A comparison of oligonucleotide and cDNA-based microarray systems

Nancy Mah,1 Anders Thelin,2 Tim Lu,1 Susanna Nikolaus,1 Tanja Kühbacher,1 Yesim Gurbuz,3 Holger Eickhoff,4 Günther Klöppel, Hans Lehra,ch, Björn Mellgård,2 Christine M. Costello,1,* and Stefan Schreiber1,*

1First Department of Medicine, Christian-Albrechts-University Kiel, D-24105 Kiel, Germany; 2AstraZeneca R&D Mömlndal, S-43183 Mömlndal, Sweden; 3Department of Pathology, Christian-Albrechts-University Kiel, D-24105 Kiel, Germany; and 4Max-Planck Institute for Molecular Genetics, D-14195 Berlin, Germany

Submitted 9 May 2003; accepted in final form 17 November 2003

Mah, Nancy, Anders Thelin, Tim Lu, Susanna Nikolaus, Tanja Kühbacher, Yesim Gurbuz, Holger Eickhoff, Günther Klöppel, Hans Lehra,ch, Björn Mellgård, Christine M. Costello, and Stefan Schreiber. A comparison of oligonucleotide and cDNA-based microarray systems. Physiol Genomics 16: 361–370, 2004. First published November 25, 2003; 10.1152/physiolgenomics.00080.2003.—Large-scale public data mining will become more common as public release of microarray data sets becomes a corequisite for publication. Therefore, there is an urgent need to clarify whether data from different microarray platforms are comparable. To assess the compatibility of microarray data, results were compared from the two main types of high-throughput microarray expression technologies, namely, an oligonucleotide-based and a cDNA-based platform, using RNA obtained from complex tissue (human colonic mucosa) of five individuals. From 715 sequence-verified genes represented on both platforms, 64% of the genes matched in “present” or “absent” calls made by both platforms. Calls were influenced by spurious signals caused by Alu repeats in cDNA clones, clone annotation errors, or matched probes that were designed to different regions of the gene; however, these factors could not completely account for the level of call discordance observed. Expression levels in sequence-verified, platform-overlapping genes were not related, as demonstrated by weakly positive rank order correlation. This study demonstrates that there is only moderate overlap in the results from the two array systems. This fact should be carefully considered when performing large-scale analyses on data originating from different microarray platforms.

IN RECENT YEARS, the use of DNA microarrays has emerged as a core technology in many research facilities. This fact is emphasized in the over 4,800 articles published in this field within the last 4 years (2000–2003), compared with about 200 citations listed in public databases (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed) in the previous decade (1990–1999). Microarray technology has been applied to such broad areas of research as tumor classification and prediction of clinical outcomes (1, 14, 23), identification of genes involved in various diseases, cellular processes, or induced responses to external stimuli (11, 13, 15), and elucidation of biological pathways (24), to name but a few.

A single microarray experiment can generate data on the expression levels of tens of thousands of genes. Generally, only a fraction of these results can be further investigated by a single workgroup. Therefore, there is a growing consensus in the scientific community of the benefit in establishing public repositories for gene expression data, analogous to freely available databases for sequence information, such as GenBank and EMBL. This endeavor has already been undertaken with the creation of public depositories for expression data, such as ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) and Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/). It is aided by the introduction of common standards for data input, annotation, information on experimental design, and data normalization, so that data is in a comprehensible format for biologists and bioinformatic specialists alike (21). Mining of the publicly available data sets can increase the confidence of a particular result (4) and conserve resources by redirecting research efforts in other directions, if meaningful data can already be extracted from existing data sets. For example, a statistical model for performing meta-analysis of independent microarray data sets has been developed using prostate cancer microarray data (17). However, as promising as meta-analysis may seem, there are few studies addressing the compatibility of data sets generated by different expression screening platforms.

Since the inception of DNA microarrays (5, 10, 18), technological advances have lead to the development of two main types of arrays, namely clone-based and oligonucleotide-based arrays (7). Both expression systems, each with different experimental designs, are routinely used to compile comparative global mRNA expression profiles of tissues or cell lines. Various technical aspects of microarrays have been examined, including probe length and composition, cross-hybridization, and hybridization effects from immobilized substrate (8, 16, 20). To date, direct comparisons between the two systems have only been carried out using publicly available data sets (9) or differential expression of defined model systems in smaller genes sets, all using fluorescent labeling methods (25). The aim of the present study was to evaluate how the performance of the two most widely used platforms compare with regard to genes identified as expressed. Since importation of data sets into the public repositories is not limited to fluorescent-labeled hybridization procedures, this study is the first to address whether fluorescent and radioactively labeled data might be comparable. To achieve our aim, total RNA extracted from human sigmoidal colon tissue was divided equally between two separate laboratories and radioactively hybridized to cDNA arrays (Human UniGene Set RZPD 1 clone set) or fluorescently hybridized to oligonucleotide arrays (Affymetrix Human Genome U95Av2 array). Results demonstrated that overlap in concordant calls made in both platforms was mod-
erate, and expression values obtained for genes represented on both arrays were not related. Incorrectly annotated clones or representation of different regions of the same gene on the arrays could not alone account for the lack of correlation between the two platforms. These results underscore a low level of compatibility between different technological platforms and raise concerns about the use of cross-platform microarray data sets to perform large-scale analysis without accounting for observed differences.

MATERIALS AND METHODS

Biopsy collection. Five patients (2 females, 3 males; age range: 55–70 yr) undergoing colonoscopy for routine cancer screening were selected for inclusion in this study (Table 1). Eight colonic mucosal biopsies were obtained from the same area in the sigmoid colon from each patient and snap-frozen in liquid nitrogen. An additional two biopsies from each patient were formalin fixed and paraffin embedded. Pathohistological examination was performed by an independent pathologist who was given no information about the patients’ health status. Clinical evaluation yielded no significant pathologic findings in this group; therefore, these patients were classified as a normal population. The study procedures were approved by the hospital ethical committee, and patients consented in writing to the additional research biopsies being taken 24 h prior to endoscopy.

Extraction of RNA. Snap-frozen biopsies were crushed to a fine powder under liquid nitrogen using a manual crusher with a Teflon head (Ronneb, Bremen, Germany), and total RNA was extracted and treated with DNase using a commercial kit (Qiagen, Hilden, Germany). RNA was eluted in RNase-free water and determined to be intact as assessed by standard formaldehyde gel electrophoresis. The presence of genomic DNA contamination was assessed by PCR amplification of GAPDH using standard PCR conditions and the following primers: GAPDH_F2, 5’- ACCACCTCCTCAAACTTTGAC-3’; GAPDH_R2, 5’-CTGTTGCGTGGACCAAACTTCTG-3’. RNA was re-treated with DNase if necessary. A complete description of the RNA extraction is available under GEO accession number GSE405. Upon confirming the quality of the RNA (intact and free of genomic DNA), 15 µg of total RNA from each of the five biological replicates was used for expression screening on either the Affymetrix or the clone-based filter platform.

Production of clone-based microarrays. Amplified PCR products from the Human UniGene Set RZPD 1 clone set (German Resource Center for Genome Research (RZPD), http://www.rzpd.de/), which consisted of ∼34,000 cDNA clones and represented 20,000 UniGene clusters (build 160), were spotted on 23 × 23-cm Hybond N nylon membranes (Amersham, Freiburg, Germany) using a Genetix robot (Genetix, München-Dornach, Germany) attached to a 384-pin gadget head (250-µm pins). Each cDNA clone was spotted in duplicate within a 6 × 6 pattern including four blank spots and two Arabidopsis thaliana (GenBank accession no. U2978S) guide spots for grid positioning. Two-hundred and fifty-eight clones representing known genes of interest were additionally spotted to complement the Human UniGene Set 1. A complete description of the clone-based microarray platform is available under GEO accession no. GPL284.

Hybridization of clone-based microarrays. Each of the five patient samples was hybridized once to the clone-based filter arrays in the following procedure. Using a commercial kit (Qiagen, 100–250 ng poly(A)⁺ RNA was isolated from 15 µg of total RNA from each patient. Reverse-transcribed target cDNA from patients was labeled with [³²P]dCTP as previously described (2), with the modification that the labeling reaction was incubated at 42°C for 1 h. RNA was hydrolyzed under alkaline conditions (0.3 M NaOH at 68°C for 20 min) and neutralized. A. thaliana cDNA (25 ng) was labeled with 50 µCi [³²P]dCTP (≈2,500 Ci/mmol, Amersham) as previously described (3). Unincorporated radiolabeled nucleotides were removed using MicroSpin G50 columns (Amersham). Prior to the first hybridization, unused filters were washed three times in 1× SSC, 0.1% SDS at room temperature for 10 min followed by three washes in 0.1× SSC, 0.1% SDS for 20 min at 80°C. Filters were prehybridized for 2 h at 50°C in hybridization solution [7% SDS, 50% formamide, 5× SSC, 2% blocking reagent (Roche Diagnostics, Mannheim, Germany), 50 mM sodium phosphate (pH 7.0)] to 100 µg/ml of A. thaliana (clone set) cDNA. Radiolabeled A. thaliana and patient cDNA were heat-denatured and cooled on ice prior to addition to prehybridized filter. Following overnight hybridization at 42°C, filters were washed three times in 1× SSC, 0.1% SDS buffer for 10 min at room temperature followed by two washes for 30 min in 0.2× SSC, 0.1% SDS at 65°C. Filters were exposed to imaging plates (BAS-MS 2325; Fujifilm, Kanagawa, Japan) for 24 h and scanned at 50-µm resolution on an imaging system (FLA-3000G, Fujifilm). There were no saturated probe data points after exposure and scanning. Image gridding was carried out using VisualGrid software (http://www.gpc-biotech.com), and the spot quantitations were imported into a custom-made Laboratory Information Management System database for normalization. A complete description of the sample labeling, hybridization, and raw output from spot quantitation software is available under GEO accession numbers GSM5994 through GSM5998.

Normalization of clone-based microarrays. Microarray data normalization was conducted in two stages: intra-array normalization and interarray normalization. Because of the comparatively large size of this membrane-based system, it was observed that some portions of the array displayed higher intensity measures than others (patchiness). Intra-array normalization removed spatial intensity measurement biases by dividing the microarray into 2,304 fields, each containing 36 spots (15 duplicated data spots, 4 background controls, and 2 positive controls for grid positioning). The minimum intensity value of the background spots in each field was used to normalize the 30 data points in that field using a simple log-transformed global mean method. Data spots for which at least one duplicate was below the filter background median were removed from analysis. Outliers (data spots where duplicate measures were not similar) were identified from a distribution of the difference between duplicates (duplicate 1 minus duplicate 2). This distribution was divided into centiles, and the tails of the distribution where centiles contained less than 10 data points were considered outliers whose values were removed from analysis.

Interarray normalization was conducted in accordance with the observation that microarray intensity distributions follow Zipf’s law (T. Lu, C. M. Costello, P. J. P. Croucher, R. Häsel, G. Deuschl, and S. Schreiber, unpublished observations). Genes on the array system were ranked according to median expression intensity over all arrays. Normalization was conducted such that the regression coefficients of the log expression intensities vs. log rank distribution for each microarray were the same as that of the median data distribution of all microarrays. A scaling factor was applied to the prenormalized data to prevent negative log values from occurring during data transforma-

Table 1. Patient data of five normal controls

<table>
<thead>
<tr>
<th>Sex</th>
<th>Age at time of study, yr</th>
<th>Histology scoring* for sigmoid colon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Patient A</td>
<td>A0, C0</td>
<td>A0, C0, L</td>
</tr>
<tr>
<td>Patient B</td>
<td>A0, C0</td>
<td>A0, C0, L</td>
</tr>
<tr>
<td>Patient C</td>
<td>A0, C0</td>
<td>A0, C0, L</td>
</tr>
<tr>
<td>Patient D</td>
<td>A0, C0</td>
<td>A0, C0, L</td>
</tr>
<tr>
<td>Patient E</td>
<td>A0, C0</td>
<td>A0, C0, L</td>
</tr>
</tbody>
</table>

*Biopsies were scored according to activity index (A0 = no activity; A1 = focal cryptitis or crypt abscess, no ulceration; A2 = disseminated cryptitis and crypt abscesses, no ulceration; A3 = ulceration), chronicity index (C0 = no chronic mucosal injury; C1 = increased lymphoplasmocytic infiltration; C2 = mucosal and submucosal fibrosis; C3 = mucosal atrophy), and additional findings (E1 = eosinophilic rich infiltrate; E2 = eosinophilic cryptitis; Ap = increased apoptosis; P = paneth cell metaplasia; L = lymphoid aggregates with germinal centers; N = nerve cell hypertrophy; M = thickened muscularis mucosa).
tion, and the same scaling factor was applied to the postnormalized data to return the values to their original magnitude. As an indicator of confidence as to whether a gene could be considered “present,” detection $P$ values were calculated for each clone probe based on the distribution of background values on each filter. Signals with a detection $P$ value < 0.05 were considered to be “present.” All relevant raw data values and normalized values are provided under GEO accession numbers GSM5994 through GSM5998.

Sequence verification of cDNA clones. Selected cDNA clones were amplified by plasmid preparation (Millipore, Billerica, MA) or PCR amplification (M13F, 5’-CGTTGTAAAAACGACGGCCAGT-3’; and M13R, 5’-TTTCACACAGGAAACACGCTATGAC-3’). Plasmids or PCR products were subjected to sequencing using BigDye Terminator chemistry on an ABI 9700 Sequencer (Applied Biosystems, Foster City, CA) using one or both of the following primers pairs: M13 For23, 5’-GACGTTGTAAAACGACGGCCAGT-3’; M13 Rev30, 5’-ATAACAATTTTCACACAGGAAACACGCTATGAG-3’; and T7, 5’-TAATACGACTCATATAGGG-3’.

Hybridization and normalization of oligonucleotide-based microarray. All five patient samples were analyzed using Affymetrix HG-U95Av2 oligonucleotide arrays, as described at http://www.affymetrix.com/products/arrays/specific/hgu95.affx. Total RNA from each sample was used to prepare biotinylated target RNA, according to the manufacturer’s protocol (http://www.affymetrix.com/support/technical/manual/expression_manual.affx). Briefly, 15 μg of total RNA was used to generate double-stranded cDNA using SuperScript reagents (Life Technologies) and a T7-linked oligo(T) primer. cRNA were synthesized using the Enzo Bioarray High Yield RNA transcript labeling kit from Affymetrix, resulting in biotinylated cRNA. Labeled cRNA were cleaned using Qiagen RNeasy kit and fragmented into 35- to 200-bp lengths using fragmentation buffer supplied by Affymetrix. Spike controls B2, bio-B, bio-C, bio-D, and Cre-x were added to the hybridization cocktail before overnight hybridization at 45°C for 16 h. Arrays were stained and washed using the EukGE-WS2 protocol (a dual staining) before being scanned on an Affymetrix scanner. None of the raw probe signals from the five hybridized arrays reached scanning saturation. Absolute analysis [Affymetrix Microarray Suite 5.0 (MAS 5.0)] was performed on the arrays using the target intensity value of 150 (global scaling, no baseline) and $\tau = 0.015$ (“$\tau$” is a default parameter that affects discrimination between match/mismatch probe pairs). Target intensity of 150 was chosen based on previous experience with Affymetrix arrays, which showed that this value typically yielded adequate scaling factors (between 0.5 and 10). In contrast to the conventional cutoff of $P < 0.05$ used for the clone-based platform, a signal for an oligonucleotide probe set was determined to be “present” if the detection $P$ value of the probe set was less than 0.06, which, for the purposes of this comparison, included “present” and “marginal” calls assigned by MAS 5.0 default criteria. The complete data sets are available under GEO accession numbers GSM5999 through GSM6003 and at http://www.mucosa.de/comparison.

Matching of clone-based PCR products to corresponding oligonucleotide probes. A detailed flowchart of the experimental setup and probe-matching procedure is outlined in Fig. 1. Two different methods were used to identify probes that theoretically represented the

Fig. 1. Flow-chart of comparison analysis. Schematic diagram shows the flow of analysis starting from hybridization of five patient samples to both microarray platforms and progressing through probe matching steps with increasing stringency.
same gene. In a more general approach, cDNA clones or Affymetrix exemplar sequences were matched to human UniGene clusters (build 160) through batch web search tools (http://genome-www5.stanford.edu/cgi-bin/SMD/source/sourceBatchSearch). As each UniGene cluster should represent a nonredundant gene (http://www.ncbi.nlm.nih.gov/UniGene/), the term “gene” and “UniGene cluster” will be used interchangeably in this paper. Probes from either system that were grouped to the same UniGene cluster were considered as “matched probes.” In cases where a gene was represented by multiple probes, gene detection was only counted once, based on the probe that gave the lowest P value, and was expressed in the highest number of patients. Using this UniGene method of matching, we identified a set of 6,645 genes to be present on both platforms.

The second approach of gene matching was more sequence specific. Full clone sequences were assembled from sequence-verified data from the 5′ and 3′ insert ends (up to 700 bp from each end). Public sequence data was used to bridge short gaps (less than 100 bp) between the ends of 5′ and 3′ sequences in cases where the clone ends did not overlap, giving rise to 272 “reconstructed” full-length clones. The corresponding probes in the Affymetrix platform were identified through BLAST homology searching in the target and probe databases available from NetAffx Analysis Center (http://www.affymetrix.com/analysis/index.affx). Probes were considered to be matched if the clone sequence had a BLAST homology score greater than 200 in the target database or if the clone sequence matched more than 10/16 probe pairs with >75% identity. All full-length clone sequences were screened for Alu repeat sequences using the Repeatmasker Web Server (http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker).

Statistical analysis of expression data. After normalization of microarray data in each respective protocol (data sets are available under GEO accession nos. GSM5994 through GSM6003), further data analysis was conducted in Microsoft Excel. The normal distribution of the mean expression data from each expression platform was tested by Kolmogorov-Smirnov test using Analyse-it software in Microsoft Excel. As the mean expression data from each platform were not normally distributed, Spearman rank order coefficients were used to determine the extent of correlation in expression levels between the two systems (19). To obtain a detection P value for each probe on the array from all five samples, individual P values for each gene were combined according to the Fisher method for combining probabilities (19).

Real-time PCR quantification. Selected genes (ACTB, CDH11, MMP1, APOA1, PBPB, CLDN4, CEACAM1, MPG, PRKCBP1, TIMM17B, MMP3, ALDOB, and MUC1) observed to be present and/or absent on the microarrays were selected for real-time PCR quantification. Primers and probes for real-time PCR were labeled with the fluorescent reporter dye 6-carboxyfluorescein (6-FAM) and the quencher dye 6-carboxytetramethylrhodamine (TAMRA) at the 5′ and 3′ ends, respectively (Eurogentec, Seraing, Belgium; or Applied Biosystems), and the sequences are supplied in Supplementary Table S1 (available at the Physiological Genomics web site). 1 Total RNA (1 μg) from each patient sample was reverse transcribed into cDNA according to manufacturer’s instructions (MultiScribe Reverse Transcriptase, Applied Biosystems). Patient cDNA was diluted 1:10, and 5 μl of the cDNA was pipetted in duplicate into a 384-well plate format. Real-time PCR was carried out using and ABI Prism 7900HT Sequence Detection System (Applied Biosystems) in the following 10-μl reaction: 0.9× Universal TaqMan Master Mix (ABI), 200 nM 6-FAM probe, 300 nM forward and reverse primers, and 5 μl cDNA diluted template. The PCR cycling profile was as follows: 2 min at 50°C, 10 min at 95°C and 40 cycles of 95°C for 15 s, 60°C for 1 min.

All patient samples for each gene assay were done in duplicate on one plate, and each gene assay was done in triplicate (except for MMP1 and TIMM17B). Real-time runs were recorded and analyzed using SDS 2.1 software (Applied Biosystems). For an absolute determination of transcript levels, the number of cycles (Ct) required to reach the threshold (set to 0.2) was taken to be a measure of transcript abundance.

RESULTS

Variation in expression signals. The overall distribution of variation in detected signals among the five patient samples was determined by calculating a scaled measure of variation (coefficient of variation) for all “present” signals in both platforms (Fig. 2). Both systems had a similar range of variation in the distribution of “present” in samples of normal individuals, as demonstrated by median and overlapping semi-quartile ranges.

To facilitate comparison between the microarray platforms, a single call (“present” or “absent”) was assigned to each probe, based on the probability of the probe’s expression signal in each of the five biological replicates. For each probe, P values were combined from each patient sample using the Fisher method, resulting in a combined detection P value. Gene probes with a combined detection P value of P < 0.05 or P < 0.06 in the clone-based or Affymetrix platforms, respectively, were defined to be called “present.” A plot of the coefficient of variation against the log10 mean expression signal illustrated the distribution of overlapping “clouds” formed by groups of “present” signals stratified by three equal groups of combined detection P values (Fig. 3). As expected, a higher range of variation was predominantly seen in samples with low expression levels within each platform. Genes called “present” in the top third percentile were expressed at higher levels than those that could be detected in the middle or bottom third percentile of combined P values.

1 The Supplementary Material for this article (Figs. S1 and S2 and Table S1) is available online at http://physiolgenomics.physiology.org/cgi/content/full/00080.2003/DC1.
To obtain an estimate of the actual numbers of genes detected by the platforms, each probe was assigned to a gene using UniGene cluster identifications (see MATERIALS AND METHODS). A total of 13,639 and 4,660 nonredundant genes were detected (based on combined detection $P$ values from all samples), representing about two-thirds of the clone-based platform and half of the Affymetrix platform, respectively.

Overlap between gene detection on both platforms. Overlap between the two microarray platforms was defined to be the proportion of calls that were concordantly called “present” or “absent” in genes represented by both platforms. From the 20,406 known and putative genes represented by the clone-based array and the 8,602 known genes represented on the Affymetrix platform, probe matching via UniGene (see MATERIALS AND METHODS) identified 6,643 genes which were represented on both systems. The 6,645 overlapping genes were classified according to the present/absent calls assigned by combined detection $P$ values (Table 2A and Fig. 1). Of the 6,645 overlapping genes on both arrays, 4,805 and 4,014 genes were called “present” in the clone-based and Affymetrix platforms, respectively, from which 3,088 genes (46%) were called “present” in both platforms. Overall, 4,002 genes (60%) were in call agreement (overlap) between the two systems. In the remaining discordant fraction, almost twice as many present calls (1,717) were made by the clone-based system compared with the Affymetrix system (926). Since discrepancies in the detection of the same genes between two different platforms could be caused by the 20–30% error rate in cDNA clone annotation (6, 22), further analysis was completed using data from only sequence-verified clones. From an in-house pool of ~2,000 partly or fully sequenced cDNA clones, 715 cDNA probes were matched to the corresponding probes on the Affymetrix array through association to UniGene clusters (Table 2B and Fig. 1). Using sequence-verified clones, the fraction of call agreement increased slightly to 64% (457 genes). Additionally, the discordant calls in the remaining 36% of genes appeared to be more evenly split between the two platforms (114 and 144 present calls in the clone-based and Affymetrix platforms, respectively) than previously observed with the non-sequence-verified genes.

As verified sequence identity only appeared to play a small role in overlap of concordant calls between the two systems, the role of probe sequence in overlap was investigated. Available fully sequenced cDNA clones (272) were matched to Affymetrix probe sets by BLAST searching. Over one-third of the probes (103) that were formerly matched to UniGene clusters no longer matched under stringent BLAST matching, due to the representation of completely different parts of the gene (i.e., probe matches did not meet cutoff criteria because probe locations were not sufficiently overlapping). About one-half of the full-length cDNA sequences (142) matched the

---

**Table 2. Distribution of UniGene cluster detection**

<table>
<thead>
<tr>
<th>Frequency of Call in Affymetrix Array</th>
<th>Frequency of call in clone-based array</th>
<th>Overlap (calls in agreement)</th>
<th>Spearman Rank Order Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present</td>
<td>Absent</td>
<td>Total</td>
</tr>
<tr>
<td>Present</td>
<td>3,088</td>
<td>1,717</td>
<td>4,805</td>
</tr>
<tr>
<td>Absent</td>
<td>926</td>
<td>914</td>
<td>1,840</td>
</tr>
<tr>
<td>Total</td>
<td>4,014</td>
<td>2,631</td>
<td>6,645</td>
</tr>
<tr>
<td></td>
<td>Present</td>
<td>Absent</td>
<td>Total</td>
</tr>
<tr>
<td>Frequency of call in clone-based array</td>
<td>384</td>
<td>114</td>
<td>498</td>
</tr>
<tr>
<td>Present</td>
<td>144</td>
<td>73</td>
<td>217</td>
</tr>
<tr>
<td>Absent</td>
<td>528</td>
<td>187</td>
<td>715</td>
</tr>
</tbody>
</table>

Assignment of UniGene cluster is based on non-sequence-verified clone annotation. *Based on median expression signal of 3,088 present calls in both platforms. †Based on median expression signal of 384 present calls in both platforms. ‡$P < 10^{-14}$, assuming a binomial distribution of probability.

---

**Fig. 3. Distribution of coefficient of variation and mean expression in present signals. Coefficient of variation and log_{10} mean expression of present signals are shown for the clone-based and Affymetrix platforms. Varying $P$ values of probe detection in the clone-based platform are colored according to top (cyan), middle (purple), or bottom (brown) third percentiles of the detection $P$ values. Similarly, the Affymetrix platform values are separated by top (dark blue), middle (pink), and bottom (yellow) third percentiles of the detection $P$ values. The dynamic range of the samples measured extended 3 logs in the clone-based system (1:564 arbitrary units) and 4 logs in the Affymetrix system (1:2,092 arbitrary units).**

---

William A. Sadow, Honoria M. Sadow, and Cheri L. Williams

Physiol Genomics • VOL 16 • www.physiolgenomics.org
same probe, and 16% (27) of the fully sequenced clones matched to different probes, for a total of 169 BLAST-matched probes (Table 3A, Fig. 1). Detailed analysis of the differences in detection of the same genes represented between platforms indicated that genes which were present in the cDNA platform were represented by cDNA clones containing Alu repeats in 33 cases (24%) of clone present calls, whereas none of the absent clone calls contained Alu sequences. When Alu-repeat-containing clone sequences were excluded from the analysis, 67% of the calls for 137 genes were in agreement (Table 3B). This would suggest that although overlap may be slightly improved by increased matching stringency, the same genes on the two systems are still not consistently detected in the remaining 33% of cases, even if the same region of the gene is represented on both arrays.

To further investigate the discordant calls from both platforms, the role of combined detection P value cutoffs was examined in the subset of probes matched by BLAST searching (137 probes, Table 3B). The distribution of the discordant calls shows that decreasing the combined P value cutoff to more stringent values did not increase the level of agreement between the platforms (Fig. 4). For example, increasing the log10 P value cutoff to −2 increases the number of genes called absent in both platforms, but also increases the number of discordant calls. At this cutoff, the net effect is an overlap of 65%. In the MAS 5.0 analysis of Affymetrix data, it is possible to influence the stringency of the calls by changing the Tau parameter. Lowering the Tau value increases the number of present calls, at the risk of creating false positives, whereas raising the Tau value decreases the number of present calls, at the risk of calling more false negatives. Decreasing the Tau value to 0.0 did increase the overlap to 72%, but increasing Tau to 0.06 decreased the overlap to 58%. These results were primarily achieved by shifting the Affymetrix P values to values above or below the cutoff (Supplementary Figs. S1 and S2). Looking at the distribution of the data points, it is clear that most of the discordant calls are not borderline present/absent calls, but rather, genes whose expression in clone-based but not Affymetrix platform, the cDNA probe sequences were chimeric with 28S rRNA. Cross-hybridization was also a potential problem for two genes that were highly polymorphic or belonged to a family of highly similar genes. However, in the remaining discordant cases, no irregularities in probe representation were observed.

**Comparison of expression levels between platforms.** Spearman rank order correlation coefficients were used to determine the relation between gene expression levels in the two platforms. Probes matched by UniGene cluster were ranked according to their median expression values in the sequence-verified subsets of probes (Fig. 1). In the first instance, the calculation was restricted to sequence-verified, UniGene-matched probes that were called “present” in both systems. The Spearman rank order coefficient (rs) was 0.131, based on 384 genes matched by UniGene cluster (Fig. 5, Table 2B). The correlation did not improve in the BLAST-matched subset expressed in clone-based but not Affymetrix platform, the cDNA probe sequences were chimeric with 28S rRNA. Cross-hybridization was also a potential problem for two genes that were highly polymorphic or belonged to a family of highly similar genes. However, in the remaining discordant cases, no irregularities in probe representation were observed.

![Fig. 4. One-hundred thirty-seven BLAST-matched genes, without Alu repeats. Data were generated using combined detection P value cutoff of 0.05 for clone-based system and P value cutoff of 0.06 and Tau = 0.015 for the Affymetrix system. Genes present on both platforms (blue diamonds), absent in both platforms (pink triangles), absent in clone-based but present in Affymetrix platform (yellow squares), and present in clone-based but absent in Affymetrix platform (cyan crosses) were plotted according to log10 of the combined detection P values.](http://physiolgenomics.physiology.org/)

Table 3. Distribution of BLAST-matched full-length clones with and without Alu repeats

<table>
<thead>
<tr>
<th>Frequency of Call in Affymetrix Array</th>
<th>Present</th>
<th>Absent</th>
<th>Total</th>
<th>Spearman Rank Order Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: BLAST-matched full-length clones with Alu repeats</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency of call in clone-based array</td>
<td>Present</td>
<td>102</td>
<td>31</td>
<td>133</td>
</tr>
<tr>
<td></td>
<td>Absent</td>
<td>23</td>
<td>13</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>125</td>
<td>44</td>
<td>169</td>
</tr>
<tr>
<td>B: BLAST-matched, full-length clones without Alu repeats</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency of call in clone-based array</td>
<td>Present</td>
<td>79</td>
<td>22</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>Absent</td>
<td>23</td>
<td>13</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>102</td>
<td>35</td>
<td>137</td>
</tr>
</tbody>
</table>

*Based on median expression signal of 102 present calls in both platforms. †Based on median expression signal of 79 present calls in both platforms. ‡P < 1.6 × 10−6, assuming a binomial distribution of probability. §P < 3.7 × 10−3, assuming a binomial distribution of probability.
(rs = −0.015), when it was further limited to 102 genes that were called present in both systems (Table 3A). In the last subset, rank order coefficients were calculated for BLAST-matched probes without Alu repeat sequences and called present in both platforms (79 genes, Table 3B). Despite elimination of cDNA probes with potential for Alu cross-hybridization and limitation to concordantly called present genes, the rank orders of the mean expression values obtained in both platforms displayed weak positive correlation (rs = 0.289). As a final test, rank order coefficients were calculated based on the rank order of individual patient expression levels within each platform for 79 genes called present in both platforms (Table 3B). In both platforms, patients were ranked by expression level for each gene, such that each of the 79 genes had its own rank order correlation. The rank order correlations ranged from strongly negative (rs = −1.00) to strongly positive (rs = 0.80), with a median and semi-quartile range of rs = −0.10 ± 0.45. These results imply that the expression values from the two technologies did not correlate with regard to median expression values or with respect to rank order of individual patient expression values within a platform.

Independent verification of transcript levels in the five patient samples was carried out by quantitative real-time PCR on 13 selected genes with varying calls between the two platforms (Fig. 6). Results indicate that for this small subset of genes, the rank order correlation of the median expression values amongst patient samples was strongly positive between the Affymetrix platform and real-time PCR (rs = 0.884). In contrast, Spearman rank order coefficients showed high correlation when only present genes were considered between clone-based and Affymetrix platforms (rs = 0.900, 5 genes), clone-based array and real-time PCR (rs = 0.733, 9 genes), and Affymetrix array and real-time PCR (rs = 0.857, 7 genes).

Independent verification of transcript levels in the five patient samples was carried out by quantitative real-time PCR on 13 selected genes with varying calls between the two platforms (Fig. 6). Results indicate that for this small subset of genes, the rank order correlation of the median expression values amongst patient samples was strongly positive between the Affymetrix platform and real-time PCR (rs = 0.884). In contrast, Spearman rank order coefficients showed high correlation when only present genes were considered between clone-based and Affymetrix platforms (rs = 0.900, 5 genes), clone-based array and real-time PCR (rs = 0.733, 9 genes), and Affymetrix array and real-time PCR (rs = 0.857, 7 genes).
and real-time PCR methods \( (r_s = 0.884) \) and only moderately to weakly positive between clone-based array and real-time PCR \( (r_s = 0.593) \) or clone-based and Affymetrix array \( (r_s = 0.384) \). When the rank order correlation was calculated only for genes called present in either Affymetrix or clone-based arrays, the correlation improved between the clone-based and Affymetrix platforms \( (r_s = 0.900) \) and between the clone-based and real-time PCR platforms \( (r_s = 0.733) \) but slightly decreased between the real-time PCR and Affymetrix platforms \( (r_s = 0.857) \).

Expression levels from real-time PCR and overall detection within the array platforms were somewhat consistent. In general, highly expressed genes tended to be present on both array platforms, and the lowly expressed genes were absent in both platforms. However, in two cases, an elevated expression signal from the clone-based platform could have been caused by a cDNA chimeric with 28S rRNA \( (MGP) \) or a spliced-in Alu repeat \( (PRKCBP1) \). A diminished expression signal from the Affymetrix platform could have been caused by a probe set which, unlike all the other 12 probe sets, was located in the middle of the gene \( (PRKCBP1) \), rather than in the last 3’ exon, which incurs a slight labeling bias over the 5’ end of the gene. Comparison of the gene sequences detected by each system showed that for each of the 13 genes, all three methods were capable of detecting the same splice variant. For the most part, the sequences used for clone-based array probes and real-time PCR probes or clone-based probes and Affymetrix probes were overlapping, but Affymetrix and real-time PCR probes were not necessarily overlapping, although they often detected parts of the same exon in over half of the genes tested.

**DISCUSSION**

Microarray-based expression screening is now widely used to quantify gene expression levels on a high-throughput basis (7). With the exponential increase of microarray experiments, and systematic gene expression profiling initiatives in the public domain, mining of data sets achieved by different platforms by meta-analyses will become inevitable. The establishment of public repositories for expression data sets offers many advantages to investigators, including 1) the gradual assembly of gene expression profiles from various tissues, cell types etc.; 2) reduction in array experiments, if sufficient data can be extracted from existing data sets; 3) increased confidence in results due to independent analysis by additional groups; and 4) improved data mining options such as the establishment of links to other genomic databases to increase knowledge of gene functions and networks. In principle, microarray platforms should detect genes in a quantitative manner. Although guidelines for submitting microarray data sets are pending (1), few studies have examined whether the results from different technology platforms are indeed compatible. In this study, the two examples from the most commonly used microarray platforms, cDNA clone- or oligonucleotide-based arrays, were compared using standardized protocols and the same pool of starting total RNA from human colonic biopsies. To our knowledge, this study is the first to compare fluorescent and radioactively labeled platforms, probed with complex human tissue, rather than more homogenous sources such as cell lines.

Both sets of microarrays showed similar distributions in terms of mean expression and coefficient of variation, with the exception that the Affymetrix mean expression values extended an extra magnitude over the clone-based data. This is due, in part, to signal amplification by streptavidin/biotin complexes used in this technology. Relative variation in probe detection for the biological replicates on both systems, as measured by coefficient of variation, did not significantly differ between the two platforms. Taken together, these descriptive parameters show that despite the different detection methods employed by the platforms (cDNA clones vs. oligonucleotide probe sets or radioactive vs. fluorescent detection), the data display similar characteristics.

Using full-length clone sequences and taking advantage of publicly released Affymetrix probe sequences, we carried out in-depth analysis of the probe sequences and their corresponding expression levels. The results clearly show that there was no rank order correlation of either 1) mean expression values from matched probes between both arrays, irrespective of the probe-matching methods (through UniGene or BLAST search of Affymetrix probe database); or 2) individual expression values for each patient within a platform. Moreover, the overlap between the two platforms was inconsistent for platform-overlapping genes; i.e., only 64–68% of full-length, matched probe sequences were concordantly called on both platforms. Lack of correlation could not be attributed to clone annotation errors or representation of different regions of the gene by array probes, suggesting the involvement of other factors.

All stages of the microarray experiment, from platform design, experimental conditions, spot quantitation, and further data processing, to the final gene expression values, may potentially contribute to differences in expression results. The Affymetrix design bases its quantitation value on the difference in fluorescence between the 16 match and mismatch probe sets, whereas the clone-based system uses the hybridization signal from a single clone spotted in duplicate. Each method has its disadvantages. The use of a difference between match and mismatch as a quantitative measure is potentially vulnerable to single base changes due to polymorphism or sequencing errors in the original sequence used for oligonucleotide design. As it is possible for a mismatched oligonucleotide to produce 5–54% signal of a perfect match (16), cross-hybridization to mismatch oligonucleotides may result in an underestimation of gene expression. On the other hand, cDNA arrays are less sensitive to single base pair changes in the probe sequence because the spotted probe is longer \( (200–1,500 \text{ bp}) \), but they are more open to cross-hybridization (i.e., gene families) and may contain latent nonspecific sequence (i.e., repetitive elements) since the cDNA clones are not completely sequenced. Compared with the Affymetrix arrays, which are directly synthesized on glass under controlled reaction conditions (12), cDNA probes may be subject to concentration variations due to printing effects and PCR amplification efficiency. The region of the gene represented by the probe may affect detected expression levels if the probes matched solely by UniGene cluster association detect different parts of the gene that are absent or poorly represented in target DNA, possibly due to different splicing or 3’ bias in target labeling. Taken together, there are many opportunities for platform-specific biases to render cross-platform data incomparable.
To date, there are few studies that specifically examine results from oligonucleotide arrays and cDNA arrays, and all studies compare fluorescently labeled platforms only. One such study evaluated the performance of cDNA and oligonucleotide arrays using 47 selected genes. Compared with quantitative real-time PCR (SYBR green), the 47 genes correlated on both array platforms in terms of direction of change (25). In a more comprehensive comparison study, both single channel (and ratio-transformed) measurements of 2,895 matched genes from 56 cell lines were analyzed using publicly available data from Affymetrix and cDNA platforms (9). There was poor correlation between the two platforms, and the authors suggested that probe-specific factors (GC-content, sequence length, average signal intensity and cross-hybridization) had differential influences in measurement. Collectively, previous studies have indicated no clear consensus as to the comparability of expression profiles from different platforms, using either ratio-transformed or single channel expression data.

This study was limited to the analysis of baseline quantitation of biological replicates and does not attempt to compare the arrays’ ability to detect changes in gene transcript levels. However, comprehensive analysis of the basic technique of hybridization alone shows that there are many factors influencing microarray compatibility at a basal level, some of which may also influence a differential gene analysis. Typically, microarrays are used to identify differentially expressed genes in an experimental setup that includes a group of arrays from two conditions (control and test condition), with the test and control samples being done on the same array (in the case of two-color hybridization) or copies of the same array (as in the case of single-channel hybridizations). Usually, the differences are expressed as fold changes or log differences for a single gene, rather than differences between different genes. Differences in a gene’s expression, not explicit levels of gene expression, are the final values reported. In theory, ratios should be a more reliable output than baseline expression signals, because the ratio of normal to control samples reduces the effect of systematic platform variables such as spotted probe concentration, printing effects, and probe sequence and length.

It is worth noting that the aim of the present study was not to definitively determine whether one array platform outperformed the other. As suggested previously (9), this might only be possible if a third source of data were available. As a cursory inquiry, we quantitated gene transcript levels of 13 genes with fluorogenic real-time PCR and found that the rank order of the expression levels displayed a strong positive correlation (r = 0.750–0.900) between all three transcript quantitation methods for genes called “present.” It would not be prudent to extrapolate these conclusions to larger data sets based on such a small number of real-time PCR assays. To determine the better platform, one would have to independently verify a large, comprehensive set of genes (possibly hundreds of genes), an endeavor not lightly undertaken by research groups, because of the high cost involved. Rather, this study pointedly compared an Affymetrix and a cDNA platform using microarray data that would typically be released upon publication. Contrary to other published comparisons, which matched cross-platform probes by gene name or homology to expressed sequence tags (ESTs), this study eliminated such variables as sequence ambiguity by basing the analysis on sequence-verified clones and fully sequenced clones, an important detail that was not adequately addressed in previous studies. No other microarray comparison has matched cross-platform microarray probes using the complete and exact probe sequence printed on the array. Additionally, this study did not restrict the analysis to a few select genes of interest (i.e., highly upregulated genes or those known to be involved in a specific process), but rather chose arbitrary genes in a non-hypothesis-driven fashion. Since large-scale analysis may involve whole gene sets, it is imperative to know whether cross-platform microarray data correlates as a whole, not just on a few select genes.

In summary, the results of this study indicate that oligonucleotide-based arrays, such as those produced by Affymetrix, and full-length clone-based arrays may be too different in experimental design to be expected to give global expression results that can be directly correlated. This suggests that microarray technologies should not be used as an absolute quantitation method and that pooling of global expression profiles from different microarray platforms for the purposes of large-scale data mining should be undertaken with caution. The observation that there is only moderate overlap and no correlation in the expression data warrants the simultaneous use of complementary approaches to obtain a complete expression profile in complex tissue.

ACKNOWLEDGMENTS

We acknowledge the technical assistance of Brigitte Mauracher and Nicola Dierkes, input from Georg Watzig and Robert Häser, and the bioinformatic support of Nikolaos Sifas and Stefan Wachter.

Present address of H. Eickhoff: Scienion AG, D-12489 Berlin, Germany.

GRANTS

This study was supported by AstraZeneca R&D Möln达尔, the German National Genome Research Network, the German Competence Network for Inflammatory Bowel Diseases, and the EU Research and Technological Development Program (Genetics of Inflammatory Bowel Disease; QLG-2-CT-2001-02161).

REFERENCES


