Review Article

Review of the Literature Examining the Correlation Among DNA Microarray Technologies

Carole L. Yauk* and M. Lynn Berndt

Environmental and Occupational Toxicology Division, Safe Environments Programme, Health Canada, Ottawa, Ontario, Canada K1A 0K9

DNA microarray technologies are used in a variety of biological disciplines. The diversity of platforms and analytical methods employed has raised concerns over the reliability, reproducibility and correlation of data produced across the different approaches. Initial investigations (years 2000–2003) found discrepancies in the gene expression measures produced by different microarray technologies. Increasing knowledge and control of the factors that result in poor correlation among the technologies has led to much higher levels of correlation among more recent publications (years 2004 to present). Here, we review the studies examining the correlation among microarray technologies. We find that with improvements in the technology (optimization and standardization of methods, including data analysis) and annotation, analysis across platforms yields highly correlated and reproducible results. We suggest several key factors that should be controlled in comparing across technologies, and are good microarray practice in general.   Environ. Mol. Mutagen. 48:380–394, 2007. © 2007 Wiley-Liss, Inc.

Key words: microarrays; toxicogenomics; gene expression

INTRODUCTION

DNA microarrays are quickly becoming standard tools in molecular biology, providing a powerful approach for the analysis of global transcriptional response. Over the past decade, microarrays have been widely used across biological disciplines and the number of published studies using the technology is still increasing (Fig. 1). As a result, the number of commercial suppliers of microarrays, associated reagents, hardware, and software continues to grow [Kawasaki, 2006; Technology Feature, 2006].

The diversity of microarray technologies and methods of data analysis have resulted in growing concern over the relationship among data obtained and published using different approaches. A number of impressive efforts have recently been made to develop standards for microarray experiments including the Minimum Information About a Microarray Experiment (MIAME) guidelines [Brazma et al., 2001; http://www.mged.org/Workgroups/MIAME/miame.html], the External RNA Controls Consortium (ERCC) [Baker et al., 2005; http://www.cstl.nist.gov/biotech/Cell&TissueMeasurements/GeneExpression/ERCC.htm], and the Microarray Quality Control project (MAQC) [http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/; Shi et al., 2006]. These projects have made significant advances toward improving the evaluation of microarray data quality and the reproducibility of results among laboratories and platforms. A large majority of journals have made submission of microarray data to publicly-available repositories and adherence to the MIAME standards compulsory for publication of experiments utilizing DNA microarrays. Adherence to established standards, alongside proven reproducibility and correlation within and between datasets produced by different microarray platforms, is essential for the usefulness of such databases. Furthermore, establishing the correlation and reproducibility among different microarray technologies is important for the validation of microarrays as robust, sensitive, and accurate detectors of differential gene expression.

Over 40 studies have been carried out since 2000 to evaluate the extent to which data produced by different microarray technologies correlate. In this review, we summarize the cross-platform studies designed to examine the correlation of gene expression profiles and differentially expressed genes among different DNA microarray technologies. The potential reasons for discrepancies reported in earlier comparative studies, and the methodological

WILEY InterScience®
DISCOVER SOMETHING GREAT

Environmental and Molecular Mutagenesis. DOI 10.1002/em

Correlation Among DNA Microarray Technologies 381

changes which led to improved correlations generally reported in more recent publications are discussed.

A BRIEF OVERVIEW OF TECHNOLOGICAL AND ANALYTICAL CHOICES

Various technical and analytical options are available for microarray experiments (Fig. 2). These options are sometimes governed by the selection of microarray platform. For example, a decision to use Affymetrix chips [http://www.affymetrix.com/; McGall et al., 1996] limits the choice of scanner, and subsequent steps through to image analysis, to those supplied by Affymetrix. The Affymetrix technology uses a combination of oligonucleotide synthesis and photolithography to position specific oligonucleotide probes in a predetermined spatial orientation. Each gene is represented by a series of different oligonucleotide probes spanning the coding region of that gene [Liu et al., 2003]. Each oligonucleotide probe is paired with a mismatch probe in which the central base in the sequence has been changed. Therefore, application of the Affymetrix system is heavily directed by the manufacturer's recommendations. However, for experiments using microarrays that are spotted on glass microscope slides, a number of alternatives are available that may contribute to variation in the data acquired (Fig. 2).

Several comprehensive reviews cover different microarray platforms and approaches [Sevenet and Cussenot, 2003; Hardiman, 2004; Stoughton, 2005; Ahmed, 2006a,b; Kawasaki, 2006]; readers are directed to these sources for a more in-depth overview of microarray technologies. Below, some typical options that can contribute to technical variation in the gene expression measurements acquired between different technologies are briefly summarized (Fig. 2).
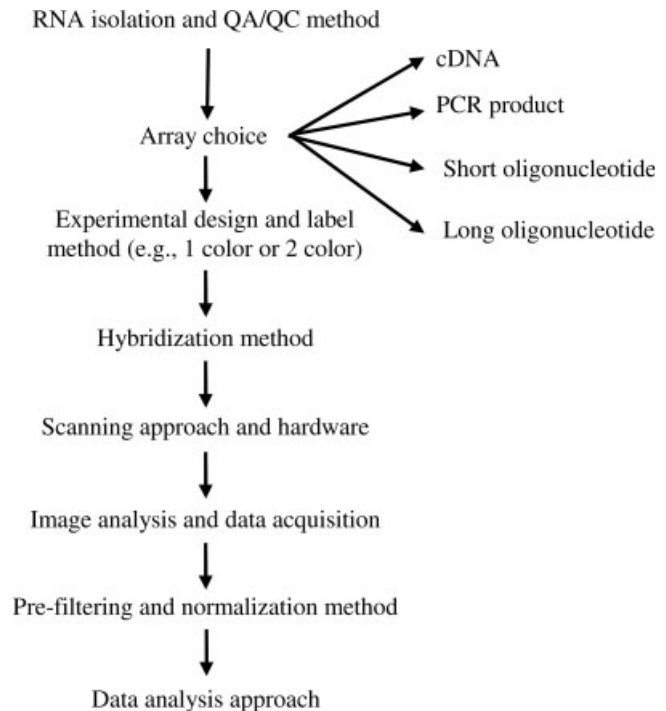
Fig. 2. Summary of choices for microarray experiments.

Probe choices for microarrays may include amplified cDNA clones, PCR gene products, or different lengths of oligonucleotides [Kawasaki, 2006]. Studies examining the correlation among microarray technologies have focused primarily on differences between probe types. However, many other factors contribute to technical variability. Methods of printing/deposition of probes onto glass slides include contact-spotting using pins, deposition by ink jet, or in situ synthesis of oligonucleotides on the slide [Hughes et al., 2001; Gao et al., 2004]. Slide surfaces may be coated with different types of matrices that gov-

ern the affinity of probe binding and affect background fluorescence [Rickman et al., 2003; Sobek et al., 2006]. Target preparation varies and may include different amounts of starting RNA, amplification, and labeling methods [Gold et al., 2004; Hardiman, 2004; Schindler et al., 2005; Singh et al., 2005; Kawasaki, 2006], all of which contribute to the type and quality of data produced. In addition, cDNA and several oligonucleotide platforms allow experiments to be carried out in one or two colors [Patterson et al., 2006]. Two color experiments may involve dye-swap, reference RNA, or loop designs [Vinciotti et al., 2005; Patterson et al., 2006]. Hybridization can be undertaken manually or using automated hybridization stations; optimization of methods is important to minimize array variability and hybridization artifacts [Yauk et al., 2005; Han et al., 2006; Yauk et al., 2006]. The scanner (high or low laser powers) and scanner settings influence background fluorescence, the number of saturated spots and the number of spots below background [Shi et al., 2005b; Timlin, 2006], and should be adjusted to maximize the linear dynamic range. Acquisition of data from images can be carried out using various algorithms through different commercial packages. The final critical steps include applying the appropriate filtering methods, evaluating microarray data quality [Shi et al., 2004], normalization [Bilban et al., 2002b; Quackenbush, 2002], and data analysis [Shi et al., 2005a; Jeffery et al., 2006]. Normalization and detection of differential gene expression are key to ensuring the accuracy and reproducibility of data across time, laboratories, and platforms, and are reviewed in detail elsewhere [Bilban et al., 2002b; Quackenbush, 2002; Armstrong and van de Wiel, 2004; Reimers, 2005; Breitling, 2006]

CORRELATION AMONG TECHNOLOGIES

Data obtained from different commercially-made and in-house microarray platforms have been compared in a large number of studies. Experiments have been carried out to determine the effective differences in accuracy (proximity to true value), sensitivity (ability to accurately detect changes at low concentrations), and specificity (to hybridize to the correct gene) among the technologies [Hardiman, 2004; van Bakel and Holstege, 2004; Draghici et al., 2006]. Intra-platform variability and reproducibility have been used as measures of the quality of the data produced for individual platforms. Several of these studies have been aimed at answering the question "which platform is the best?" The answer to this question is inarguably experiment-specific. A more relevant question is 'which platforms generate comparable and reproducible data?'

In the remaining sections, experiments investigating the reproducibility of data among DNA microarray technolo-

gies and the correlation among these data with respect to expression profiles and the identification of differentially expressed genes are reviewed.

EARLY STUDIES: 2000–2003

For the sake of simplicity, the discussion has been separated into early (years 2000–2003) and later (years 2004 to present) studies. Comparative studies began to change around 2004 when investigators began to: apply larger sample sizes, include more microarray platforms, examine relationships among laboratories, employ more sophisticated bioinformatics approaches, and generally find datasets to be more correlated. In addition to being limited in scope (i.e. comparing only 2–3 technologies or using small sample sizes), the early studies focused heavily on the comparison of cDNA microarrays to other technologies.

Table I summarizes all of the experiments that we identified that examined cross-platform performance prior to 2004. Of 13 studies, 8 produced results supporting the reproducibility and concordance of data across different microarray technologies. These findings should be interpreted with caution, as they are based on the authors' conclusions, and the term "agree" is somewhat ambiguously defined. At a time when expectations for the prospects of DNA microarrays were extremely high, these negative findings were discouraging. For example, a widely cited article by Tan et al. [2003] examined gene expression of technical and biological replicates of RNA on Codelink (Amersham oligonucleotide), Affymetrix, and Agilent cDNA arrays. Although internal consistency was high, Pearson correlation coefficients were moderate to poor across technologies (0.48–0.59). Similarly, a comprehensive investigation of 56 NCI-cell lines using 2 array technologies (Affymetrix and cDNA) showed poor correlations between the datasets [Kuo et al., 2002]. These findings led researcher to conclude that the general outlook for comparing across laboratories and platforms was bleak [Hardiman, 2004]. It was hypothesized that the discrepancies in these studies arose from intrinsic differences in the properties of the arrays, as well as the processing and analysis of the data. As a result, it was suggested that data from microarray analyses be interpreted with caution.

Subsequent studies revisited both of the datasets described earlier (an excellent demonstration of the utility of making microarray data publicly-available) and were able to reanalyze the data. Reanalysis revealed significantly improved correlations, providing insight into the basis for discrepancies found among the technologies. Shi et al. [2005a] examined the Tan et al. [2003] data in more detail and found that intra-platform consistency was generally low, suggesting that experimental protocols may

Environmental and Molecular Mutagenesis. DOI 10.1002/em

Correlation Among DNA Microarray Technologies   383

TABLE I. Studies Examining the Correlation Among Microarray Technologies from 2000 to 2003

Publication	Platforms	Probe ID	Validation	Author's conclusion
Kane et al. [2000]	Operon 50mer, PCR probes	Sequence similarity	None	Agreement
Hughes et al. [2001]	Agilent oligo, cDNA	Sequence similarity; not clearly specified	None	Agreement
Yuen et al. [2002]	Affymetrix, custom cDNA	Sequence similarity (47 genes studied, all sequence confirmed on cDNA platform)	Quantitative (Q) RT-PCR	Agreement
Kuo et al. [2002]	Affymetrix, cDNA	Sequence similarity: For each cDNA probe a best matching probe set was identified by sequence alignment.	None	Disagreement. Poor correlation between technologies. 56 cell lines from the NCI-60 cell lines studied; independent microarray experiments compared from two labs using different materials and protocols. Cells were cultured independently and all were processed separately. Observed that cross-hybridization of genes to various probes reduced correlation of cDNA arrays to the oligonucleotide arrays. Low abundant transcripts performed poorly. Subsequent re-analysis and improvement by sequence matching [Carter et al., 2005]
Kothapalli et al. [2002]	Affymetrix, Incyte cDNA	Sequence similarity: A subset of clones were sequence verified	Northern	Disagreement. Noted that a large proportion of cDNAs were incorrectly annotated. Discrepancies were related to probes
Li et al. [2002]	Affymetrix, Incyte cDNA	Unigene or Genbank	QRT-PCR	Disagreement. Conclusion based primarily on sensitivity and specificity–Affymetrix found 218 genes differentially expressed versus 4 for cDNA
Barczak et al. [2003]	Affymetrix, Operon 70mer	Unigene ID	None	Agreement
Carter et al. [2003]	Agilent oligonucleotide, cDNA	Sequence matched	QRT-PCR	Agreement
Lee et al. [2003]	Affymetrix, cDNA	UniGene	QRT-PCR	Agreement
Wang et al. [2003]	cDNA, Custom oligo	Sequence similarity	RT-PCR	Agreement
Rogojina et al. [2003]	Affymetrix, Clontech cDNA	Genbank ID	QRT-PCR and Q-immunoblot	Disagreement. Small sample size; examined list of genes that arbitrarily had 1.7-fold or greater change
Tan et al. [2003]	Affymetrix, Agilent cDNA, Amersham 30mer	Genbank ID	None	Disagreement. Correlations in expression level and significant gene expression were divergent. Subsequent reanalysis and improvement by Shi et al. [2005a] (see Table II)
Baum et al. [2003]	Febiot Geniom®, Affymetrix, cDNA	Not described	QRT-PCR	Agreement

not have been optimized for the array platform used. Furthermore, by applying more appropriate statistical tests (examining ratios instead of absolute measurements of gene expression) they were able to significantly increase the correlation coefficients obtained in comparisons of the technologies. Shi et al. [2005a] concluded that a combination of low intra-platform consistency and poor choice of data analysis procedures were the cause for discordance among the datasets, rather than inherent technical differences among the platforms as suggested by Tan et al. Data produced by microarray hybridization of RNA from NCI-60 cell lines [Kuo et al., 2002] were re-evaluated by stringent sequence mapping [Carter et al., 2005] of matched probes. By redefining probe sets, a substantially higher level of cross-platform consistency and correlation was found. The authors concluded that by using probes targeting overlapping transcript sequence regions a greater level of concordance can be obtained compared to using UniGene ID or other sequence-matching approaches. It should be noted that the study by Kuo et al. was carried out in two different laboratories using cells cultured independently, rather than using the same RNA samples matched for both platforms. Therefore, real biological variability will cause differences in the two datasets produced.

These early studies were key to identifying potential sources of discrepancies between microarray datasets, highlighting the need to investigate this issue in more detail. As microarray technologies, annotation, and techniques for analysis continue to be refined, a number of important sources of error and data misinterpretation have been identified in these early studies and are summarized in subsequent sections.

LATER STUDIES: 2004 TO PRESENT

In 2003, discrepancies in the literature led us to carry out our own cross-platform evaluation [Yauk et al., 2004]. Gene expression from three replicates of three different RNA sources (mouse whole lung, mouse lung cell line, and Stratagene Universal mouse reference RNA) were evaluated with six different technologies encompassing different reporter systems (short oligonucleotides, long oligonucleotides, and cDNAs), labeling techniques, and hybridization protocols. We were unable to match probes through sequences, because not all platform providers made sequence information available (probes were matched by UniGene ID). By applying rigorous filtering and normalizations, and using an adequate sample size, we found that the top performing platforms exhibited low levels of technical variability which resulted in an increased ability to detect differential expression, and that biology, rather than technology, accounted for the majority of variation in the data when normalized ratios were

examined. Subsequent studies have confirmed that with improved technologies, annotations, statistical rigor, and experimental design, the data from different microarray platforms are highly comparable.

Table II summarizes the studies identified that examined the correlation among microarray technologies within the last 3 years (2004 to present). Among the 32 studies that we identified, only three concluded that microarray platforms do not correlate well (<10%). Careful examination of these studies reveals some potential errors that the authors may have made in reaching these conclusions. The remaining studies show a moderate to high level of correlation among technologies.

Studies With Poor Correlation

Mah et al. [2004] examined RNA expression profiles of human colonoscopy samples using [^{33}P]dCTP labeling and hybridization to probes generated from a human cDNA clone set spotted on nylon filters, compared to data generated from human Affymetrix HG-U95Av2 chips. The authors found weak correlations using Spearman rank order coefficients on normalized signal intensities between the two systems for sequence-matched probes. Examination of absolute expression of genes fails to account for the major role that different probe sequences and location will play in resulting signal intensity. The same transcript can produce different signal intensities for different probes; even over-lapping probe sequences targeting the same transcript can produce different signal intensities [Draghici et al., 2006]. More recent studies emphasize that examination of log ratios rather than expression intensities will greatly increase the observed correlation coefficients [Park et al., 2004]. While analyses based on signal intensities are appropriate for case-control study designs using a single platform, the use of signal intensities is not appropriate for cross-platform comparisons. Therefore, the measurement of ratios may have yielded increased correlation in the Mah et al. analysis.

Severgnini et al. [2006] compared Amersham Codelink and Affymetrix microarrays through hybridization of four samples (Human breast cancer cell line MDA-MB-231; two treated and two untreated samples). The authors carried out a hierarchical cluster analysis on genes matched by LocusLink IDs on normalized expression levels and found poor correlation coefficients and clustering. Again, this negative result likely reflects analysis on normalized signal intensities rather than on ratios. Investigation of differentially expressed genes was carried out independently for both platforms (rather than on the genes in common only, filtered for poor/saturated/absent genes). The authors found 105 genes differentially expressed on Affymetrix, and 42 on Codelink, 9 of which were found in common. A more appropriate analysis would have been to examine differential expression on the filtered set of

Environmental and Molecular Mutagenesis. DOI 10.1002/em

Correlation Among DNA Microarray Technologies 385

TABLE II. Studies Examining the Correlation Among Microarray Technologies from 2004 to 2006

Publication	Platforms	Probe ID	Validation	Authors' conclusion
Yauk et al. [2004]	Affymetrix, Agilent cDNA, Agilent oligo, Codelink (Amersham), Mergen, NIA cDNA	Unigene ID	None	Agreement Good platforms correlate well. Expression profiles clustered by biology rather than technology. cDNA platforms less sensitive than oligonucleotide
Woo et al. [2004]	Affymetrix, in-house spotted cDNA and oligonucleotide	MGI identifiers	None	Agreement Good concordance in expression level and statistical significance between Affymetrix and oligonucleotide arrays; cDNA arrays showed poor concordance with other platforms
Jurata et al. [2004]	Affymetrix, Agilent cDNA	Unigene ID	Real time RT-PCR	Moderate agreement Gene changes overlapping between the two platforms were co-directional; RT-PCR validation rates were similar
Park et al. [2004]	Affymetrix, Agilent cDNA, custom oligonucleotides from three different sources printed by Agilent	Unigene ID	Real time RT-PCR	Agreement Expression level correlation low, but log ratios high; correlation is stronger for highly expressed genes
Mecham et al. [2004a]	Affymetrix, Agilent cDNA	Sequence matched	None	Agreement Cross-platform analysis greatly improved by sequence matching
Mah et al. [2004]; Two different labs	Affymetrix, cDNA	Unigene (sequence verified)	Real time RT-PCR	Disagreement Poor correlation when matched by expression level
Jarvinen et al. [2004]	Affymetrix, Agilent cDNA, custom-cDNA	Unigene ID	None	Moderate agreement Good correlations for commercial, whereas the correlations between the custom-made and either commercial platforms were lower. Discrepant findings due to clone errors, old annotations, or unknown causes
Shippy et al. [2004]	Affymetrix, Codelink (Amersham)	Unigene ID	Real time RT-PCR	Agreement After noise adjustment (use present genes)
Ulrich et al. [2004]; ILSI/HESI collaboration	Affymetrix, Clontech, Incyte, NIEHS, Molecular Dynamics, PHASE-1	Comparison of pathways	Real time RT-PCR	Agreement Correlation of the biological pathways involved in response to toxicant exposure
Stec [2005]	Affymetrix, Millennium Pharmaceuticals cDNA	UniGene ID, then sequence matched using BLAST	None	Moderate agreement Increased significantly after sequence matching; discrepant correlations between Affymetrix and cDNA measurements could be explained by probe sequence differences
Irizarry et al. [2005]; Lab-lab comparison (10 labs)	Affymetrix (5 labs), cDNA (3 labs), 2-color oligonucleotide (Qiagen 70mer; analyzed in 2 labs)	Unigene, locuslink, RefSeq	Real time RT-PCR	Agreement Among best performing laboratories; increased data quality with more stringent pre-processing
Dobbin et al. [2005]; Cross-laboratory comparison	Affymetrix (inter- and intra-laboratory correlation)	None required	None	Agreement Intra-laboratory correlation was only slightly stronger the inter-laboratory. Samples clustered by biology rather than laboratory

TABLE II. Continued

Publication	Platforms	Probe ID	Validation	Authors' conclusion
Larkin et al. [2005]	Affymetrix, TIGR cDNA	Sequence mapped TIGR	Real time RT-PCR	Agreement. Biological treatment had a greater effect on gene expression than platform for 90% of the genes
Bammler et al. [2005]; Lab-lab comparison 7 labs, 12 platforms	Affymetrix, Agilent Amersham (Codelink), Compugen, Operon, 2 custom oligo, 5 custom cDNA,	Transcripts matched using NIA mouse index	None	Moderate agreement. Standardized protocols and data analysis required
Pylatuik and Fobert [2005]	Affymetrix, Genomic Amplicon arrays, Operon oligo	Locus ID (Arabidopsis)	Northern blot	Moderate agreement. Signal intensity-dependant
Shi et al. [2005a]	Tan et al., 2003; dataset	Genbank acc. No	N/A	Agreement. Concluded that the quality of the original dataset was poor and inappropriate methods were applied for analysis. Alternate analysis had 10× >concordance
Barnes et al. [2005]	Affymetrix, Illumina Beadarrays	Sequence matched using BLAST [Kent, 2002]	None	Agreement. For genes with high expression and concordance improved for probes that were verified to target same transcript
Gwinn et al. [2005]	Affymetrix, Amersham (Codelink), cDNA	Amersham (locuslink ID) Affymetrix (GenBank)–links made via IDs given by company; genes of interest used probe sequence	Real time RT-PCR	Disagreement. Each platform yielded unique gene expression profiles
Petersen et al. [2005]	Affymetrix, in-house cDNA, in-house Operon oligonucleotide	Unigene	Real time RT-PCR	Agreement. High concordance for significant expression ratios and 1.5 to 2-fold changes (93–99%)
Carter et al. [2005]	Affymetrix, Stanford cDNA	Sequence matched	None	Agreement. Re-examination of NCI-60 cell line data. Overlapping probes correlate well
Wang et al. [2005]; 5 data sets were either collected from a public source or generated in house	Affymetrix, Agilent oligo, cDNA	UniGene ID	None	Agreement. Data were more consistent between two commercial platforms and less consistent between custom arrays and commercial arrays; expression at the gene level exhibited an acceptable level of agreement. Lab and sample effect was greater than platform effect
Schlingemann et al. [2005]	Affymetrix, in-house long oligo	Unigene ID	Real time RT-PCR	Agreement. Similar profiles and strong correlations were found for the 2 platforms
Warnat et al. [2005]	6 different cDNA and oligo array studies previously published from several laboratories	Unigene ID	None	Agreement. Integrated raw microarray data from different studies for supervised classifications. More platforms better for predictive analysis
Ali-Seyed et al. [2006]	Affymetrix, Applied Biosystems	Promoter analysis	Real time RT-PCR	Agreement. AB more sensitive and more correlated with RT-PCR

(continued)

Environmental and Molecular Mutagenesis. DOI 10.1002/em

Correlation Among DNA Microarray Technologies 387

TABLE II. *Continued*

Publication	Platforms	Probe ID	Validation	Authors' conclusion
Severgnini et al. [2006]	Affymetrix, Amersham (codelink)	Locuslink ID	Real time RT-PCR	Disagreement Only 9 genes found to be differentially expressed in common out of 42 (Affymetrix) and 105 (Codelink) in total
de Reynies et al. [2006]	Affymetrix, GE Healthcare (Amersham), Agilent	Sequence mapped	Real time RT-PCR	Moderate agreement 1 color more precise than 2 color; Affymetrix and Agilent were more concordant based on detection of differential genes
Wang et al. [2006]	Agilent, Applied biosystems	Sequence matched (BLAST)	Real time RT-PCR	Agreement 1375 genes confirmed with RT-PCR
Kuo et al. [2006]; Lab-lab comparison as well	Affymetrix, Amersham, Mergen, ABI, custom cDNA, MGH, MWG, Agilent, Compugen, Operon	Probes sequence matched within 1 exon (Unigene, LocusLink, RefSeq, Refseq exon)	Real time RT-PCR	Strong agreement Commercial better than in-house, 1-color better than 2-color
Shi et al. [2006]; Lab to lab comparison (3 sites); used commercial RNA sources; MAQC	Affymetrix, Agilent (1 and 2 color), Applied Biosystems, Eppendorf, GE Healthcare, Illumina, in-house spotted Operon oligonucleotide	Probes sequence mapped to RefSeq and to AceView using 3′ probe for genes with multiple oligonucleotides	Real time RT-PCR; TaqMan® (Roche Molecular Systems); StaRT-PCR and QuantiGene carried out by Canales et al. [2006]	Strong agreement Intra-platform consistency across test sites and high inter-platform concordance with respect to differentially expressed genes; high correlation between QRT-PCR values and microarray results
Guo et al. [2006]; Included 2 test sites for Affymetrix; MAQC	Affymetrix, Agilent, Applied Biosystems, GE Healthcare	Probes sequence mapped to RefSeq	None	Agreement High inter-site and cross-platform concordance in the detection of differential gene expression using fold change rankings; fold change ranking outperforms other analysis methods
Canales et al. [2006]; MAQC	Affymetrix, Agilent, Applied Biosystems, Eppendorf, GE Healthcare, Illumina	Probes sequence mapped to RefSeq	Real time RT-PCR; TaqMan® (997 genes), StaRT-PCR (205 genes), and QuantiGene (244 genes)	Agreement High correlation between gene expression values and microarray results. Main variable was probe sequence and target location

genes in common only. Therefore, a combination of inappropriate statistical analysis and small sample size may have contributed to the negative findings in this study.

Lastly, [Gwinn et al., 2005] found minimal similarity between Affymetrix, Amersham Codelink and NCI cDNA platforms when they analyzed three technical replicates of a human cell culture exposed to benzo[a]pyrene compared to three technical replicates of a control sample. We suggest a few potential reasons for low correlations found in this study. First, probes were generally matched using gene information provided by the manufacturer. In the latter part of the study, the authors re-examined their data and carried out some sequence investigation but the results were not fully presented and appear to be inconclusive. Other potential factors contributing to the poor correlation observed include: (a) normalizations and analyses were not carried out for all platforms combined, but were carried out individually within a platform; (b) differential expression was investigated using a small sample size ($n = 3$) for a subtle toxicological effect (e.g., Affymetrix only found 23 genes differentially expressed); (c) signal log ratios (SLR) were arbitrarily defined as differentially expressed if ± 0.6 (with no measure of variability or statistic presented), while similarity to the other platform was arbitrarily defined as within ± 0.2 SLR for the same gene on another platform; (d) it is unclear what level of filtering was applied to examine the correlation among the genes that were in common among the platforms.

Studies With Moderate to High Correlation

Since 2004, the vast majority (29/32) of technical papers comparing microarray platforms have generated results that show a moderate to high degree of correlation among the technologies. Several of these studies have been very comprehensive encompassing many microarray platforms analyzed in both one and two colors, employing different probe types spotted both in-house and commercially, and using data from the same samples analyzed in several different laboratories. By fine-tuning approaches and analyses, these studies identified methods to yield increased correlation among laboratories and platforms. Below, we present general findings that have resulted in increasing our understanding of how microarray platforms relate to each other, and we discuss a few of the more comprehensive studies.

Several studies have demonstrated that using a sequence-driven rather than an annotation-driven approach to analyzing data from different platforms yields improved correlation among technologies [Mecham et al., 2004a; Carter et al., 2005; Kuo et al., 2006]. Annotation of microarray platforms has improved greatly over the past several years, but errors in annotation will continue to affect analyses until genomes are completely validated and curated. In addition to potentially matching incorrect probes due to errors in annotation, sequence-driven matching improves correlation by ensuring that probe pairs are examining similar gene regions. The re-examination of the NCI cancer cell lines [Carter et al., 2005] using sequence-driven probe matching, described earlier, exemplifies the importance of ensuring the appropriate comparisons of probes/genes are made in cross-platform analyses. Similarly, Stec et al. [2005] compared platforms using either UniGene identifiers or by sequence matching using BLAST alignments. They found higher correlations when the Affymetrix probe identifiers were sequence-matched to ensure they fell within the cDNA probes. Mecham et al. [2004a] also found significantly improved correlation for Affymetrix compared to cDNA platforms using sequence matching. Sequence matching eliminates errors introduced by mis-annotation, and potential discrepancies introduced by probes aligning with multiple family members or alternative transcripts. Probes targeting regions of a gene in close proximity (e.g. within the same exon) are more likely to have highly correlated expression ratios [Canales et al., 2006; Kuo et al., 2006].

A number of the later studies highlight the importance of removing unreliable data from experiments prior to analysis, termed filtering. These studies generally found that probes for genes with strong expression signal tended to give more highly correlated results than those with weaker signals [Park et al., 2004; Shippy et al., 2004; Barnes et al., 2005; Pylatuik and Fobert, 2005; Kuo et al., 2006]. Signal within the background range is highly variable and contributes to much of the noise observed in microarray datasets [Bilban et al., 2002a; Park et al., 2004; Shippy et al., 2004; Draghici et al., 2006; Kuo et al., 2006]. Most commercially-available image-acquisition programs now have implemented algorithms to flag poor quality, low signal, and saturated spots. Filtering methods applied to microarray datasets prior to analysis increases the correlation among technologies [Pounds and Cheng, 2005; Kuo et al., 2006].

Optimization and standardization of protocols ensures that data produced within a technology is reproducible. Intra-platform reproducibility is obviously required before inter-platform relationships can be evaluated. A number of the later studies concluded that concordance was high among the best performing laboratories, platforms, or for commercial compared to in-house microarrays [Jarvinen et al., 2004; Yauk et al., 2004; Bammler et al., 2005; Irizarry et al., 2005; Wang et al., 2005; Kuo et al., 2006]. These results were likely due to the optimization of protocols within laboratories that routinely use a technology, technical expertise acquired in laboratories that use the platform routinely, and increased standardization through use and development of commercially-available microarrays compared to in-house microarrays. Improvements in methodology, the development of quality control standards

Environmental and Molecular Mutagenesis. DOI 10.1002/em

Correlation Among DNA Microarray Technologies   389

and references, and the implementation of standards for data analysis will improve the relationship among data produced by different microarray platforms [Bammler et al., 2005].

Recently, a number of large-scale efforts have produced comprehensive studies evaluating microarray performance across technologies and laboratories. Members of the Toxicogenomics Research Consortium [Bammler et al., 2005] examined data produced by seven laboratories and 12 microarray platforms. Each laboratory was provided with aliquots of two different RNA samples (one liver RNA sample and one mixture of tissues). They found that correlation across platforms and laboratories was generally poor. However, by implementing standardized protocols for RNA labeling, hybridization, filtering, processing, data acquisition, and normalization, increased reproducibility was obtained. Unsurprisingly, raw intensity values correlated poorly. The highest levels of reproducibility obtained were between laboratories using commercial arrays and applying standardized protocols. This analysis yielded median correlation coefficients of 0.87–0.92. The consortium concluded that the microarray platform has a large effect on the variability in the data, and standardization is required to generate data that are reproducible across laboratories. However, the group also noted that reproducibility among platforms was generally very high when analyses were carried out on biological categories identified by gene ontology analysis.

Irizarry et al. [2005] examined microarray data produced by 10 different laboratories from three different platforms using the same RNA samples. Measurement of relative expression (e.g. ratios), rather than absolute measures of gene expression were found to correlate well among the best performing laboratories. The authors emphasize the importance of experience and expertise with a platform before a laboratory can produce accurate and reproducible data, and that laboratory effect can be a strong variable (>platform effect). Furthermore, the authors stress the importance of pre-processing (normalization) before making any cross-platform comparisons.

Kuo et al. [2006] examined correlation among five different platforms encompassing both cDNA and oligonucleotide microarrays, one and two-color hybridizations, commercial and in-house chips and including results generated in two different laboratories. They matched probes at the gene id, gene, and exon level using UniGene, LocusLink, RefSeq, and RefSeq exon. Data mapped through probe sequences were more correlated than through other identifiers. Log ratios showed high correlation (0.63–0.92) for all platforms except academic cDNA and Compugen. Spot quality filtering had a strong positive effect on correlation coefficients. Inter-laboratory Pearson and Spearman correlations for log_2 ratios were high within platforms (0.79 for Mergen; 0.89 for Affymetrix; 0.93 for Amersham). Quantitative RT-PCR was carried out for 160 genes and agreed well with the microarray platforms, although RT-PCR had a larger dynamic range. The authors concluded that with stringent preprocessing and sequence matching, consistency and reproducibility among platforms and laboratories was good for highly expressed genes and variable for genes with lower expression.

A large scale real-time PCR validation experiment was conducted by Wang et al. [2006]. The authors used TaqMan® gene expression to evaluate the performance of Agilent and Applied Biosystems (AB) microarrays for 1375 genes. The authors compared log_2 fold-changes and found that the dynamic range was greatest for RT-PCR, followed by AB and then Agilent. Despite differences in the dynamic range, moderate to strong correlations of fold change were found for AB ($R^2 = 0.71$–0.75) and Agilent (0.45–0.52). The estimated range of fold changes (in log_2 scale) was from −10 to 10 for TaqMan®, −4 to 6 for AB, and −2 to 2 for Agilent, indicating ratio compression for microarray platforms. Ratio compression was expected because of various technical limitations (e.g., narrower dynamic range, signal saturation, and cross-hybridizations). In the analysis of differential expression, the authors noted that sensitivity and specificity were highest for genes with high and medium expression levels, compared to those with low expression levels.

The MicroArray Quality Control (MAQC) project evaluated inter- and intra-platform reproducibility in a series of papers [Canales et al., 2006; Guo et al., 2006; Patterson et al., 2006; Shi et al., 2006; Shippy et al., 2006]. The project was led by US Food and Drug Administration scientists and involved 137 participants from 51 organizations. Shi et al. [2006] presented data evaluating five replicates of two distinct, high-quality RNA samples from four titration pools using seven microarray platforms (each platform was evaluated at three independent tests sites). Probe sequences were mapped to the RefSeq human mRNA database [http://www.ncbi.nlm.nih.gov/RefSeq/; Pruitt and Maglott, 2001; Pruitt et al., 2005] and to the AceView database [Thierry-Mieg and Thierry-Mieg, 2006]. The relative expression between matched probes was examined. Rank correlations of the log ratios were in good agreement between all sites, with a median of 0.87 (lowest was $R = 0.69$). Generally, differentially expressed genes showed an overlap of at least 60%, with many comparisons yielding 80% or more between platforms, and 90% within platforms (between sites). An average overlap of 89% was found between test sites using the same platform and 74% across one-color microarray platforms. The Affymetrix, Agilent, and Illumina platforms showed correlation values of 0.90 to TaqMan® assays, while GE Healthcare and NCI had an average of 0.84. The results were validated using two additional quantitative gene expression platforms [Canales et al., 2006] that also showed high concordance. In addition,

toxicogenomics data generated from rats exposed to aristolochic acid, riddelliine, and comfrey and analyzed using four different microarray platforms were evaluated [Guo et al., 2006]. These data showed high concordance in inter-laboroatory and cross-platform comparisons. The results of the MAQC project provide strong support for inter-platform consistency and reproducibility and support the use of microarray platforms for the quantitative characterization of gene expression.

SUMMARY OF FACTORS LEADING TO HIGHER CORRELATION AMONG TECHNOLOGIES

Later studies resolved many of the issues surrounding the lack of correlation found in earlier studies. Sources of error in the early cross-platform microarray experiments can be divided into problems resulting from the platform and protocols, and those that result from the experimental design or method of analysis.

Platform Issues

One of the most important problems that arose in early studies was incorrect annotation of probes on the various microarray platforms. For many cDNA platforms, sequencing of clones from the libraries spotted revealed that a large number were incorrect or contaminated [Halgren et al., 2001; Taylor et al., 2001; Kothapalli et al., 2002; Jarvinen et al., 2004; Kuo et al., 2006]. Errors in annotation were not exclusive to cDNA platforms. For example, Mecham et al. [2004b] examined mammalian Affymetrix microarrays and found that greater than 19% of the probes on each platform did not correspond to their appropriate mRNA reference sequence. Dai et al. [2005] investigated Affymetrix probe information and concluded that the original probe set definitions were inaccurate, and many previous conclusions derived from GeneChip analyses could be significantly flawed. Harbig et al. [2005] re-annotated the Affymetrix U133 plus 2.0 arrays using BLAST matching against documented and postulated human transcripts. They redefined ~37% of the probes and identified more than 5,000 probesets that detected multiple transcripts. In addition to ensuring that probes detect the correct gene, with improvements in annotation and subsequent probe refinement, fewer probes on commercial arrays will hybridize to multiple splice variants, show cross-hybridization to other genes in the same family and hybridize to nonspecific probes. Therefore, as a result of errors in annotation, early studies that matched genes based on the annotation provided by the manufacturer, or by the cDNA clone set provider, were examining a large portion of incorrectly matched gene sets.

Errors in annotation continue to be an issue that affects every microarray technology. However, major improvements have been made as more sequence information is curated, validated, and annotated in high-quality databases such as Refseq [http://www.ncbi.nlm.nih.gov/RefSeq/]. In addition, in early studies probe sequence was not available and users had to trust manufacturer gene identification. Today, a large portion of microarray platform providers make all probe sequence information available; in addition, MIAME guidelines require submission of probe sequences for each spot on a microarray [http://www.mged.org/Workgroups/MIAME/miame.html]. Cross-checking probe sequence annotation is an important first step for validation of expression changes for any gene.

Other platform issues that relate to potential discrepancies in cross-platform comparisons result from sub-optimal printing, labeling, hybridizing, and washing methods in early studies [Kuo et al., 2006]. In general, several of the early studies suffered from lack of technical expertise with microarrays and more specifically, with one of the platforms in their comparison. Poor quality data will be generated when inexperienced technicians carry out the hybridization and/or sub-optimal protocols are used. The realization and control of the effect of environmental influences, such as ozone [Fare et al., 2003] on the fluorescent chemicals used, have also resulted in improved acquisition of data. Finally, the general quality of printing of both cDNA and oligonucleotide microarrays has improved significantly over the past 5 years.

In summary, methodological and platform improvements have been made over the past several years that have resulted in a decrease in the observed technical variability and resulted in superior performance. The result has been a general increase in the measured correlation among DNA microarray technologies.

Experimental Design and Analysis Issues

Many of the studies described in Table I suffer from flaws in experimental design. In some studies, data were generated in different laboratories at different times using different samples [Kuo et al., 2002]. To directly evaluate the correlation among technologies, the exact same RNA sample should be used across all experiments [Kawasaki, 2006]. Biological variability and tissue heterogeneity will significantly contribute to variance between the datasets. In addition, many of the early studies did not apply a large enough sample size, including both technical and biological replicates, to arrive at the conclusions drawn. Lastly, to investigate differential expression, samples that are sufficiently distinct should be examined [Kuo et al., 2006].

An important pre-processing step involves filtering microarray image data for poor quality, saturated, and low-signal spots. Poor quality spots and saturated signal do not accurately represent the expression of a gene. As discussed earlier, signals near background and reaching saturation do not provide accurate or reliable measures of

Environmental and Molecular Mutagenesis. DOI 10.1002/em

Correlation Among DNA Microarray Technologies 391

gene expression. Stringent filtering methods were not routinely applied to data in early microarray studies and would have greatly improved the quality of these datasets.

Appropriate statistical tools, including normalization, clustering, and identification of differentially expressed genes need to be applied in any microarray experiment. Microarray normalizations and statistical analyses have changed over time and current methods are superior to those applied in the early studies. In addition, microarrays should be normalized both within and between the technologies, incorporating a normalization approach across all of the data in the experiment [Kuo et al., 2006]. A major finding of the MAQC consortium was that the correct tools need to be applied to identify differentially expressed genes [Guo et al., 2006]. Surprisingly, the authors found that traditional parametric analyses and other microarray-tailored analyses may not derive comparable gene lists from alternative technologies. The authors suggest that gene lists generated by fold-change ranking were more reproducible than those using other methods. More work needs to be done to determine the most accurate and reproducible methods for deriving lists of differentially expressed genes from different technologies.

Early studies examining overall expression level of genes ignored the influence of probe position and sequence on the derived signal intensity [Draghici et al., 2006]. Improved correlations were generated when relative ratios were compared rather than absolute measures of gene expression in the later studies. Therefore, the measurement of ratios (to a control or reference sample) rather than signal intensities is now generally applied in cross-platform analyses. In addition, microarrays do not provide quantitative measures, and are therefore not very precise or accurate. As a result, absolute magnitude of a change should not be compared across platforms. Rather, emphasis should be placed on the direction of change [Kawasaki, 2006].

All of the above factors contributed to the discrepancies observed in the first studies examining the correlation of expression profiles across DNA microarray technologies. Subsequent changes in methods yielded improved correlation metrics, as described in the remaining sections.

CONCLUSIONS

In general, microarray platforms and associated technologies and tools have improved greatly over the past decade. As potential sources of error and reasons for discrepancies between technologies are uncovered, the relationship among gene expression data produced using the different platforms is becoming more clear. Some key points include: (a) probe sequence will affect measured intensity; (b) relative ratios are more comparable than absolute measures; (c) annotation problems still complicate analysis and genes should be evaluated at the

sequence level; (d) stringent filtering leads to more reproducible and comparable measurement of gene expression; (e) normalization and method of data analysis will affect the derived gene expression profiles; (f) validation using an alternative method is required. The laboratory and platform effect remains a major issue and comparisons need to be drawn carefully. Ensuring the appropriate experimental design before making comparisons between datasets is critical to acquiring meaningful correlations. Several of the above points (in particular (d), (e), and (f)) are good microarray practice in general, and should apply to any experiments employing this technology.

The development of standardized protocols for everything from RNA labeling to data handling will also improve the measured correlation between platforms and laboratories. Increased automation will lead to lower technical variability and result in higher correlation among technologies [Yauk et al., 2005]. The development of internal and external controls will facilitate evaluation of data quality [van Bakel and Holstege, 2004; Yauk et al., 2006]. Implementation of standards and references will lead to a better understanding of the relationship among the gene expression measures from different technologies [Andersen and Foy, 2005; Kawasaki, 2006]. Decreasing intra-platform variability is an important first step towards ensuring that microarrays produce robust and reproducible data.

In conclusion, the vast majority of papers published over the past several years support a high degree of correlation among microarray technologies. Evaluation of gene expression using alternative approaches (e.g. quantitative real-time PCR) also supports the conclusion that microarrays provide reliable and reproducible measures of transcript levels and profiles. When data are acquired and handled correctly measures of gene expression are highly correlated. This review provides a framework identifying several key features of general good microarray practice, as well as identifying critical mechanisms to ensure that data produced by different microarray technologies are comparable.

REFERENCES

Ahmed FE. 2006a. Microarray RNA transcriptional profiling. I. Platforms, experimental design and standardization. Expert Rev Mol Diagn 6:535–550.

Ahmed FE. 2006b. Microarray RNA transcriptional profiling. II. Analytical considerations and annotation. Expert Rev Mol Diagn 6:703–715.

Environmental and Molecular Mutagenesis. DOI 10.1002/em

392   Yauk and Berndt

Ali-Seyed M, Laycock N, Karanam S, Xiao W, Blair ET, Moreno CS. 2006. Cross-platform expression profiling demonstrates that SV40 small tumor antigen activates Notch, Hedgehog, and Wnt signaling in human cells. BMC Cancer 6:54.

Andersen MT, Foy CA. 2005. The development of microarray standards. Anal Bioanal Chem 381:87–89.

Armstrong NJ, van de Wiel MA. 2004. Microarray data analysis: From hypotheses to conclusions using gene expression data. Cell Oncol 26:279–290.

Baker SC, Bauer SR, Beyer RP, Brenton JD, Bromley B, Burrill J, Causton H, Conley MP, Elespuru R, Fero M, et al. 2005. The external RNA controls consortium: A progress report. Nat Methods 2: 731–734.

Bammler T, Beyer RP, Bhattacharya S, Boorman GA, Boyles A, Bradford BU, Bumgarner RE, Bushel PR, Chaturvedi K, Choi D, Cunningham ML, Deng S, Dressman HK, Fannin RD, Farin FM, Freedman JH, Fry RC, Harper A, Humble MC, Hurban P, Kavanagh TJ, Kaufmann WK, Kerr KF, Jing L, Lapidus JA, Lasarev MR, Li J, Li YJ, Lobenhofer EK, Lu X, Malek RL, Milton S, Nagalla SR, O'Malley JP, Palmer VS, Pattee P, Paules RS, Perou CM, Phillips K, Qin LX, Qiu Y, Quigley SD, Rodland M, Rusyn I, Samson LD, Canales RD, Luo Y, Willey JC, Austermiller B, Barbacioru CC, Boysen C, Hunkapiller K, Jensen RV, Knight CR, Lee KY, Ma Y, Maqsodi B, Papallo A, Peters EH, Poulter K, Ruppel PL, Samaha RR, Shi L, Yang W, Zhang L, Goodsaid FM. 2005. Standardizing global gene expression analysis between laboratories and across platforms. Nat Methods 2:351–356.

Barczak A, Rodriguez MW, Hanspers K, Koth LL, Tai YC, Bolstad BM, Speed TP, Erle DJ. 2003. Spotted long oligonucleotide arrays for human gene expression analysis. Genome Res 13:1775–1785.

Barnes M, Freudenberg J, Thompson S, Aronow B, Pavlidis P. 2005. Experimental comparison and cross-validation of the Affymetrix and Illumina gene expression analysis platforms. Nucleic Acids Res 33:5914–5923.

Baum M, Bielau S, Rittner N, Schmid K, Eggelbusch K, Dahms M, Schlauersbach A, Tahedl H, Beier M, Guimil R, et al. 2003. Validation of a novel, fully integrated and flexible microarray benchtop facility for gene expression profiling. Nucleic Acids Res 31:e151.

Bilban M, Buehler LK, Head S, Desoye G, Quaranta V. 2002a. Defining signal thresholds in DNA microarrays: Exemplary application for invasive cancer. BMC Genomics 3:19

Bilban M, Buehler LK, Head S, Desoye G, Quaranta V. 2002b. Normalizing DNA microarray data. Curr Issues Mol Biol 4:57–64.

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, et al. 2001. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 29:365–371.

Breitling R. 2006. Biological microarray interpretation: The rules of engagement. Biochim Biophys Acta 1759:319–327.

Canales RD, Luo Y, Willey JC, Austermiller B, Barbacioru CC, Boysen C, Hunkapiller K, Jensen RV, Knight CR, Lee KY, et al. 2006. Evaluation of DNA microarray results with quantitative gene expression platforms. Nat Biotechnol 24:1115–1122.

Carter MG, Hamatani T, Sharov AA, Carmack CE, Qian Y, Aiba K, Ko NT, Dudekula DB, Brzoska PM, Hwang SS, Ko MS. 2003. In situ-synthesized novel microarray optimized for mouse stem cell and early developmental expression profiling. Genome Res 13:1011–1021.

Carter SL, Eklund AC, Mecham BH, Kohane IS, Szallasi Z. 2005. Redefinition of Affymetrix probe sets by sequence overlap with cDNA microarray probes reduces cross-platform inconsistencies in cancer-associated gene expression measurements. BMC Bioinformatics 6:107

Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, Bunney WE, Myers RM, Speed TP, Akil H, Shippy R, Harris SC, Zhang L, Mei N, Chen T, Herman D, Goodsaid FM, Hurban P, Phillips KL, Xu J, Deng X, Sun YA, Tong W, Dragan YP, Shi L. 2005. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res 33:e175.

de Reynies A, Geromin D, Cayuela JM, Petel F, Dessen P, Sigaux F, Rickman DS. 2006. Comparison of the latest commercial short and long oligonucleotide microarray technologies. BMC Genomics 7:51

Dobbin KK, Beer DG, Meyerson M, Yeatman TJ, Gerald WL, Jacobson JW, Conley B, Buetow KH, Heiskanen M, Simon RM, et al. 2005. Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays. Clin Cancer Res 11(2, Part 1):565–572.

Draghici S, Khatri P, Eklund AC, Szallasi Z. 2006. Reliability and reproducibility issues in DNA microarray measurements. Trends Genet 22:101–109.

Fare TL, Coffey EM, Dai H, He YD, Kessler DA, Kilian KA, Koch JE, LeProust E, Marton MJ, Meyer MR, et al. 2003. Effects of atmospheric ozone on microarray data quality. Anal Chem 75:4672–4675.

Gao X, Gulari E, Zhou X. 2004. In situ synthesis of oligonucleotide microarrays. Biopolymers 73:579–596.

Gold D, Coombes K, Medhane D, Ramaswamy A, Ju Z, Strong L, Koo JS, Kapoor M. 2004. A comparative analysis of data generated using two different target preparation methods for hybridization to high-density oligonucleotide microarrays. BMC Genomics 5:2.

Guo L, Lobenhofer EK, Wang C, Shippy R, Harris SC, Zhang L, Mei N, Chen T, Herman D, Goodsaid FM, et al. 2006. Rat toxicogenomic study reveals analytical consistency across microarray platforms. Nat Biotechnol 24:1162–1169.

Gwinn MR, Keshava C, Olivero OA, Humsi JA, Poirier MC, Weston A. 2005. Transcriptional signatures of normal human mammary epithelial cells in response to benzo[a]pyrene exposure: A comparison of three microarray platforms. Omics 9:334–350.

Halgren RG, Fielden MR, Fong CJ, Zacharewski TR. 2001. Assessment of clone identity and sequence fidelity for 1189 IMAGE cDNA clones. Nucleic Acids Res 29:582–588.

Han T, Melvin CD, Shi L, Branham WS, Moland CL, Pine PS, Thompson KL, Fuscoe JC. 2006. Improvement in the reproducibility and accuracy of DNA microarray quantification by optimizing hybridization conditions. BMC Bioinformatics 7(Suppl. 2):S17.

Harbig J, Sprinkle R, Enkemann SA. 2005. A sequence-based identification of the genes detected by probesets on the Affymetrix U133 plus 2.0 array. Nucleic Acids Res 33:e31.

Hardiman G. 2004. Microarray platforms—Comparisons and contrasts. Pharmacogenomics 5:487–502.

http://www.affymetrix.com/.

http://www.cstl.nist.gov/biotech/Cell&TissueMeasurements/GeneExpression/ERCC.htm.

http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/.

http://www.mged.org/Workgroups/MIAME/miame.html.

http://www.ncbi.nlm.nih.gov/RefSeq/.

Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM, Ziman M, Schelter JM, Meyer MR, Kobayashi S, Davis C, Dai H, He YD, Stephaniants SB, Cavet G, Walker WL, West A, Coffey E, Shoemaker DD, Stoughton R, Blanchard AP, Friend SH, Linsley PS. 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat Biotechnol 19:342–347.

Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, Frank BC, Gabrielson E, Garcia JG, Geoghegan J, Germino G, Griffin C, Hilmer SC, Hoffman E, Jedlicka AE, Kawasaki E, Martinez-Murillo F, Morsberger L, Lee H, Petersen D, Quackenbush J, Scott A, Wilson M, Yang Y, Ye SQ, Yu W. 2005. Multiple-laboratory comparison of microarray platforms. Nat Methods 2:345–350.

Jarvinen AK, Hautaniemi S, Edgren H, Auvinen P, Saarela J, Kallioniemi OP, Monni O. 2004. Are data from different gene expression microarray platforms comparable? Genomics 83:1164–1168.

Environmental and Molecular Mutagenesis. DOI 10.1002/em

Correlation Among DNA Microarray Technologies   393

Jeffery IB, Higgins DG, Culhane AC. 2006. Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. BMC Bioinformatics 7:359.

Jurata LW, Bukhman YV, Charles V, Capriglione F, Bullard J, Lemire AL, Mohammed A, Pham Q, Laeng P, Brockman JA, Altar CA. 2004. Comparison of microarray-based mRNA profiling technologies for identification of psychiatric disease and drug signatures. J Neurosci Methods 138:173–188.

Kane MD, Jatkoe TA, Stumpf CR, Lu J, Thomas JD, Madore SJ. 2000. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. Nucleic Acids Res 28:4552–4557.

Kawasaki ES. 2006. The end of the microarray tower of babel: Will universal standards lead the way? J Biomol Tech 17:200–206.

Kent WJ. 2002. BLAT—The BLAST-like alignment tool. Genome Res 12:656–664.

Kothapalli R, Yoder SJ, Mane S, Loughran TP Jr. 2002. Microarray results: How accurate are they? BMC Bioinformatics 3:22.

Kuo WP, Jenssen TK, Butte AJ, Ohno-Machado L, Kohane IS. 2002. Analysis of matched mRNA measurements from two different microarray technologies. Bioinformatics 18:405–412.

Kuo WP, Liu F, Trimarchi J, Punzo C, Lombardi M, Sarang J, Whipple ME, Maysuria M, Serikawa K, Lee SY, McCrann D, Kang J, Shearstone JR, Burke J, Park DJ, Wang X, Rector TL, Ricciardi-Castagnoli P, Perrin S, Choi S, Bumgarner R, Kim JH, Short GF, 3rd, Freeman MW, Seed B, Jensen R, Church GM, Hovig E, Cepko CL, Park P, Ohno-Machado L, Jenssen TK. 2006. A sequence-oriented comparison of gene expression measurements across different hybridization-based technologies. Nat Biotechnol 24:832–840.

Larkin JE, Frank BC, Gavras H, Sultana R, Quackenbush J. 2005. Independence and reproducibility across microarray platforms. Nat Methods 2:337–344.

Lee JK, Bussey KJ, Gwadry FG, Reinhold W, Riddick G, Pelletier SL, Nishizuka S, Szakacs G, Annereau JP, Shankavaram U, Lababidi S, Smith LH, Gottesman MM, Weinstein JN. 2003. Comparing cDNA and oligonucleotide array data: Concordance of gene expression across platforms for the NCI-60 cancer cells. Genome Biol 4:R82.

Li J, Pankratz M, Johnson JA. 2002. Differential gene expression patterns revealed by oligonucleotide versus long cDNA arrays. Toxicol Sci 69:383–390.

Liu G, Loraine AE, Shigeta R, Cline M, Cheng J, Valmeekam V, Sun S, Kulp D, Siani-Rose MA. 2003. NetAffx: Affymetrix probesets and annotations. Nucleic Acids Res 31:82–86.

Mah N, Thelin A, Lu T, Nikolaus S, Kuhbacher T, Gurbuz Y, Eickhoff H, Kloppel G, Lehrach H, Mellgard B, Costello CM, Schreiber S. 2004. A comparison of oligonucleotide and cDNA-based microarray systems. Physiol Genomics 16:361–370.

McGall G, Labadie J, Brock P, Wallraff G, Nguyen T, Hinsberg W. 1996. Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. Proc Natl Acad Sci USA 93:13555–13560.

Mecham BH, Klus GT, Strovel J, Augustus M, Byrne D, Bozso P, Wetmore DZ, Mariani TJ, Kohane IS, Szallasi Z. 2004a. Sequence-matched probes produce increased cross-platform consistency and more reproducible biological results in microarray-based gene expression measurements. Nucleic Acids Res 32:e74.

Mecham BH, Wetmore DZ, Szallasi Z, Sadovsky Y, Kohane I, Mariani TJ. 2004b. Increased measurement accuracy for sequence-verified microarray probes. Physiol Genomics 18:308–315.

Park PJ, Cao YA, Lee SY, Kim JW, Chang MS, Hart R, Choi S. 2004. Current issues for DNA microarrays: Platform comparison, double linear amplification, and universal RNA reference. J Biotechnol 112:225–245.

Patterson TA, Lobenhofer EK, Fulmer-Smentek SB, Collins PJ, Chu TM, Bao W, Fang H, Kawasaki ES, Hager J, Tikhonova IR, Walker SJ, Zhang L, Hurban P, de Longueville F, Fuscoe JC, Tong W, Shi L, Wolfinger RD. 2006. Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. Nat Biotechnol 24:1140–1150.

Petersen D, Chandramouli GV, Geoghegan J, Hilburn J, Paarlberg J, Kim CH, Munroe D, Gangi L, Han J, Puri R, Staudt L, Weinstein J, Barrett JC, Green J, Kawasaki ES. 2005. Three microarray platforms: An analysis of their concordance in profiling gene expression. BMC Genomics 6:63.

Pounds S, Cheng C. 2005. Statistical development and evaluation of microarray gene expression data filters. J Comput Biol 12:482–495.

Pruitt KD, Maglott DR. 2001. RefSeq and LocusLink: NCBI gene-centered resources. Nucleic Acids Res 29:137–140.

Pruitt KD, Tatusova T, Maglott DR. 2005. NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 33(Database issue): D501–D504.

Pylatuik JD, Fobert PR. 2005. Comparison of transcript profiling on Arabidopsis microarray platform technologies. Plant Mol Biol 58: 609–624.

Quackenbush J. 2002. Microarray data normalization and transformation. Nat Genet 32(Suppl.):496–501.

Reimers M. 2005. Statistical analysis of microarray data. Addict Biol 10:23–35.

Rickman DS, Herbert CJ, Aggerbeck LP. 2003. Optimizing spotting solutions for increased reproducibility of cDNA microarrays. Nucleic Acids Res 31:e109.

Rogojina AT, Orr WE, Song BK, Geisert EE Jr. 2003. Comparing the use of Affymetrix to spotted oligonucleotide microarrays using two retinal pigment epithelium cell lines. Mol Vis 9:482–496.

Schindler H, Wiese A, Auer J, Burtscher H. 2005. cRNA target preparation for microarrays: Comparison of gene expression profiles generated with different amplification procedures. Anal Biochem 344:92–101.

Schlingemann J, Habtemichael N, Ittrich C, Toedt G, Kramer H, Hambek M, Knecht R, Lichter P, Stauber R, Hahn M. 2005. Patient-based cross-platform comparison of oligonucleotide microarray expression profiles. Lab Invest 85:1024–1039.

Sevenet N, Cussenot O. 2003. DNA microarrays in clinical practice: Past, present, and future. Clin Exp Med 3:1–3.

Severgnini M, Bicciato S, Mangano E, Scarlatti F, Mezzelani A, Mattioli M, Ghidoni R, Peano C, Bonnal R, Viti F, Milanesi L, De Bellis G, Battaglia C. 2006. Strategies for comparing gene expression profiles from different microarray platforms: Application to a case-control experiment. Anal Biochem 353:43–56.

Shi L, Tong W, Goodsaid F, Frueh FW, Fang H, Han T, Fuscoe JC, Casciano DA. 2004. QA/QC: Challenges and pitfalls facing the microarray community and regulatory agencies. Expert Rev Mol Diagn 4:761–777.

Shi L, Tong W, Fang H, Scherf U, Han J, Puri RK, Frueh FW, Goodsaid FM, Guo L, Su Z, Han T, Fuscoe JC, Xu ZA, Patterson TA, Hong H, Xie Q, Perkins RG, Chen JJ, Casciano DA. 2005a. Cross-platform comparability of microarray technology: Intra-platform consistency and appropriate data analysis procedures are essential. BMC Bioinformatics 6 (Suppl. 2):S12.

Shi L, Tong W, Fang H, Scherf U, Han J, Puri RK, Frueh FW, Goodsaid FM, Guo L, Su Z, Han T, Fuscoe JC, Xu ZA, Patterson TA, Hong H, Xie Q, Perkins RG, Chen JJ, Casciano DA. 2005b. Microarray scanner calibration curves: Characteristics and implications. BMC Bioinformatics 6(Suppl. 2):S11.

Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, Baker SC, Collins PJ, de Longueville F, Kawasaki ES, Lee KY, Luo Y, Sun YA, Willey JM, Setterquist RA, Fischer GM, Tong W, Dragan YP, Dix DJ, Frueh FW, Goodsaid FM, Herman D, Jensen RV, Johnson CD, Lobenhofer EK, Puri RK, Schrf U, Thierry-Mieg J, Wang C, Wilson M, Wolber PK, Zhang L, Slikker W, Jr. 2006. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol 24:1151–1161.

Environmental and Molecular Mutagenesis. DOI 10.1002/em

394   Yauk and Berndt

Shippy R, Sendera TJ, Lockner R, Palaniappan C, Kaysser-Kranich T, Watts G, Alsobrook J. 2004. Performance evaluation of commercial short-oligonucleotide microarrays and the impact of noise in making cross-platform correlations. BMC Genomics 5:61.

Shippy R, Fulmer-Smentek S, Jensen RV, Jones WD, Wolber PK, Johnson CD, Pine PS, Boysen C, Guo X, Chudin E, Sun YA, Willey JC, Thierry-Mieg J, Thierry-Mieg D, Setterquist RA, Wilson M, Lucas AB, Novoradovskaya N, Papallo A, Turpaz Y, Baker SC, Warrington JA, Shi L, Herman D. 2006. Using RNA sample titrations to assess microarray platform performance and normalization techniques. Nat Biotechnol 24:1123–1131.

Singh R, Maganti RJ, Jabba SV, Wang M, Deng G, Heath JD, Kurn N, Wangemann P. 2005. Microarray-based comparison of three amplification methods for nanogram amounts of total RNA. Am J Physiol Cell Physiol 288:C1179–C1189.

Sobek J, Bartscherer K, Jacob A, Hoheisel JD, Angenendt P. 2006. Microarray technology as a universal tool for high-throughput analysis of biological systems. Comb Chem High Throughput Screen 9:365–380.

Stec J, Wang J, Coombes K, Ayers M, Hoersch S, Gold DL, Ross JS, Hess KR, Tirrell S, Linette G, Hortobagyi GN, Fraser Symmans W, Pusztai L. 2005. Comparison of the predictive accuracy of DNA array-based multigene classifiers across cDNA arrays and Affymetrix GeneChips. J Mol Diagn 7:357–367.

Stoughton RB. 2005. Applications of DNA microarrays in biology. Annu Rev Biochem 74:53–82.

Tan PK, Downey TJ, Spitznagel EL Jr, Xu P, Fu D, Dimitrov DS, Lempicki RA, Raaka BM, Cam MC. 2003. Evaluation of gene expression measurements from commercial microarray platforms. Nucleic Acids Res 31:5676–5684.

Taylor E, Cogdell D, Coombes K, Hu L, Ramdas L, Tabor A, Hamilton S, Zhang W. 2001. Sequence verification as quality-control step for production of cDNA microarrays. Biotechniques 31:62–65.

Technology Feature. 2006. Microarrays: Table of suppliers. Nature 442:1071–1072.

Thierry-Mieg D, Thierry-Mieg J. 2006. AceView: A comprehensive cDNA-supported gene and transcripts annotation. Genome Biol 7 (Suppl. 1):S12.1–14.

Timlin JA. 2006. Scanning microarrays: Current methods and future directions. Methods Enzymol 411:79–98.

Ulrich RG, Rockett JC, Gibson GG, Pettit SD. 2004. Overview of an interlaboratory collaboration on evaluating the effects of model hepatotoxicants on hepatic gene expression. Environ Health Perspect 112:423–427.

van Bakel H, Holstege FC. 2004. In control: Systematic assessment of microarray performance. EMBO Rep 5:964–969.

Vinciotti V, Khanin R, D'Alimonte D, Liu X, Cattini N, Hotchkiss G, Bucca G, de Jesus O, Rasaiyaah J, Smith CP, Kellam P, Wit E. 2005. An experimental evaluation of a loop versus a reference design for two-channel microarrays. Bioinformatics 21: 492–501.

Wang HY, Malek RL, Kwitek AE, Greene AS, Luu TV, Behbahani B, Frank B, Quackenbush J, Lee NH. 2003. Assessing unmodified 70-mer oligonucleotide probe performance on glass-slide microarrays. Genome Biol 4:R5.

Wang H, He X, Band M, Wilson C, Liu L. 2005. A study of inter-lab and inter-platform agreement of DNA microarray data. BMC Genomics 6:71.

Wang Y, Barbacioru C, Hyland F, Xiao W, Hunkapiller KL, Blake J, Chan F, Gonzalez C, Zhang L, Samaha RR. 2006. Large scale real-time PCR validation on gene expression measurements from two commercial long-oligonucleotide microarrays. BMC Genomics 7:59.

Warnat P, Eils R, Brors B. 2005. Cross-platform analysis of cancer microarray data improves gene expression based classification of phenotypes. BMC Bioinformatics 6:265.

Woo Y, Affourtit J, Daigle S, Viale A, Johnson K, Naggert J, Churchill G. 2004. A comparison of cDNA, oligonucleotide, and Affymetrix GeneChip gene expression microarray platforms. J Biomol Tech 15:276–284.

Yauk CL, Berndt ML, Williams A, Douglas GR. 2004. Comprehensive comparison of six microarray technologies. Nucleic Acids Res 32:e124.

Yauk C, Berndt L, Williams A, Douglas GR. 2005. Automation of cDNA microarray hybridization and washing yields improved data quality. J Biochem Biophys Methods 64:69–75.

Yauk CL, Williams A, Boucher S, Berndt LM, Zhou G, Zheng JL, Rowan-Carroll A, Dong H, Lambert IB, Douglas GR, Parfett CL. 2006. Novel design and controls for focused DNA microarrays: Applications in quality assurance/control and normalization for the Health Canada ToxArray. BMC Genomics 7:266.

Yuen T, Wurmbach E, Pfeffer RL, Ebersole BJ, Sealfon SC. 2002. Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays. Nucleic Acids Res 30:e48.