

Patient-based cross-platform comparison of oligonucleotide microarray expression profiles

Joerg Schlingemann^{1,*}, Negusse Habtemichael^{2,*}, Carina Ittrich³, Grischa Toedt¹, Heidi Kramer¹, Markus Hambek⁴, Rainald Knecht⁴, Peter Lichter¹, Roland Stauber² and Meinhard Hahn¹

¹Division of Molecular Genetics, Deutsches Krebsforschungszentrum, Heidelberg, Germany;

²Chemotherapeutisches Forschungsinstitut Georg-Speyer-Haus, Frankfurt am Main, Germany; ³Central Unit Biostatistics, Deutsches Krebsforschungszentrum, Heidelberg, Germany and ⁴Department of Otorhinolaryngology, Universitätsklinik, Johann-Wolfgang-Goethe-Universität Frankfurt, Frankfurt, Germany

The comparison of gene expression measurements obtained with different technical approaches is of substantial interest in order to clarify whether interplatform differences may conceal biologically significant information. To address this concern, we analyzed gene expression in a set of head and neck squamous cell carcinoma patients, using both spotted oligonucleotide microarrays made from a large collection of 70-mer probes and commercial arrays produced by *in situ* synthesis of sets of multiple 25-mer oligonucleotides per gene. Expression measurements were compared for 4425 genes represented on both platforms, which revealed strong correlations between the corresponding data sets. Of note, a global tendency towards smaller absolute ratios was observed when using the 70-mer probes. Real-time quantitative reverse transcription PCR measurements were conducted to verify expression ratios for a subset of genes and achieved good agreement regarding both array platforms. In conclusion, similar profiles of relative gene expression were obtained using arrays of either single 70-mer or multiple short 25-mer oligonucleotide probes per gene. Although qualitative assessments of the expression of individual genes have to be made with caution, our results indicate that the comparison of gene expression profiles generated on these platforms will help to discover disease-related gene signatures in general.

Laboratory Investigation (2005) 85, 1024–1039. doi:10.1038/labinvest.3700293; published online 23 May 2005

Keywords: comparative study; expression profiling; head and neck cancer; oligonucleotide microarray; oligonucleotide probes; reproducibility of results; statistics and numerical data

Today, researchers can choose from a broad variety of methods for global transcriptional profiling. Among the different technical approaches, microarray technology has gained a premier position. In principle, microarrays can be produced either by robotic printing ('spotting') of DNA on a chemically modified glass surface,¹ or by *in situ* synthesis of oligonucleotides via custom phosphoramidite chemistry using either photolithography on a silane-

reacted quartz substrate² or ink-jet technology on a hydrophobic glass support.³

Spotted arrays usually contain cDNA-specific PCR amplicons (cDNA arrays), ranging from several hundred to a few thousand basepairs in size. Generally, no more than one amplicon is used to probe a given gene. Although they are technically challenging and require both optimized protocols⁴ and workflow,⁵ cDNA arrays are typically produced by individual research groups or core facilities. Alternatively, they can be purchased from several commercial suppliers. But after the discovery of frequent discrepancies in the annotation of cDNA clones,⁶ investigators began to realize potential drawbacks of this highly advocated technology. *In situ* synthesis of oligonucleotide probes requires sophisticated equipment for photolithography and solid phase chemistry, which is usually too complex

Correspondence: Dr M Hahn, PhD, Division of Molecular Genetics, Deutsches Krebsforschungszentrum, INF 280, D-69120 Heidelberg, Germany and Dr R Stauber, PhD, Chemotherapeutisches Forschungsinstitut Georg-Speyer-Haus, Paul-Ehrlich-Str. 42-44, D-60596 Frankfurt am Main, Germany.
E-mails: m.hahn@dkfz.de, stauber@em.uni-frankfurt.de

*These two authors contributed equally to this study.

Received 21 February 2005; revised 8 April 2005; accepted 9 April 2005; published online 23 May 2005

and elaborate for an academic environment. A widespread commercial implementation of this technology is the Affymetrix GeneChip platform,² which currently uses 11–16 pairs (11 for the arrays used in this study) of perfect-match and single-base-mismatch 25-mer oligonucleotides for each gene. Recently, large collections of longer oligonucleotides (50–80 bases), produced by established suppliers using conventional phosphoramidite chemistry, have become increasingly popular as probes for spotted DNA arrays. Technical advantages of oligonucleotide arrays include a constant DNA concentration across all spots and biophysically optimized sequences, reducing secondary structures, avoiding repetitive sequence motives and providing a fixed range for both T_m and length. All this accounts for more uniform, stable and predictable hybridization conditions. The overall costs for long oligonucleotide arrays will often be lower when labor and other costs associated with cDNA libraries, such as replication, amplification or sequence verification, are regarded.

Considering this diversity of approaches and the resulting technical differences, researchers are highly interested in the general accuracy and reliability of microarray data and the cross-platform comparability. Several independent methods like Northern blotting or real-time quantitative reverse transcription-PCR (RQ-PCR) have been used to validate microarray results for a small number of transcripts. Generally, there was a good agreement between the corresponding values, affirming the ability to accurately profile gene expression with array-based approaches.

Former studies also compared global expression measurements between cDNA arrays and short oligonucleotide arrays^{7,8} or SSH.⁹ Recently, Barczak *et al*¹⁰ compared results between spotted arrays of 70-mer oligonucleotides and *in situ* synthesized Affymetrix GeneChip arrays. Using RNA of a cell line and a commercial reference RNA, they found strong correlations of the corresponding data sets. Despite these studies clarifying some fundamental questions, there still remains considerable uncertainty regarding the comparability of data from clinical specimens. As this lack of understanding

constitutes a barrier, which keeps researchers from an immense amount of potentially valuable information (via efficient integration of microarray data generated on different array platforms), we conducted a comparison with tumor samples from clinical practice, which evaluates cross-platform reproducibility in a practical setting.

Materials and methods

Experimental Design

To assess the degree of concordance between expression profiles obtained with either spotted oligonucleotide microarrays made from a large collection of 70-mer probes or commercial arrays produced by *in situ* synthesis of sets of multiple 25-mer oligonucleotides per gene, we analyzed relative gene expression in a set of six human head and neck squamous cell carcinoma (HNSCC) samples vs either healthy control mucosa ($n = 4$) or lymph node metastases ($n = 2$) of the respective patients as the reference (Table 1). For the spotted 70-mer arrays, relative expression levels were calculated by averaging the normalized \log_2 -ratios of two replicate two-color hybridizations per patient, one performed with inverse assignment of fluorophores (dye swap). This procedure was used to eliminate dye-related signal correlation bias.^{11,12} For the commercial 25-mer arrays, relative expression levels were derived by subtracting normalized \log_2 -transformed probe-level data (fluorescence intensities) of two single-color hybridizations per patient, corresponding to the respective tumor and reference tissue.

Patient Characteristics and Tissue Samples

Tissue samples from six patients were obtained during 1998–2002 from patients undergoing surgical resection at the Department of Otorhinolaryngology, JW-Goethe Universität Frankfurt. All cases were diagnosed histopathologically as HNSCC and staged according to the TNM classification of malignant tumors,¹³ based on criteria recommended by the ‘Union International contre le Cancer’ (UICC) (Table 1).

Table 1 Patient and disease characteristics

Patient	Primary site	Age (years)	Sex	pT	pN	PM	Grading	Samples analyzed ^a
160	Hypopharynx	48	M	3	1	0	2	PT/N
171	Hypopharynx	58	M	3	2a	0	2	PT/M
173	Oropharynx	56	M	3	2	0	2	PT/N
180	Hypopharynx	57	M	2	3	0	2	PT/N
186	Hypopharynx	47	F	2	2	0	2	PT/N
205	Oropharynx	49	M	3	1	0	2	PT/M

^aAll cases were diagnosed histopathologically as HNSCC and staged according to the TNM classification of malignant tumors. The indicated tissues were used for gene expression profiling.

N: normal mucosa, PT: primary HNSCC, M: lymph node metastasis.

The study protocol was approved by the local ethics committee after obtaining the patients' informed consent to participate in the study, and was processed anonymously. Grade 2 HNSCC specimens, corresponding healthy control mucosa surrounding the tumor and lymph node metastases were surgically resected, immediately frozen in liquid nitrogen and stored at -80°C . The neoplastic specimens contained $>80\%$ tumor tissue and $<10\%$ necrotic debris.

RNA Extraction

Frozen tissue samples (30–50 mg) were combined with 1 ml Trizol (Invitrogen, Karlsruhe, Germany) and dispersed using an Ultra-Turrax T25 tissue homogenizer (IKA Werke, Staufen, Germany). Total RNA was extracted according to the recommendations given by the Trizol protocol and further purified on RNeasy Mini spin columns (Qiagen, Hilden, Germany). Integrity and purity of total RNA were assessed on a Bioanalyzer 2100 (Agilent Technologies, Boeblingen, Germany) using a RNA 6000 Nano LabChip Kit (Agilent) according to the manufacturer's instructions.

Preparation and Postprocessing of Spotted Oligonucleotide Arrays

Synthetic 70-mer oligonucleotides ('Human Genome Oligo Set Version 2.1'; consisting of 21 329 oligonucleotides representing human genes and transcripts plus 24 controls, as well as 'Human Genome Oligo Set Version 2.1 Upgrade', consisting of 5462 human 70-mer probes) were purchased from Operon Technologies (Cologne, Germany) and dissolved in FBNC spotting buffer (Formamide, Betaine, NitroCellulose)⁴ at $40\ \mu\text{M}$, using a MiniTrak robotic liquid handling system (Perkin Elmer, Rodgau-Juegesheim, Germany). DNA spotting was performed in duplicates on Nexterion Slide E epoxysilane coated slides (SCHOTT Nexterion, Mainz, Germany) using a VersArray ChipWriter Pro microarrayer (Bio-Rad, Munich, Germany) equipped with Stealth SMP3 Micro Spotting Pins (Telechem, Sunnyvale, USA). Spot centers were $129\ \mu\text{m}$ apart. DNA adhesion to the glass surface was accomplished by 1 h incubation at 60°C , followed by UV irradiation ($2 \times 120\ \text{mJ}/\text{cm}^2$ at 254 nm) in a Stratalinker Model 2400 UV illuminator (Stratagene). Just prior to hybridization, slides were washed for 2 min in 0.2% SDS (w/v), 2 min in ddH₂O at room temperature and 2 min in boiling ddH₂O (95°C), followed by 3 min centrifugation at 2000 r.p.m.

Target Preparation for Spotted Oligonucleotide Arrays

Fluorescent antisense cDNA was prepared from $2\ \mu\text{g}$ total RNA as described elsewhere.¹² Briefly, the

mRNA content was linearly amplified by *in vitro* transcription ('T7 amplification'). The resulting aRNA was subsequently converted to cDNA and labelled by dye-dUTP incorporation using Klenow fragment.

Hybridization of Spotted Oligonucleotide Arrays

Corresponding dye-labelled cDNA samples were combined and purified on Microcon YM-30 filter columns (Millipore, Eschborn, Germany), as previously described.¹⁴ For blocking of repetitive sequence elements, $25\ \mu\text{g}$ C₀t-1 DNA (Roche Diagnostics), $25\ \mu\text{g}$ poly-A RNA (Sigma) and $75\ \mu\text{g}$ yeast tRNA (Sigma) were added before the final washing step. Purified, dye-labeled cDNA was mixed with $120\ \mu\text{l}$ UltraHyb hybridization buffer (Ambion), agitated for 30–60 min at 60°C , then for 10 min at 70°C on a thermo mixer and subsequently applied to preheated (60°C) microarrays mounted in a GeneTAC Hybridization Station (Genomic Solutions, Ann Arbor, USA). Hybridizations were performed for 16 h at 42°C with gentle agitation. Thereafter, the arrays were automatically washed at 36°C with (i) $0.5 \times \text{SSC}$, 0.1% (w/v) SDS for 5 min; (ii) $0.05 \times \text{SSC}$, 0.1% (w/v) SDS for 3 min; (iii) $0.05 \times \text{SSC}$ for 2 min. Flow time was set to 40 s, respectively. Immediately after completion of the final washing step, the arrays were unmounted, immersed in $0.05 \times \text{SSC}$, 0.1% (w/v) Tween 20 and dried by centrifugation in 50 ml Falcon tubes (30 s at 500, 1000 and 1500 r.p.m., respectively, followed by a final step of 90 s at 2000 r.p.m.).

Scanning, Image Analysis and Data Processing of Spotted Oligonucleotide Arrays

Hybridized microarrays were scanned at $5\ \mu\text{m}$ resolution and variable PMT voltage to obtain maximal signal intensities with $<0.1\%$ probe saturation, a count ratio of 0.8–1.2 (Cy5/Cy3) and maximal congruence of histogram curves, using a GenePix 4000B microarray scanner (Axon Instruments, Union City, USA). Subsequent image analysis was performed with the corresponding software GenePix Pro 5.0. Spots not recognized by the software were excluded from further considerations. Result files containing all relevant scan data were further processed using the open source statistical software environment R (<http://www.r-project.org>)¹⁵ together with libraries (packages) of the Bioconductor project (<http://www.bioconductor.org>).¹⁶ Raw fluorescence intensities were normalized applying variance stabilization.¹⁷ To eliminate low-quality data, the data points were ranked according to spot homogeneity, as assayed by the ratio of median-to-mean fluorescence intensity, the ratio of spot-to-local background intensity and the standard deviation of the logarithmic ratios ($\log_2 \text{Cy5}/\text{Cy3}$) between spot replicates. Those data points ranked among the

lower 20% were removed from the data set. Genes that could not be quantified in more than 33% of all experiments after filtering were excluded as well. To combine the data of dye swap experiments, the \log_2 -transformed intensity ratios of one array were inverted and averaged with the corresponding values of the other array.

Target Preparation, Hybridization and Data Analysis for Affymetrix GeneChip Arrays

Total RNA (5 μg) was used to prepare biotinylated cRNAs for hybridization, following the guidelines given in the Affymetrix GeneChip Expression Analysis Technical Manual.¹⁸ cRNA clean-up was performed on RNeasy Mini filters (Qiagen). In all, 10 μg of fragmented, labelled cRNA were hybridized to Affymetrix HG U133A arrays (Affymetrix, Santa Clara, CA, USA) using standard conditions (16 h, 45°C). Arrays were washed and stained in a Fluidics Station 400 (Affymetrix) and scanned on a Gene Array Scanner 2500 (Agilent), as recommended by Affymetrix. Raw fluorescence intensities from all hybridizations were normalized applying variance stabilization¹⁷ with additional scaling. Additionally, *MAS5*¹⁹ as well as *gcRMA*²⁰ expression values were calculated.

Matching of Oligonucleotide Probe Sequences

The Bioconductor **R** package *AnnBuilder*²¹ and GenBank accession numbers, provided by Affymetrix and Operon, were used to map probe sequences to corresponding UniGene clusters (build #175). Microarray data were only used if the Affymetrix probe set and the Operon probe corresponded to the same UniGene cluster from the intersection of both platforms ($n = 4425$). For simplicity, if probe sets (Affymetrix) mapped to multiple UniGene clusters or if several probes (Operon) or probe sets (Affymetrix) mapped to the same UniGene, they were excluded from further analyses.

Statistical Analyses

To investigate the linear relationship between data points in Figure 2, regression lines were determined by minimizing the sum of squares of the Euclidean distance of points to the fitted line ('orthogonal regression'), as there is no clear assignment of dependent and independent variables. Correlations were estimated using the Pearson correlation coefficient together with its 95% confidence interval. An optional filtering procedure additionally excluded those data points considered unreliable as they correspond to probes associated with signal intensities less than two standard deviations above local background for at least one channel of the pair of Operon chips or to probe sets with mean \log_2

expression values below the median for all probe sets of the pair of GeneChips.^{10,22} We chose this strategy in order to extract high-quality data from both array platforms as a sound foundation for quantitative comparisons. More sophisticated filtering based on variance rather than absolute expression levels was not applied due to the deliberate shortage of replicates. Expression ratios of genes with a signal close to the background (low abundance) in only one of the two investigated conditions are clearly significant in a biological context. They were, however, considered less appropriate to this comparative study, as their results were expected to carry an increased and mathematically inevitable degree of variation not caused by characteristics of the investigated platforms.

Identification of differentially expressed genes was performed by empirical Bayes inference for paired data.²³ Moderated *t*-statistics, based on shrinkage of the estimated sample variance towards a pooled estimate and corresponding *P*-values, were calculated using the Bioconductor **R** package *limma*.²⁴ *P*-values were adjusted according to the method proposed by Benjamini and Hochberg²⁵ to control the false discovery rate at a level of 10%. The magnitude of the effects as well as the corresponding *P*-values are illustrated as volcano plots.²⁶

To remove systematic variation resulting from the different technical approaches of the investigated array platforms or differences in sample handling procedures between the two labs participating in this study, 'Distance Weighted Discrimination (DWD)'²⁷ was performed on normalized \log_2 -ratios from both array platforms, using Matlab software freely available at <https://genome.unc.edu/pubsup/dwd/>. Further details about cross-platform adjustment of microarray data can be obtained at <http://genome.med.unc.edu:8080/caBIG/DWDNCI60.htm> and <http://genome.med.unc.edu:8080/caBIG/paper1.pdf>. Identification of differentially expressed genes and DWD were only performed for spots with quantified \log_2 -ratios in all four primary HNSCC vs normal mucosa experiments.

Gene ontology (GO)²⁸ data mining was performed using the GOCharts functionality of the 'Database for Annotation, Visualization and Integrated Discovery (DAVID)',²⁹ which is available at <http://david.niaid.nih.gov/david/>. Overrepresentation analysis was carried out with the software application 'Expression Analysis Systematic Explorer (EASE)',³⁰ downloaded from <http://david.niaid.nih.gov/david/ease.htm>.

Real-Time PCR Analysis

For selected genes (Table 2), changes in mRNA levels detected in microarray experiments were evaluated by reverse transcription (RT) and quantitative real-time PCR analysis, using the iCycler (BioRad, Munich, Germany). In all, 1 μg of total

Table 2 Primer sequences used for RQ-PCR

Gene symbol	Forward primer (5' → 3')	Reverse primer (5' → 3')
<i>OSF2</i>	ATTAGGCTTGGCATCTGCTC	CTCGCGGAATATGTGAATCG
<i>GMDS</i>	GCGCTCATCACCGGTATCAC	CTCTGGGCTCCAAGGTTGTAG
<i>TMPRSS2</i>	TCCTGACGCAGGCTTCCAAC	CGAACACACCGATTCTCGTCC
<i>BGN</i>	TGGTTCACTGCTCCGACCTG	GGATCTCCACCAGGTGGTTC

RNA was converted to cDNA using Superscript II reverse transcriptase (Invitrogen) and oligo(dT) primer, according to the manufacturer's specifications. PCR reaction mixtures consisted of 12.5 μ l of 2 \times iQTM SYBR[®] Green Supermix (Abgene, Hamburg, Germany), 0.5 μ l of each 10 μ M target primer and 1 μ l diluted cDNA template (1:10) in a reaction volume of 25 μ l. Thermal cycling conditions comprised an initial denaturation step of 15 min at 95°C, 40 cycles of 30 s at 95°C and 30 s variable annealing/elongation temperature, depending on the respective set of target primers. dsDNA-specific fluorescence was measured at the end of each extension phase. Product-specific amplification was confirmed by a melting curve analysis. The relative expression ratio (*R*) of a target gene was calculated using the equation

$$R = \frac{(E_{\text{target}})^{\Delta CP_{\text{target}}(\text{control-sample})}}{(E_{\text{ref}})^{\Delta CP_{\text{ref}}(\text{control-sample})}}$$

based on its real-time PCR efficiencies (*E*) and the crossing point (*CP*) differences of sample vs a control, and expressed in comparison to a reference gene.³¹ The target gene expression was normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH).

Accession Numbers

All relevant data from this study are available from GEO³² (<http://www.ncbi.nlm.nih.gov/geo>) under the accession numbers GPL96 and GPL1384 (for the array platform), GSM29702-GSM29705, GSM29747-GSM29758, GSM29808-GSM29813, GSM29818 and GSM29820 (for expression data from individual arrays) as well as GSE1722 (for the experimental series).

Results

Probe Matching

We analyzed the gene expression profiles of 12 specimens obtained from six head and neck cancer patients (Table 1). Four primary HNSCC were assayed vs corresponding healthy mucosa and another two primary HNSCC vs corresponding lymph node metastases of the respective patients.

This analysis was performed both on *in situ*-synthesized Affymetrix HG-U133A arrays, containing

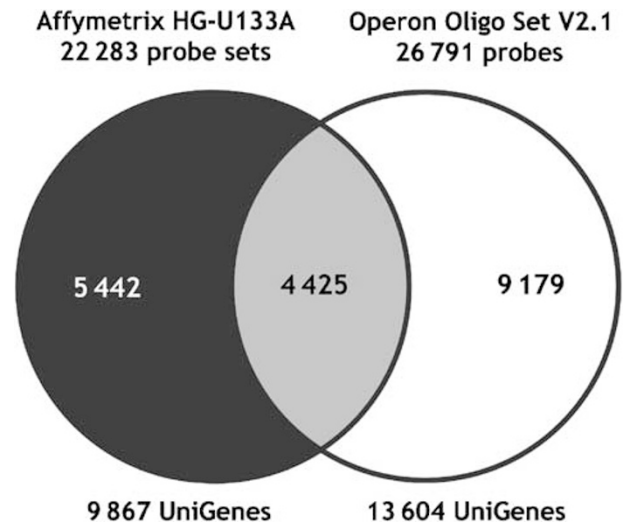


Figure 1 Intersection of probes or probe sets from the different array platforms. Probe sequences were mapped to UniGene clusters (build #175), based on GenBank accession numbers provided by the manufacturers. A total of 9867 UniGene clusters were found for the probe sets of the HG-U133A arrays, while 13 604 were retrieved for the Operon arrays. A total of 4425 genes were represented on both array types.

22 283 sets of 25-mer probes, and on spotted long oligonucleotide arrays containing 26 791 70-mer probes of the Operon Human Genome Oligo Set Version 2.1 and Version 2.1 Upgrade. A total of 9867 UniGene clusters were found for the probe sets of the HG-U133A arrays, while 13 604 were retrieved for the Operon arrays, using GenBank accession numbers provided by the manufacturers. A total of 4425 genes were represented on both platforms, as identified by consistent assignment of UniGene clusters to the corresponding probes or probe sets. We used this large set of genes as a basis for comparing expression data from the two array systems (Figure 1).

Intraplatform Reproducibility of Expression Ratios

For the platform of spotted long oligonucleotide arrays, correlations of expression ratios measured on individual arrays were $r=0.99$ for identically repeated hybridizations and $r=-0.98$ for dye swap hybridizations repeated with inverse assignment of fluorophores (data not shown). Similar correlations

had been reported for the Affymetrix system.³³ Hence, both array platforms provide highly reproducible measurements of gene expression profiles, which is an essential prerequisite for the success of a cross-platform comparison.

Cross-Platform Reproducibility of Expression Ratios

Normalized \log_2 -transformed absolute signal intensities were calculated for the arrays from both platforms using variance stabilization by *vsu*.¹⁷ For GeneChip arrays, \log_2 expression ratios were obtained by subtracting \log_2 -transformed absolute signal intensities of the two respective arrays from each patient. For the spotted long oligonucleotide array, \log_2 -ratios from two-color dye swap hybridizations were inverted and averaged. To ensure that the observed effects were not due to characteristics of the data processing algorithm, we repeated the

analyses of GeneChip arrays using background correction and normalization by *gcRMA*²⁰ as well as the *MAS5* algorithm.¹⁹ For all patients, there was a clear correlation between differential expression measurements made with either array type ($r=0.56$ – 0.76), and the correlation improved substantially ($r=0.61$ – 0.85) when measurements from probes with low-intensity signals were excluded (Figure 2). Except for patients 160 and 186, the respective regression lines all showed a slope clearly smaller than 1, indicating that, on average, absolute log ratios obtained on the Operon long oligonucleotide platform were lower than the corresponding values measured with Affymetrix arrays. The changes in correlation were marginal when *gcRMA* was used to normalize the GeneChip results. *MAS5* yielded lower correlations with unfiltered data, but the results were similar to those of *vsu* or *gcRMA* when filtered data were used (Table 3).

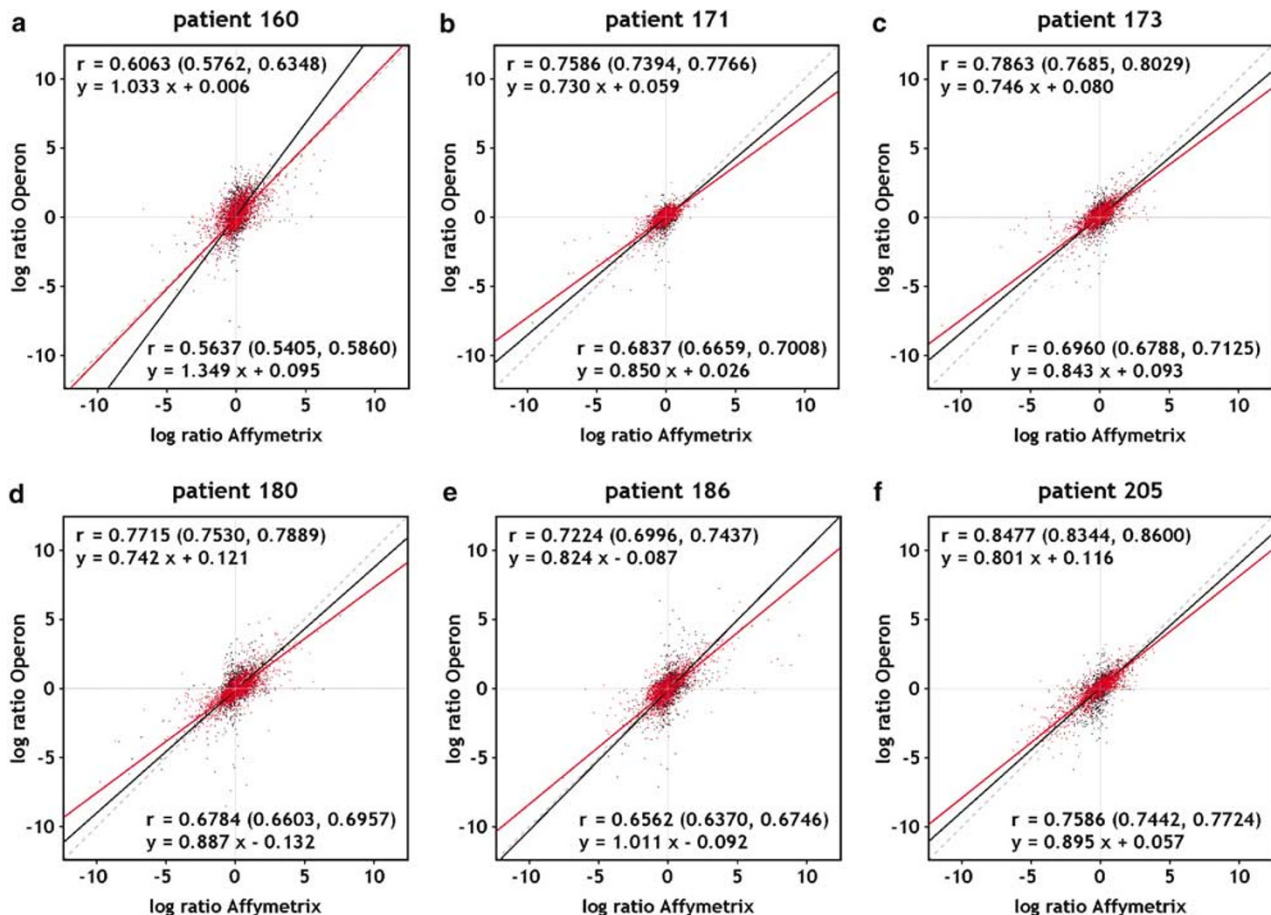


Figure 2 Scatter plots comparing normalized, \log_2 -transformed expression ratios of spotted long oligonucleotide arrays to ratios obtained with Affymetrix GeneChip short oligonucleotide arrays. For the spotted arrays, normalized ratio data from dye swap experiments were combined. For Affymetrix arrays, the ratios of normalized intensity values from corresponding arrays were used. Hybridized targets were derived from (a) patient 160, (b) patient 171, (c) patient 173, (d) patient 180, (e) patient 186 and (f) patient 205. Orthogonal regression analysis was performed to derive the regression lines shown in black (unfiltered data) and red (filtered data) as well as their respective linear equations (shown in the lower part of the plots for unfiltered data and in the upper part for filtered data). Pearson's correlation coefficients and their associated 95% confidence intervals are listed as well. Dashed lines through origin with slope 1 are displayed to accentuate the reduced slope. For panels (a)–(f), calculations were based on 3472, 3595, 3600, 3569, 3522 and 3474 data points for the unfiltered data sets as well as 1796, 2011, 1889, 1954, 1816 and 1866 data points for the filtered data sets.

Table 3 Correlation of gene expression ratios obtained with either Affymetrix GeneChip arrays or Operon long oligonucleotide arrays

Patient	Affy <i>vsn</i> vs Operon <i>vsn</i> ^a	Affy <i>MAS5</i> vs Operon <i>vsn</i> ^b	Affy <i>gcRMA</i> vs Operon <i>vsn</i> ^c
160	0.564 (0.606)	0.456 (0.571)	0.544 (0.585)
171	0.684 (0.759)	0.453 (0.707)	0.671 (0.749)
173	0.696 (0.786)	0.553 (0.783)	0.710 (0.795)
180	0.678 (0.772)	0.577 (0.763)	0.681 (0.777)
186	0.656 (0.722)	0.561 (0.712)	0.664 (0.734)
205	0.759 (0.848)	0.636 (0.837)	0.762 (0.841)

^aGeneChip results were normalized by variance stabilization (*vsn*), ^b*MAS5*, and ^c*gcRMA*. Values obtained upon removal of low-intensity signals are given in parentheses.

Table 4 Correlation of gene expression ratios before and after systematic bias correction by DWD

Patient	Affy <i>vsn</i> vs Operon <i>vsn</i> ^a	Affy <i>MAS5</i> vs Operon <i>vsn</i> ^b	Affy <i>gcRMA</i> vs Operon <i>vsn</i> ^c
160	<i>r</i> : 0.550 (0.704) [0.735] <i>s</i> : 1.343 (1.289) [1.126]	<i>r</i> : 0.431 (0.577) [0.693] <i>s</i> : 0.755 (0.899) [1.280]	<i>r</i> : 0.525 (0.678) [0.703] <i>s</i> : 1.137 (1.237) [1.083]
173	<i>r</i> : 0.686 (0.848) [0.893] <i>s</i> : 0.865 (0.906) [0.867]	<i>r</i> : 0.544 (0.701) [0.886] <i>s</i> : 0.539 (0.677) [0.911]	<i>r</i> : 0.707 (0.850) [0.896] <i>s</i> : 0.745 (0.853) [0.811]
180	<i>r</i> : 0.697 (0.785) [0.816] <i>s</i> : 0.852 (0.878) [0.828]	<i>r</i> : 0.577 (0.680) [0.806] <i>s</i> : 0.644 (0.747) [0.940]	<i>r</i> : 0.696 (0.778) [0.821] <i>s</i> : 0.767 (0.838) [0.805]
186	<i>r</i> : 0.667 (0.771) [0.830] <i>s</i> : 0.937 (0.956) [0.874]	<i>r</i> : 0.568 (0.680) [0.819] <i>s</i> : 0.647 (0.744) [0.895]	<i>r</i> : 0.674 (0.767) [0.824] <i>s</i> : 0.739 (0.809) [0.726]

^aGeneChip results were normalized by variance stabilization (*vsn*), ^b*MAS5*, and ^c*gcRMA*. Values obtained upon bias correction by DWD are given in parentheses. Values in brackets were derived after additionally removing low-intensity signals.
r: Pearson correlation. *s*: slope of the respective orthogonal regression lines.

Systematic Bias Correction by 'DWD'

As the samples were processed at different institutions and assayed using different array platforms and protocols, considerable systematic biases were expected to be manifested in the data sets as differences in gene expression patterns. In order to identify and adjust systematic biases imposed by characteristics of the different array platforms, we used the method of DWD.²⁷ Following this procedure, there was a clear improvement in the correlations of relative expression measurements (Table 4). As before, correlations obtained after normalization by *MAS5* were lower than the respective values generated with *vsn* or *gcRMA*, unless low-intensity signals were excluded from the analyses. The respective orthogonal regression lines showed little if any change in slope when the data from both platforms had been normalized by the *vsn* algorithm. Moderate changes were detected upon DWD in case *gcRMA* had been used to normalize Affymetrix data, whereas normalization by *MAS5* tended to cause more severe variation. On average, the slopes were closest to 1 when *vsn* was used to normalize the Affymetrix data and deviated the most from 1 upon normalization by *MAS5*. As expected, systematic bias correction by DWD shifted the slopes towards one in almost all cases (Table 4).

Significant Differences and Similarities

A different approach to detect differences between the two array systems is to compare the sets of genes identified as differentially expressed. In Figure 3, volcano plots²⁶ show the log₂-ratios of those 2861 genes consistently detected in the four primary HNSCC vs normal mucosa experiments and their respective *P*-values. The two platforms identified similar numbers of differentially expressed genes (≥ 2 -fold difference), both regarding raw *P*-values ($P \leq 0.001$) or FDR-adjusted *P*-values (adj. $P \leq 0.1$).²⁵ There were 45 genes identified as differentially expressed in all tumor samples on the Affymetrix platform, 53 were scored on Operon arrays, and the intersection contained 21 genes discovered on both systems (Table 5, Figure 4a). Plotting of the corresponding mean log ratios (Figure 4b) revealed that even genes scored by only one of the systems generally showed the same direction, but not the same degree of differential expression on the other. GO data mining²⁹ for 'biological process' (at level 3) assigned the majority of annotated genes from each platform to cell growth and/or maintenance as well as various metabolic pathways (Figure 5). However, using the software EASE,³⁰ which performs a statistical analysis of the GO categories assigned to the differentially expressed genes, accounting for

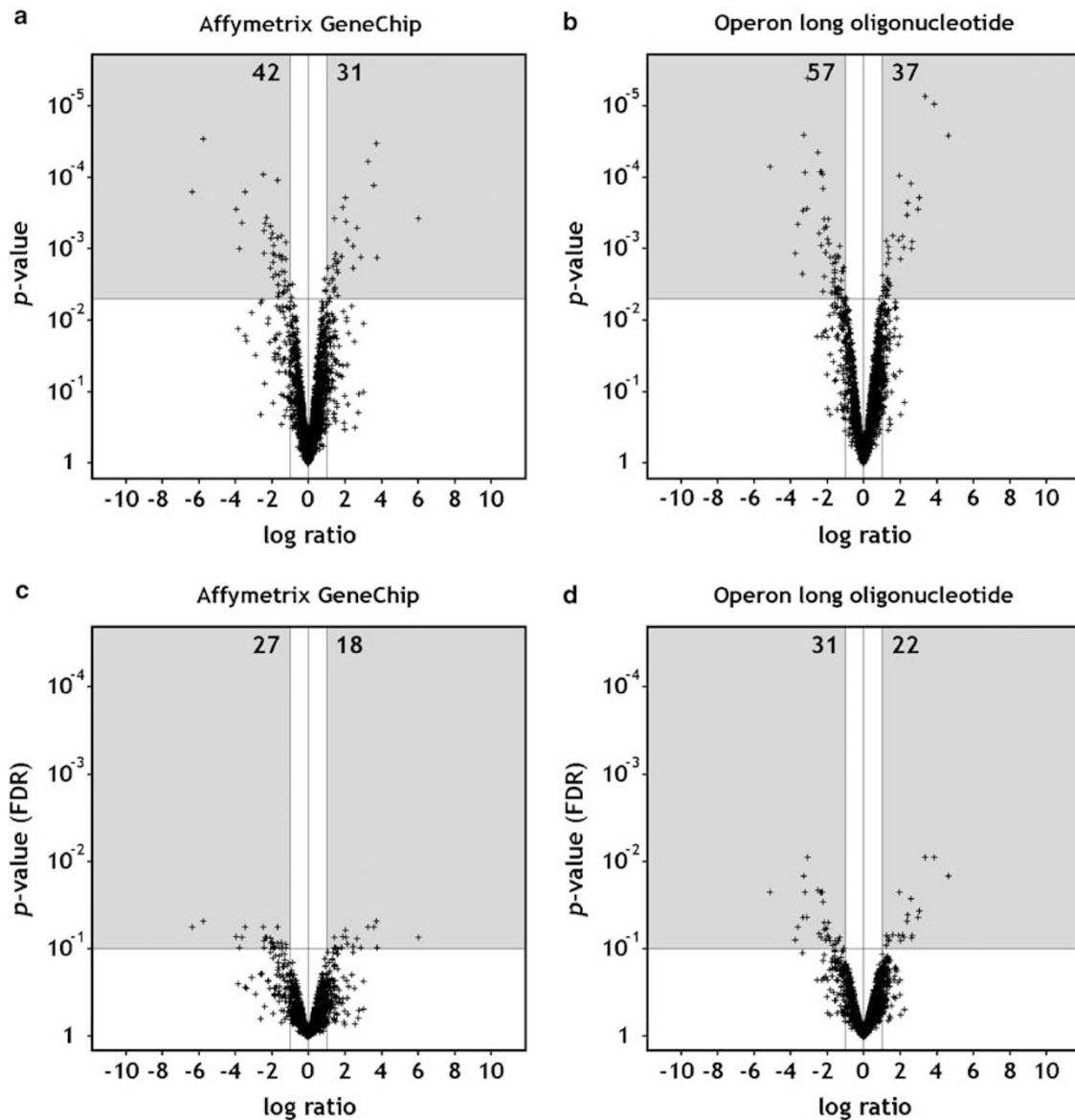


Figure 3 Volcano plots of P -values against \log_2 -transformed expression ratios. Mean \log_2 -ratios of tumor vs reference samples are shown on the x-axis. The corresponding P -values of significance, derived by empirical Bayes inference (a, b) or empirical Bayes inference and subsequent adjustment to control the FDR (c, d), are displayed on the y-axis. Results are shown for those 2861 genes consistently detected in the four primary HNSCC vs normal mucosa experiments. The plots were segmented to illustrate the relation of statistical significance ($P \leq 0.005$, adj. $P \leq 0.1$) to significance based on a two-fold change criterion. Only genes indicated by spots in the upper left and right segments of the plots satisfy both criteria, their numbers explicitly shown. Genes located in the lower left and right segments display a large fold-change but fail to achieve statistical significance. Genes found in the middle segments show no relevant difference of expression, with (upper segments) or without (lower segments) additional statistical significance associated with this observation.

the distribution of GO categories in the list of all analyzed genes to find those categories that are the most overrepresented (and can therefore be described as ‘themes’), revealed a trend towards components of the extracellular matrix for both of the platforms. Furthermore, genes involved in lipid metabolism were significantly overrepresented only among the differentially expressed genes identified on the Affymetrix system, whereas the Operon platform additionally detected genes engaged in ion binding (Table 6).

RQ-PCR Analysis

For a small subset of genes, we verified differential expression measurements by RQ-PCR analysis (Figure 6). There was good qualitative agreement between the values determined by either GeneChip arrays, Operon arrays or RQ-PCR. All platforms showed the same direction of regulated gene expression. However, the magnitude of differential expression differed considerably depending on both the experimental approach and the algorithm

Table 5 Genes scored as differentially expressed with either one or both evaluated platforms

UniGene	Gene name	Gene symbol	OMIM	Log ₂ Affy ^a	P-value Affy ^b	log ₂ Operon ^a	P-value Operon ^b	Δ exp Affy ^c	Δ exp Operon ^c
Hs.136348	osteoblast specific factor 2 (fascin I-like)	<i>OSF-2</i>		6.01	0.07	4.67	0.01	+	+
Hs.443625	collagen, type III, alpha 1	<i>COL3A1</i>	120180	1.05	0.19	3.87	0.01	–	+
Hs.75823	ALL1-fused gene from chromosome 1q	<i>AF1Q</i>	604684	3.78	0.10	2.99	0.04	+	+
Hs.28792	inhibin, beta A (activin A, activin AB alpha polypeptide)	<i>INHBA</i>	147290	3.72	0.05	–0.11	0.85	+	–
Hs.232115	collagen, type I, alpha 2	<i>COL1A2</i>	120160	3.60	0.06	0.58	0.39	+	–
Hs.528321	collagen, type V, alpha 1	<i>COL5A1</i>	120215	2.44	0.09	3.38	0.01	+	+
Hs.437173	collagen, type IV, alpha 1	<i>COL4A1</i>	120130	3.28	0.06	3.05	0.04	+	+
Hs.409602	sulfatase 1	<i>SULF1</i>		2.88	0.10	2.59	0.03	+	+
Hs.372679	Fc fragment of IgG, low affinity IIIa, receptor for (CD16)	<i>FCGR3A</i>	146740	1.10	0.40	2.67	0.07	–	+
Hs.434488	chondroitin sulfate proteoglycan 2 (versican)	<i>CSPG2</i>	118661	2.66	0.08	2.43	0.04	+	+
Hs.821	biglycan	<i>BGN</i>	301870	2.15	0.09	2.65	0.07	+	+
Hs.83354	lysyl oxidase-like 2	<i>LOXL2</i>	606663	0.10	0.86	2.39	0.05	–	+
Hs.435795	insulin-like growth factor binding protein 7	<i>IGFBP7</i>	602867	0.87	0.23	2.20	0.07	–	+
Hs.118893	Melanoma associated gene	<i>D2S448</i>	600134	2.11	0.23	2.15	0.07	–	+
Hs.408096	fragile X mental retardation, autosomal homolog 1	<i>FXR1</i>	600819	2.06	0.07	1.21	0.14	+	–
Hs.102308	potassium inwardly-rectifying channel, subfamily J, member 8	<i>KCNJ8</i>	600935	1.33	0.13	2.04	0.08	–	+
Hs.15099	Rho-related BTB domain containing 1	<i>RHOBTB1</i>	607351	2.03	0.06	0.88	0.17	+	–
Hs.122645	laminin, beta 1	<i>LAMB1</i>	150240	1.89	0.07	1.96	0.02	+	+
Hs.81988	disabled homolog 2, mitogen-responsive phosphoprotein (<i>Drosophila</i>)	<i>DAB2</i>	601236	1.48	0.26	1.93	0.07	–	+
Hs.235935	nephroblastoma overexpressed gene	<i>NOV</i>	164958	1.82	0.10	1.40	0.13	+	–
Hs.85195	myeloid leukemia factor 1	<i>MLF1</i>	601402	1.64	0.10	1.20	0.16	+	–
Hs.246875	DRE1 protein	<i>DRE1</i>		1.31	0.23	1.61	0.07	–	+
Hs.436708	Kruppel-like factor 7 (ubiquitous)	<i>KLF7</i>	604865	1.55	0.10	0.35	0.49	+	–
Hs.278469	taste receptor, type 2, member 14	<i>TAS2R14</i>		1.52	0.10	0.01	1.00	+	–
Hs.7753	calumenin	<i>CALU</i>	603420	1.48	0.15	1.27	0.07	–	+
Hs.528298	Sec23 homolog A (<i>S. cerevisiae</i>)	<i>SEC23A</i>		1.43	0.10	–0.16	0.78	+	–
Hs.433452	HEG homolog	<i>HEG</i>		1.42	0.07	0.77	0.23	+	–
Hs.179657	plasminogen activator, urokinase receptor	<i>PLAUR</i>	173391	1.15	0.23	1.41	0.08	–	+
Hs.16530	chemokine (C-C motif) ligand 18 (pulmonary and activation-regulated)	<i>CCL18</i>	603757	1.20	0.28	1.40	0.08	–	+
Hs.370774	ankyrin repeat and BTB (POZ) domain containing 2	<i>ABTB2</i>		0.65	0.36	1.38	0.07	–	+
Hs.312419	origin recognition complex, subunit 3-like (yeast)	<i>ORC3L</i>	604972	1.36	0.11	1.37	0.08	–	+
Hs.462693	zinc finger protein 22 (KOX 15)	<i>ZNF22</i>	194529	0.90	0.35	1.27	0.09	–	+
Hs.130958	ribonuclease/angiogenin inhibitor	<i>RNH</i>		0.18	0.78	–1.06	0.10	–	+
Hs.1321	coagulation factor XII (Hageman factor)	<i>F12</i>	234000	–0.02	0.98	–1.28	0.07	–	+
Hs.434933	regulator of G-protein signalling 12	<i>RGS12</i>	602512	0.26	0.67	–1.37	0.08	–	+
Hs.112028	misshapen/NIK-related kinase	<i>MINK</i>		–0.46	0.51	–1.39	0.08	–	+
Hs.5215	integrin beta 4 binding protein	<i>ITGB4BP</i>	602912	–1.41	0.10	–1.11	0.18	+	–
Hs.211556	ELOVL family member 6, elongation of long chain fatty acids	<i>ELOVL6</i>		–1.48	0.08	–0.60	0.32	+	–
Hs.40968	heparan sulfate (glucosamine) 3-O-sulfotransferase 1	<i>HS3ST1</i>	603244	–0.11	0.85	–1.53	0.08	–	+
Hs.132853	enthoprotin	<i>ENTH</i>	607265	–1.56	0.10	–0.70	0.55	+	–
Hs.528666	RAR-related orphan receptor A	<i>RORA</i>	600825	–1.56	0.10	–1.45	0.14	+	–
Hs.15519	oxysterol binding protein-like 2	<i>OSBPL2</i>	606731	–1.66	0.06	–1.03	0.17	+	–
Hs.91139	solute carrier family 1, member 1	<i>SLC1A1</i>	133550	–1.68	0.08	–1.34	0.14	+	–
Hs.378738	AHNAK nucleoprotein (desmoyokin)	<i>AHNAK</i>		–1.71	0.20	–1.51	0.09	–	+



Table 5 Continued

<i>UniGene</i>	<i>Gene name</i>	<i>Gene symbol</i>	<i>OMIM</i>	<i>Log₂ Affy^a</i>	<i>P-value Affy^b</i>	<i>log₂ Operon^a</i>	<i>P-value Operon^b</i>	<i>Δ exp Affy^c</i>	<i>Δ exp Operon^c</i>
Hs.393239	sterol-C4-methyl oxidase-like	<i>SC4MOL</i>	607545	-1.81	0.10	-0.52	0.61	+	-
Hs.77870	hypothetical protein FLJ12750	<i>FLJ12750</i>		-1.66	0.17	-1.83	0.08	-	+
Hs.212787	Microtubule associated serine/threonine kinase family member 4	<i>MAST4</i>		-1.80	0.12	-1.89	0.07	-	+
Hs.105435	GDP-mannose 4,6-dehydratase	<i>GMDS</i>	602884	-1.91	0.13	-1.68	0.09	-	+
Hs.82237	tripartite motif-containing 29	<i>TRIM29</i>		0.13	0.83	-1.93	0.05	-	+
Hs.166311	SAM and SH3 domain containing 1	<i>SASH1</i>	607955	-1.42	0.11	-1.96	0.07	-	+
Hs.1588	4-aminobutyrate aminotransferase	<i>ABAT</i>		-1.97	0.08	-0.94	0.18	+	-
Hs.424551	integral type I protein	<i>P24B</i>		-2.02	0.08	-1.56	0.08	+	+
Hs.434243	KIBRA protein	<i>KIBRA</i>		-1.21	0.10	-2.04	0.06	+	+
Hs.90797	O-acyltransferase (membrane bound) domain containing 2	<i>OACT2</i>		-1.61	0.16	-2.06	0.08	-	+
Hs.437043	KIAA0540 protein	<i>KIAA0540</i>		-1.43	0.10	-2.13	0.06	+	+
Hs.446429	prostaglandin D2 synthase 21kDa (brain)	<i>PTGDS</i>	176803	-1.93	0.10	-2.15	0.05	+	+
Hs.5541	ATPase, Ca++ transporting, ubiquitous	<i>ATP2A3</i>	601929	-1.89	0.09	-2.23	0.02	+	+
Hs.356726	scinderin	<i>SCIN</i>		-0.76	0.32	-2.26	0.02	-	+
Hs.118747	solute carrier family 15 (H+/peptide transporter), member 2	<i>SLC15A2</i>	602339	-2.30	0.07	-2.13	0.11	+	-
Hs.430324	annexin A9	<i>ANXA9</i>	603319	-1.80	0.34	-2.30	0.07	-	+
Hs.206501	hypothetical protein from clone 643	<i>LOC57228</i>		-2.35	0.07	-2.36	0.02	+	+
Hs.169238	fucosyltransferase 3 (galactoside 3(4)-L-fucosyltransferase)	<i>FUT3</i>	111100	-2.43	0.08	-1.59	0.10	+	-
Hs.348350	dehydrogenase/reductase (SDR family) member 1	<i>DHRS1</i>		-2.43	0.10	-2.20	0.03	+	+
Hs.257697	programmed cell death 4 (neoplastic transformation inhibitor)	<i>PDCD4</i>		-2.45	0.06	-1.21	0.27	+	-
Hs.282975	carboxylesterase 2 (intestine, liver)	<i>CES2</i>	605278	-1.92	0.10	-2.49	0.02	+	+
Hs.31130	transmembrane 7 superfamily member 2	<i>TM7SF2</i>	603414	-2.08	0.07	-3.05	0.01	+	+
Hs.436657	clusterin (complement lysis inhibitor, apolipoprotein J)	<i>CLU</i>	185430	-2.10	0.28	-3.11	0.04	-	+
Hs.167218	BarH-like homeobox 2	<i>BARX2</i>	604823	-1.23	0.09	-3.28	0.01	+	+
Hs.134478	RecQ protein-like 5	<i>RECQL5</i>	603781	0.26	0.65	-3.31	0.04	-	+
Hs.439309	transmembrane protease, serine 2	<i>TMPRSS2</i>	602060	-3.44	0.06	-3.19	0.02	+	+
Hs.298023	aquaporin 5	<i>AQP5</i>	600442	-1.44	0.36	-3.58	0.06	-	+
Hs.272813	dual oxidase 1	<i>DUOX1</i>	606758	-1.28	0.28	-3.72	0.08	-	+
Hs.103944	mucin 7, salivary	<i>MUC7</i>	158375	-3.77	0.10	-0.62	0.32	+	-
Hs.116651	epithelial V-like antigen 1	<i>EVA1</i>	604873	-3.96	0.07	-1.97	0.22	+	-
Hs.438862	EPS8-like 1	<i>EPS8L1</i>		-3.63	0.07	-5.11	0.02	+	+
Hs.226391	anterior gradient 2 homolog (<i>Xenopus laevis</i>)	<i>AGR2</i>	606358	-5.75	0.05	-1.33	0.14	+	-
Hs.13775	homeodomain-only protein	<i>HOP</i>	607275	-6.35	0.06	-2.42	0.07	+	+

Differentially expressed genes were selected from a subset ($n = 2861$) consistently detected in the four primary HNSCC vs normal mucosa experiments.

^a \log_2 -transformed expression ratios were averaged within each platform.

^bCorresponding P -values of significance derived by empirical Bayes inference and subsequent adjustment to control the FDR.

^cTo be scored as differentially expressed (Δexp), genes had to satisfy both statistical significance ($P \leq 0.1$) and significance based on a two-fold change criterion.

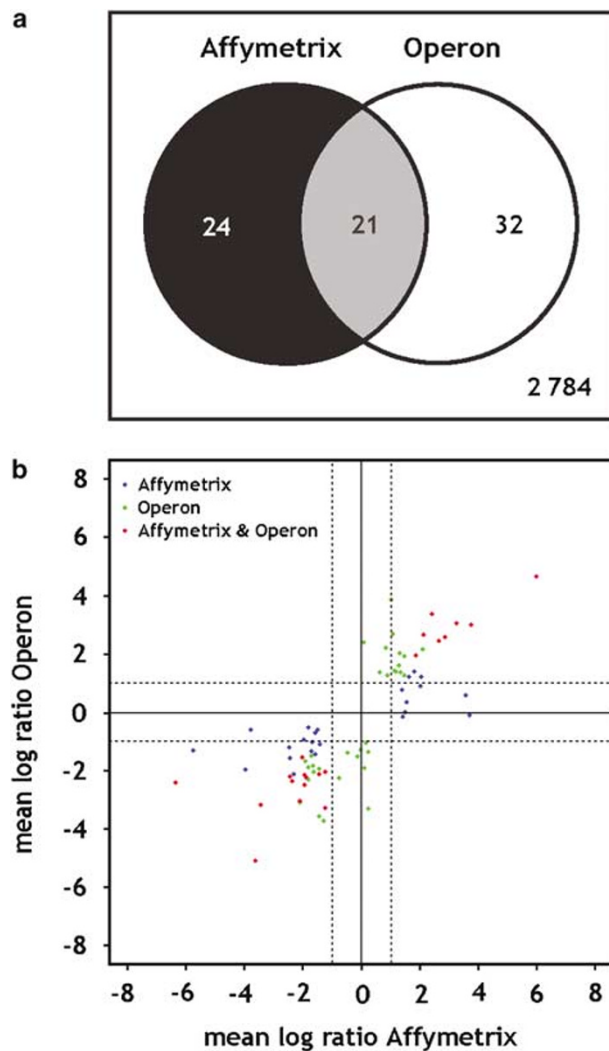


Figure 4 Summary of genes scored as differentially expressed with either one or both evaluated platforms. (a) Venn diagram showing subsets of genes that exhibit a significant differential expression with either technology, taken from a pool that contained only those genes that could repeatedly be quantified in the four primary HNSCC vs normal mucosa experiments. (b) log-log plot illustrating the relationship of the log-ratios for the 77 genes shown in (a).

applied for normalization. Firstly, GeneChip intensity measurements were transformed by variance stabilization (*vsn*), which was also used for the long oligonucleotide arrays and derives an approximately constant variance along the complete intensity range.¹⁷ Normalization was additionally accomplished employing the *MAS5* algorithm from the current version of the Affymetrix Microarray Suite software package.¹⁹ At least for the small number of genes and patients shown here, there is a tendency for higher ratios with *vsn* normalization.

Discussion

The comparison of gene expression measurements obtained with different technical approaches or

different implementations of a proven technology is of considerable interest to researchers from all fields of the biological and biomedical sciences. Several studies have addressed this topic, with rather heterogeneous results.

Recently, Mah *et al*⁸ compared absolute expression levels quantified on Affymetrix short oligonucleotide and radioactively labeled cDNA-based filter arrays. The expression values from the two technologies showed merely poor correlations. Tan *et al*³⁴ evaluated the performance of three commercial microarray platforms and found only modest correlations when comparing both absolute and relative gene expression measurements. Strikingly, log₂-ratios from the two platforms using short oligonucleotide probes and biotinylated cRNA targets (Affymetrix and Amersham; $r=0.52$) did not correlate better with each other than with those of cDNA arrays (Agilent; $r=0.53$ or $r=0.59$). In a comparison of Affymetrix GeneChip arrays and two different collections of 70-mer oligonucleotides, Barczak *et al*¹⁰ found moderate correlations of corresponding signal intensities ($r=0.56$ – 0.60), but strong correlations of respective relative expression values ($r=0.80$ without filtering, $r=0.83$ – 0.89 after exclusion of probes or probe sets with low signal intensities). Similarly, Shippy *et al*³⁵ described improved correlations between expression measurements from Affymetrix GeneChip and Amersham CodeLink arrays upon removal of genes within platform noise ($r=0.62$ vs $r=0.79$). Measuring relative gene expression values on Affymetrix short oligonucleotide arrays, commercial (Agilent) and custom-made, sequence-validated cDNA arrays, Järvinen *et al*³⁶ observed reasonable correlations of log₂-ratios. Interestingly, the correlation between the two different cDNA platforms ($r=0.73$) was weaker than the correlations between the commercial or custom-made cDNA arrays and the Affymetrix system ($r=0.84$ and 0.76 , respectively). A recent study by Tan *et al*,³⁴ showing very poor correlation between Affymetrix, Amersham and Agilent arrays, came to broad public attention³⁷ and raised general concerns regarding the comparability of expression data across labs and platforms.

A different approach to review the possibility for meaningful translation of microarray data is meta-analysis of extensive data sets of similar type, produced in different labs and on different platforms.^{38–40} Since many additional parameters such as classification of the samples or individual laboratory practices influence the outcome of these studies, the results are rather inconclusive concerning comparability on the technological level. Generally, at least common patterns and/or groups of genes could be confirmed.

Aside from meta-analyses, the above-mentioned studies were usually based on data generated with homogeneous cell lines and by averaging over several technical replicates. We intended to increase the practical significance of our evaluation by the

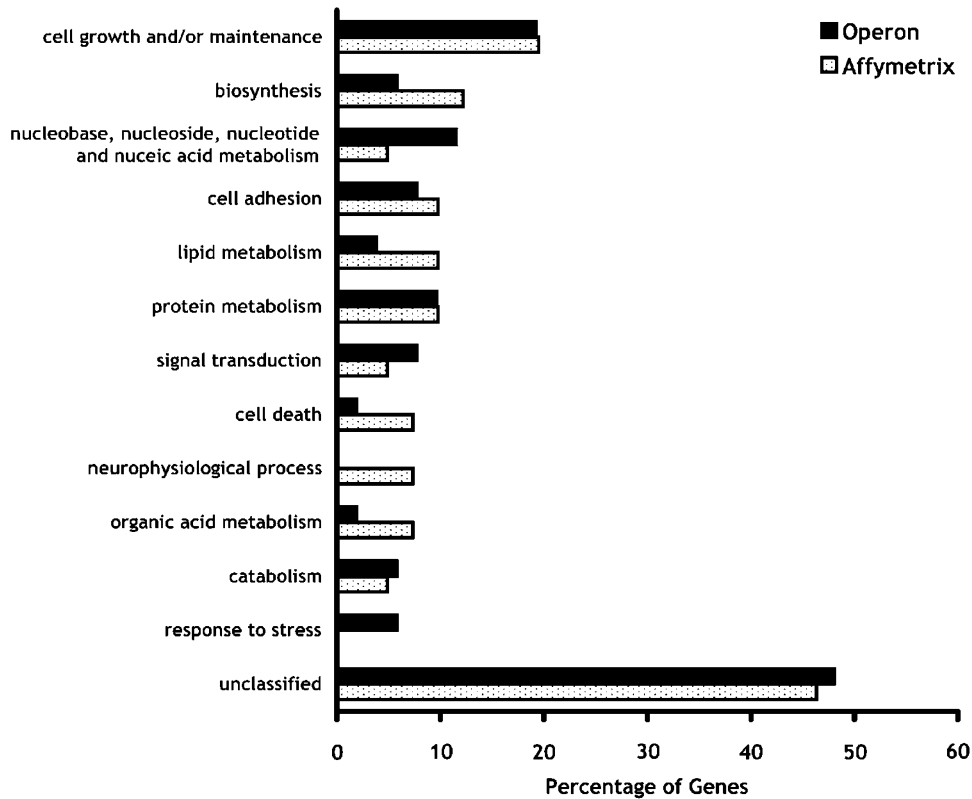


Figure 5 GO data mining. The 45 regulated genes detected with the Affymetrix system as well as the 52 regulated genes found with Operon arrays were characterized according to their biological process classification in the GO database (at level 3). Roughly half of the genes did not have a GO classification at this level. The majority of the remaining genes were involved with cell growth and/or maintenance as well as various metabolic pathways.

use of clinical samples in combination with modest technical replication (two single arrays per patient for the Affymetrix platform and two dye swap replicates per patient for the spotted oligonucleotide arrays). Additionally, we kept the protocols for target preparation as comparable as possible. Since the Affymetrix platform utilizes biotinylated cRNA generated by *in vitro* transcription (IVT), we decided to employ a linear, IVT-based amplification and labelling protocol for the spotted oligonucleotide arrays instead of the usual dye-labelling by reverse transcription. Recently, we showed that this novel procedure generates highly reproducible expression profiles with down to 2 ng of starting material.¹² We could also demonstrate that the correlation of expression ratios obtained with spotted oligonucleotide arrays is higher between replicate amplified sample pairs than between amplified and RT-labelled sample pairs or replicate RT-labelled sample pairs. Accordingly, one can expect that consistent target amplification would also be beneficial, if expression ratios are to be compared across platforms. Comparative studies that do not account for this consideration might introduce additional systematic bias, resulting in reduced agreement between platforms.

To match the probes from the two platforms, we used accession numbers provided by Affymetrix

and Operon and mapped them to the current version of the UniGene database. Although provided by both manufacturers, transcript identifiers from the RefSeq collection^{41,42} were not chosen for matching the platforms, since reference sequences can change through consolidation of the database. Recently, Mecham *et al*⁴³ showed that up to 50% of Affymetrix probes do not have a matching sequence in the current version of RefSeq. Despite these considerations, platform matching by RefSeq identifiers yielded approximately similar and partly even improved results in terms of cross-platform correlation (9922 genes could be assigned as represented on both platforms, correlations of unfiltered log₂-ratios were $r=0.66$ - $r=0.81$; data not shown). Evolution of the UniGene database (accession numbers that were removed due to misalignment or retraction by their submitters; UniGene clusters that were retired as they could be joined or split) and the associated loss of cross-references may also explain why we identified less genes common to both array types than previously reported by Barczak *et al*.¹⁰ We also decided against matching by GenBank accession numbers, since corresponding probes and probe sets can be annotated by different accession numbers of the same UniGene, causing this procedure to exclude large amounts of potentially useful information.

Table 6 EASE overrepresentation analysis of the genes listed in Table 5

System ^a	Gene category ^b	List hits ^c	List total ^d	Population hits ^e	Population total ^f	EASE score ^g	LH in PH (%) ^h	LH in Δ exp (%) ⁱ
<i>Affymetrix</i> ^j								
GO cellular component	Extracellular matrix	6	34	55	2348	0.00084	10.91	14.63
GO molecular function	Extracellular matrix structural constituent	4	37	14	2410	0.0010	28.57	9.76
GO biological process	Cell adhesion	6	37	99	2386	0.015	6.06	14.63
GO biological process	Lipid metabolism	6	37	112	2386	0.025	5.36	14.63
GO biological process	Lipid biosynthesis	4	37	44	2386	0.027	9.09	9.76
GO biological process	Cellular lipid metabolism	5	37	78	2386	0.028	6.41	12.20
GO cellular component	Endoplasmic reticulum	5	34	99	2348	0.048	5.05	12.20
GO cellular component	Extracellular matrix (sensu Metazoa)	3	34	26	2348	0.050	11.54	7.32
GO biological process	Fatty acid metabolism	3	37	26	2386	0.057	11.54	7.32
GO biological process	Steroid metabolism	3	37	27	2386	0.061	11.11	7.32
GO molecular function	Extracellular matrix structural constituent c. t. s. ^k	2	37	5	2410	0.073	40.00	4.88
GO cellular component	Extracellular region	4	34	75	2348	0.086	5.33	9.76
GO cellular component	Collagen	2	34	7	2348	0.094	28.57	4.88
<i>Operon</i> ^j								
GO cellular component	Extracellular matrix	7	45	55	2348	0.00045	12.73	13.46
GO molecular function	Extracellular matrix structural constituent	4	48	14	2410	0.0022	28.57	7.69
GO molecular function	Cation binding	10	48	192	2410	0.010	5.21	19.23
GO molecular function	Metal ion binding	10	48	225	2410	0.027	4.44	19.23
GO molecular function	Ion binding	10	48	225	2410	0.027	4.44	19.23
GO biological process	Cell adhesion	6	44	99	2386	0.031	6.06	11.54
GO molecular function	Calcium ion binding	6	48	104	2410	0.049	5.77	11.54
GO cellular component	Extracellular region	5	45	75	2348	0.050	6.67	9.62
GO molecular function	Scavenger receptor activity	2	48	3	2410	0.057	66.67	3.85
GO cellular component	Extracellular	5	45	79	2348	0.058	6.33	9.62
GO cellular component	Extracellular matrix (sensu Metazoa)	3	45	26	2348	0.084	11.54	5.77
GO biological process	Organismal physiological process	8	44	214	2386	0.085	3.74	15.38
GO cellular component	Golgi apparatus	5	45	93	2348	0.094	5.38	9.62
GO molecular function	Extracellular matrix structural constituent c. t. s. ^k	2	48	5	2410	0.094	40.00	3.85

^aSystem: the system of categorizing genes, in this case the GO classification type.

^bGene category: the specific category of genes within the classification system, in this case the GO category of the superordinate GO classification type (different levels are possible).

^cList hits (LH): number of genes in the list of differentially expressed genes that belong to the respective GO category.

^dList total: number of differentially expressed genes that could be annotated within the respective GO classification system.

^ePopulation hits (PH): number of genes in the list of all analyzed genes ($n = 2861$) belonging to the respective GO category.

^fPopulation total: number of analyzed genes with annotation data in the respective GO classification system.

^gEASE score: The upper bound of the distribution of Jackknife Fisher exact probabilities given the LH, List Total, PH and Population Total. Categories with the lowest EASE score are significantly overrepresented in the list of differentially expressed genes.

^hLH in PH: percentage of differentially expressed genes belonging to the respective category in the group of all analyzed genes in this category.

ⁱLH in Δ exp: percentage of differentially expressed genes belonging to the respective category in the group of all differentially expressed genes.

^jDifferentially expressed genes from each platform were analyzed separately.

^kc.t.s.: conferring tensile strength.

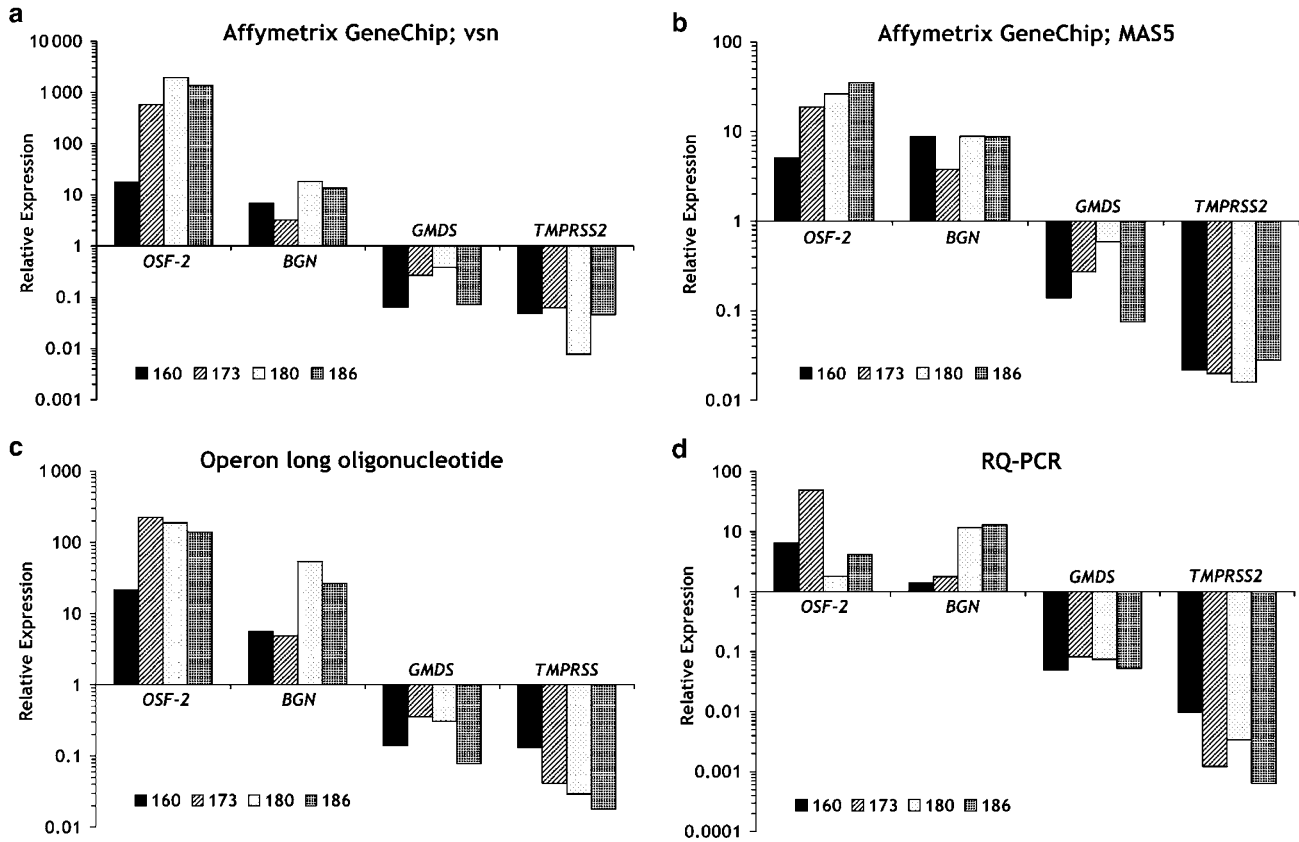


Figure 6 Comparison of relative gene expression for the genes *OSF2*, *GMD5*, *TMPRSS2* and *BGN*. Expression ratios were determined for tumor vs control tissue of the indicated patients, using either Affymetrix GeneChip arrays (a, b), Operon long oligonucleotide arrays (c) or real-time quantitative PCR analysis (d). Affymetrix ratios were either normalized by variance stabilization (a) or the MAS5 algorithm (b).

When comparing unfiltered, \log_2 -transformed expression ratios of individual patients, obtained with either GeneChip short oligonucleotide arrays or spotted long oligonucleotide arrays, we detected considerable variations in the degree of correlation (Figure 2). As reported previously,^{10,44} these correlations improved after the exclusion of probes and probe sets associated with low signal intensities. This observation might, at least in part, be attributed to variations in the performance of individual array experiments.

Systematic bias adjustment by DWD could further improve the correlations between expression ratios (Table 4). DWD is an advanced method for the adjustment of various systematic differences across microarray experiment subpopulations, including sample source, batch and platform effects,²⁷ which facilitates the merging of different data sets. DWD uses an approach similar to that of support vector machines (SVM),⁴⁵ but delivers improved performance in the context of high-dimensional, low sample size (HDLSS) data such as those obtained by microarray analyses. Both methods aim at finding a hyperplane in high-dimensional space, which separates defined subpopulations of data as completely as possible. The essential difference is that, while SVM tries to maximize the minimum distance

(margin) of all the data to the separating plane, DWD works by maximizing the sum of the inverse distances. In this way, all data points have an influence on the result (optimized position of the hyperplane), and data piling at the margins is avoided, a problem associated with the minimum distance criterion of SVM. After determination of the DWD direction vector, all data points of each subpopulation are projected onto the direction given by this vector. Finally, data points from each subpopulation are shifted in the DWD direction by subtracting the DWD direction vector multiplied by their projected means, thereby effectively removing systematic variation while preserving any variation in the DWD direction not caused by systematic effects. Applied to our data, the DWD approach clearly and consistently improved cross-platform correlations while shifting the slopes of corresponding regression lines towards 1 (Table 4). The latter effect was minimal in case both data sets had been normalized by the same algorithm (*vs**n*), as this procedure not unexpectedly yielded slopes closest to one even before DWD. A slope close to one implies that genes are more likely to yield similar results (regardless of differential expression) on both of the investigated platforms. Further improvements of DWD performance can be anticipated for more

extensive data sets, and we are confident that this method could greatly enhance agreement in future comparative studies.

Our goal was to compare reliable measurements from both systems, both of which can be regarded as detecting overlapping but different subsets of the actual set of differentially expressed genes. This was confirmed by EASE overrepresentation analysis,³⁰ which revealed that some of the differentially expressed genes could be assigned to the same 'theme' on both platforms, whereas others were exclusive to one of the platforms (Table 6). On each array system, approximately 50 genes were consistently and repeatedly scored as differentially expressed, and the intersection of these groups contained 21 common genes (Figure 4a). The majority of genes restricted to one of the platforms showed no sufficient degree and/or significance, but at least the same direction of regulated expression on the other platform (Figure 4b). Therefore, it does not matter if a clinical study uses Affymetrix or Operon long oligonucleotide arrays, as long as these are used consistently and combined with high quality control standards throughout the whole investigation.

For a subset of genes, we verified microarray-derived expression ratios by RQ-PCR and found good qualitative agreement between the two array platforms and the PCR-based method (Figure 6).

We have shown that, overall, expression profiles obtained with either long (Operon) or multiple short (Affymetrix) oligonucleotide microarrays display a reasonable correlation, with variable concordance of individual genes. Based on patient samples, we obtained results that are in good agreement with previous studies that utilized cell line-derived RNA. Projecting these findings to a larger series of array experiments, one could expect to obtain similar albeit not identical results, concerning, for example, a hierarchical clustering or a gene expression signature, with either of the two investigated platforms. On the level of individual genes and quantitative precision, however, our results reaffirm that microarrays have to be considered a screening technology and that their data should be regarded with caution. This should be kept in mind particularly when comparing data from different array platforms. Recently, important progress has been made to facilitate this transfer of information. Guidelines provided by the 'Microarray Gene Expression Data Society (MGED)' (<http://www.mged.org>), which developed the 'Minimum Information About a Microarray Experiment (MIAME)' specifications,⁴⁶ assist researchers in the annotation of their microarray experiments. Further improvement is provided by public microarray repositories, which facilitate the publication and sharing of properly annotated gene expression data. Statistical methods like DWD²⁷ can further improve the comparability of microarray data sets, since systematic biases arising from platform-specific parameters, such as measure-

ment precision (reproducibility), accuracy (regarding the 'true' values), specificity and sensitivity or differences in protocol performance, can be properly weighted and adjusted accordingly. The utility of future array studies could further improve if the 'External RNA Control Consortium (ERCC)' is successful in its effort to standardize controls for the calibration of microarray experiments. But ultimately, meaningful comparison, translation and integration of expression data will be impaired as long as industrial standards are missing for the production of arrays as well as for the design of array probes.

Acknowledgements

We thank Axel Benner for helpful discussions and critical reading of the manuscript as well as Bernd Groner for excellent support. Bjoern Tews and Silke Deckert are gratefully acknowledged for their skillful technical assistance. Our study was supported by grants of the Bundesministerium für Bildung und Forschung (FKZ KR-S07T13 to RS; NGFN-1, 01 GR 0101 to PL; NGFN-2, 01 GS 0460 and 01 GR 0418 to PL and MH) as well as the EU MolTools project (LSHG-CT-2004-503155 to PL).

References

- 1 Schena M, Shalon D, Davis RW, *et al*. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;270:467–470.
- 2 Lockhart DJ, Dong H, Byrne MC, *et al*. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 1996;14:1675–1680.
- 3 Hughes TR, Mao M, Jones AR, *et al*. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* 2001;19:342–347.
- 4 Wrobel G, Schlingemann J, Hummerich L, *et al*. Optimization of high-density cDNA-microarray protocols by 'design of experiments'. *Nucleic Acids Res* 2003;31:e67.
- 5 Kokocinski F, Wrobel G, Hahn M, *et al*. QuickLIMS: facilitating the data management for DNA-microarray fabrication. *Bioinformatics* 2003;19:283–284.
- 6 Knight J. When the chips are down. *Nature* 2001;410:860–861.
- 7 Yuen T, Wurmbach E, Pfeffer RL, *et al*. Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays. *Nucleic Acids Res* 2002;30:e48.
- 8 Mah N, Thelin A, Lu T, *et al*. A comparison of oligonucleotide and cDNA-based microarray systems. *Physiol Genomics* 2004;16:361–370.
- 9 Schlingemann J, Hess J, Wrobel G, *et al*. Profile of gene expression induced by the tumour promoter TPA in murine epithelial cells. *Int J Cancer* 2003;104:699–708.
- 10 Barczak A, Rodriguez MW, Hanspers K, *et al*. Spotted long oligonucleotide arrays for human gene expression analysis. *Genome Res* 2003;13:1775–1785.

- 11 Cox WG, Beaudet MP, Agnew JY, *et al*. Possible sources of dye-related signal correlation bias in two-color DNA microarray assays. *Anal Biochem* 2004;331:243–254.
- 12 Schlingemann J, Thuerigen O, Ittrich C, *et al*. Effective transcriptome amplification for expression profiling on sense-oriented oligonucleotide microarrays. *Nucleic Acids Res* 2005;33:e29.
- 13 UICC. TNM Classification of Malignant Tumours, 6th edn. John Wiley & Sons: New York, 2002.
- 14 Schena M, Shalon D, Heller R, *et al*. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci USA* 1996; 93:10614–10619.
- 15 R Development Core Team. R: A Language and Environment for Statistical Computing, 2nd edn. R Foundation for Statistical Computing: Vienna, 2004.
- 16 Gentleman RC, Carey VJ, Bates DM, *et al*. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;5:R80.
- 17 Huber W, Von Heydebreck A, Sultmann H, *et al*. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* 2002;18Suppl. 1:S96–S104.
- 18 Affymetrix. GeneChip Expression Analysis Technical Manual, 5th edn. Affymetrix: Santa Clara, 2004.
- 19 Affymetrix. Microarray Suite User Guide, 5th edn. Affymetrix: Santa Clara, 2001.
- 20 Wu Z, Irizarry R, Gentleman R, *et al*. A model-based background adjustment for oligonucleotide expression arrays. *J Am Stat Assoc* 2004;99:909–917.
- 21 Zhang J, Carey V, Gentleman R. An extensible application for assembling annotation for genomic data. *Bioinformatics* 2003;19:155–156.
- 22 Yang MC, Ruan QG, Yang JJ, *et al*. A statistical method for flagging weak spots improves normalization and ratio estimates in microarrays. *Physiol Genomics* 2001;7:45–53.
- 23 Smyth GK, Yang YH, Speed T. Statistical issues in cDNA microarray data analysis. *Methods Mol Biol* 2003;224:111–136.
- 24 Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004;3:Article 3.
- 25 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 1995;57:289–300.
- 26 Wolfinger RD, Gibson G, Wolfinger ED, *et al*. Assessing gene significance from cDNA microarray expression data via mixed models. *J Comput Biol* 2001;8:625–637.
- 27 Benito M, Parker J, Du Q, *et al*. Adjustment of systematic microarray data biases. *Bioinformatics* 2004;20:105–114.
- 28 Ashburner M, Ball CA, Blake JA, *et al*. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25:25–29.
- 29 Dennis Jr G, Sherman BT, Hosack DA, *et al*. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 2003;4:P3.
- 30 Hosack DA, Dennis Jr G, Sherman BT, *et al*. Identifying biological themes within lists of genes with EASE. *Genome Biol* 2003;4:R70.
- 31 Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 2001;29:e45.
- 32 Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 2002;30: 207–210.
- 33 Baugh LR, Hill AA, Brown EL, *et al*. Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res* 2001;29:E29.
- 34 Tan PK, Downey TJ, Spitznagel Jr EL, *et al*. Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res* 2003;31: 5676–5684.
- 35 Shippy R, Sendera TJ, Lockner R, *et al*. Performance evaluation of commercial short-oligonucleotide microarrays and the impact of noise in making cross-platform correlations. *BMC Genomics* 2004;5:61.
- 36 Jarvinen AK, Hautaniemi S, Edgren H, *et al*. Are data from different gene expression microarray platforms comparable? *Genomics* 2004;83:1164–1168.
- 37 Marshall E. Getting the noise out of gene arrays. *Science* 2004;306:630–631.
- 38 Rhodes DR, Barrette TR, Rubin MA, *et al*. Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res* 2002;62: 4427–4433.
- 39 Sorlie T, Tibshirani R, Parker J, *et al*. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci USA* 2003;100:8418–8423.
- 40 Mitchell SA, Brown KM, Henry MM, *et al*. Interplatform comparability of microarrays in acute lymphoblastic leukemia. *BMC Genomics* 2004;5:71.
- 41 Pruitt KD, Katz KS, Sicotte H, *et al*. Introducing RefSeq and LocusLink: curated human genome resources at the NCBI. *Trends Genet* 2000;16:44–47.
- 42 Pruitt KD, Maglott DR. RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res* 2001;29: 137–140.
- 43 Mecham BH, Wetmore DZ, Szallasi Z, *et al*. Increased measurement accuracy for sequence-verified microarray probes. *Physiol Genomics* 2004;18:308–315.
- 44 Baum M, Bielau S, Rittner N, *et al*. Validation of a novel, fully integrated and flexible microarray bench-top facility for gene expression profiling. *Nucleic Acids Res* 2003;31:e151.
- 45 Boser BE, Guyon IM, Vapnik VN, *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. ACM Press: Pittsburgh, 1992, pp 144–152.
- 46 Brazma A, Hingamp P, Quackenbush J, *et al*. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001;29:365–371.