.
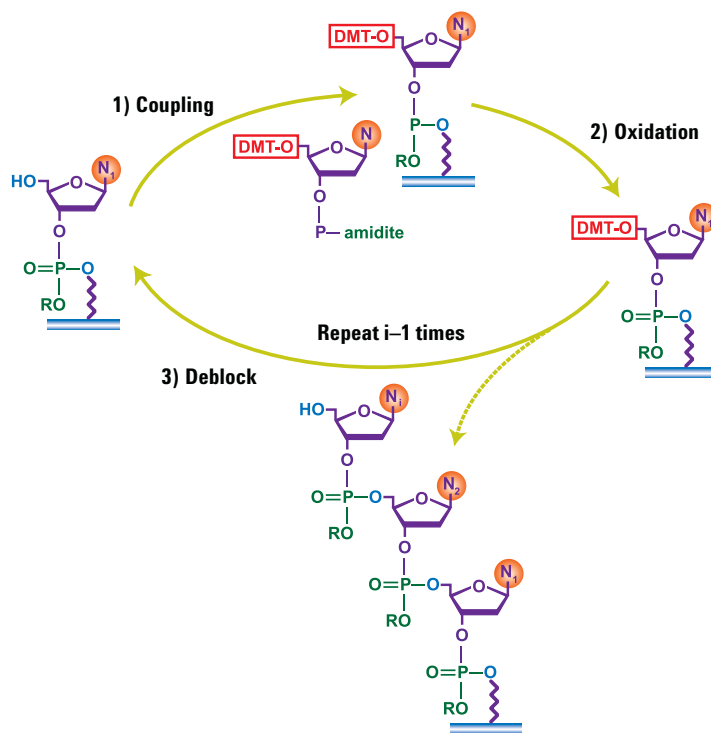

Figure 1. Phosphoramidite DNA synthesis cycle.

The other two steps, oxidation and deblocking, occur in a flow cell, which exposes the entire array to uniform doses of the same reagents, using a novel flood chemistry process.

Other commercial array manufacturers spatially control the *deblocking* step of synthesis using light chemistry. This requires four synthesis cycles to create each layer of nucleotides on the microarray, since each type of nucleoside, A, C, G, and T, must be separately deblocked and coupled. In contrast, Agilent's ink-jet method prints all four nucleosides simultaneously, requiring only one synthesis cycle per layer.

Achieving high sequence fidelity

The most important factor in achieving high sequence fidelity is maximizing the per-cycle synthesis yield. Because making an oligonucleotide requires many synthesis cycles, even small losses in each cycle compound to make a huge difference in the final percentage of full-length products formed. Agilent's typical per-cycle yield is 99.5%, compared to a yield of 98% for conventional processes.[1] In addition, depurination, an undesirable side reaction of conventional methods, reduces those per-cycle yields further.
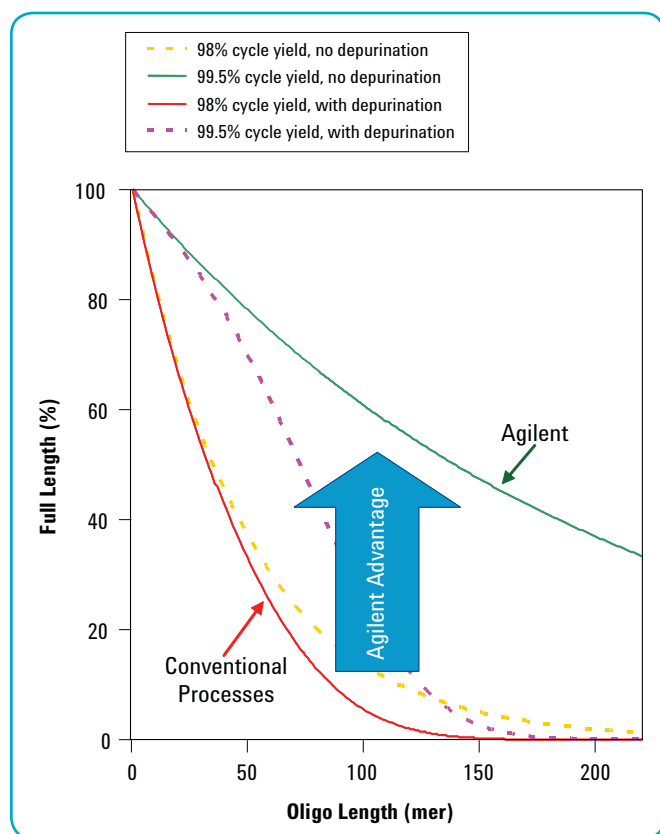

Figure 2. Calculated yield of full-length oligonucleotides as a function of length, at two different cycle yields, with and without depurination.

Figure 2 shows the calculated percentage of full-length products expected as a function of oligonucleotide length for different cycle yields, with and without depurination. Agilent's inkjet process has a 99.5% per-cycle yield with no depurination, producing the most full-length products at any oligonucleotide length. For 60-mer oligonucleotides—a probe length commonly used in gene expression experiments—the Agilent process would produce approximately three times as much full-length product as conventional DNA synthesis processes.

There are two main reasons for these differences in cycle yield and sequence fidelity. First, with conventional chemistry, the deblocking reaction is reversible. Because of this, deblocked nucleotides sometimes become re-blocked before the next nucleotide can be added, resulting in a deletion. Second, with light-activated DNA synthesis, the deblock step is a first order reaction with a defined reaction half-life and therefore comparatively lengthy light exposure times are required to drive this reaction to completion. As seen in **Table 1**, which shows the calculated efficiency of a first-order reaction as a function of half-life count, the reaction is 50.0% complete after one half-life. During the second half-life, half of the remaining reactants form products, making the overall reaction 75.0% complete. Up to ten half-lives are required before the reaction is essentially fully complete, at 99.9%, and is finally useful for oligonucleotide synthesis.

Because each half-life can be on the order of tens of seconds, and four entire synthesis cycles are required for each nucleotide layer on the microarray, very long light exposure times are required. Digital light processors or micromirrors can be used to direct light-based synthesis. Micromirrors contrast ratios are typically low, so long exposure times contribute to leakage of light to adjacent features on the microarray, creating improper activation. Both long exposure times and light leakage lead to random deletions and insertions, lowering the sequence fidelity and signal-to-noise ratio.

In comparison, the reactant droplets produced by Agilent's inkjet process provide a multiple-million-fold excess of reagents at each feature location. The vast excess of reagents drives the reaction forward, making it both instantaneous and complete. This produces higher reaction yields and better probe sequence fidelity.

The results (unpublished) of a study by Dr. Gregory Hannon at Cold Spring Harbor Laboratory **(Table 2)**[2] confirm a per-cycle yield of 99.5% for the Agilent method, and demonstrate the quality of oligonucleotides produced. In the study, 150-mer oligonucleotides from a specially-constructed Agilent microarray were subjected to PCR, cloning, and sequencing. Out of the 384 total sequences analyzed, 181 (47%) were a perfect match with the intended targets. The calculated per-cycle yield required to produce this result with 150-mer oligonucleotides is 99.5%, and it corresponds to an error rate of only one in 250–300 nucleotides. With conventional processes, oligonucleotides are typically restricted to a maximum length of approximately 100 nucleotides, and it is nearly impossible to obtain 150-mers commercially. This study shows that Agilent has surmounted the standard 100-mer ceiling, and that even much longer oligonucleotides can be produced with high sequence fidelity.

Half-life count	Efficiency (%)
1	50.0%
2	75.0%
3	87.5%
4	93.8%
5	96.9%
6	98.4%
7	99.2%
8	99.6%
9	99.8%
10	99.9%

Table 1. Predicted efficiency of a first-order reaction after 1–10 half-lives.

First-Order Reaction

$$A + \text{ light (h}\upsilon) \rightarrow B + C$$
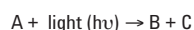
Legend of variable (including half-life)

Where A is molecule to be light deblocked and B and C are reactants

$$-\frac{d\,[A]}{dt} = k[A]$$

Where [A] is concentration of reactant and k is rate constant (light intensity dependent)

$$t_{1/2} = \frac{\ln(2)}{k}$$

Where $t_{1/2}$ is half-life and k is rate constant (light intensity dependent)

Perfect match	181 (47%)
1 mismatch	49
2 mismatches	15
3 mismatches	16
4 mismatches	9
5 mismatches	6
6 mismatches	10
7 mismatches	15
8 mismatches	8
9 mismatches	5
10 mismatches	5
More than 10 mismatches	49
Total number of sequences	**384**

Table 2. Probe sequence analysis results.

DNA quality and dynamic range

Errors in microarray probe sequences both decrease signal intensity and increase noise in expression measurements. Nucleic acid hybridization is specific and cooperative, and any error in probe sequence severely shifts the equilibrium away from favorable binding of the intended target, reducing the intensity of the true signal. At the same time, because nucleic acid samples are highly complex, probe errors also increase non-specific binding, increasing noise. Both of these effects worsen as the number of errors increases. **Figure 3** depicts a theoretical feature on a microarray, and how probe errors both decrease signal intensity and increase non-specific binding.

In two control experiments, nucleotide deletions in probes severely affected signal intensity. The first experiment used a microarray containing seven different sets of 60-mer probes, each designed to hybridize to a known human target. Within
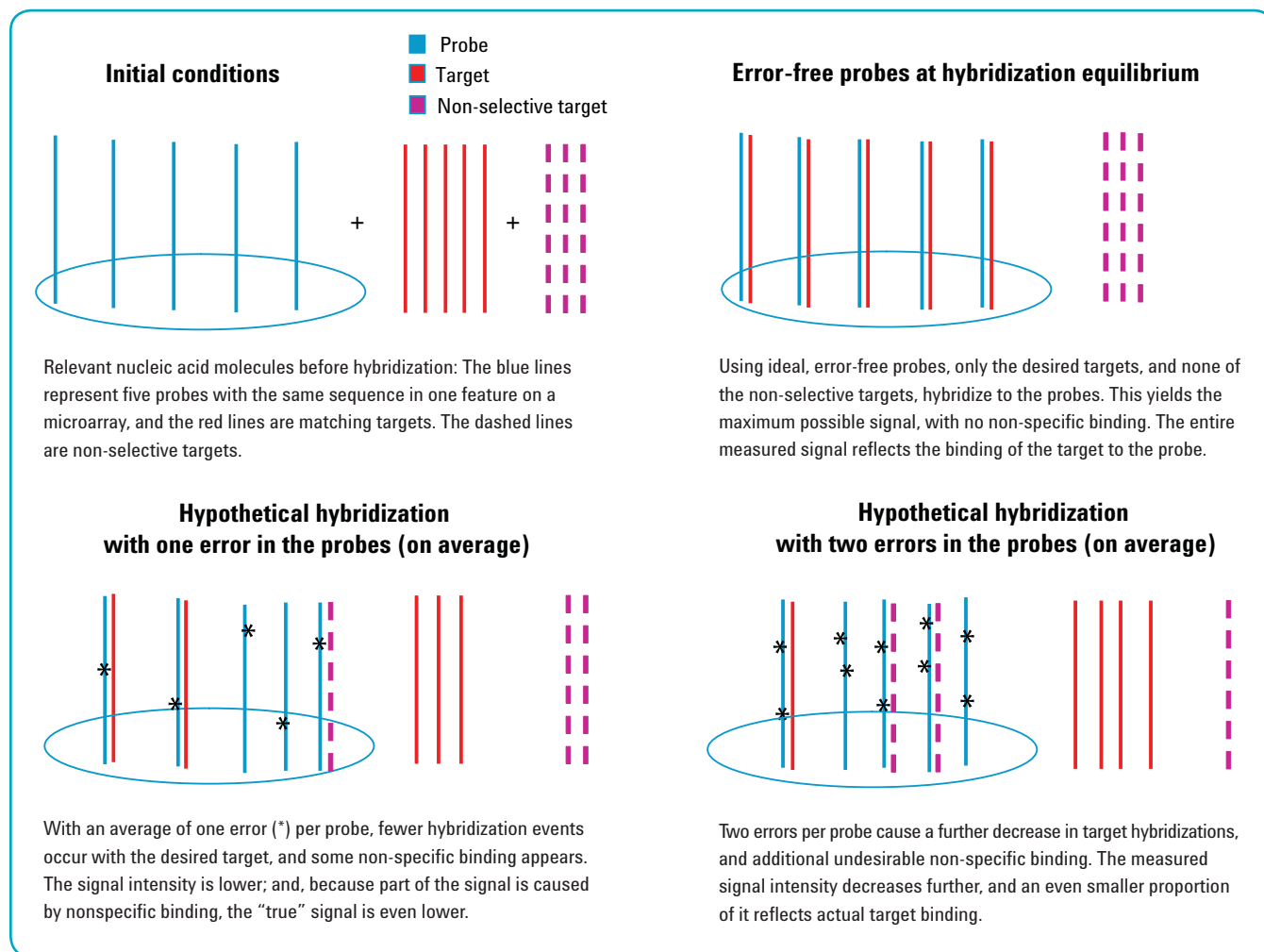

Initial conditions

Probe
Target
Non-selective target

Relevant nucleic acid molecules before hybridization: The blue lines represent five probes with the same sequence in one feature on a microarray, and the red lines are matching targets. The dashed lines are non-selective targets.

Error-free probes at hybridization equilibrium

Using ideal, error-free probes, only the desired targets, and none of the non-selective targets, hybridize to the probes. This yields the maximum possible signal, with no non-specific binding. The entire measured signal reflects the binding of the target to the probe.

Hypothetical hybridization with one error in the probes (on average)

With an average of one error (*) per probe, fewer hybridization events occur with the desired target, and some non-specific binding appears. The signal intensity is lower; and, because part of the signal is caused by nonspecific binding, the "true" signal is even lower.

Hypothetical hybridization with two errors in the probes (on average)

Two errors per probe cause a further decrease in target hybridizations, and additional undesirable non-specific binding. The measured signal intensity decreases further, and an even smaller proportion of it reflects actual target binding.

Figure 3. Probe errors decrease signal and increase non-specific binding.
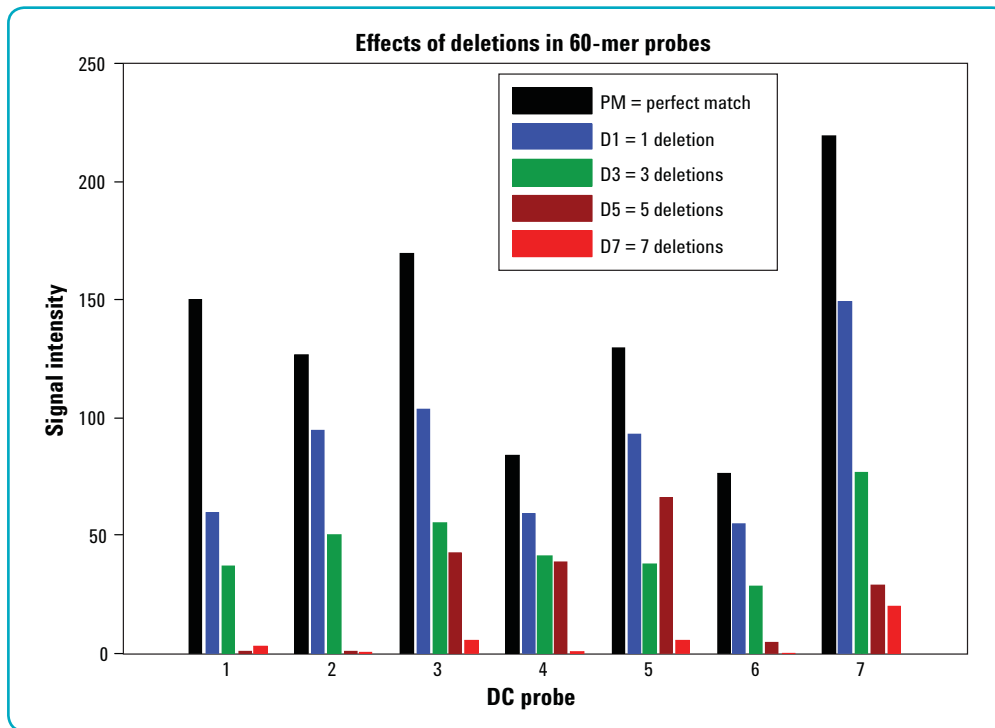
Figure 4. Effects of deletions in seven sets of 60-mer probes. Human DNA targets were hybridized to 60-mer probes with 0 (PM), 1, 3, 5, and 7 deletions.
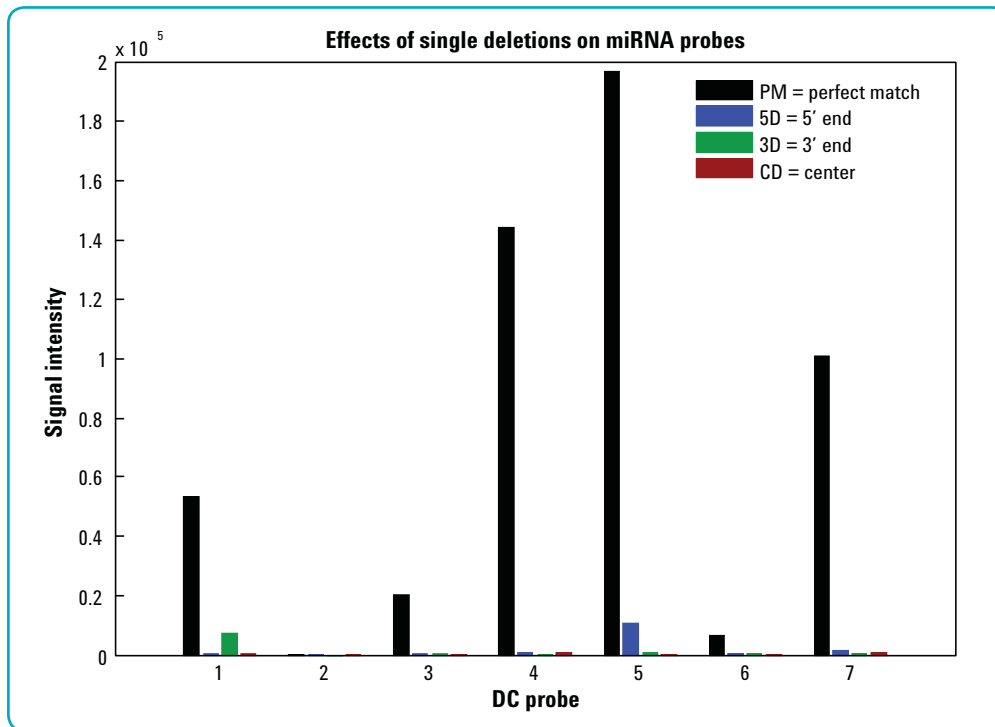

Figure 5. Effects of single deletions in seven sets of 16–25mer probes. End-labeled, synthetic miRNA targets 16–25 nucleotides in length were hybridized to an array of perfect match and single-deletion probes. Signal intensity is shown for perfect match probes (black bars) and probes with a single deletion near the 5′ end (blue bars), the 3′ end (green bars), or the center (maroon bars) of the probe.

each set were probes that perfectly matched a target, or that contained from one to seven deletions. Labeled human DNA was then hybridized to the probes. Signal intensity measurements **(Figure 4)** showed that one deletion causes a significant decrease in intensity of approximately 20–50%, and that additional deletions further degrade the signal.

The second experiment **(Figure 5)** studied the effect of single deletions on signal intensity using shorter probes. A microarray was constructed with seven sets of 16- to 25-mer probes, each set containing perfect match probes, and single-deletion probes with the deletion located in the center, near the 3′ end, or near the 5′ end of the probe. End-labeled synthetic miRNA 16–25 nucleotides in length was hybridized to the array, and signal intensities were measured. In every case, the deletion, without regard to its location, essentially eliminated the detectable signal. Both of these experiments demonstrate a strong relationship between sequence fidelity and signal strength.

An additional control experiment was performed to measure the extent and linearity of the dynamic range achieved using the Agilent platform. A labeled "spike-in" mixture of target oligonucleotides was hybridized to an Agilent 4-pack human

gene expression microarray. The spike-in mixture, a collection of ten different probes at concentrations collectively spanning a million-fold range, contained sequences matching the probes in the standard quality control grid on the microarray. Results of the experiment **(Figure 6)** show a linear (slope = 0.95, R2 = 1.00) relationship between spike-in concentration and signal intensity spanning more than five orders of magnitude, with a total dynamic range of almost six orders of magnitude. This experiment confirms that the Agilent platform can deliver accurate relative quantitation across a significantly wide range of expression levels compared to other microarray platforms. Such performance is enabled by the high sequence fidelity of Agilent oligo probes and their high signal-to-noise ratio. It should also be noted that all Agilent microarrays contain a standard quality control grid, and that Agilent provides the same spike-in mixture used in this experiment as a means to measure and test dynamic range with every sample run on the Agilent platform.

60-mer oligonucleotides optimize sensitivity and specificity

One published study evaluated the relative sensitivity and specificity of different lengths of oligonucleotides from 20mers to 60mers.[3] As shown in **Figure 7**, the areas shaded in red identify the oligo lengths that lead to optimal sensitivity and specificity. The conclusion of this study was that "60-base oligonucleotides hybridized at 30–32% formamide represented a practical compromise between maximal sensitivity and specificity." Because of the relatively high cycle yields of Agilent oligonucleotides printing technology, Agilent arrays are uniquely suited to produce oligos of 60 nucleotides or longer with the highest fidelity.

Dynamic range and biological data

While the fidelity of the oligonucleotide synthesis process and the dynamic range of microarray signal intensity are important technical metrics, the ultimate measure of a microarray platform is its ability to produce high-quality biological data.
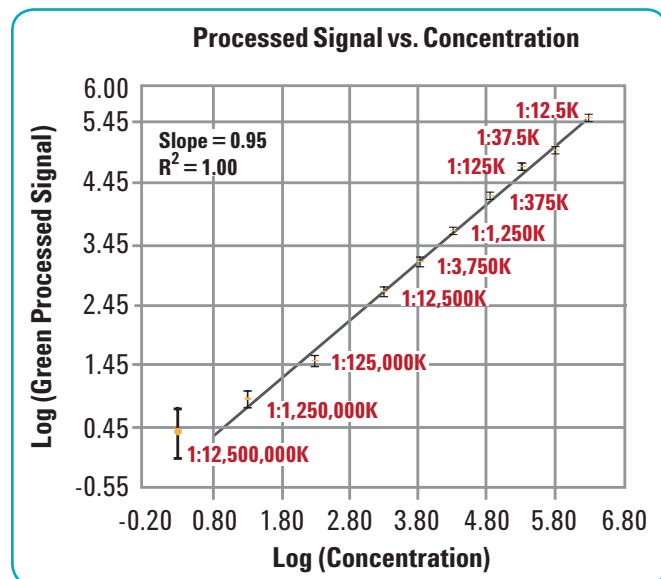

Figure 6. Linear relationship of spike-in concentration to signal intensity, measured on the Agilent platform.

In 2006, the FDA Microarray Quality Control (MAQC) Project released a broad-based performance comparison of microarrays from all major suppliers.[4, 5] One part of the study was a comparison of microarray-derived gene expression levels to reference levels for the same genes measured by TaqMan®, a highly accurate, industry-standard quantitative real-time PCR gene expression assay. The results **(Figure 8)** indicate that the Agilent expression platform had the highest overall concordance and correlation (slope = 0.88, R^2 = 0.89–0.90) with real-time PCR, showing among the highest slopes and lowest amounts of signal compression. Another part of the MAQC study **(Figure 9)** compared the number of differentially expressed genes detected by each platform. Out of approximately 12,000 differentially expressed genes common to all platforms, the Agilent platform was able to resolve approximately 20% more genes than the others.

The Agilent platform resolves more differentially expressed genes than other microarray platforms because of its superior dynamic range. The histogram in **Figure 10** shows the distribution of signal intensities using an Agilent 4x44K Whole Human Genome gene expression microarray. Relative intensities of the 41,000 features are well distributed over six orders of mag-

nitude. Percentages in the figure give the proportion of total features with intensities within each order of magnitude. In **Figure 11** two-thirds (almost 29,000) of the features can still be resolved, even when the histogram is restricted to intensities well above the noise floor.

In addition, signal intensity distributions obtained using the Agilent platform were compared to those from a competing platform **(Figure 12)**. Stratagene Universal Human Reference RNA, a reference standard containing total RNA from ten different human cell lines, was hybridized to an Agilent 4x44K Whole Human Genome gene expression microarray. Signal intensities were distributed over five orders of magnitude above the noise floor. FDA MAQC results for a microarray from a competing platform using the same sample showed signal intensities that were compressed into only two and a half orders of magnitude. Percentages in the figure again give the proportion of total features with intensities within each order of magnitude. The extra dynamic range afforded by the Agilent platform significantly extends the number of genes that can be identified as differentially expressed. When these data are expressed as side-by-side scatter plots **(Figure 13)**, the advantage of Agilent's large dynamic range stands out. On the Agilent
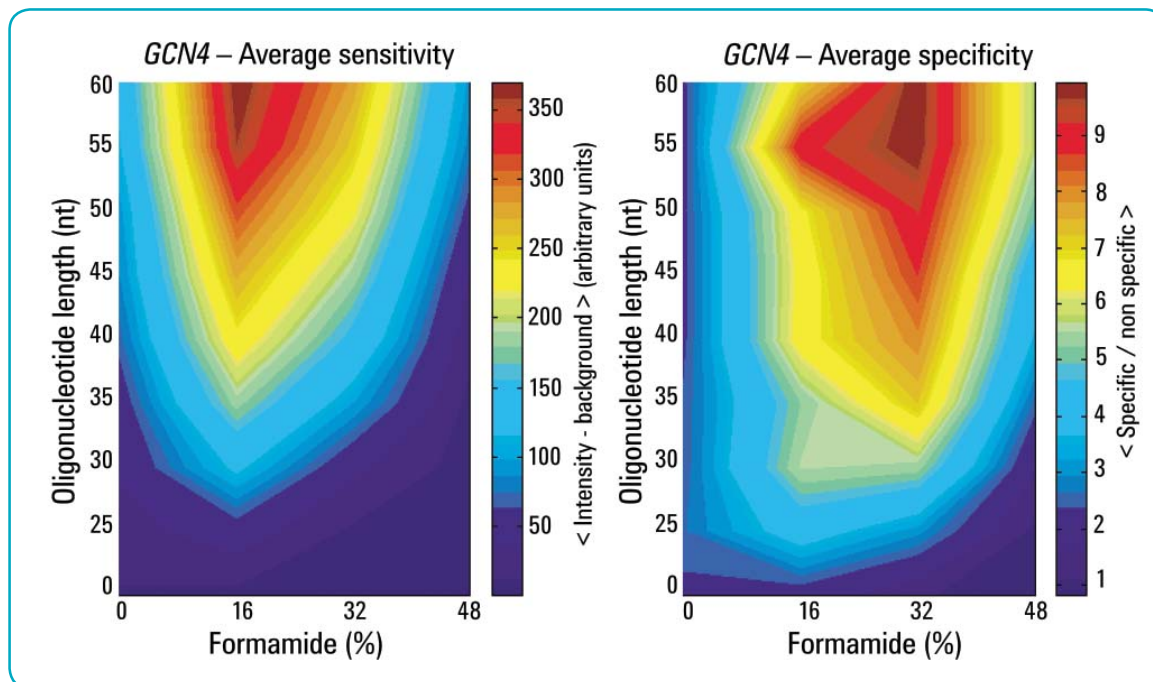

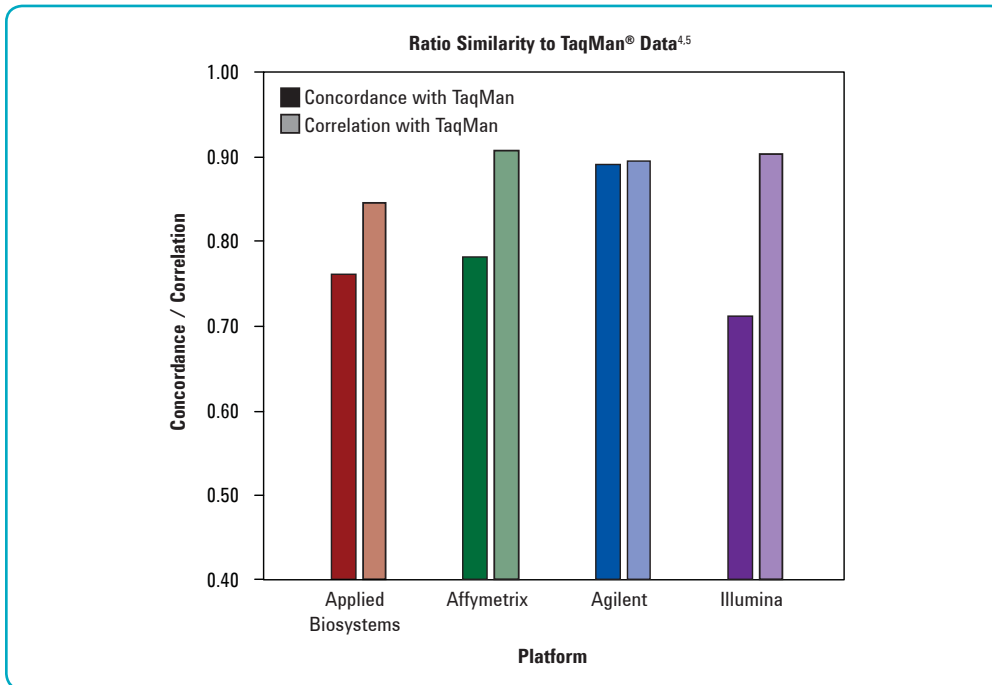Figure 7. Dependence of sensitivity and specificity on oligonucleotide length and hybridization stringency.[3]

Figure 8. Ratio similarity to TaqMan® real-time PCR assay. Accurate detection of differential expression is measured by concordance (slope or compression effect) and correlation (tightness of scatter) to an orthogonal gold-standard method. The Agilent microarray platform clearly exhibits the best relationship to the TaqMan® assay amongst the platforms tested. The ideal result is a value of 1.00 for both concordance and correlation.
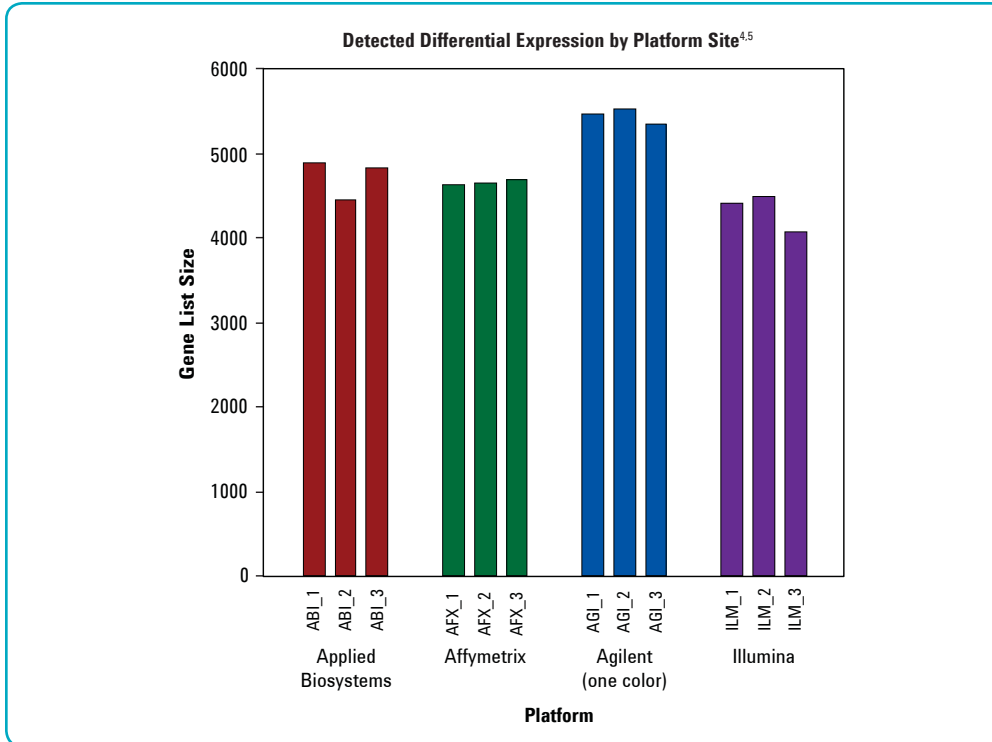

Figure 9. Detection of differentially expressed genes. Agilent's whole genome microarrays have been designed to maximize detection of differential expression. Every Agilent test site detected more differential expression than any other platform site.
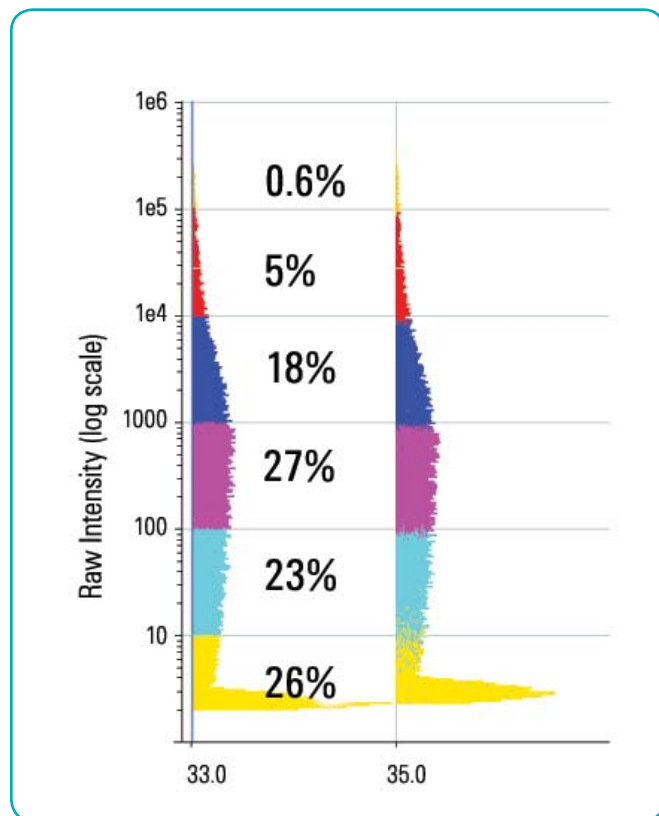
Figure 10. Histogram showing distribution of signal intensities using an Agilent Whole Human Genome 4x44K microarray. Signals from all human probes are represented, n = 41,000.

Figure 11. Signal intensity histogram from Figure 10 restricted to genes well above the noise floor. 28,948 genes are still represented.

platform, at least 62% of the genes can be distinguished well above the noise floor. Using the competing platform, the intensities are severely compressed, and only about 31% of the genes can be distinguished above the noise.

Surface chemistry

In addition to the high-fidelity DNA synthesis, other factors are also important in producing wide dynamic range and accurate biological data. For example, the microarray substrate's surface is a potential source of non-specific binding that can increase background noise levels. Agilent employs a hydrophobic, density-optimized array substrate surface that minimizes background non-specific binding, and boosts sensitivity at the lower end of the intensity scale.

Scanner resolution

Another consideration is how well the array scanner resolution is matched with the size of array features. If feature sizes are small, and similar in size to the scan pixel dimensions, scanned feature edges are ill-defined, and some scan pixels report va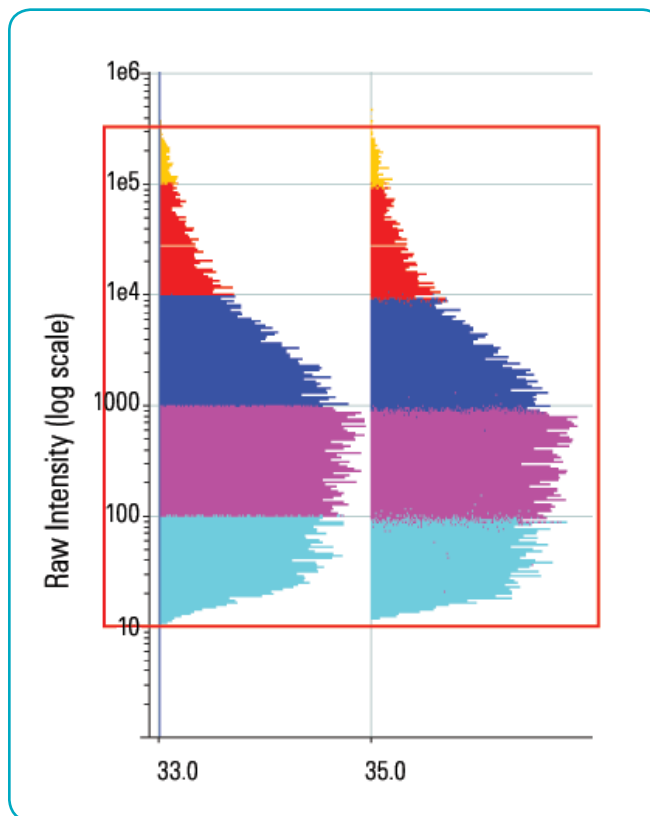lues that are actually a mixture of background and feature intensities. In this case, it is unclear which pixels actually represent the true signal. On the Agilent platform, feature sizes are large compared with scan pixel dimensions, resulting in clearly-defined feature edges; and, with many pixels per feature, statistical calculations become possible, providing significance for the reported value.

Continuing advancements

The infrastructure surrounding the microarray platform is also important. Agilent, for instance, has a broad range of experience and expertise in nucleic acid chemistry, from making hundreds of millions of different oligonucleotides in femtomole quantities daily for microarray applications, to synthesizing kilogram quantities of DNA for therapeutic use. In addition, Agilent maintains an industrial-scale manufacturing facility, with a class 10,000 clean room, that provides very high production capacity, offering consistent supplies of commercial arrays with industry-leading oligonucleotide quality.

Finally, the Agilent platform's core inkjet oligonucleotide synthesis technology leverages the continuing advances in the broader piezoelectric inkjet printing industry. As droplets
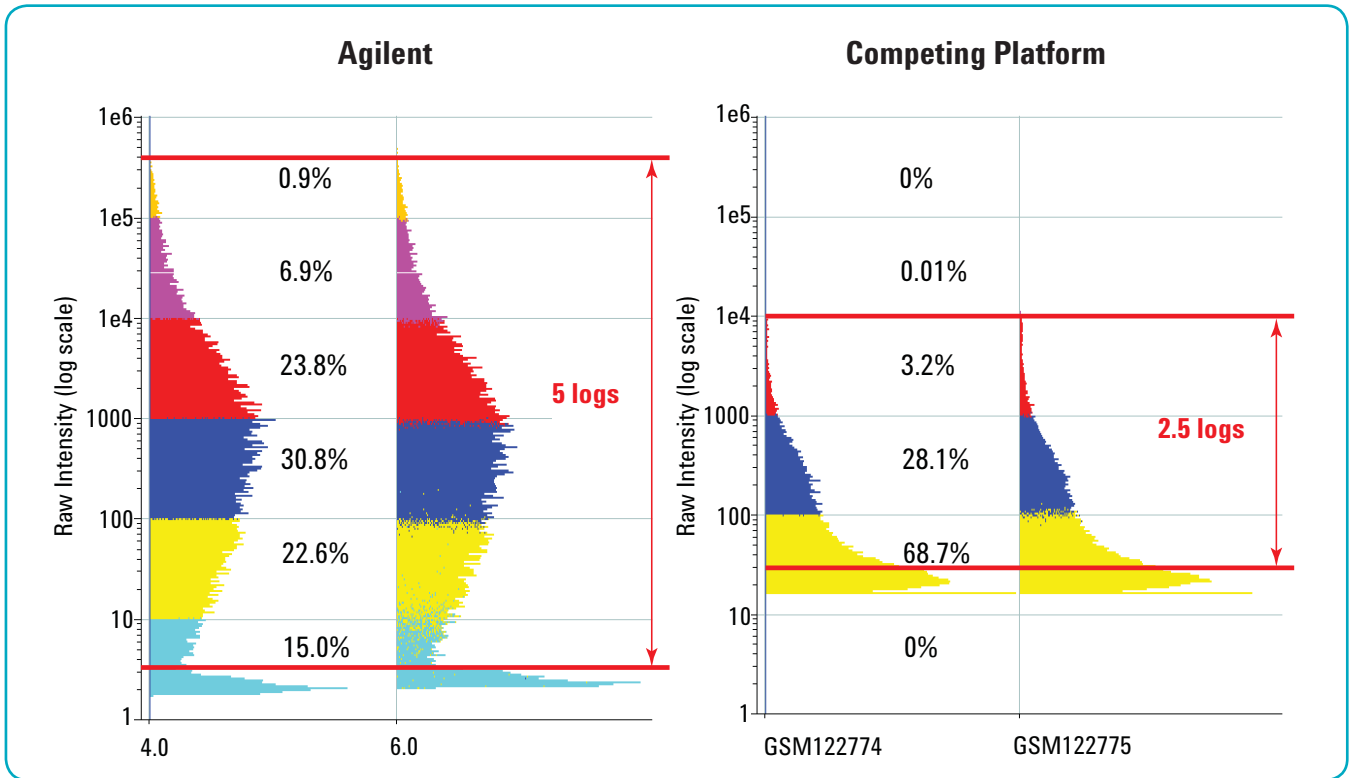
Figure 12. Signal intensity histograms obtained by hybridizing Stratagene Universal Reference RNA to an Agilent 4x44K Whole Human Genome microarray (n = 41,000) (left), and an array from a competing platform (n = 54,613) (right). Red arrows indicate the usable dynamic range. Note the severe compression observed on the competing platform.
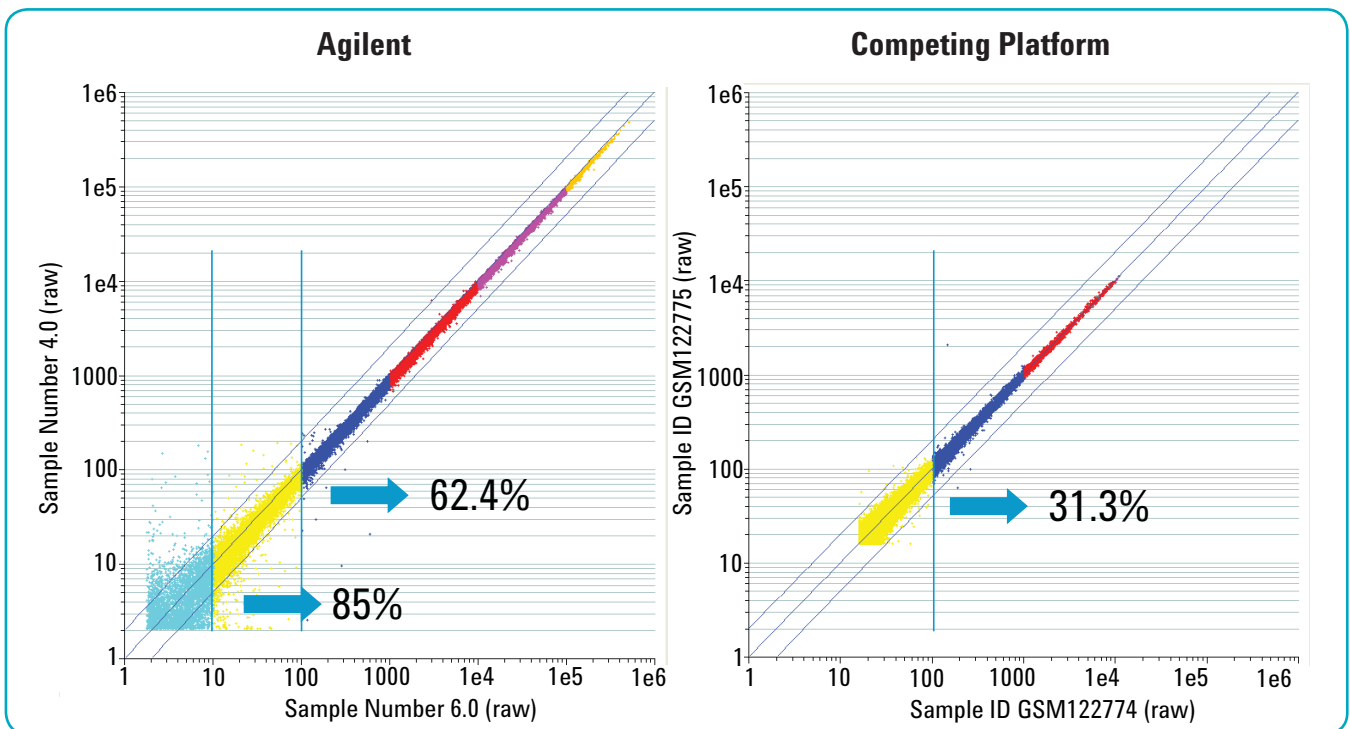

Figure 13. Scatter plots representing the same data as in Figure 12. Stratagene Universal Reference RNA was hybridized to an Agilent 4x44K Whole Human Genome microarray (n = 41,000) (left), and an array from a competing platform (n = 54,613) (right). Percentages give the proportion of total genes that can be resolved above the approximate noise floors indicated by the vertical blue lines.

become smaller, feature densities can increase—up to one million features can now be printed on a single 1″ x 3″ microarray slide. As piezoelectric printing technology evolves, higher density arrays will enable new applications for microarray-based research.

Conclusion

Regardless of probe synthesis method, probe sequence fidelity is a key factor controlling the dynamic range of signal intensity in microarray experiments. Using higher-quality probes produces greater dynamic range, which in turn results in more useful biological data. Agilent's inkjet oligonucleotide synthesis process has distinct advantages over conventional and light-based chemistries, including a higher per-cycle yield that dramatically increases the sequence fidelity of probe synthesis, resulting in over five orders of magnitude of dynamic range. In gene expression experiments, the increased dynamic range enables the detection of significantly more genes above background levels. Further, in side-by-side comparisons, the Agilent platform produced the best overall concordance and correlation with real-time PCR reference standards. These results, and other considerations discussed in this application note, confirm that the Agilent platform is the system of choice for accurate, microarray-based measurements of gene expression levels.

References

1. The cycle yield refers to the combined yield of the three reactions for the synthesis cycle, i.e., coupling, oxidation, and deblocking. While coupling yields can typically be over 99.5%, the deblocking reaction is reversible in conventional synthesis and suffers from the photochemistry limitations in light-activated syntheses. Hence, deblocking is typically the largest source of error in DNA synthesis, even though attention is usually given to coupling yield. The issue is compounded by the fact that deblocking creates errors that are still reactive during subsequent cycles and they therefore result in random deletions.

2. Data not published, personal communication.

3. Hughes T.R., et al. (2001) **Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer.** *Nat. Biotechnol.* 19(4), 342-7.

4. MAQC Consortium, Shi, L., et al. (2006) **The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements.** *Nat. Biotechnol.* 24(9), 1151-61. View online at http://www.nature.com/nbt/journal/v24/n9/full/nbt1239.html

5. View supplementary information (Tables S7, S12 & S13) online at http://www.nature.com/nbt/journal/v24/n9/extref/nbt1239-S8.pdf

About Agilent Technologies

Agilent Technologies is a leading supplier of life science research systems that enable scientists to understand complex biological processes, determine disease mechanisms, and speed drug discovery. Engineered for sensitivity, reproducibility, and workflow productivity, Agilent's life science solutions include instrumentation, microfluidics, software, microarrays, consumables, and services for genomics, proteomics, and metabolomics applications.

Buy online:
www.agilent.com/chem/store

Find an Agilent customer center in your country:
www.agilent.com/chem/contactus

U.S. and Canada
1-800-227-9770
agilent_inquiries@agilent.com

Asia Pacific
adinquiry_aplsca@agilent.com

Europe
info_agilent@agilent.com