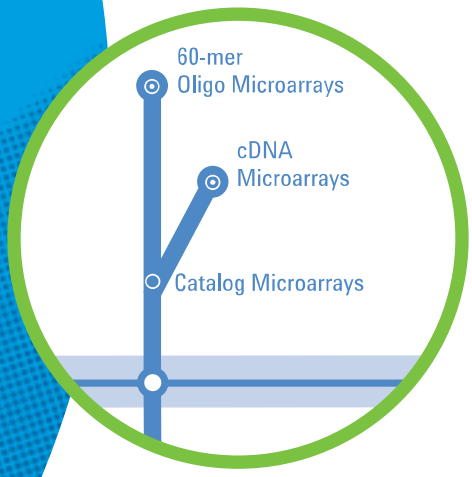
Performance comparison of Agilent's 60-mer and 25-mer *in situ* synthesized oligonucleotide microarrays

Stephanie B. Fulmer-Smentek, Ph.D.

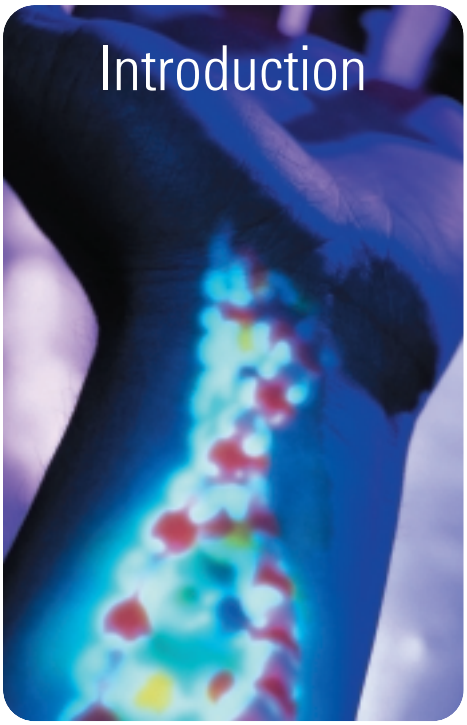


Agilent Technologies

Abstract

Microarrays of oligonucleotide probes are a predominate tool for gene expression profiling, with probe length as a key consideration in the design of a gene expression microarray. This application note describes experiments designed to compare the performance of short (25-mer) and long (60-mer) *in situ* synthesized oligonucleotide DNA microarrays. The 25-mer microarray used in this analysis was an iteratively optimized microarray design with one 25-mer probe or probe/control pair for each yeast open reading frame (ORF). The 60-mer microarray was generated from first round designed probes without optimization and also consisted of one probe per ORF.

In general, 60-mer microarrays tended to have higher sensitivity, with an average lower limit of detection estimated to be approximately 0.001 pM as compared to approximately 0.008 pM for 25-mers. Reproducibility of log ratio values, system noise and accuracies of log ratio determination were comparable between the two microarray types. Comparison of log ratio measurements from a complex gene expression experiment also demonstrated overall concordance in direction of expression. These results demonstrate the clear advantages of using 60-mer oligonucleotide microarrays for gene expression profiling. The sensitivity of the 60-mer microarrays is significantly improved without sacrificing any other aspect of their performance.



Introduction

Agilent Technologies produces *in situ* synthesized oligonucleotide microarrays that enable the flexible custom design of high quality oligonucleotide microarrays of 25 or 60 bases in length. Traditionally, 60-mers are believed to be more sensitive, due to the larger area available for hybridization. 60-mers are also more tolerant of sequence mismatches, which results in simplified analysis of highly polymorphic regions through the use of longer probes. 25-mers, on the other hand, are viewed as being more specific and allow for a measurement of specificity due to the inclusion of matched control probes. Previous comparisons of performance between the two microarray formats have been carried out under conditions more ideally suited to one or the other format[1]. In this note, we describe an experimental comparison between microarrays of 25-mers and 60-mers each used according to their specific recommended protocols[2].

The 25-mer microarray used in this analysis was an optimized 25-mer design, consisting of 1 probe or probe/control pair per open reading frame (ORF) for most ORFs in the yeast genome. Each 25-mer probe on the microarray has a corresponding centrally deleted ("deletion control") probe, which acts as a control for nonspecific hybridization. The microarray design was experimentally optimized by choosing the best 25-mer probe for each gene following at least 2 rounds of experimental iteration. Probes were chosen based on sensitivity and specificity criteria as determined from experimental data. Only about one third of the probes on the microarray required deletion control probes; for the remaining two thirds of the probes, the deletion control was removed from the final design because the hybridization signal associated with these probes had minimal effect on the final result under the experimental conditions tested. The removal of these deletion control probes allowed for the inclusion of an increased number of experimental probes to accommodate all yeast genes on a single microarray.

The 60-mer microarray used in these experiments also consisted of one probe per ORF for most ORFs in the yeast genome. Unlike the 25-mer microarray, these probes were not experimentally optimized. The 60-mer probes represented a "first round" probe design, using Agilent Technologies' standard probe design algorithm, which includes optimizing for base composition and selection for minimal homology within the yeast transcriptome. Due to the length of these probes, use of simple deletion control probes as a measure of specificity was generally uninformative and therefore not included in the design.

These experiments were performed using yeast as a model system. The use of a relatively simple genome provides several advantages for this type of analysis. First, the availability of defined knockout strains of yeast allows for a straightforward measurement of sensitivity and accuracy of log ratio in a complex background through the use of spiked-in transcripts for the knocked out gene. These knockout/spike-in experiments resemble a normal gene expression experiment. Second, the relatively small size of the yeast genome allows for analysis of the entire genome on a single small format microarray. Finally, the completely sequenced yeast genome and well annotated gene sequences allow for simplified analysis of data without the complications present in more complex genomes due to unknown gene sequences and complexity.

This note describes the results of experiments conducted to compare the performance of Agilent's 25-mer and 60-mer *in situ* synthesized oligonucleotide microarrays. These results yield a clearer understanding of the performance characteristics of each type of microarray, as well as the trade-offs inherent in choosing oligonucleotides of a given length.

Methods

Microarrays

The 25-mer microarray was an iteratively optimized design, with one probe or probe/deletion control pair for each of 6,146 open reading frames of *Saccharomyces cerevisiae* (SGD/February, 2001). Probes were selected using Agilent's standard probe design algorithms, which include selection of probes with optimal predicted melting temperature, minimal self-hybridization potential and limited potential for cross-hybridization. Probes were optimized around sensitivity and specificity criteria. For the 60-mer microarray, a single 60-mer oligonucleotide probe was designed for each of 6349 open reading frames (SGD/April, 2001) of *S. cerevisiae* using standard probe design algorithms. These methods select probes with optimal base composition and limited cross-hybridization potential. 543 features consisted of control probes and the remainder of the microarray was filled with 1563 features of replicated yeast probes. Data represented here reflects only the first replicate of each yeast probe. For both the 60-mer and 25-mer microarrays, probes were synthesized using SurePrint inkjet *in situ* synthesis at Agilent's manufacturing facility on an 8455 (89x95) feature microarray format.

Sporulation experiment

A wild-type yeast strain, R1165, was obtained from Rosetta Inpharmatics and was grown overnight in synthetic complete media (SC, Complete Minimal Glucose Broth, Teknova, Inc, Half Moon Bay, CA) to a concentration of approximately 3 x10^7 cells/ml. Approximately 10 mL (for the control sample, SC) or 30 ml (for the sporulation sample, Spo) were centrifuged at 500 x g at 4ºC. The supernatants were removed and the cell pellets were resuspended in 250 mL prewarmed SC or Spo media (1% Potassium Acetate, 0.1% yeast extract, 0.05% Dextrose). Cultures were grown at 30ºC for approximately 8 hours prior to harvesting. Total RNA was harvested using the RNeasy Maxi Kit (Qiagen, Valencia, CA) using enzymatic lysis, following manufacturer recommended procedures. PolyA+ RNA was isolated from the total RNA using one round of Oligotex (Qiagen) purification, including an additional wash following RNA binding, following manufacturers recommended conditions. Integrity of the total and polyA+ RNA was monitored using the Agilent 2100 bioanalyzer, RNA 6000 LabChip kit and either the Eukaryotic Total RNA or mRNA assay. Ribosomal RNA contamination of between 10-30% detected in the polyA+ samples did not interfere with subsequent analyses.

Synthetic Target	Spike-in ratio (R:G)	25 mer Target Conc. (pM)		60 mer Target Conc. (pM)	
		Red	Green	Red	Green
YEL009c	1:3	1	3	0.2	0.6
YEL009c	5:1	3	0.6	1	0.2
YDR345c	1:3	3	9	1	3
YDR345c	5:1	15	3	5	1

Table 1: Concentrations and ratios of synthetic spike-in transcripts

RNA labeling

Complex RNA from yeast knockouts or sporulation experiments was labeled with either cyanine 3 or cyanine 5-CTP using Agilent's Fluorescent Linear Amplification Kit, following recommended procedures. 200 ng poly A+ RNA was used as a starting material. Amplification yields, determined by UV spectroscopy, ranged from 200 (for sporulation samples) to 400 fold (for knockout RNA). Synthetic transcripts (YDR345c and YEL009c) were generated by PCR addition of a T7 promoter to a PCR product containing the entire open reading frame (yeast ORFs from Research Genetics). *In vitro* transcription, in the presence of cyanine 3 or cyanine 5- CTP, was carried out using 0.25 pmol of the described template PCR product and reagents from Agilent's Fluorescent Linear Amplification Kit. Synthetic targets were included in the hybridization experiments at various concentrations and ratios. Labeled synthetic targets were verified as full length by analysis on the Agilent 2100 bioanalyzer using the RNA 6000 LabChip kit.

Microarray Hybridization

All hybridizations were carried out using a Robbin's Scientific Hybridization Oven, Agilent's Rotator and Microarray Hybridization Chambers, and *in situ* Hybridization Kit. For 60-mer microarrays, a 2X target mix was generated containing 0.125 µg cyanine 3 and 0.125 µg cyanine 5-labeled yeast knockout cRNA, appropriate concentrations of synthetic targets, and 25 µl of Agilent's 10X control solution in a final volume of 125 µL. A similar mix was prepared for 25-mer microarrays, with 1.25 µg cyanine 3 and 1.25 µg cyanine 5-labeled yeast knockout cRNA. The remaining ingredients were the same. The samples were fragmented by addition of 5 µL 25X fragmentation buffer followed by incubation at 60°C for 30 min. Samples were then moved to ice, and fragmentation was stopped by addition of 125 µL of Agilent's 2X *in situ* hybridization buffer. Microarray chambers were filled and the microarrays hybridized for 17 hours at 60°C with mixing in the hybridization oven.

Microarrays were disassembled under 60°C wash solution 1 (6x SSC, 0.005% Triton X-102) and were washed in Wash Solution 1 at room temperature for 10 minutes, followed by 5 minutes in Wash Solution 2 (0.1x SSC, 0.005% Triton X-102) at 0-4°C. Microarrays were then dried using filtered nitrogen and scanned on Agilent's dual-laser DNA Microarray Scanner (G2565BA). Data were extracted using Agilent's Feature Extraction software (G2567AA), and were loaded into Microsoft® Access for analysis and visualized with either Microsoft Excel or Spotfire DecisionSite.

These experiments were designed to assess the performance of Agilent's 25-mer and 60-mer *in situ* synthesized oligonucleotide microarrays. The specific metrics analyzed were sensitivity, reproducibility, noise, and accuracy of log ratio measurements. In addition, a series of experiments were designed to compare the performance of 25-mer and 60-mer microarrays for the analysis of a complex gene expression experiment. Experiments were performed using either a defined system of complex cRNAs from yeast knockout strains coupled with synthetic transcripts of the knocked out gene (described in Table 1), or with a complex gene expression experiment of yeast grown under conditions of sporulation compared to control growth conditions. The data generated from these experiments highlight the differences in performance of 25-mer and 60-mer microarrays.

Sensitivity

Sensitivity was measured in two different ways. First, the linear dynamic range for both microarray formats was determined. Shown in Figure 1, the linear dynamic range for both 25-mers and 60-mers extends over at least 2 orders of magnitude, from 0.25 pM to 25 pM for 25-mers and from 0.05 pM to 5 pM for 60-mers. Target concentrations tested were from 0.25 pM to 50 pM for 25-mers and from 0.05 pM to 17.5 pM for 60-mers. For each spike-in experiment, the synthetic targets were spiked in at concentrations within the linear range for the microarray format. Also evident in the data shown in Figure 1 is the effect of saturation on the dynamic range. At high target concentrations, the signal for 60-mers flatten out at values consistent with scanner saturation. On the other hand, 25-mers appear to be biochemically saturated well below the signal values for scanner

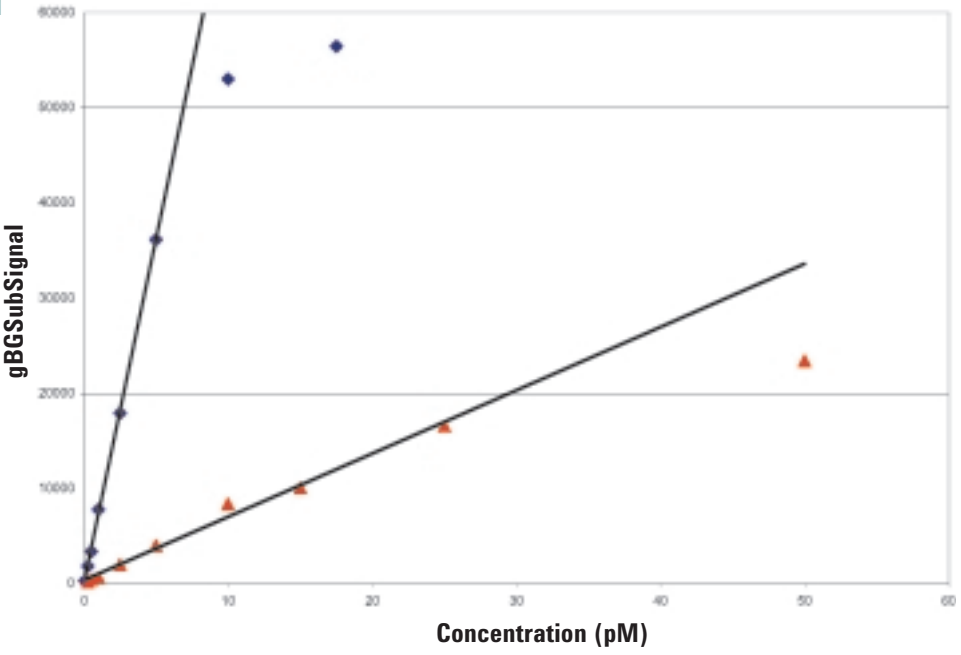
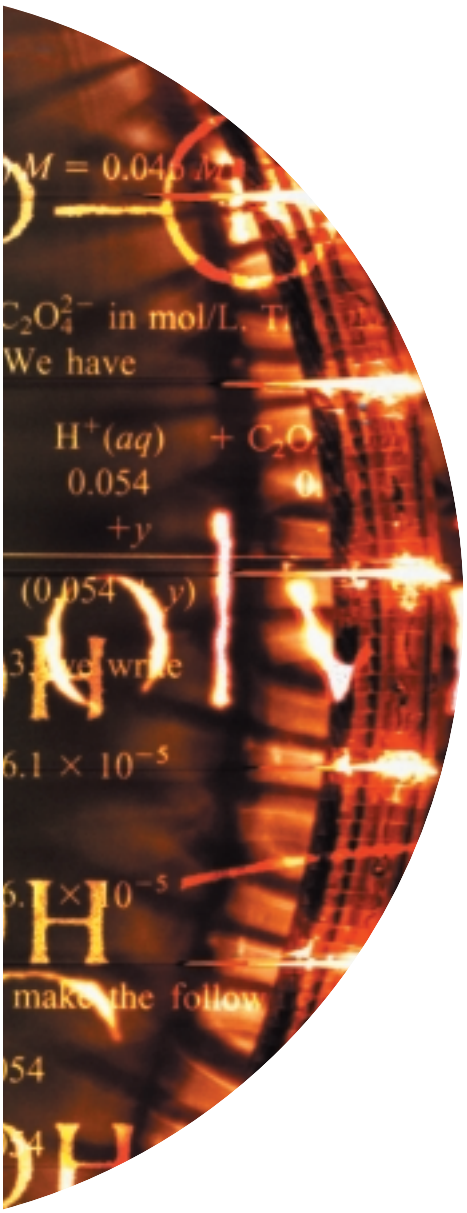


Figure 1: Titration of synthetic targets. Green background subtracted signal for either 25-mers (in red) or 60-mers (in blue) is shown as a function of the spiked-in target concentration in the final hybridization solution. Averaged data is shown for at least 2 replicate microarrays per concentration. Lines represent the linear curve fit through the first concentrations (0.25 pM- 25 pM for 25-mers and 0.05 pM- 5 pM for 60-mers). R^2 correlation coefficients were 0.9901 (25-mers) and 0.9996 (60-mers).



saturation. This possible biochemical saturation leads to a decreased dynamic range, as the entire dynamic range of the scanner is not utilized. In addition, the higher slope of the 60-mers dose-response curve suggests that 60-mers will be more sensitive to smaller fold changes in gene expression because the same change in target concentration leads to a greater change in signal for the 60-mers than for the 25-mers.

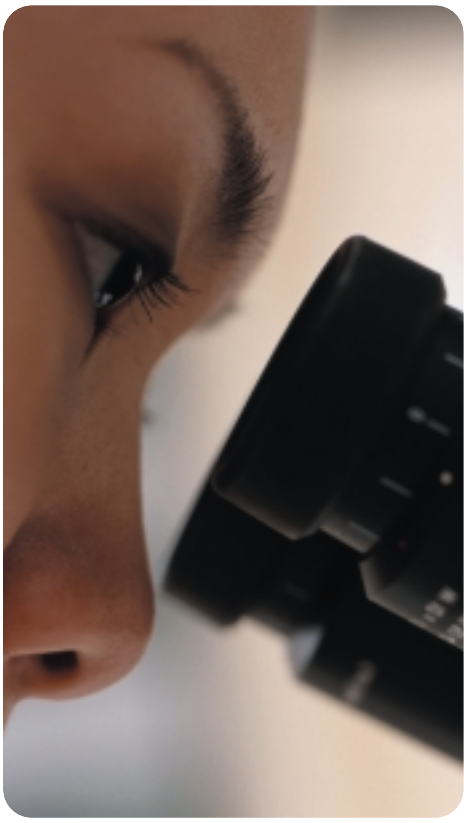
Sensitivity was also measured as lower limits of detection (LLD), which were calculated for both microarray formats from the detection of synthetic transcripts in 4 different spike-in experiments (Table 1). The lower limit of detection was calculated by defining the lower limit of detectability as 3 standard deviations over average background, as measured from a pool of negative control features. Table 2 shows the lower limits of detection from each spike-in experiment. Overall, 60-mers showed higher sensitivity than 25-mers, as measured by the LLD. The LLD calculated in the 60-mer experiments ranged from **five to eight fold lower** than those calculated for 25-mers. This sensitivity may be unique to microarrays manufactured using Agilent's SurePrint *in situ* synthesis process, which achieves high nucleotide coupling efficiency using standard phosphordiamide chemistry[3]. Microarrays synthesized by other manufactures may not achieve this level of sensitivity with either 60-mers or 25-mers.

Other performance parameters

The four spike-in experiments described above were also used to demonstrate the performance of 60-mer and 25-mers for other performance metrics, which are summarized in Table 3. For reproducibility, system noise and accuracy of detection of log ratios, 60-mers and 25-mers performed comparably. These results suggest that the increased sensitivity of 60-mer oligonucleotide microarrays is not at the expense of other performance metrics.

	25 mer Target Conc. (pM)		25 mer LLD (pM)		60 mer Target Conc.(pM)		60 mer LLD (pM)	
Experiment	Red	Green	rLLD	gLLD	Red	Green	rLLD	gLLD
YEL009c, 1:3	1	3	0.0065	0.0055	0.2	0.6	0.0007	0.0023
YEL009c, 5:1	3	0.6	0.0046	0.0045	1	0.2	0.0008	0.0017
YDR345c, 1:3	3	9	0.0098	0.0242	1	3	0.0013	0.0032
YDR345c, 5:1	15	3	0.0124	0.0133	5	1	0.0010	0.0025
Average			0.0083	0.0119			0.0010	0.0024

Table 2: Calculated lower limits of detection for spiked-in synthetic target for both microarray formats



Detection of differential gene expression.

The final measure of microarray performance compared the results from a complex gene expression experiment, hybridized to both microarray platforms. The gene expression from yeast cells grown under conditions designed to stimulate sporulation ("Spo") was compared to the expression from yeast grown under control conditions (growth in synthetic complete media, "SC") using both 25-mer and 60-mer yeast microarrays. For each microarray format, four microarrays were hybridized with cyanine 5-labeled Spo cRNA and cyanine 3-labeled SC cRNA, and four microarrays were hybridized with the dye polarity swapped. For each polarity, the data were averaged and 99.9% confidence intervals were calculated. Data were not included in the analysis if probes for a given gene failed to produce signal in either channel classified as well above background by the Agilent Feature Extraction software for both formats. The data were averaged across the 8 microarrays, and calls of significant gene expression made based on the 99.9% confidence intervals of both halves of the dye swap.

	Sensitivity (ave LLD, pM)		Reproducibility (median SD of LogR)	System Noise (SD of Log Ratio in SvS)	Accuracy (% spiked-in ratio)			
	Red	Green			A	B	C	D
60-mers	**0.0010**	**0.0024**	0.018	0.051	91	105	66	63
25 mers	**0.0083**	**0.0119**	0.016	0.042	82	138	68	77

Table 3: Summary of performance metrics for 25-mer and 60-mer microarrays, demonstrating similarity of performance for all metrics except sensitivity. Sensitivity data is presented as the average LLD (red and green) across all spike-in experiments. Individual experiment data is shown in Table 2. Reproducibility is shown as the median standard deviation around the log ratio for one spike-in experiment (YDR345c, 5:1). System noise is represented by the standard deviation of average log ratio values (averaged across 4 microarrays) for all probes in a single spike-in experiment, with the spiked in probe disregarded (essentially a self comparison experiment). Accuracy is represented as the % of the spiked in ratio that was determined (Observed/Expected ratio). Values shown are for the 4 different spike-in experiments (A: YDR345c,5:1; B: YDR345c,1:3; C: YEL005c,5:1; D, YEL009c,1:3).

Comparisons between log ratios for the two microarray formats are shown in Figure 2, in which data were removed for probes not demonstrating a log ratio significantly different from zero ("unchanged") in both microarray formats. The figure demonstrates that the log ratio measurements between the two microarray formats are fairly similar. Overall, a majority of genes (4541/6049) correlated between the two formats, with 2035 determined to be unchanged, 1126 found to be significantly overexpressed, and 1380 found to be significantly underexpressed in both formats. Five genes demonstrated significant log ratio changes in opposite directions between the two formats ("anticorrelated"); of those genes, all but one gene either generated very low signal in both formats or did not show gene expression difference greater than 1.5 fold in both formats. For the remaining genes, 484 were determined to be significantly over or underexpressed on 25-mer but not on 60-mer microarrays, and 1019 were determined to be significantly over or underexpressed on 60-mer, but not 25-mer microarrays. An additional 75 genes could not be compared because they were undetectable with the 25-mer microarrays, although all gave clearly detectable signal with the 60-mer microarrays.

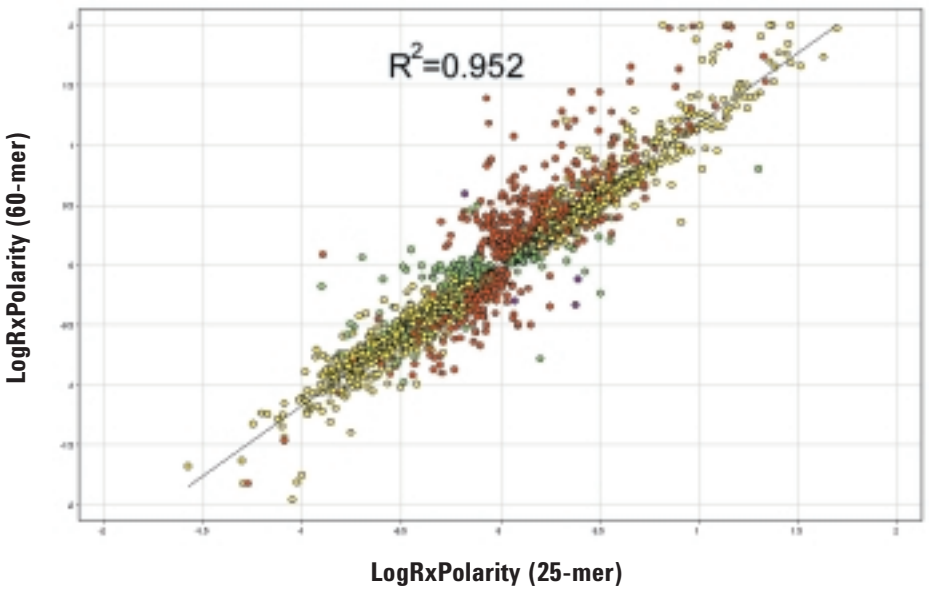
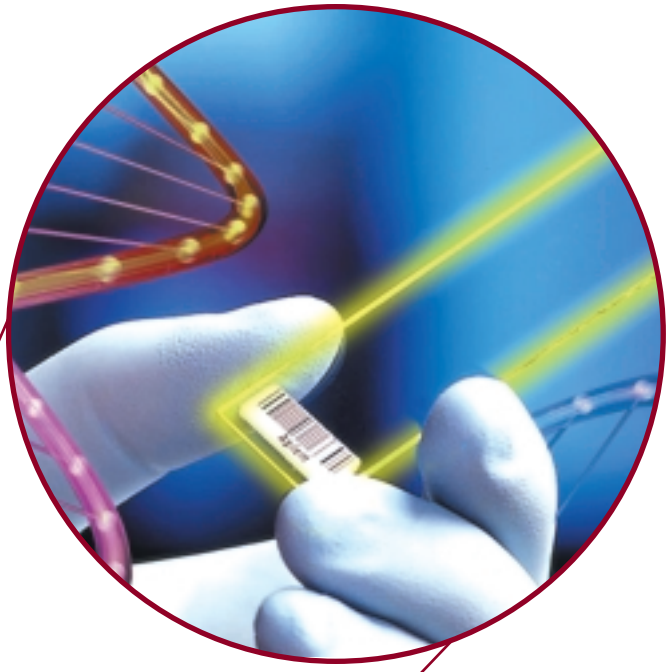
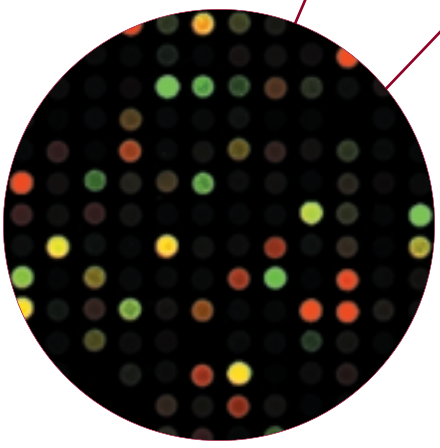
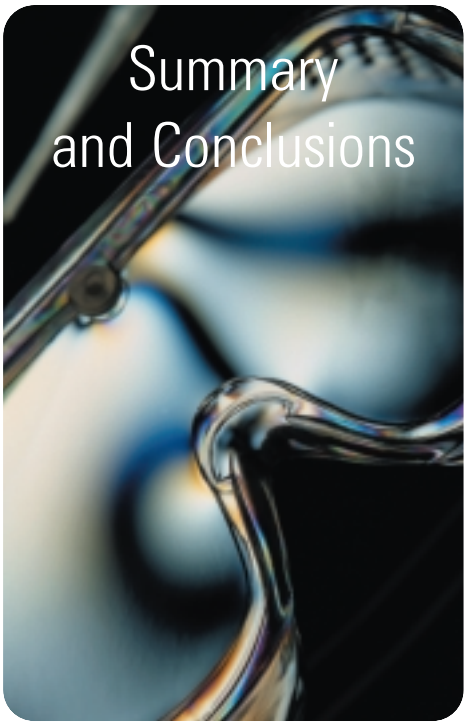


Figure 2: Comparison of log ratio measurements between the two microarray probe lengths. Each data point represents a single yeast gene, with the x value representing the log ratio determined by the 25-mer microarray and the y value indicating the log ratio determined by the 60-mer microarray. Data points are not shown if genes were not significantly differentially expressed for either of the 2 probe lengths. Points in yellow are correlated and significantly different from zero for both lengths. Data points shown in green are significant for 25-mers only. Red data points are significant for 60-mers only. Points in purple are anticorrelated between the two probe lengths. The straight line fit, with R^2 correlation coefficient, is shown. The slope of the line is 1.18.

Additionally, the log ratio values tended to be further from zero on the 60-mer microarrays than on the 25-mer microarrays. This difference is seen in the slope of the straight line fit for the graph shown in Figure 2. If both the 25-mer and 60-mer microarrays showed the same log ratio the slope would be equal to one. The slope of the graph is 1.18, illustrating that the log ratio values tend to be farther from zero for the 60-mer microarrays. This difference suggests that the 60-mers detect differences in gene expression with a broader dynamic range, which may result from the higher sensitivity displayed by the 60-mer microarray format.







Comparison of the performance of 25-mers and 60-mers reveals that microarrays of 60-mers are approximately 5 to 8 fold more sensitive than 25-mer microarrays. In all other aspects of performance, 25-mer and 60-mer microarrays are comparable. However, the work involved to design and create them is not comparable. The 25-mer microarrays used in this study were from an iteratively optimized design, whereas the 60-mer microarrays resulted from a single pass probe design scheme. These results illustrate that considerably less time and work is involved in the design of 60-mer microarrays to achieve a similar level of performance as optimized 25-mer microarrays.

Additionally, 60-mers are generally more forgiving of sequencing errors or polymorphisms than 25-mers. As sequence knowledge grows at an increasing pace, the advantages of 60-mers become more substantial, creating a powerful way to study gene expression associated with new genomes, new sequence information, or new functional sequence annotation. Accurate and reproducible 60-mer microarrays can be created and used more quickly once new sequence data is released, thus accelerating the pace of research with out sacrificing the quality of the results.

The data presented here demonstrate the performance of Agilent Technologies *in situ* synthesized 25-mer and 60-mer oligonucleotide microarrays. This performance may not be available from other microarray manufacturers. In combination with our probe design services, Agilent offers unprecedented flexibility in the design of DNA microarrays, allowing fast turn around of both microarrays and experiments. Using this powerful technology, scientists can design high quality microarrays around experiments without sacrificing time, or significantly increasing experimental cost.

References

1. Hughes, et al. Nature Biotech. 2001, 19, 342-347
2. Fulmer-Smentek, Agilent Technologies, Jan 1, 2002, Publication number 5988-5022
3. Beaucage, S. L.; Caruthers, M. H. Tetrahedron Letters 1981, 22, 1859-1862

www.agilent.com/chem/dnasupport

Print Date: June 1, 2003
Publication Number: 5988-5977EN

Agilent Technologies