A Versatile Assay for High-Throughput Gene Expression Profiling on Universal Array Matrices

Jian-Bing Fan,[1,3] Joanne M. Yeakley,[1] Marina Bibikova,[1] Eugene Chudin,[1] Eliza Wickham,[1] Jing Chen,[1] Dennis Doucet,[1] Philippe Rigault,[1] Baohong Zhang,[1] Richard Shen,[1] Celeste McBride,[1] Hai-Ri Li,[2] Xiang-Dong Fu,[2] Arnold Oliphant,[1] David L. Barker,[1] and Mark S. Chee[1]

[1]*Illumina, Inc., San Diego, California 92121, USA;* [2]*Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California 92093, USA*

We report a flexible, sensitive, and quantitative gene-expression profiling system for assaying more than 400 genes, with three probes per gene, for 96 samples in parallel. The cDNA-mediated annealing, selection, extension and ligation (DASL) assay targets specific transcripts, using oligonucleotides containing unique address sequences that can hybridize to universal arrays. Cell-specific gene expression profiles were obtained using this assay for hormone-treated cell lines and laser-capture microdissected cancer tissues. Gene expression profiles derived from this assay were consistent with those determined by qRT-PCR. The DASL assay has been automated for use with a bead-based 96-array matrix system. The combined high-throughput assay and readout system is accurate and efficient, and can cost-effectively profile the expression of hundreds of genes in thousands of samples.

Microarray technology has been widely used in large-scale genomic research, especially in whole-genome gene expression profiling (Lockhart et al. 1996; Lipshutz et al. 1999, Hughes et al. 2001; Holloway et al. 2002; Nuwaysir et al. 2002). Comparison of gene expression patterns in different cell types, developmental stages, and disease states should enable the discovery of characteristic gene-expression patterns that can be associated with functionally important states. For example, microarray-based tumor classification (Golub et al. 1999; Perou et al. 2000; Welsh et al. 2001), as well as treatment response and clinical outcome prediction (Dhanasekaran et al. 2001; West et al. 2001; van 't Veer et al. 2002) have been demonstrated in many cancer types. Now, the challenges are in validating molecular profiles derived from genome-wide microarray experiments in large clinical sample sets, and translating these findings to the clinic (Chuaqui et al. 2002). Such a transition would have profound implications for both basic research and clinical medicine, but is currently hampered by a lack of suitable technologies. Whole-genome arrays are expensive and cumbersome for large numbers of samples. Although quantitative RT–PCR (qRT-PCR) has been in routine use to assay one or a few genes, more flexible and high-throughput validation approaches are required to enable the screening of hundreds of potential biomarkers in large sets of clinical samples.

We previously developed a ligation-based RNA assay for parallel analysis of ~100 mRNA splicing variants (Yeakley et al. 2002). The assay, dubbed RASL (for RNA-mediated annealing, selection, and ligation), permits analysis of mRNA transcripts without prior RNA purification or cDNA synthesis. Here, we describe improvements to the RASL assay for broader RNA-profiling applications. The new DASL assay differs in important respects from the RASL assay. First, the DASL assay converts the RNA target to cDNA, without sacrificing sensitivity. Using a cDNA template offered technical advantages such as enabling analysis of significantly degraded mRNAs, and facilitated the automation of the assay. Second, the DASL assay uses locus-specific oligonucleotide extension and ligation, which resulted in increased specificity compared with the oligonucleotide ligation scheme used in RASL.

The DASL assay is currently multiplexed to detect over 1200 sequence targets simultaneously, and is carried out on 96 samples in parallel, using a matrix of optical fiber bundles (the Sentrix array matrix; Oliphant et al. 2002). Because of its multiplexed assay format, universal-array readout, and high sample throughput, the DASL assay system fills a gap between the existing RNA-profiling technologies of qPCR and gene-specific oligonucleotide or cDNA microarrays, and is complementary to those approaches.

RESULTS

The DASL Assay

A diagram illustrating the biochemistry of the DASL assay is shown in Figure 1. First, total RNA is converted to cDNA using biotinylated primers (both oligo-d$(T)_{18}$ and random hexamers). The biotinylated cDNA is then attached to a streptavidin solid support and assay oligonucleotides are annealed to their target sequences in the cDNA. A pair of oligonucleotides is annealed to each target site. Three to 10 sites are targeted per gene, and up to 1536 oligonucleotide pairs are multiplexed together in a single reaction. High locus specificity is achieved in two ways, first, by the requirement that both members of an oligonucleotide pair must hybridize in close proximity for an assay signal to be generated, and second, by the removal of excess and mishybridized oligonucleotides by washing after the annealing step.

Annealed upstream oligonucleotides are extended and ligated to their corresponding downstream oligonucleotides to create PCR templates that can be amplified with common primers (P1 and P2). In contrast, the RASL assay is based on locus-specific oligonucleotide ligation (Yeakley et al. 2002). We have found that locus-specific oligonucleotide extension and ligation increases signal-to-noise ratio (Abravaya et al. 1995; Fan et al. 2003). Another advantage is that placement of a gap between the

[3]Corresponding author.
E-MAIL jfan@illumina.com; FAX (858) 202-4680.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.2167504.

14:878–885 ©2004 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/04; www.genome.org

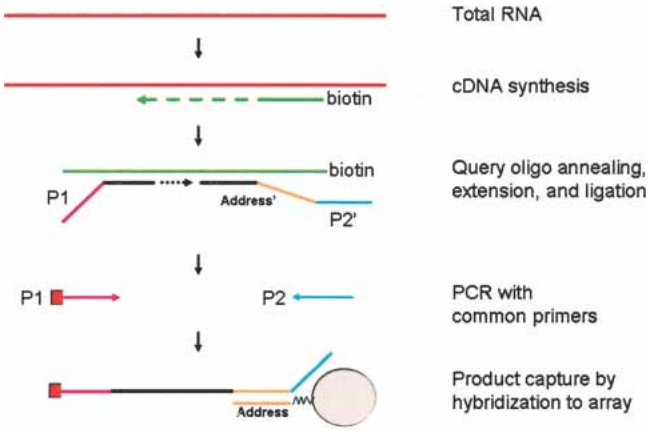

Figure 1 DASL: a cDNA-based assay for RNA profiling. RNA is converted to cDNA, using biotinylated oligo-d$(T)_{18}$ and random hexamers, and immobilized to a streptavidin-coated solid support. Two oligonucleotides are designed to interrogate each target site on the cDNA. The upstream oligonucleotide consists of two parts, the gene-specific sequence and a universal PCR primer sequence (P1) at the 5′-end. The downstream oligonucleotide consists of three parts, the gene-specific sequence, a unique address sequence which is complementary to a capture sequence immobilized on the array, and a universal PCR primer sequence (P2′) at the 3′-end. A single address sequence is uniquely associated with a single target site. The upstream oligonucleotide hybridizes to the targeted cDNA site, extends and ligates to its corresponding downstream oligonucleotide to create a PCR template that can be amplified with universal PCR primers (P1 and P2). The PCR products, which are fluorescently labeled by incorporation of the 5′-labeled primer P1, are hybridized to capture sequences on the beads in the array, and fluorescence intensity is measured for each bead.

upstream and downstream oligonucleotides (typically 1 to 20 bases) provides flexibility in positioning the probes to avoid unfavorable sequences. Both RASL and DASL formats use one PCR primer pair to amplify all of the targets and generate amplicons of ~100 bp. This uniformity results in a relatively unbiased amplification of the PCR template population (Yeakley et al. 2002).

Impact of Multiplexing Level and RNA Input

To assess the impact of different multiplexing levels, we performed experiments with query oligos targeting 281, 665, and 1141 sequence sites. The lower multiplexed oligonucleotide pools represented subsets of the 1141-plex pool. Our results showed that all of the 1141-plex-specific genes were detected only in the 1141-plex reaction, but not in the 281-plex and 665-plex reactions, indicating that the assay is very specific. Consistent expression profiles were obtained with all three multiplexing levels for the shared genes as follows: 281-plex versus 1141-plex ($R^2 = 0.95$) and 665-plex versus 1141-plex ($R^2 = 0.98$). Currently, the number of sequences that can be measured is limited by the number of unique address sequences on each array.

We have systematically measured DASL assay performance as a function of RNA input. Using an oligonucleotide pool targeting 270 mouse genes related to immunological pathways, we assayed total RNAs isolated from two mouse cell lines, EL-4 (T-cell) and A-20 (B-cell). Assay performance was assessed at 1212-plex with RNA input at 100, 75, 50, 25, 10, and 5 ng of total RNA from each cell line. As summarized in Table 1, decreasing the input RNA from 100 to 5 ng resulted in a ~10% reduction (78% vs. 70%) in gene expression detectability and approximately twofold reduction (95% vs. 41%) of genes for which smaller than a twofold difference could be detected between the two cell lines. In general, adequate performance was obtained with 25 ng of total RNA under these conditions.

Because the assay output depends on the relative abundance of the PCR templates (ligated query oligonucleotides), the minimum amount of RNA needed as input to the assay varies according to the expression levels of the targeted transcripts. Therefore, assay sensitivity depends on which transcripts are targeted by the query oligonucleotide pool.

Assay Dynamic Range and Differential Expression Detection

We derived dose-response relationships by spiking different amounts of six synthetic RNAs into a background of mouse total RNA. The six synthetic RNAs were pooled using a permuted matrix experimental design, and each pool was spiked into a background of 100 ng of mouse total RNA. The six synthetic RNAs were transcribed in vitro from plasmids containing prokaryotic gene sequences (*cat*, *cre*, *gus*, *lacZ*, *luc*, and *bla*, respectively) with attached poly$(A)_{30}$ tails. The spiked RNAs were assayed along with the 270 mouse immunology genes (see above). The amount of the synthetic RNAs spiked into the DASL assay ranged from 1×10^4 to 1×10^7 molecules, with a threefold difference for a given RNA in adjacent pools. As a control, the *bla* RNA was kept constant in all pools at 3×10^5 molecules. Each experiment was done in six replicates in order to generate statistics. This experimental design allowed us to measure assay performance with regard to dose-response and fold-difference detection over a range of target concentrations. As shown in Figure 2, the DASL assay has a dynamic range of at least 2.5 orders of magnitude (from 3×10^4 to 1×10^7 molecules or more), in which an average of 1.8-fold difference detection (range 1.3–2.4) was obtained with 95% confidence.

In interpreting these data, it is important to realize that the assay output for one target site is a function of the fraction of the amplifiable pool (total PCR template) that comprises that target. We use a fixed PCR cycle number for all samples, and the PCR reagents are shared among all amplicons in each reaction. As a result, the limit of detection for a given transcript in the DASL assay depends on its abundance relative to other transcripts being targeted in the assay. Given this caveat, data from RNA dilution studies suggest that the assay can quantitatively detect $<1 \times 10^4$ RNA molecules at ~1200-plex levels. If fewer targets are measured, better sensitivity can be achieved (see Discussion).

Table 1. DASL Performance as a Function of RNA Input

RNA input	100 ng	75 ng	50 ng	25 ng	10 ng	5 ng
% genes "on"	78	76	74	73	70	70
% genes with <2-fold difference detectable	95	80	84	82	59	41
Correlation (R^2) compared to 100 ng	1.000	0.986	0.983	0.970	0.946	0.918

Varying levels of total RNAs from A-20 and EL-4 cells were input to DASL assays. Values shown for detection and reproducibility are the average of those for each cell line. The two cell lines were compared for the proportion of genes differentially expressed as a function of input.

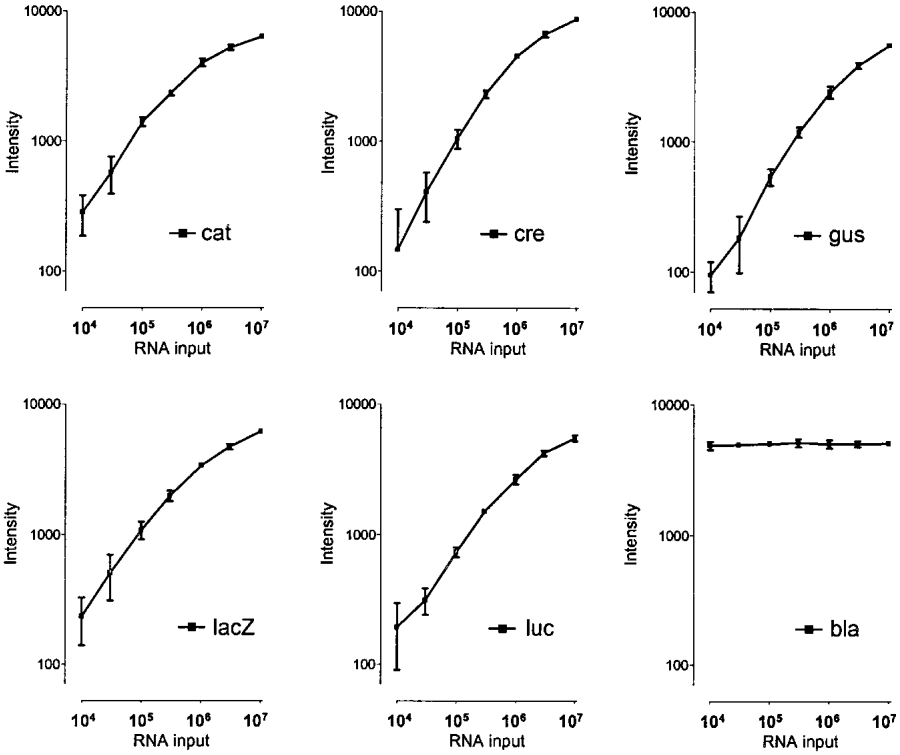

Figure 2 DASL assay performance, dynamic range and differential expression detection. Intensity data from six synthetic RNAs (*cat*, *cre*, *gus*, *lacZ*, *luc*, and *bla*) are shown as log intensity vs. log RNA input (molecules) in the DASL assay. The RNA input is given as molecules of spiked transcript per DASL reaction. The error bars represent the 95% confidence interval of intensity values for six replicate assays.

DASL Assay Is Consistent With qRT–PCR and Other Array-Based Analysis

Quantitative RT–PCR was used to validate the array experiments. The abundance of 22 housekeeping and cell-specific genes was determined in three total RNA samples (A-20, EL-4, and a 1 : 1 mixture of the two) using both methods. The DASL assay was carried out at 1212-plex and compared with single-plex real time RT–PCR with SYBR Green detection. Figure 3 shows the correlation between the log-fold differences in transcript abundance determined by qRT-PCR and DASL. Although there is some variation, the overall correlation between the two methods shows that the DASL assay can reliably report differences in expression between samples. It is worthwhile to point out that the logarithmic-fold difference in transcript abundance as determined by qRT–PCR was larger than that determined by DASL. This type of underestimating bias has been reported previously for both oligonucleotide arrays and cDNA arrays (Yuen et al. 2002).

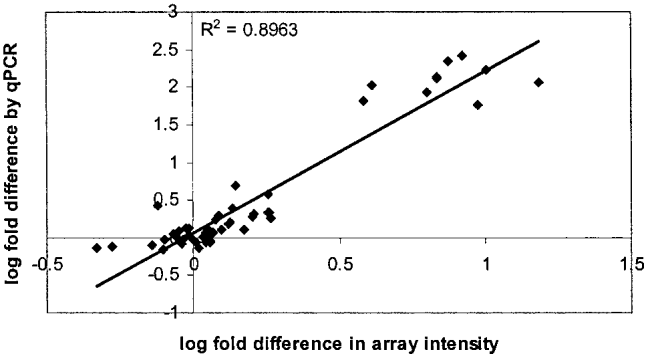

Figure 3 Correlation of DASL with qRT–PCR. The logarithmic fold difference in abundance in pairwise comparisons between A-20, EL-4, and a 1 : 1 mix of the two was estimated for expressed genes in both the DASL assay (fold difference in array intensity, *x*-axis) and qRT–PCR (fold difference in abundance derived from crossover threshold, *y*-axis).

In further experiments comparing the DASL assay to other expression techniques, we have compared expression profiles obtained with the DASL assay to those obtained with standard gene-specific probe arrays for the 270 mouse immunology genes, using a common set of RNA samples. Similar results (Speirman rank correlation = 0.83) were obtained with the two assay platforms with regard to the fold differences detected among the samples.

Cell-Specific Response to Hormone Treatment

To determine the sensitivity of the DASL assay to biologically relevant changes in expression, 212 human prostate-related genes (Dhanasekaran et al. 2001; Welsh et al. 2001) were monitored in prostate cell lines treated with androgen or a solvent control. Consistent expression patterns among the cell lines were observed in the technical replicates ($R^2 = 0.99$) and biological replicates ($R^2 = 0.98$). As shown in Figure 4A, there was no detectable impact of androgen on gene expression in cell lines DU-145, RWPE-1, and PC-3. LAPC-4 cells (Fig. 4A, bottom, left) showed a very limited response; four of the 212 genes exceeded a threshold for a significant difference (95% confidence limit). However, in LNCaP cells, 29 genes were down-regulated, and 44 genes were up-regulated as a result of androgen treatment (Fig. 4A, bottom, right). These results are consistent with the known insensitivity of PC-3 cells to androgen and the sensitivity of LNCaP cells to androgen treatment (Mitchell et al. 2000).

The assay intensity data were used to cluster the individual samples on the basis of their expression patterns for the 212 prostate-related genes. As shown in Figure 4B, differences in expression patterns measured in the DASL assay were sufficient to cluster samples from the same cell line together, regardless of androgen response (agglomerative coefficient = 0.98). For samples from the same cell type +/− androgen treatment, two patterns were observed. For cell lines DU-145, RWPE-1, PC-3, and LAPC-4, clustering did not distinguish samples according to treatment as is indicated by the insignificant cluster distance, that is, the Height shown in Figure 4B. In contrast, the androgen-sensitive cell line LNCaP showed a clear segregation. Furthermore, the LNCaP cell line was most similar in expression to the normal prostate tissue samples, consistent with published reports (Mitchell et al. 2000). Therefore, the results of the DASL analysis were concordant with previously observed biological differences, indicating that the assay is sufficiently reproducible and sensitive to sample differences to allow useful comparisons between biological samples.

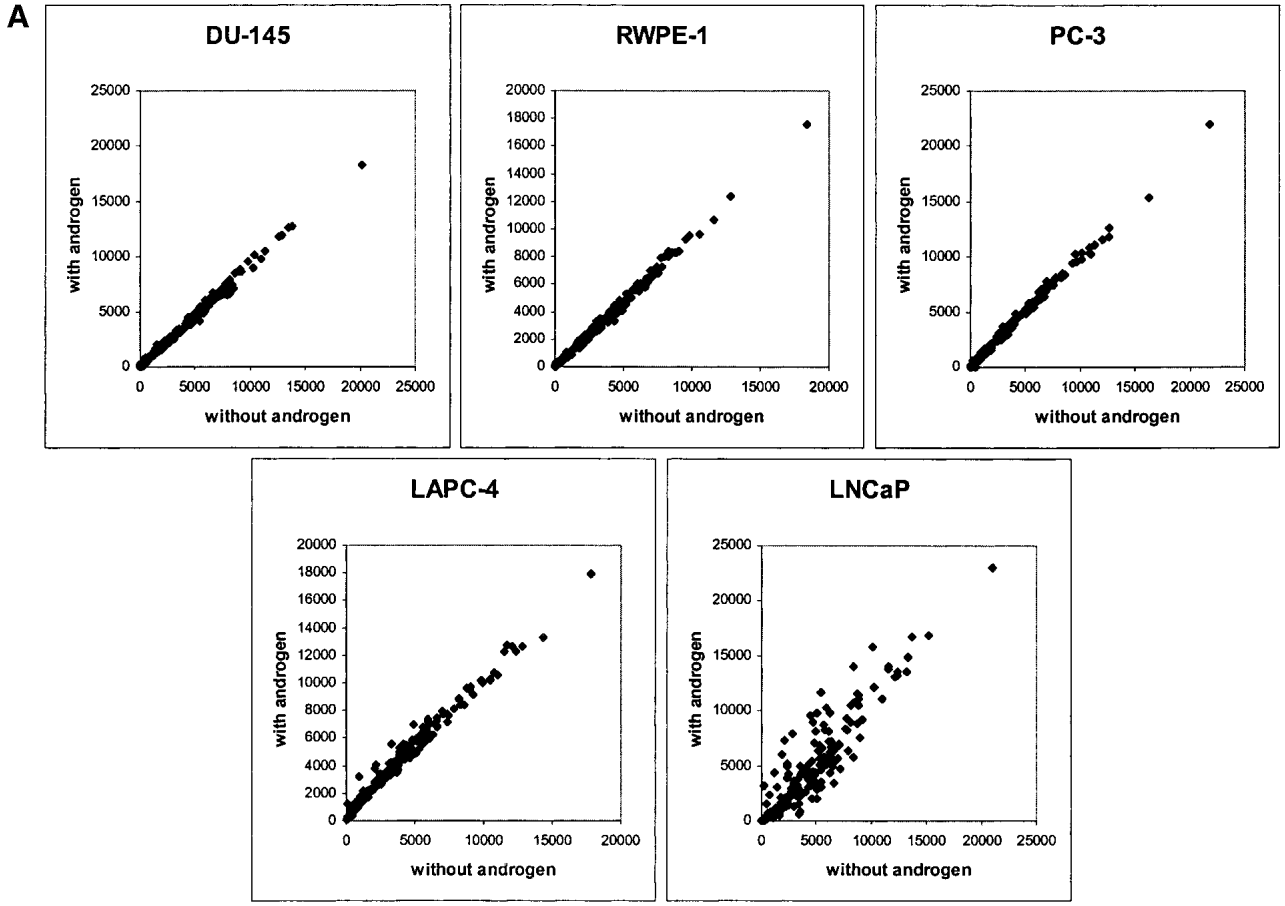

Figure 4 (Continued on next page)

Analysis of Laser-Capture Microdissected Samples

We measured gene expression of the 212 prostate-related genes in laser-capture microdissected (LCM) human prostate samples to determine whether expression data could be reliably derived from such samples (Fig. 5). LCM samples containing different cell numbers were dissected from nearby areas of a sample of benign prostatic hyperplasia. For each LCM sample, three replicate experiments were carried out, each with independent cDNA synthesis, DASL assay processing, and array hybridization. Reproducible results were obtained from as few as 22 microdissected cells' worth of RNA (R^2 = 0.97, 0.96, 0.90, and 0.94 for 176, 88, 44, and 22 cells, respectively). In addition, gene expression profiles obtained with 176, 88, 44, and 22 LCM cells correlate well (R^2 = 0.91, 0.87, and 0.76 for 176 vs. 88, 176 vs. 44, and 176 vs. 22, respectively). Further, 72% of the transcripts detected in the 176-cell samples were also detectable in the 22-cell samples. The two highest expressors in these samples were prostate-specific antigen and prostatic acid phosphatase, consistent with published reports (Mitchell et al. 2000). Varied expression levels were observed for these two genes among the triplicates with lower input (44 cells and 22 cells), consistent with the increased noise observed with low RNA inputs (Fig. 2).

Gene Expression Profiling in Degraded RNA Samples

The DASL assay uses random priming in the cDNA synthesis, and therefore does not depend on an intact poly(A) tail for T7-oligo-d(T) priming (Phillips and Eberwine 1996). In addition, the assay requires a relatively short target sequence of about 50 nucleotides for query oligonucleotide annealing, suggesting that the assay may perform well with significantly degraded RNAs. To test this possibility, RNA aliquots isolated from PC-3 and LNCaP prostate cell lines were heated at 95°C for 5, 10, 20, 40, and 60 min. The RNA was substantially degraded by heating, as shown in the Bioanalyzer traces shown in Figure 6. All RNA samples were converted to cDNA and analyzed by both the DASL assay and qPCR. The fraction of detectable transcripts remaining after degradation was estimated using qRT-PCR with three different primer pairs targeting ~180 bp fragments in the housekeeping genes EF1A1, TAX1BP1, and AGPAT1. As shown in Table 2, incubation at 95°C for 1 h resulted in the loss of ~99.9% of amplifiable species. The same cDNAs were used in a DASL assay measuring the expression of the prostate genes, and highly reproducible results were obtained, with good correlation between technical replicates (Table 2). More than 87% of the genes were detected in the PC-3 sample incubated at 95°C for 1 h compared with the intact RNA sample (Table 2). Correlation between expression profiles in intact RNA and degraded RNA samples was lower, with R^2 ranging from 0.9 to 0.6, probably due to different rates of RNA degradation for different transcripts. Very similar results were obtained for the LNCaP RNA (data not shown).

DISCUSSION

We have developed an automated method for gene expression profiling of >1200 sequence targets (400 genes at 3 probes per gene), for 96 samples in parallel. The assay has a 2.5 log dynamic range, over which an average 1.8-fold difference in RNA abun-

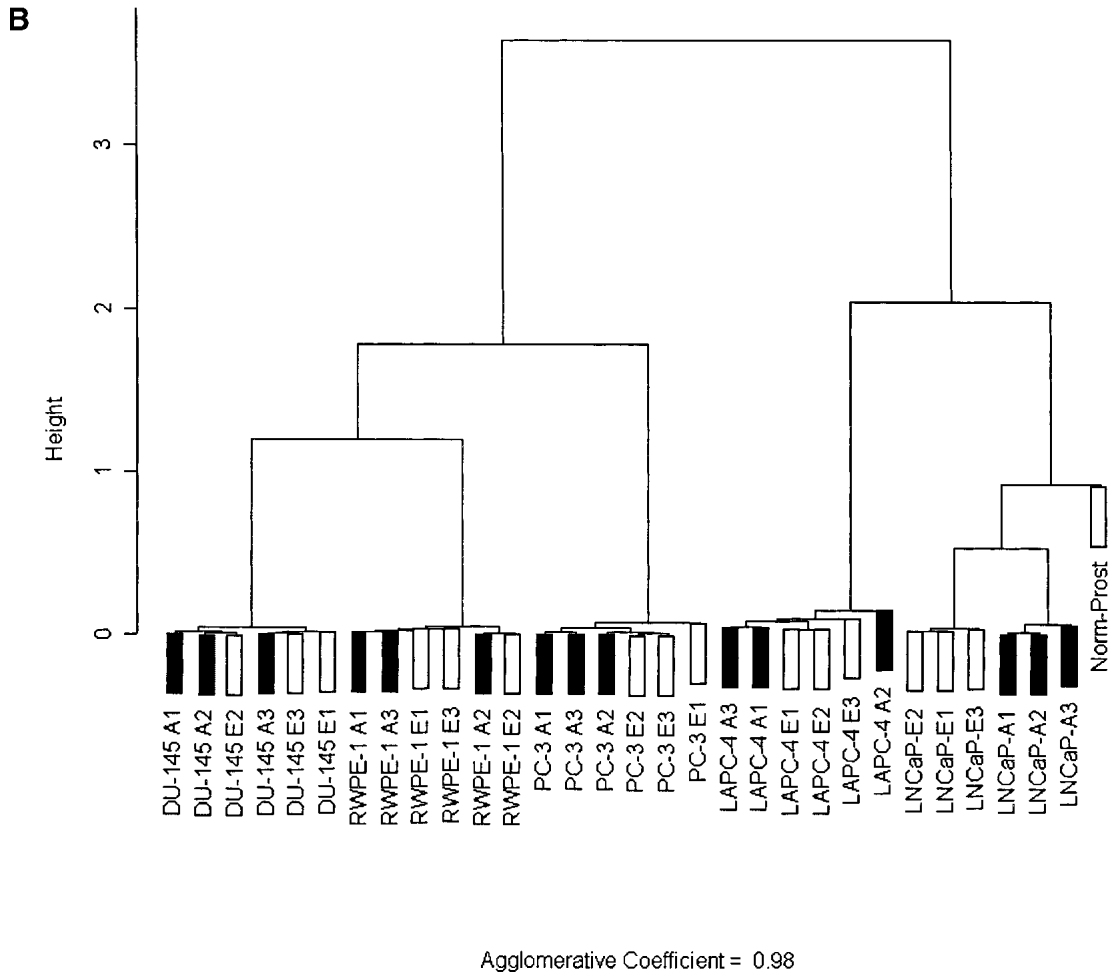

Figure 4 Cell-specific responses to androgen treatment. (*A*) Expression of 212 human prostate-related genes was monitored in prostate cell lines treated with androgen or vehicle. Data for each condition were plotted to illustrate the impact of treatment with androgen (*y*-axis) vs. vehicle (*x*-axis). (*B*) Agglomerative clustering using data from the 212 genes. For each cell line, three randomly selected biological replicates are presented for both androgen-treated (filled box) and untreated (open box) conditions. The distance between subclusters (*y*-axis, the height) measures the divergence of their expression patterns.

dance can be detected with 95% confidence. The assay is highly sensitive, permitting microarray analysis of <100 cells. Furthermore, the DASL assay on the Sentrix array matrix platform allows high-throughput monitoring of hundreds of genes in hundreds to many thousands of samples. We have used this assay for several human, mouse, *Arabidopsis*, and maize (Shou et al. 2004) studies, monitoring as many as 400 genes. In one *Arabidopsis* pathogen response study, gene expression analyses of 388 virus-responsive genes were performed on 540 RNA samples isolated from a variety of wild-type, mutant, and transgenic plants at different developmental stages. This study could have readily been scaled to thousands of samples, had that many been available.

In this study, we systematically designed 3, 4, 5, and up to 10 probes for various sets of genes, and assessed the effect of the probe number on assay quantitation. Our results showed that three optimally designed probes performed comparably to four or more probes with regard to their ability to detect expressed genes as well as differential expression among samples. Expression profiles generated with three probes correlate well with those generated with four probes ($R^2 = 0.99$). Further lowering the probe number negatively impacted assay reproducibility. Three probes per gene also allow monitoring of probe concordance in reporting gene expression changes.

Several expression assays with similar features have been described, and the published results support the capability of the DASL assay to profile gene expression (Hsuih et al. 1996; Nilsson et al. 2001) as well as alternative splicing (Yeakley et al. 2002) and allele-specific expression (Baner et al. 2003; M. Bibikova, unpubl.). Compared with other expression analysis methods, the DASL assay is most like a multiplexed version of RT–PCR, in that both assays utilize cDNA synthesis and PCR amplification. However, DASL offers much higher multiplexing capacity due to the use of common PCR primers, and the array-based readout. A key aspect of the DASL assay design is the incorporation of an address sequence, so that labeled products can be read out on a universal array. This approach offers substantial flexibility, because changing the array content is accomplished simply by reassigning the address sequences. As a result, one can refine gene sets iteratively, if desired, because no custom beads or arrays need to be developed.

Another important aspect of the DASL assay is that its sensitivity depends on the gene set. Thus, the limit of detection is more related to the distribution of targeted transcripts than to

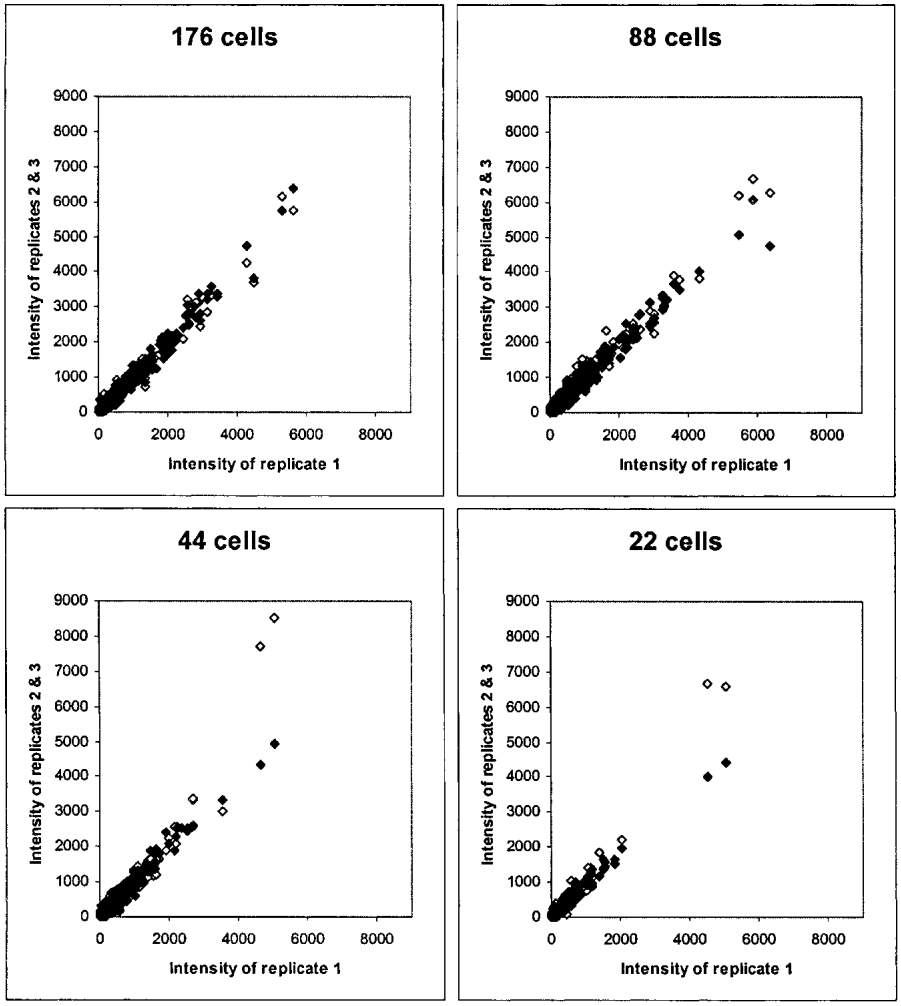

Figure 5 Reproducible gene-expression profiling of laser-captured samples. Total RNAs from laser-captured samples from different numbers of cells were assayed in triplicate. Cell number indicates the number of cells' worth of RNA input to each DASL assay. Correlation between individual replicates is shown by plotting the intensities for two replicates against the intensities of the third.

the mass of input RNA. Nonetheless, as a first approximation for ~400 genes, the assay requires fairly small amounts of starting material; 25 ng of total RNA from complex transcriptomes is sufficient, significantly less than the 5 to 10 µg amounts required by many other microarray-based methods. This level of sensitivity should be sufficient for DASL analyses of RNAs from cell-based screens in 96-well microtiter plates. Multiplexed gene expression analysis has emerged as a viable and efficient approach for high-throughput drug target validation and drug discovery (Johnson et al. 2002). We are developing protocols for expression profiling of cell lysates using the DASL assay, and preliminary data suggest that this approach should be feasible.

Given the competition for PCR reagents in the amplification step, in principle it is possible to construct query oligonucleotide sets that exceed the assay's dynamic range. For example, if a query oligonucleotide set targeted very rare transcripts together with highly abundant species, competition during PCR could squelch the rare transcripts' signal. One possibility for extending the dynamic range could be a strategy in which normalization of differentially expressed genes' representation is accomplished using different PCR primers for abundant transcripts. Our preliminary results using different T_ms for PCR primers targeting abundant transcripts and varying the annealing temperature during PCR showed that this strategy improved assay sensitivity for the less abundant transcripts (data not shown). Of course, using such a strategy requires prior knowledge about expression levels, but if necessary, query oligonucleotide pools could be designed to accommodate vastly different expression levels.

Because the DASL assay targets specific transcripts, it should be insensitive to the complexity of the untargeted RNAs. Consistent with this idea, a cell-dilution experiment showed that cell-specific expression could be detected in RNA from a mixture of one cell in 100 using the DASL assay. Further, selective targeting of transcripts specific to one cell type allowed the detection of one cell against the background of 1000 cells of a different type. These data suggest that DASL can be used to detect and analyze different cell types within a heterogeneous mixture.

A particularly attractive feature of the DASL assay is its tolerance of RNA degradation. The input RNA can be degraded to an average of 100–200 nucleotide fragments and still give robust results (Fig. 6). We have shown that the DASL assay can be used on LCM samples. The assay can also be used on RNA samples isolated from formalin-fixed, paraffin-embedded tissues, with good reproducibility ($R^2 = 0.95$; M. Bibikova, D. Talantov, E. Chudin, J. Yeakley, Y. Wang, and J. Fan, in prep.). Formalin-fixed archival tissues represent an invaluable resource for gene expression analysis, as they are the most widely available materials for studies of human disease. The ability to analyze gene expression in archived tissues will allow not only prospective but also retrospective analyses, because clinical follow up is already available. Therefore, the DASL assay used with the Sentrix array matrix may facilitate research in correlating gene expression profiles with clinical parameters, with the aim of developing biomarkers for use in disease classification and treatment.

METHODS

Assay Probe Design and Test

As shown in Figure 1, for each targeted site, two locus-specific sequences are designed for a T_m of from 57 to 62°C as described previously (Fan et al. 2003). Each target site is designed not to cross exon boundaries. Query oligonucleotides are tested for assay performance using a standard set of samples, including pooled reference RNAs (Stratagene or BD Clontech) for the species of interest. Serial dilutions of RNAs test the dose-response performance of query oligonucleotides.

RNA Samples

RNA samples were isolated from mouse cell lines (EL-4, T-cell and A-20, B-cell) and human prostate cancer cell lines (DU-145,

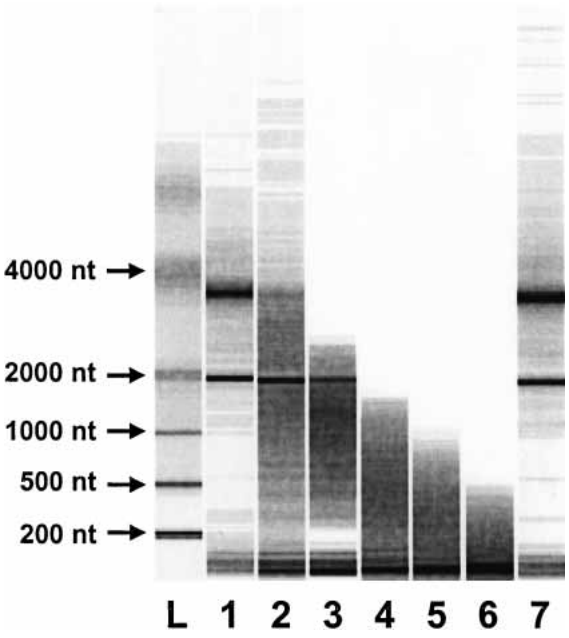

Figure 6 Agilent Bioanalyzer image of PC-3 RNA samples after incubation at 95°C for various periods of time. (L) RNA size ladder; (lanes *1–7*) intact RNA; (lanes *2–6*) 5, 10, 20, 40, and 60 min incubation at 95°C.

RWPE-1, PC-3, LAPC-4, and LNCaP), using a standard TriZol method (Invitrogen). Human prostate samples for LCM were obtained through a collaboration between the UCSD Cancer Center and X.-D. Fu. Synthetic RNAs were transcribed in vitro from plasmids containing prokaryotic gene sequences with attached poly(A) tails.

Real-Time Quantitative RT–PCR

Real-time quantitative RT–PCR analyses were performed on the ABI Prism 7900HT sequence detection system (Applied Biosystems). PCR primers were designed to amplify ~150 bp fragments. Each reaction contained 5 µL of SYBR Green PCR Master Mix (Applied Biosystems), 1 µL of cDNA template, and 250 nM each forward and reverse primer in a total reaction volume of 10 µL. The PCR consisted of an initial enzyme activation step at 95°C for 12 min, followed by 40 cycles of 95°C for 20 sec, 54°C for 20 sec, and 72°C for 30 sec. To assess the final PCR product, a melting curve was generated using a ramp from 60 to 95°C (Applied Biosystems).

cDNA Synthesis and DASL Process

RNA samples were normalized by OD_{260}. Quality testing included analysis by capillary electrophoresis using a Bioanalyzer (Agilent). For most experiments, the following processes were performed on a Tecan Genesis Workstation 150 (Tecan). First, a 20-µL reverse transcription reaction containing a reaction mix (MMC; Illumina) and total RNA (up to 1 µg) was incubated at room temperature for 10 min, and then at 42°C for 1 h. After cDNA synthesis, the remainder of the assay was identical to the GoldenGate assay (Oliphant et al. 2002; Fan et al. 2003), using Illumina-supplied reagents and conditions (BeadLab User Manual, Illumina). Briefly, the cDNA was immobilized on paramagnetic beads and washed. For most experiments, 10% of the immobilized cDNA was processed through the remaining steps. Pooled query oligonucleotides were annealed to the cDNA under a controlled hybridization program, and then washed to remove excess or mishybridized oligonucleotides. Hybridized oligonucleotides were then extended and ligated to generate amplifiable templates. A PCR reaction was performed with fluorescently labeled universal PCR primers. Single-stranded PCR products were prepared by denaturation, then hybridized to a Sentrix array matrix (Yeakley et al. 2002; Fan et al. 2003). The array hybridization was conducted under a temperature gradient program, and arrays were imaged using a BeadArray Reader 1000 scanner (Illumina; Barker et al. 2003). Cy3 labeling was used for all expression analyses.

Array Image Processing and Signal Extraction

Image analysis and data extraction software were as described previously (Oliphant et al. 2002; Galinsky 2003). Briefly, each sequence type is represented by an average of 30 beads on the array. Bead signals are computed with weighted averages of pixel intensities, and local background is subtracted. Sequence-type signal is calculated by averaging corresponding bead signals with outliers removed (using median absolute deviation).

Array Data Normalization

Each oligonucleotide pool contains sequence types designated as negative and positive controls. Negative controls are designed not to have any significant complementarity to naturally occurring sequences. Positive controls are designed to target synthetic RNAs spiked into the sample in fixed quantity. Two major algorithms were used for data normalization. The first, the cubic spline method, was used when the number of up-regulated and down-regulated genes was approximately the same for each point of the signal range (Workman et al. 2002). The normalization uses quantiles of sequence-type signals to fit smoothing B-splines. With the second algorithm, the positive controls method, normalization coefficients (a,b) were computed for each array using iteratively reweighted least-squares fit $y_v = ay_x + b$. Here, y_v, y_x are vectors of intensities of probes corresponding to positive controls on virtual (average of all arrays) and given arrays, respectively.

Expression Analysis

Detection *P*-values were computed using a normal model based on signals of negative controls. For differential expression scores, we dynamically constructed an error model assuming that rank invariant probes (<3% in relative rank change between condition and reference groups) are not differentially expressed. In addition, probes corresponding to negative controls were selected as not differentially expressed. For selected probes, we defined a vector $\overrightarrow{S}$ – of separation values for probes identified in the previous step. Here, elements of $\overrightarrow{S}$ are defined as follows:

$$S_i = \frac{|\mu_{ref} - \mu_{cond}|}{\sqrt{\frac{\sigma^2_{ref}}{n_{ref}} + \frac{\sigma^2_{cond}}{n_{cond}}}}$$

Table 2. DASL Performance as a Function of RNA Degradation Level

Duration of treatment at 95°C	0 min	5 min	10 min	20 min	40 min	60 min
% RNA detectable by RT–PCR	100	53.02	22.90	5.47	0.75	0.13
Correlation between technical replicates, R^2	1.00	1.00	1.00	0.99	0.98	0.94
Correlation with undegraded RNA, R^2	1.00	0.91	0.83	0.77	0.71	0.61
% of genes detected, compared to undegraded RNA	100	99	98	97	94	88

where μ denotes mean probe signal in the reference and condition groups, respectively. The σ denotes standard deviation associated with μ. When either the reference or condition groups contain replicate arrays, σ was computed across replicates. Otherwise, the standard deviation of individual bead intensities carrying a particular probe was used. A score for an individual probe was computed as $\mathrm{sgn}(\mu_{ref} - \mu_{cond}) \times 10 \times \log_{10}(p)$, where p is given by a normal model $N(\mu_S, \sigma_S(\bar{I}) + r)$ with μ_S being the median of $\bar{S}$, r being a regularization constant, and intensity-dependent variance σ_S being determined through the exponential fit of $|S_i - \mu_S| \sim A\exp(\lambda I)$. To compute a differential expression score for the gene, we computed scores for all probes corresponding to a given gene, and reported an average score. In addition, we reported a concordance score equal to

$$\frac{|n_+ - n_-|}{|n_+ + n_-|}$$

where $n_{+/-}$ is the number of probes showing up/down-regulation. To determine minimal resolvable fold change, we used piecewise linear approximation of intensity versus concentration. Concentration levels were considered resolvable if corresponding ranges of intensities did not overlap.

Clustering Algorithm

A matrix of correlation coefficients between array signals was computed. Agglomerative nesting was applied using the Agnes function in the R package with Ward's method.

ACKNOWLEDGMENTS

We thank Steven Barnard, Scott Butler, Diping Che, Todd Dickinson, Francisco Garcia, Todd Rubano, Chanfeng Zhao, and Lixin Zhou for their dedicated efforts and contributions to the development of the bead chemistry, imaging systems, and image analysis tools, bioinformatics, and process automation, and Tim McDaniel, Shawn Baker, and Ken Kuhn for useful discussions. This work was supported in part by grants from the National Institutes of Health (R33 CA88351 and R44 CA83398).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Abravaya, K., Carrino, J.J., Muldoon, S., and Lee, H.H. 1995. Detection of point mutations with a modified ligase chain reaction (Gap-LCR). *Nucleic Acids Res.* **23:** 675–682.

Baner, J., Isaksson, A., Waldenstrom, E., Jarvius, J., Landegren, U., and Nilsson, M. 2003. Parallel gene analysis with allele-specific padlock probes and tag microarrays. *Nucleic Acids. Res.* **31:** e103.

Barker, D.L., Theriault, G., Che, D., Dickinson, T. Shen, R., and Kain, R. 2003. Self-assembled random arrays: High-performance imaging and genomics applications on a high-density microarray platform. *Proc. SPIE* **4966:** 1–11.

Chuaqui, R.F., Bonner, R.F., Best, C.J., Gillespie, J.W., Flaig, M.J., Hewitt, S.M., Phillips, J.L., Krizman, D.B., Tangrea, M.A., Ahram, M., et al. 2002. Post-analysis follow-up and validation of microarray experiments. *Nat. Genet.* **32:** S509–S514.

Dhanasekaran, S.M., Barrette, T.R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K.J., Rubin, M.A., and Chinnaiyan, A.M. 2001. Delineation of prognostic biomarkers in prostate cancer. *Nature* **412:** 822–826.

Fan, J.B., Oliphant, A., Shen, R., Kermani, B., Garcia, F., Gunderson, K.L., Hansen, M., Steemers, F., Butler, B.L., Deloukas, P., et al. 2003. Highly parallel SNP genotyping. *Cold Spr. Harb. Symp. Biol.* **68:** (in press)

Galinsky, V.L. 2003. Automatic registration of microarray images II: Hexagonal grid. *Bioinformatics* **19:** 1832–1836.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., et al. 1999. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* **286:** 531–537.

Holloway, A.J., van Laar, R.K., Tothill, R.W., and Bowtell, D.D. 2002. Options available—from start to finish—for obtaining data from DNA microarrays II. *Nat. Genet.* **32:** S481–S489.

Hsuih, T.C.H., Park, Y.N., Zaretsky, C., Wu, F., Tyagi, S., Kramer, F.R., Sperling, R., and Zhang, D.Y. 1996. Novel, ligation-dependent PCR assay for detection of hepatitis C virus in serum. *J. Clin. Microbiol.* **34:** 501–507.

Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., et al. 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol.* **19:** 342–347.

Johnson, P.H., Walker, R.P., Jones, S.W., Stephens, K., Meurer, J., Zajchowski, D.A., Luke, M.M., Eeckman, F., Tan, Y., Wong, L., et al. 2002. Multiplex gene expression analysis for high-throughput drug discovery: screening and analysis of compounds affecting genes overexpressed in cancer cells. *Mol. Cancer Ther.* **1:** 1293–1304.

Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., and Lockhart, D.J. 1999. High density synthetic oligonucleotide arrays. *Nat. Genet.* **21:** 20–24.

Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., et al. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14:** 1675–1680.

Mitchell, S., Abel, P., Ware, M., Stamp, G., and Lalani, E. 2000. Phenotypic and genotypic characterization of commonly used human prostatic cell lines. *British Journal of Urology Int.* **85:** 932–944.

Nilsson, M., Antson, D.-O., Barbany, G., and Landegren, U. 2001. RNA-templated DNA ligation for transcript analysis. *Nucleic Acids Res.* **29:** 578–581.

Nuwaysir, E.F., Huang, W., Albert, T.J., Singh, J., Nuwaysir, K., Pitas, A., Richmond, T., Gorski, T., Berg, J.P., Ballin, J., et al. 2002. Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res.* **12:** 1749–1755.

Oliphant, A., Barker, D.L., Stuelpnagel, J.R., and Chee, M.S. 2002. BeadArray technology: Enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques* **32:** S56–S61.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., et al. 2000. Molecular portraits of human breast tumours. *Nature* **406:** 747–752.

Phillips, J. and Eberwine, J.H. 1996. Antisense RNA amplification: A linear amplification method for analyzing the mRNA population from single living cells. *Methods* **10:** 283–288.

Shou, H., Bordallo, P., Fan, J.B., Yeakley, J.M., Bibikova, M., Sheen, J., and Wang, K. 2004. Expression of an active tobacco mitrogen activated protein kinase kinase kinase enhances freezing tolerance in transgenic maize. *Proc. Natl. Acad. Sci.* **101:** 3298–3303.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., et al. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415:** 530–536.

Welsh, J.B., Zarrinkar, P.P., Sapinoso, L.M., Kern, S.G., Behling, C.A., Monk, B.J., Lockhart, D.J., Burger, R.A., and Hampton, G.M. 2001. Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. *Proc. Natl. Acad. Sci.* **98:** 1176–1181.

West, M., Blanchette, C., Dressman, H., Huang, E., Ishida, S., Spang, R., Zuzan, H., Olson, J.A., Marks, J.R., and Nevins, J.R. 2001. Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc. Natl. Acad. Sci.* **98:** 11462–11467.

Workman, C., Jensen, L.J., Jarmer, H., Berka, R., Gautier, L., Nielser, H.B., Saxild, H.H., Nielsen, C., Brunak, S., and Knudsen, S. 2002. A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol.* **3:** RESEARCH0048.

Yeakley, J.M., Fan, J.B., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M.S., and Fu, X.D. 2002. Profiling alternative splicing on fiber-optic arrays. *Nat. Biotechnol.* **20:** 353–358.

Yuen, T., Wurmbach, E., Pfeffer, R.L., Ebersole, B.J., and Sealfon, S.C. 2002. Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays. *Nucleic Acids Res.* **30:** e48.

Received November 13, 2003; accepted in revised form February 11, 2004.