

Research article

Open Access

Application of a correlation correction factor in a microarray cross-platform reproducibility study

Kellie J Archer*^{1,4}, Catherine I Dumur², G Scott Taylor³, Michael D Chaplin⁴, Anthony Guiseppi-Elie³, Geraldine Grant⁵, Andrea Ferreira-Gonzalez² and Carleton T Garrett²

Address: ¹Department of Biostatistics, Virginia Commonwealth University, 730 East Broad St., Richmond, VA, USA, ²Department of Pathology, Virginia Commonwealth University, Richmond, VA, USA, ³Center for Bioelectronics, Biosensors and Biochips, School of Engineering, Virginia Commonwealth University, Richmond, VA, USA, ⁴Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, VA, USA and ⁵Molecular and Microbiological Department, George Mason University, Manassas, VA, USA

Email: Kellie J Archer* - kjarcher@vcu.edu; Catherine I Dumur - cidumur@vcu.edu; G Scott Taylor - staylor@illumina.com; Michael D Chaplin - chaplinmd@vcu.edu; Anthony Guiseppi-Elie - guiseppi@clemsun.edu; Geraldine Grant - ggrant1@gmu.edu; Andrea Ferreira-Gonzalez - agonzale@vcu.edu; Carleton T Garrett - ctgarret@vcu.edu

* Corresponding author

Published: 15 November 2007

Received: 23 May 2007

BMC Bioinformatics 2007, **8**:447 doi:10.1186/1471-2105-8-447

Accepted: 15 November 2007

This article is available from: <http://www.biomedcentral.com/1471-2105/8/447>

© 2007 Archer et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Recent research examining cross-platform correlation of gene expression intensities has yielded mixed results. In this study, we demonstrate use of a correction factor for estimating cross-platform correlations.

Results: In this paper, three technical replicate microarrays were hybridized to each of three platforms. The three platforms were then analyzed to assess both intra- and cross-platform reproducibility. We present various methods for examining intra-platform reproducibility. We also examine cross-platform reproducibility using Pearson's correlation. Additionally, we previously developed a correction factor for Pearson's correlation which is applicable when X and Y are measured with error. Herein we demonstrate that correcting for measurement error by estimating the "disattenuated" correlation substantially improves cross-platform correlations.

Conclusion: When estimating cross-platform correlation, it is essential to thoroughly evaluate intra-platform reproducibility as a first step. In addition, since measurement error is present in microarray gene expression data, methods to correct for attenuation are useful in decreasing the bias in cross-platform correlation estimates.

Background

Previous microarray gene expression studies have examined within-platform reproducibility among different generations of the Affymetrix GeneChip [1,2] and among cDNA-based array platforms [3,4]. Subsequently, several cross-platform reproducibility studies have been reported,

many of which examined either the consistency of intensities or the consistency with which different platforms identify genes significantly differently expressed [5-18]. Results from another large cross-platform study, the MicroArray Quality Control (MAQC) project, led by the US Food and Drug Administration with 51 participating

universities and major biotechnology companies, have also been reported [19-24]. Some of these early studies demonstrated poor cross-platform correlations. For example, among 384 genes commonly declared present in a cDNA-based microarray and the Affymetrix HG-U95Av2 GeneChip platform, the Spearman correlation was only 0.131. Other cross-platform studies also reported low cross-platform correlations [5,8]. In addition, in a study examining three microarray platforms in ten laboratories, correlations between Affymetrix and two-channel arrays ranged from 0.13 – 0.57 [25]. More recent research has demonstrated that poor correlations may be observed when at least one platform under examination suffers from low intra-platform reproducibility or when a poor data analytic method is applied [26].

Most of these studies estimated Pearson's correlation as a means of assessing cross-platform reproducibility. That is, we consider X and Y to be microarray gene expression values from two different platforms, and ρ_{XY} is estimated. However, for microarray data, both random variables X and Y are subject to measurement error. It is well known that the fluorescent intensities from the scanned microarray images are proxies for the true underlying gene expression values [27]. Therefore, microarray gene expression values are measured with error. When examining cross-platform correlation, inconsistencies in measured intensities can be due to systematic platform biases as well as random intra-platform variability. Statistical methods that account for measurement error (ME), such as regression calibration, have been applied in a variety of scenarios to correct for the known bias caused by ME in parameter estimation [28]. In a recent review, the authors stated that within the next 5 years, "calibration methods will be introduced to systematically correct ratio underestimation by microarray technology" [29]. We have undertaken such an effort to account for the random intra-platform variability by developing a "disattenuated" correlation estimate [30] which accounts for random intra-platform variation in both X and Y , and demonstrate its use in measuring cross-platform correlation.

Microarray hybridizations were performed using three different technologies, each in a different laboratory. The Affymetrix (Affy) HG-U133A GeneChip was utilized in the Virginia Commonwealth University's (VCU) Division of Molecular Diagnostics Laboratory. A custom-designed oligonucleotide microarray designed specifically to interrogate genes more commonly expressed in brain tissue was used in VCU's School of Engineering's Center for Bioelectronics, Biosensors and Biochips (C3B). The C3B microarray platform comprises 10,000 genes represented by 3' fifty-mer oligonucleotides (MWG Biotech) that were spotted in duplicate. Finally, a cDNA microarray spotted with full and partial length PCR probes (Research Genet-

ics/Invitrogen) was used in George Mason University's (GMU) Center for Biomedical Genomics and Informatics.

Each laboratory designed a small experiment to assess intra-platform quality control. Each laboratory used the same lot of reference RNA, the Stratagene Total Human RNA, for hybridizing a set of technical replicates for a process variability study. These 'self-self' hybridizations permit meaningful assessments of reproducibility since, under ideal circumstances such as that the same experimental conditions exist among platforms and that there are no probe-binding affinity effects, each gene across the set of chips should exhibit linearly related gene expression intensities across platforms. Although the RNA hybridized was from the same lot, the study designs and protocols differed from lab to lab. A description of each experiment can be found in the Methods section of this paper.

Results

Within-platform comparisons

Prior to estimating cross-platform correlations, we performed a thorough examination of intra-platform reproducibility, as recommended [29]. Since the Stratagene Total Human RNA was used as both the experimental and reference sample, the expected \log_2 ratio for all genes is 1, so that no correlation is expected when comparing two arrays in terms of the \log_2 ratio. Therefore for two channel arrays, we restricted attention to intensities from one channel as well as to the post-normalized intensities from that same channel. For the Affymetrix GeneChip, intensities were highly correlated across the set of three technical replicates for all expression summary methods (Table 1 and Figure 1). The GMU arrays were strongly correlated, though the C3B arrays were not highly correlated (Figures 2 and 3).

The weighted kappa statistics indicated that the Affymetrix platform had the highest agreement among ranked intensities (Table 2), followed by the GMU array which also exhibited good agreement among the technical replicates when considering the ranked gene intensities. The weighted kappa statistics for C3B platform suggested the

Table 1: Average correlation for the Affymetrix, C3B, and GMU Stratagene Technical Replicates dataset for various expression summary methods.

Platform	Expression Summary Method	Average Correlation
Affymetrix (N = 22,283)	MAS 5.0	0.9955
	RMA	0.9994
	GC-RMA	0.9998
C3B (N = 21,168)	Channel 1 foreground	0.6560
	Print-tip loess normalized Ch1	0.6593
GMU (N = 21,168)	Cy5 foreground	0.6236
	Print-tip loess normalized Cy5	0.8475

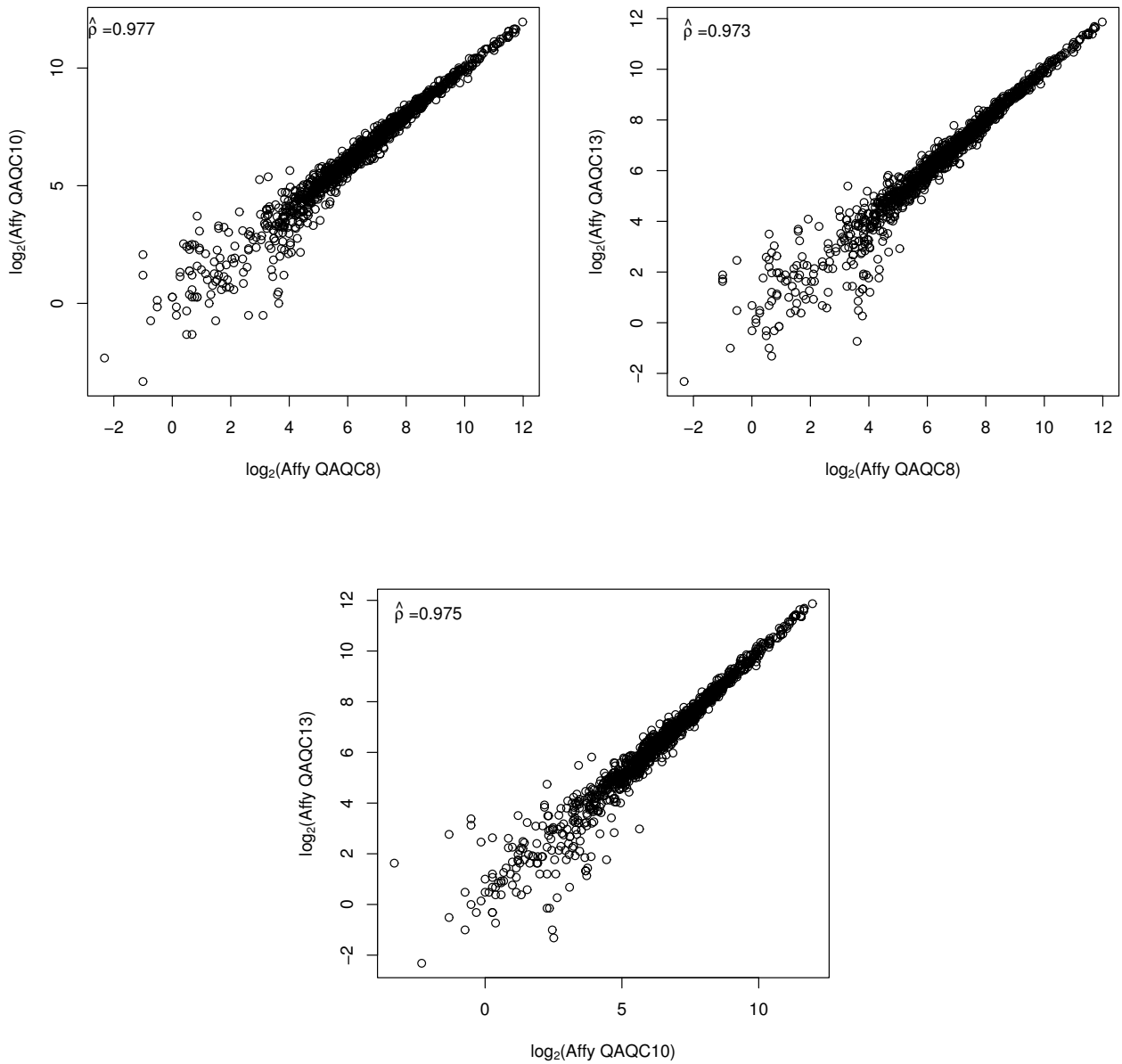


Figure 1
Affymetrix. Pairwise scatterplots and Pearson's correlation for Affymetrix GeneChips (MAS5 summaries) restricted to the 1,288 genes in common among the three platforms.

ranked intensities from the three technical replicates were not in agreement, yielding an insignificant p-value for two of the array comparisons. A similar conclusion, that the Affymetrix platform followed by the GMU array demonstrated the highest reproducibility, with low reproducibility among the C3B arrays, was noted upon examination of the proportion of invariant features (Table 3). Although intra-platform reproducibility varied among the three platforms studied, all platforms yield gene expression

intensities that are subject to some degree of measurement error.

Cross-platform comparisons

For the GMU array the 21,168 spots correspond to 19,894 distinct clones, with the feature name of each spot denoted by Unigene ID. There were 2,744 Affy probe sets that matched a GMU Unigene ID. Among these, 145 Unigene IDs were interrogated by more than one probe set. After restricting attention to unique clones and probes sets

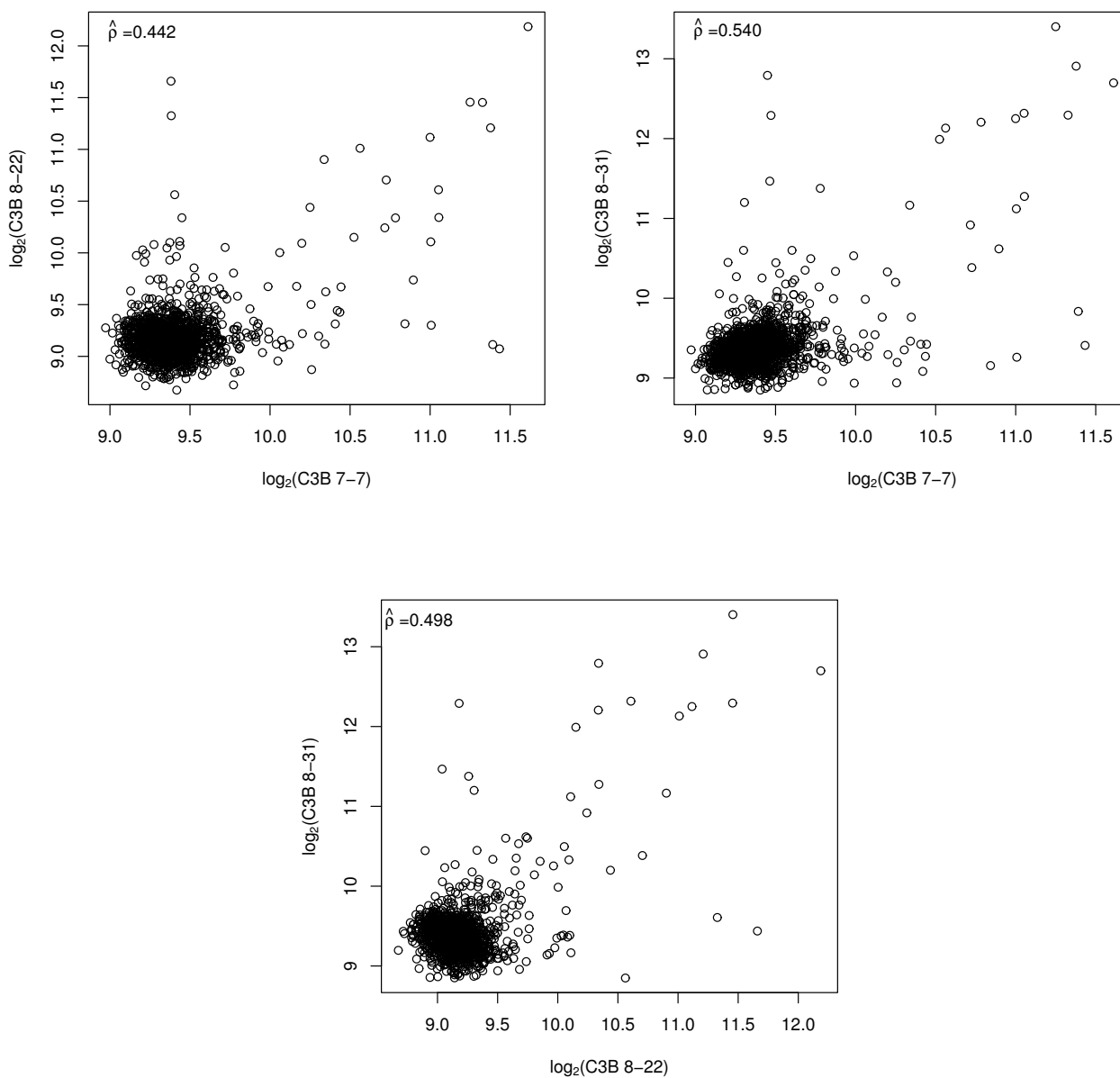


Figure 2
C3B. Pairwise scatterplots and Pearson's correlation for C3B arrays restricted to the 1,288 genes in common among the three platforms.

there were 2,587 unique probe sets/clones in common to GMU and the Affy platforms. For the C3B arrays, since its design is essentially two identical subarrays laid out in duplicate with the feature name of each spot denoted by RefSeqID, the average expression for each RefSeqID was calculated prior to merging the spots with the Affymetrix probe sets. That is, the 21,168 long oligos correspond to 10,040 distinct genes. For the C3B array, there were 9,000 distinct RefSeqIDs were interrogated by at least one Affymetrix probe set meeting our criteria. Once the data

from the two different 2-channel arrays were merged to the Affymetrix GeneChip data (i.e., GMU-Affy and C3B-Affy), these two resulting datasets were then merged by Affymetrix probe set ID, resulting in 1,288 common probe sets/spots among the three platforms.

Not accounting for measurement error, the average Pearson correlations ($\bar{\rho}_w$) of the log transformed Affymetrix GeneChip expression and C3B array expression are reported in Table 4 for MAS 5.0, RMA, and GC-RMA

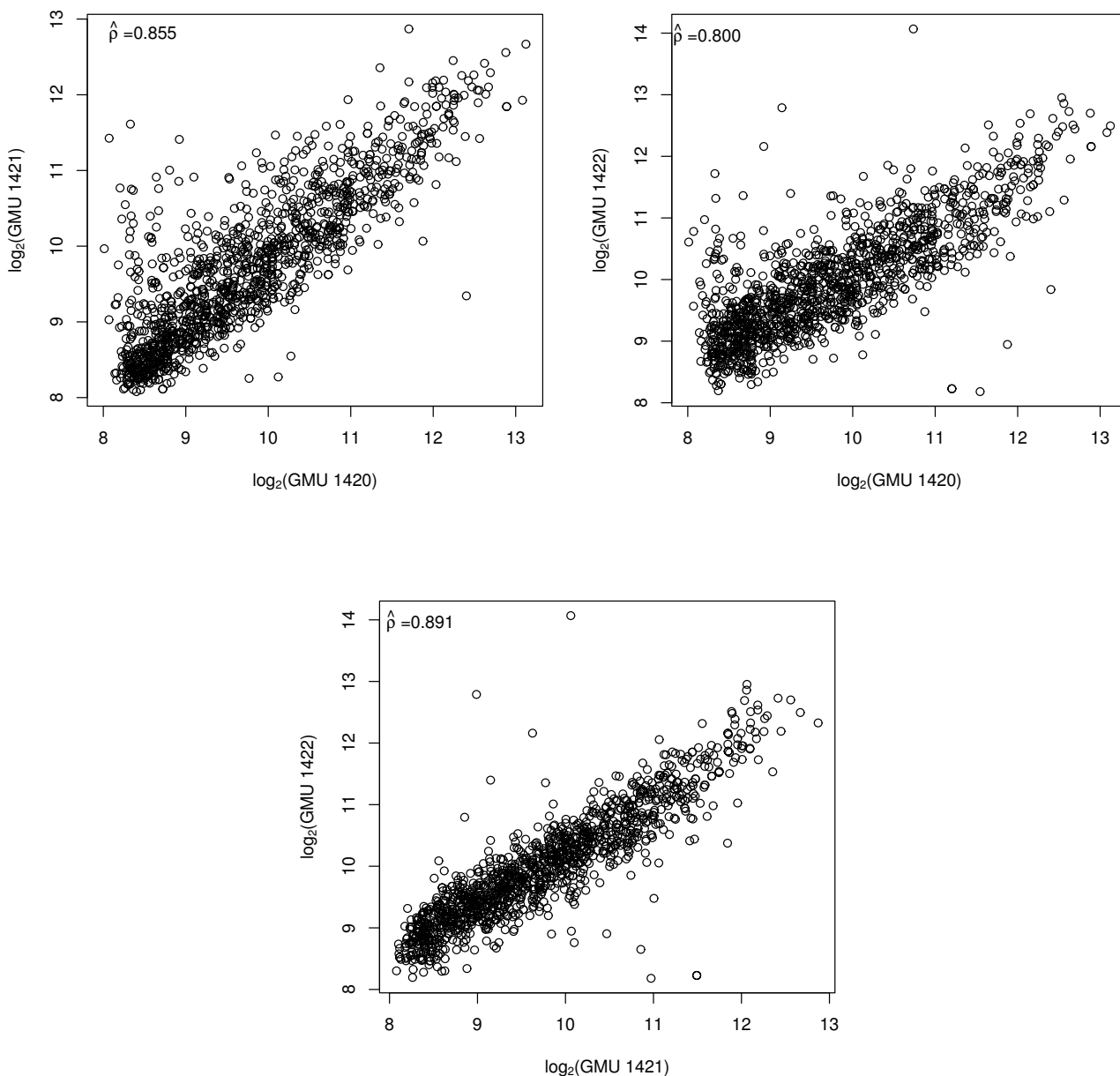


Figure 3
GMU. Pairwise scatterplots and Pearson's correlation for GMU arrays restricted to the 1,288 genes in common among the three platforms.

expression summaries as 'naïve' estimates of correlation. In addition, the disattenuated correlations ($\tilde{\rho}$), obtained when considering that the C3B and Affy gene intensities are subject to measurement error, are also reported. Noting that the attenuation for the C3B arrays is 0.386, that is, over half of the variability is attributed to measurement error, the disattenuated correlations estimated using measurement error models are substantially higher, irre-

spective of the Affymetrix expression summary method used. This suggests that previous use of Pearson's correlation under-estimated true underlying cross-platform correlations. That is, the effect of the presence of random intra-platform variation is degraded performance on the apparent cross-platform correlation. Therefore, by removing random intra-platform variation through measurement error methodology, the cross-platform correlation will go up.

Table 2: Observed agreement and p-value for each pairwise comparison within each platform using the weighted kappa statistic. Print-tip loess normalized Cy5 intensities were used for both two-channel arrays; MAS5.0 expression summaries were used for Affymetrix GeneChips.

Chips compared	Observed Agreement	P-value
GMU		
1420 v 1421	71.89%	< 0.0001
1420 v 1422	66.31%	< 0.0001
1421 v 1422	75.41%	< 0.0001
C3B		
7-7 v 8-22	28.18%	1.00
7-7 v 8-31	51.82%	< 0.0001
8-22 v 8-31	29.14%	1.00
Affymetrix		
QAQC8 v QAQC10	89.75%	< 0.0001
QAQC8 v QAQC13	89.54%	< 0.0001
QAQC10 v QAQC13	89.57%	< 0.0001

The average Pearson correlations ($\bar{\rho}_w$) of the log transformed Affymetrix GeneChip expression and GMU array expression are also reported in Table 4 for MAS 5.0, RMA, and GC-RMA expression summaries, as well as the disattenuated correlations ($\tilde{\rho}$). The attenuation for the GMU arrays is 0.824, therefore the disattenuated correlations estimated using measurement error models are larger than their corresponding naïve estimates, though not as markedly in comparison to the C3B arrays. This is due to the higher reliability among the GMU expression intensities.

Discussion

In this paper, both intra- and cross-platform reproducibility was examined for the Affymetrix and two dual channel microarrays (C3B and GMU). We applied various methods for examining within-platform reproducibility including Pearson's correlation, the weighted kappa, and percent of invariant genes. We also examine cross-platform reproducibility using Pearson's correlation. We previously demonstrated the effectiveness of applying a correlation correction factor via a small simulation study and demonstrated its application in estimating gene-specific correlations. In this paper we demonstrated its use in

Table 3: Frequency and percent of invariant features from each platform (P < 0.0001). Print-tip loess normalized Cy5 intensities were used for both two-channel arrays; MAS5.0 expression summaries were used for Affymetrix GeneChips.

	Invariant N, (%)	Not Invariant N, (%)
Affymetrix	11,732 (52.65%)	10,551 (47.35%)
C3B	594 (2.81%)	20,574 (97.19%)
GMU	3,036 (14.34%)	18,132 (85.66%)

Table 4: Cross-platform average Pearson correlations ($\bar{\rho}_w$) and disattenuated cross-platform correlations ($\tilde{\rho}$) for Stratagene Technical Replicate Dataset using MAS 5.0, RMA, and GC-RMA Affy expression summaries.

Platform	MAS 5.0		RMA		GC-RMA	
	$\bar{\rho}_w$	$\tilde{\rho}$	$\bar{\rho}_w$	$\tilde{\rho}$	$\bar{\rho}_w$	$\tilde{\rho}$
C3B	0.240	0.391	0.210	0.338	0.168	0.270
GMU	0.384	0.428	0.348	0.383	0.399	0.440

estimating cross-platform reproducibility. We note that correcting for measurement error by estimating the "disattenuated" correlation removes the bias or attenuation inherent in cross-platform correlation estimates. Specifically, to the extent that random intra-platform variation is present, the effect is degraded performance on the apparent cross-platform correlation. Therefore, by removing random intra-platform variation through measurement error methodology, the cross-platform correlation will go up.

Due to the increased public availability of gene expression microarray data through Gene Expression Omnibus [31] and ArrayExpress [32], researchers are increasingly interested in methods that integrate the results from various microarray studies performed on similar types of samples [33-37]. A careful understanding of variability due to platform-specific bias and random intra-platform variability will help investigators select methods for integrating cross-platform results. Specifically, the amount of attenuation for a specific platform could be used as a platform-specific quality measure and incorporated into a meta-analytic framework [38]. Moreover, gene-specific attenuation factors could be used to adjust for quality in a gene-wise fashion in such models.

A major application of DNA microarray technology is differential gene expression profiling, or the detection of the differences in expression levels of genes between two different types of samples. Some have argued that the consistency of the differences via fold-change or ratio is a more relevant metric for assessing cross-platform comparability than intensities from a single channel. However, to estimate the correlation between fold-changes from two platforms, two different samples are needed. We therefore plan to use data from the MAQC project to examine cross-platform fold-change correlations. In addition, it has been suggested that a more relevant metric is not agreement in the identification of individual differentially expressed genes, but rather whether consistent and accurate predictions of sample class is obtained from the

platforms being compared [39]. This metric should be included in such cross-platform studies as well.

Previous researchers demonstrated that single and two channel microarrays yield consistent results, and concluded that the selection of which technology to use is not necessarily a critical factor in the design of a microarray study [20]. Here we demonstrate the critical need to thoroughly evaluate intra-platform reproducibility, a finding which has been noted by others [26]. In this study, we examined two dual channel platforms and the Affymetrix platform. While the C3B and GMU platforms are not widely used by the microarray research community, they do represent a class of microarrays that are commonly used, two channel custom spotted/home brewed arrays. Thus, we believe these results are of general interest to those who use both commercial and custom designed arrays. While the C3B two channel platform had poor reproducibility, the GMU two channel and Affymetrix platforms had good reproducibility. We repeated the intra-platform analysis using the following three sets of randomly selected Affymetrix GeneChips (6, 12, 2), (5, 16, 14), and (5, 2, 3) and the intra-platform Affymetrix results were consistently reproducible with what is presented in this paper. This high reproducibility of the Affymetrix GeneChip data has also been reported by other investigators [14,40]. These data have proven useful in selecting a platform for studying biological specimens being collected by our tissue bank. We recommend that prior to performing expensive microarray hybridizations using irreplaceable biological specimens procured from clinical studies, a thorough assessment of intra-platform reproducibility be conducted.

One limitation of this study is that platform is completely confounded with laboratory technician and protocol, that is, the platform-specific sequence of reactions, scanner, procedures and events involved in the production of microarray data. It was previously noted that there is a high positive correlation between technician experience and intra-platform correlation [25]. This is consistent with our findings, whereby a first year graduate student performed the C3B hybridizations ($\bar{\rho} = 0.656$), while the GMU and Affy hybridizations were performed by Ph.D. faculty members ($\bar{\rho} = 0.848$ and $\bar{\rho} = 0.996$, respectively). Future studies that control for external factors that may influence intra-platform reliability are warranted.

In calculating cross-platform correlation, we assumed that the correlation estimated using the 1288 matching probes across the three platforms are representative of expected correlation of genes in the human genome that could be represented on the platforms. Examination of

absolute tag counts for the Stratagene Total Human RNA obtained using Serial Analysis of Gene Expression data (available from GEO #GSM1734) revealed that the intensity distribution of the 1,288 genes in common among the three platforms is not representative of the range of expected values (Figures 4, 5, 6, 7). Thus the commonly invoked procedure of estimating cross-platform consistency using only probes in common to all platforms is demonstrated to suffer from bias related to genomic coverage and probe annotation. Future studies comparing commercially available and custom designed arrays need to take this into consideration.

Conclusion

When estimating cross-platform correlation, it is essential to thoroughly evaluate intra-platform reproducibility as a first step. We also note that the commonly invoked procedure of estimating cross-platform consistency using only probes in common to all platforms is demonstrated to suffer from bias related to genomic coverage and probe annotation. Future studies comparing commercially available and custom designed arrays need to take this into consideration. Moreover, to the extent that random intra-platform variation is present, the effect is degraded performance on the apparent cross-platform correlation. Therefore, by removing random intra-platform variation through measurement error methodology, the cross-platform correlation will go up. Methods to correct for atten-

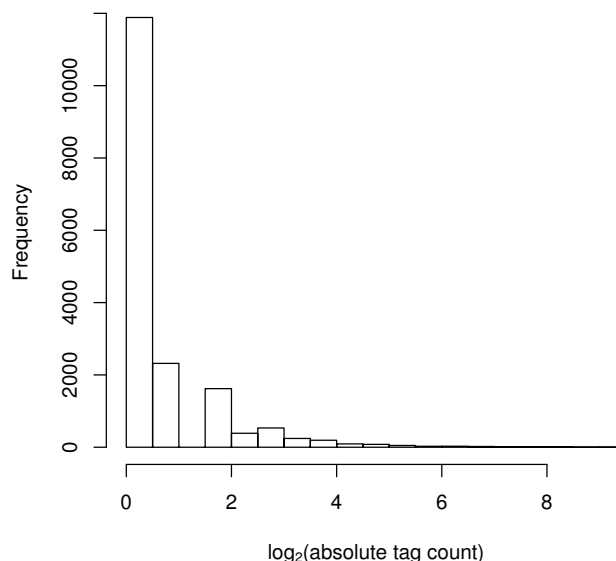


Figure 4
Histogram of log₂ absolute tag counts from SAGE.
 Histogram of log₂ absolute tag counts from Serial Analysis of Gene Expression using the Stratagene Total Human RNA for the 14000 unique tags. Data available from GEO Accession #GSM1734.

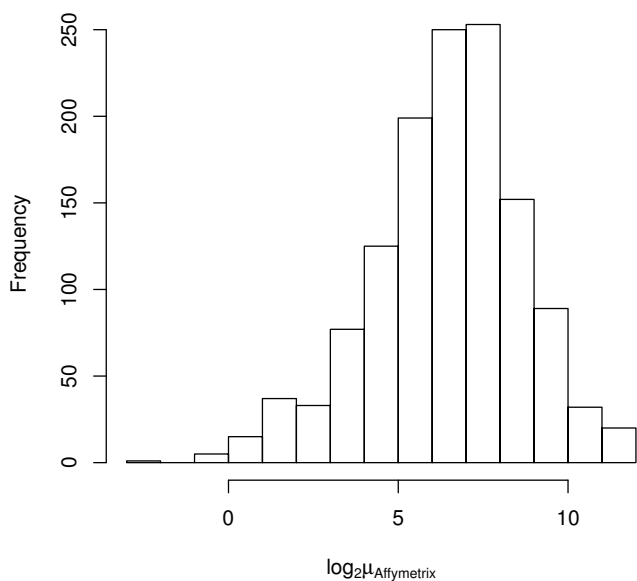


Figure 5
Histogram of log₂ average Affymetrix MAS5 signal. Histogram of log₂ average Affymetrix MAS5 signal for the Stratagene Total Human RNA using the 1,288 genes in common among the three platforms.

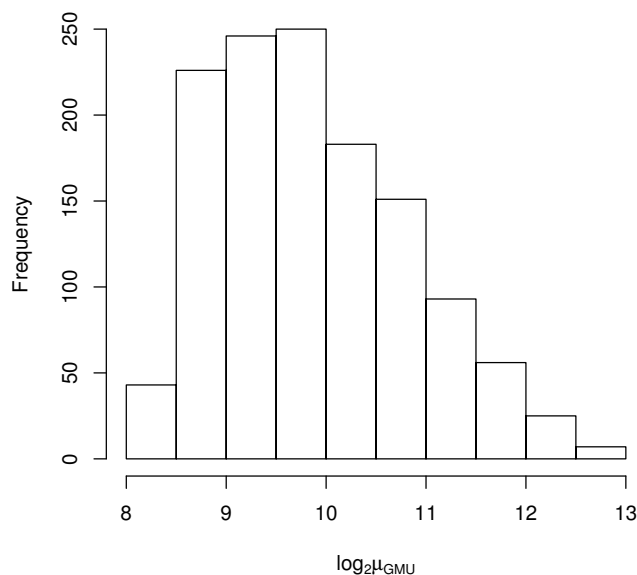


Figure 7
Histogram of log₂ average GMU signal. Histogram of log₂ average GMU signal for the Stratagene Total Human RNA using the 1,288 genes in common among the three platforms.

uation, such as that presented, are thus useful in decreasing such a bias in cross-platform correlation estimates. Platform-specific attenuation estimates may subse-

quently be used as a platform-specific quality measure and incorporated into a meta-analytic framework.

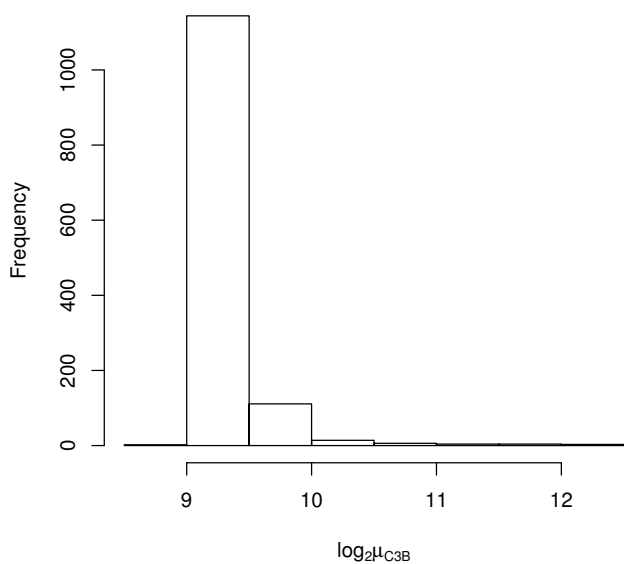


Figure 6
Histogram of log₂ average C3B signal. Histogram of log₂ average C3B signal for the Stratagene Total Human RNA using the 1,288 genes in common among the three platforms.

Methods

Stratagene Technical Replicates Dataset

Previously, each laboratory designed a small experiment to assess intra-platform quality control. Each laboratory used the same lot of reference RNA, the Stratagene Total Human RNA, for hybridizing a set of technical replicates for a process variability study. These 'self-self' hybridizations permit meaningful assessments of reproducibility since, under ideal circumstances such as that the same experimental conditions exist among platforms and that there are no probe-binding affinity effects, each gene across the set of chips should exhibit linearly related gene expression intensities across platforms. Although the RNA hybridized was from the same lot, the study designs and protocols differed from lab to lab.

The Affy platform was assessed using an unbalanced three-factor design using 16 technical replicates [41]. The same reference RNA sample was examined in 16 different chips run on two days in four different modules of the Affymetrix fluidics workstation. Fresh fragmented cRNAs were hybridized to the first four GeneChips on Day 1 while frozen fragmented cRNAs were hybridized to remaining four GeneChips on Day 1 and to all eight GeneChips processed on Day 2. To eliminate operator

variations, the same person completed the synthesis and hybridization of all 16 chips. The images were scanned at a 6 μm resolution using the Agilent G2500A Technologies Gene Array scanner. The full set of 16 Affymetrix GeneChips is publicly available [42].

At GMU, the RNA was amplified using the MessageAmp aRNA Kit (Ambion). The amplified RNA (aRNA) was quantified and its quality was monitored by agarose gel and average size by the Agilent 2100 Bioanalyzer. The same amount of aRNA (4 μg) were labeled with Cy3 and Cy5 according to the The Institute for Genomic Research protocol and hybridized to three Human I chips. For each chip, the Stratagene Total Human RNA served as both the experimental and reference sample [43]. The ScanArray Express HT confocal laser scanner with settings at 75% of photomultiplier tube, 75% of laser power, and 10 μm of pixel resolution was used. Images were acquired by ScanArray Express 2.0 software and processed with QuantArray software.

The C3B laboratory assessed quality of their fabricated microarray using a fractional factorial design. The factors investigated were cDNA labeling strategy (3 levels: Dye conjugated nucleotide, aminoallyl, and Genesphere dendimer labeling), input total RNA concentration ratio (3 levels: 1:1, 1:2, 1:4), hybridization time (2 levels: 4 and 16 hours), hybridization buffer (3 levels: Genesphere, MWG, and Amersham buffer), and production lot (2 levels: lot 7 and 9). Due to the expense of microarray production and hybridization, a fractional factorial design, rather than the full factorial design, was used. Therefore, all combinations of experimental conditions were not included. Specifically, by assuming that high-order interactions are negligible, information regarding the main effects and low-order interactions may be obtained by running only a fraction of the complete factorial design. Since we were interested in examining the effects of hybridization buffer (3 levels), RNA input ratio (3 levels), labeling strategy (3 levels), hybridization time (2 levels), and lot (2 levels), we were initially interested in a $3^3 \times 2^2$ design. However, due to the expense involved in running a full factorial microarray experiment, a 2^{8-2} fractional factorial design was adopted with defining relation is $I = ABCDG = ABEFH = CDEFGH$. This resolution V design permits estimation of all main effects and two-factor interactions under the assumption that three-way and higher order interaction terms may be ignored. Thus our experiment required 64 C3B arrays to be hybridized given the factors and levels of interest. Again, for each array the Stratagene Total Human RNA served as both the experimental and reference sample. Hybridized arrays were scanned with ScanArray Express microarray scanner (Perkin Elmer) at 80% laser power, 70% PMT gain, and 5 μm scan resolution. Spot

intensities were acquired from the images using QuantArray software.

The analyses conducted in the current study were restricted to an equal number of chips by platform to ensure one technology did not dominate the results simply because of having a larger sample size. Three arrays were hybridized at GMU, so a random sample of size 3 was taken from the 16 Affy hybridized samples. These three GeneChips were QAQC8.CEL (Day 1 Frozen), QAQC10.CEL (Day 2 Frozen), and QAQC13.CEL (Day 2 Frozen). The three replicates selected from the C3B fractional factorial study were chosen based on 'optimal' hybridization conditions identified from the fractional factorial experiment. Specifically, the number of genes found to be significantly different from the analysis of variance model was used as the metric estimating the relative influence of each main and two-factor interaction term. The level of each factor having the smallest number of genes differentially expressed was considered optimal. The three C3B chips used in this study were hybridized using the same buffer (Amersham), ratio of input experimental and control samples (1:1), and labeling method (Aminoallyl Post RT). The chips differed with respect to lot number and hybridization time, though these factors were found to not significantly influence the resulting intensities in the larger study.

Normalization

Since single-channel arrays measure expression intensities on an absolute scale whereas two-channel arrays measure expression intensities on a ratio-metric scale, we first investigated intra-platform reproducibility using different methods for calculating gene expression to aid in our determination of how to best transform the intensities from the three platforms to a similar scale. In addition, since the objective included an assessment of platform-specific reproducibility across the set of available technical replicates, methods for within-array normalization rather than methods that simultaneously normalize the data across all arrays, were applied in a platform-specific fashion.

For the two-channel arrays, we employed a commonly used procedure of normalizing the spot-level intensities on the array using print-tip loess regression and the subsequently analyzing the normalized spot-level intensities [44]. The use of normalized spot intensities has removed the systematic sources of variability (or at least, reduced) attributed to technical artifacts of no interest, such as deposition differences, differences in labeling efficiencies, print-tip differences etc. Specifically, due to spot differences attributed to deposition gain, print-tip, and dye effects noted among two-channel arrays, each two-chan-

nel array (C3B and GMU) was normalized by estimating the corrections \hat{M}_i for spots $i = 1, \dots, G$ by fitting print-tip loess regression models to the $M_i = \log_2(\text{channel } 1_i / \text{channel } 2_i)$ (log difference) on $A_i = (\log_2(\text{channel } 1_i) + \log_2(\text{channel } 2_i))/2$ (log average) [45]. Probe intensities were then adjusted by $M_i^{norm} = M_i - \hat{M}_i$, therefore, M_i^{norm} represents the normalized log ratios [46]. In addition, to enforce an absolute expression measure, the normalized ratios were subsequently transformed to yield the channel 1 normalized intensities by $x_i^{norm} = 2^{\frac{M_i^{norm}}{2}}$ [44]. Background was estimated by the Quantarray software as the mean intensity among those pixels within the masked area between the 5th and 20th percentile of intensities for a given spot. Since simple background subtraction has been demonstrated to increase spot-level variability [47], no background correction was applied.

The Affymetrix GeneChip Operating System (GCOS) was used to calculate expression summaries with a target intensity of 100 using the Microarray Suite version 5.0 (MAS 5.0) method [48]. For completeness, we also estimated expression using the robust multiarray average (RMA) [49] and GC-RMA methods [50], although these methods normalize and estimate probe set expression summaries utilizing data across the entire set of GeneChips and therefore may overestimate reproducibility. All normalization and expression summary methods were performed using the R software [51] and relevant Bioconductor packages [52].

Identifying common genes across platforms

The RESOURCERER annotation and cross-reference database [53] was developed to help investigators identify genes commonly interrogated by different microarray platforms. Other software tools such as MergeMaid [54], GeneHopper [55], MatchMiner [56], and ProbeMatchDB [57] have been developed for a similar purpose. Recent research has demonstrated improved cross-platform correlations when spots are matched by sequence rather than by gene identifiers [58-60].

Therefore, probe sets and spots with common sequences to all three platforms were retained for analysis using the following method. First, the GCG program 'netfetch' was used to obtain the NCBI GenBank records for spot IDs on the GMU and C3B microarray platforms. The perfect match (PM) probe level sequence data for the Affymetrix HG-U133A GeneChip was downloaded from the Affymetrix website (06/14/2005). BLASTN (v2.2.10) was used to query the Affymetrix probe sequences against the C3B

sequences. Thereafter, all probe sets for which at least 60% of the probes reported low e-scores values ($E < 0.000001$) for the same spot were retained as matches. This threshold was determined considering the breakdown bound of the Tukey biweight estimator used in the MAS 5.0 expression summary algorithm. M-estimators with symmetric ψ -function have breakdown bound close to 50%. Therefore, probe sets for which > 60% of its PM probes specifically interrogated the same RefSeqID were retained. For the C3B microarray, each RefSeqID is spotted two times on the array. For the intra-platform reliability study (Stratagene dataset), average spot intensity per RefSeqID was retained as C3B gene expression. For the Affymetrix GeneChips, when multiple probe sets interrogated the same transcript, first, that probe set with the maximum proportion of probes with $E < 0.000001$ was retained; when two or more probe sets had the same proportion, then the most 3' probe set was retained, defined by the probe set with maximum stop query sequence location among probes within a GenBank ID; when both quantities were the same, the probe set was randomly selected.

This process was completed separately for the Affy-C3B and Affy-GMU platform pairs. These two resulting datasets were merged by Affymetrix probe set ID, resulting in a dataset containing only genes in common to all three platforms.

All raw microarray files used in this study are publicly available [61].

Intra-platform analyses

It has been suggested that poor cross-platform correlation is likely a result of low intra-platform consistency [26]. Therefore, prior to estimating cross-platform reproducibility and gene-specific reliability, intra-platform reproducibility for three different microarray platforms was examined. After normalization and calculation of gene expression summaries, within-platform correlation was estimated using average Pearson correlation for the $K = 3$ chips. In addition, reproducibility was examined by comparing the proportion of invariant genes across the set of technical replicates within a platform. Specifically, for spot $i = 1, \dots, G$, the ranked expression for the k^{th} replicate of platform l is denoted by R_{ikl} . We then identified the rank difference for each spot i within platform l as $\Delta_{ij} = \text{abs}(\text{argmax}_{ij}(R_{ikl}) - \text{argmin}_{ij}(R_{ikl}))$. A gene was designated as 'invariant' for platform l using the indicator $I(\Delta_{ij}/G \leq 0.05)$. As an example, this would correspond to permitting the rank to shift by no more than 1,114 when 22,283 genes are spotted on the array. Statistical tests of hypothesis comparing the proportions of invariant genes across platforms were conducted using a chi-square test.

Finally, the weighted kappa statistic was estimated by first grouping gene expression intensities into 25 approximately equal-sized classes based on their ranked intensities, γ_i . A weighted kappa statistic was used to allow a smaller penalty of misclassification among closely related classes, where the weights were taken to be $w_{rc} = (1 - 0.1 \times |r - c|)$ when $|r - c| < 10$ and 0 otherwise.

Attenuation

When fitting a linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

for observed random variables x_i and y_i on observations $i = 1, \dots, n$, it is assumed $x_i \sim N(\mu_x, \sigma_x^2)$, $\varepsilon_i \sim N(0, \sigma_e^2)$ which is independent of x_i , and x_i is measured without error [62]. Using the formulas for estimating Pearson's correlation and the slope parameter β_1 , Pearson's correlation can be shown to be

$$\hat{\rho}(x, y) = \frac{\hat{\sigma}_x}{\hat{\sigma}_y} \hat{\beta}_1.$$

Therefore, Pearson's correlation measures the strength of the linear relationship between X and Y.

For a general problem, suppose x_i cannot be measured precisely but rather is measured with error. Denote the error-prone measurements $x_{i_w} = x_i + u_i$ where $u_i \sim (0, \sigma_u^2)$.

It is well known that fitting the model

$$y_i = \beta_0 + \beta_{1*} x_{i_w} + \varepsilon_i$$

using the error-prone values x_{i_w} leads to the attenuated estimate β_{1*} for β_1 [28]. That is, the slope parameter is biased. Therefore, when fitting a simple linear regression model using the error prone measurements x_{i_w} , the least-squares estimate is

$$\beta_{1*} = \lambda \beta_1,$$

where β_1 is the true slope parameter describing the relationship between y_i and x_i and λ is the attenuation factor. The attenuation factor is given by

$$\lambda = \sigma_x^2 / (\sigma_x^2 + \sigma_u^2) < 1$$

and is used to estimate β_1 when measurement error is present in both X and Y [28].

Estimating cross-platform correlation

From the intra-platform results, it is clear that microarray gene expression data is subject to measurement error. When estimating cross-platform correlation, let X and Y represent the random variables for two different platforms, known to be measured with error. That is, $X_{i_w} = X_i + u_i$ where $X_i \sim N(\mu_x, \sigma_x^2)$ and $u_i \sim (0, \sigma_u^2)$ while $Y_{i_w} = Y_i + v_i$ where $Y_i \sim N(\mu_y, \sigma_y^2)$, $v_i \sim (0, \sigma_v^2)$. The average Pearson's correlation ($\bar{\rho}_w$), which is not corrected for measurement error, can be estimated as

$$\bar{\rho}_w = \sum_{j=1}^3 \hat{\rho}(\bar{x}_w, \gamma_{j_w}) / 3,$$

where \bar{x}_w is the average \log_2 Affymetrix intensities and γ_{j_w} is C3B or GMU expression. However, a more appropriate measure, the "disattenuated" correlation [30], can be calculated as

$$\rho_w = \lambda_p \times \rho$$

where

$$\lambda_p = \frac{\sigma_x \sigma_y}{\sqrt{\sigma_x^2 + \sigma_u^2} \sqrt{\sigma_y^2 + \sigma_v^2}}.$$

This estimate adjusts for the bias present in estimating the correlation when measurement error is present. Estimates for σ_x , σ_u , σ_y and σ_v were fit using the regression calibration rcal function in Stata version 9 [63]. In estimating σ_u^2 and σ_v^2 , the repeated measurements were assumed to be unbiased for the true gene expression values. Moreover, any missing value was treated as missing at random. Previous investigators have reported high reproducibility estimates for Affymetrix expression values [14,40], therefore, we were primarily interested in estimating the correlation between Affymetrix and the custom designed arrays (C3B and GMU) that we have used in various cancer genomics projects. The disattenuated correlation, $\bar{\rho}$, and average Pearson correlation, $\bar{\rho}_w$, were estimated separately for the GMU and C3B platforms relative to Affymetrix.

Authors' contributions

KJA performed the statistical analyses and drafted the manuscript. CID, AFG, and CTG designed and performed the MDX Affymetrix quality control study. GST and TGE designed and performed the C3B quality control study. GMG designed and performed the GMU quality control study. MDC performed the BLAST search and assisted with merging the cross-platform data. All authors read and approved the final manuscript.

Acknowledgements

This research was supported by the Commonwealth Technology Research Fund (CTRF #SE2002 02) and the Center for Bioelectronics, Biosensors and Biochips.

References

- Hwang KB, Kong SW, Greenberg SA, Park PJ: **Combining gene expression data from different generations of oligonucleotide arrays.** *BMC Bioinformatics* 2004, **5**:159.
- Nimgaonkar A, Sanoudou D, Butte AJ, Haslett JN, Kunkel LM, Beggs AH, Kohane IS: **Reproducibility of gene expression across generations of Affymetrix microarrays.** *BMC Bioinformatics* 2003, **4**:27.
- Yue H, Eastman PS, Wang BB, Minor J, Doctolero MH, Nuttall RL, Stack R, Becker JW, Montgomery JR, Vainer M, Johnston R: **An evaluation of the performance of cDNA microarrays for detecting changes in global mRNA expression.** *Nucleic Acids Research* 2001, **29**(8):e41.
- Yang IV, Chen E, Hasseman JP, Liang W, Frank BC, Wang S, Sharov V, Saeed AI, White J, Li J, Lee NH, Yeatman TJ, Quackenbush J: **Within the fold: assessing differential expression measures and reproducibility in microarray assays.** *Genome Biology* 2002, **3**:research0062.1-0062.12.
- Kuo W, Jenssen T, Butte A, Ohno-Machado L, Kohane I: **Analysis of matched mRNA measurements from two different microarray technologies.** *Bioinformatics* 2002, **18**:405-412.
- Yuen T, Wurmbach E, Pfeffer RL, Ebersole BJ, Sealfon SC: **Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays.** *Nucleic Acids Research* 2002, **30**:1-9.
- Barczak A, Rodriguez MW, Hanspers K, Koth LL, Tai YC, Bolstad BM, Speed TP, Erle DJ: **Spotted long oligonucleotide arrays for human gene expression analysis.** *Genome Research* 2003, **13**:1775-1785.
- Tan P, Downey T, Spitznagel E, Xu P, Fu D, Dimitrov D, Lempicki R, Raaka B, Cam M: **Evaluation of gene expression measurements from commercial microarray platforms.** *Nucleic Acids Research* 2003, **31**:5676-5684.
- Rogojina AT, Orr WE, Song BK, Geisert EE: **Comparing the use of Affymetrix to spotted oligonucleotide microarrays using two retinal pigment epithelium cell lines.** *Molecular Vision* 2003, **9**:482-496.
- Petersen D, Chandramouli G, 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: **Three microarray platforms: an analysis of their concordance in profiling gene expression.** *BMC Genomics* 2005, **6**:63.
- Parrish ML, Wei N, Duenwald S, Tokiwa GY, Wang Y, Holder D, Dai H, Zhang X, Wright C, Hodor P, Cavet G, Phillips RL, Sun BI, Fare TL: **A microarray platform comparison for neuroscience applications.** *Journal of Neuroscience Methods* 2004, **132**:57-68.
- Martinez-Murillo F, Hoffman E: **Comparison of spotted cDNA arrays and Affymetrix oligonucleotide arrays: High concordance under stringent parameters.** *American Journal of Human Genetics* 2001, **69**:468.
- Woo Y, Affourtit, Daigle S, Viale A, Johnson K, Naggert J, Churchill G: **A comparison of cDNA, oligonucleotide, and Affymetrix GeneChip gene expression microarray platforms.** *Journal of Biomolecular Techniques* 2004, **15**:276-284.
- Yauk C, Berndt L, Williams A, Douglas G: **Comprehensive comparison of six microarray technologies.** *Nucleic Acids Research* 2004, **32**:e124.
- Park PJ, Cao YA, Lee SY, Kim JW, Chang MS, Hart R, Choi S: **Current issues for DNA microarrays: platform comparison, double linear amplification, and universal RNA reference.** *Journal of Biotechnology* 2004, **112**:225-245.
- Mah N, Thelin A, Lu T, Nikolaus S, Kühbacher T, Gurbuz Y, Eickhoff H, Klöppel G, Lehrach H, Mellgard B, Costello CM, Stefan S: **A comparison of oligonucleotide and cDNA-based microarray systems.** *Physiological Genomics* 2004, **16**:361-370.
- Lee J, Bussey K, Gwadry F, Reinhold W, Riddick G, Pelletier S, Nishizuka S, Szakacs G, Annereau J, Shankavaram U, Lababidi S, Smith L, Gottesman M, Weinstein J: **Comparing cDNA and oligonucleotide array data: concordance of gene expression across platforms for the NCI-60 cancer cells.** *Genome Biology* 2003, **4**:R82.
- Larkin JE, Frank BC, Gavras H, Quackenbush J: **Independence and reproducibility across microarray platforms.** *Nature Methods* 2005, **2**:337-344.
- Shi L, Reid L, Jones W, Shippy R, Warrington J, Baker S, Collins P, de Longueville F, Kawasaki E, Lee K, Luo Y, Sun Y, Willey J, Setterquist R, Fischer G, Tong W, Dragan Y, Dix D, Frueh F, Goodsaid F, Herman D, Jensen R, Johnson C, Lobenhofer E, Puri R, Schrf U, Thierry-Mieg J, Wang C, Wilson M, Wolber P, Zhang L, Amur S, Bao W, Barbacioru C, Lucas A, Bertholet V, Boysen C, Bromley B, Brown D, Brunner A, Canales R, Cao X, Cebula T, Chen J, Cheng J, Chu T, Chudin E, Corson J, Corton J, Croner L, Davies C, Davison T, Delenstarr G, Deng X, Dorris D, Eklund A, Fan X, Fang H, Fulmer-Smentek S, Fuscoe J, Gallagher K, Ge W, Guo L, Guo X, Hager J, Haje P, Han J, Han T, Harbottle H, Harris S, Hatchwell E, Hauser C, Hester S, Hong H, Hurban P, Jackson S, Ji H, Knight C, Kuo W, LeClerc J, Levy S, Li Q, Liu C, Liu Y, Lombardi M, Ma Y, Magnuson S, Maqodi B, McDaniel T, Mei N, Myklebost O, Ning B, Novoradovskaya N, Orr M, Osborn T, Papallo A, Patterson T, Perkins R, Peters E, Peterson R, Philips K, Pine P, Pusztai L, Qian F, Ren H, Rosen M, Rosenzweig B, Samaha R, Schena M, Schroth G, Shchegrova S, Smith D, Staedtler F, Su Z, Sun H, Szallasi Z, Tezak Z, Thierry-Mieg D, Thompson K, Tikhonova I, Turpaz Y, Valanar B, Van C, Walker S, Wang S, Wang Y, Wolfinger R, Wong A, Wu J, Xiao C, Xie Q, Xu J, Yang W, Zhang L, Zhong S, Zong Y, Jr WS, MAQC Consortium: **The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements.** *Nature Biotechnology* 2006, **24**(9):1151-1160.
- Patterson T, Lobenhofer E, Fulmer-Smentek S, Collins P, Chu T, Bao W, Fang H, Kawasaki E, Hager J, Tikhonova I, Walker S, Zhang L, Hurban P, de Longueville F, Fuscoe J, Tong W, Shi L, Wolfinger R: **Performance comparison of one-color and two-color platforms within the Microarray Quality Control (MAQC) project.** *Nature Biotechnology* 2006, **24**(9):1140-1150.
- Canales R, Luo Y, Willey J, Austermliller B, Barbacioru C, Boysen C, Hunkapiller K, Jensen R, Knight C, Lee K, Ma Y, Maqodi B, Papallo A, Peters E, Poulter K, Ruppel P, Samaha R, Shi L, Yang W, Goodsaid F: **Evaluation of DNA microarray results with quantitative gene expression platforms.** *Nature Biotechnology* 2006, **24**(9):1115-1122.
- Shippy R, Fulmer-Smentek S, Jensen RV, Jones WD, Wolber PK, Johnson CD, Pine PS, Boysen C, Guo X, Chudin E, Sun YA, Wiley 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: **Using RNA sample titrations to assess microarray platform performance and normalization techniques.** *Nature Biotechnology* 2006, **24**(9):1123-1131.
- Tong W, Lucas AB, Shippy R, Fan X, Fang H, Hong H, Orr MS, Chu TM, Guo X, Collins PJ, Sun YA, Wang SJ, Bao W, Wolfinger RD, Shchegrova S, amd Janet A, Warrington LG, Shi L: **Evaluation of external RNA controls for the assessment of microarray performance.** *Nature Biotechnology* 2006, **24**(9):1132-1139.
- Guo L, Lobenhofer EK, Wang C, 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: **Rat toxicogenomic study reveals analytical consistency across microarray platforms.** *Nature Biotechnology* 2006, **24**(9):1162-1169.
- Irizarry R, Warren D, Spencer F, Biswal S, Frank B, Gabrielson E, Garcia J, Geoghegan J, Germino G, Griffin C, Hilmer S, Hoffman E, Jedlicka A, Kawasaki E, Kim I, Morsberger L, Lee H, Peterson D, Quackenbush J, Scott A, Wilson M, Yang Y, Ye S, Tyu W: **Multiple-laboratory comparison of microarray platforms.** *Nature Methods* 2005, **2**:345-350.

26. Shi L, Tong W, Fang H, Scherf U, Han J, Puri R, Fruech F, Goodsaid F, Guo L, Su Z, Han T, Fuscoe J, Xu Z, Patterson T, Hong H, Xie Q, Perkins R, Chen J, Casciano D: **Cross-platform comparability of microarray technology: Intra-platform consistency and appropriate data analysis procedures are essential.** *BMC Bioinformatics* 2004, **6(Suppl 2)**:S212.
27. Shi L, Tong W, Su Z, Han T, Han J, Puri RK, Fang H, Frueh FW, Goodsaid FM, Guo L, Branham WS, Chen JJ, Xu ZA, Harris SC, Hong H, Xie Q, Perkins RG, Fuscoe JC: **Microarray scanner calibration curves: characteristics and implications.** *BMC Bioinformatics* 2005, **6(Suppl 2)**:S11.
28. Carroll R, Ruppert D, Stefanski L, Crainiceanu C: *Measurement Error in Nonlinear Models: A Modern Perspective* New York: Chapman & Hall; 2006.
29. Shi L, Tong W, Goodsaid FM, Fruech FW, Fang H, Han T, Fuscoe JC, Casciano DA: **QA/QC: challenges and pitfalls facing the microarray community and regulatory agencies.** *Expert Review of Molecular Diagnostics* 2004, **4**:761-777.
30. Archer KJ, Dumur CI, Taylor GS, Chaplin MD, Guiseppi-Elie A, Buck GA, Grant GM, Ferreira-Gonzalez A, Garrett CT: **A disattenuated correlation estimate when variables are measured with error: Illustration estimating cross-platform correlations.** *Statistics in Medicine* 2007 in press. doi: 10.1002/sim2984.
31. Barrett T, Suzek T, Troup D, Wilhite S, Ngau W, Ledoux P, Rudnev D, Lash A, Fujibuchi W, Edgar R: **NCBI GEO: mining millions of expression profiles – database and tools.** *Nucleic Acids Research* 2005, **33**:D562-D566.
32. Parkinson H, Sarkans U, Shojatalab M, Abeygunawardena N, Contrino S, Coulson R, Farne A, Lara G, Holloway E, Kapushesky M, Lilja P, Mukherjee G, Oezcimen A, Rayner T, Rocca-Sera P, Sharma A, Sansone S, Brazma A: **ArrayExpress-a public repository for microarray gene expression data at the EBI.** *Nucleic Acids Research* 2005, **33**:D553-D555.
33. Rhodes D, Barrette T, Rubin M, Ghosh D, Chinnaiyan A: **Meta-analysis of microarrays: Interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer.** *Cancer Research* 2002, **62**:4427-4433.
34. Rhodes D, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan A: **Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression.** *Proceedings of the National Academy of Science* 2004, **101**:9309-9314.
35. Grützmann R, Boriss H, Ammerpohl O, Lüttges J, Kalthoff H, Schackert H, Klöppel G, Saeger H, Pilarsky C: **Meta-analysis of microarray data on pancreatic cancer defines a set of commonly dysregulated genes.** *Oncogene* 2005, **24(32)**:5079-5088.
36. Ghosh D, Barrette T, Rhodes D, Chinnaiyan A: **Statistical issues and methods for meta-analysis of microarray data: a case study in prostate cancer.** *Functional and Integrative Genomics* 2003, **3**:180-188.
37. Shen R, Ghosh D, Chinnaiyan AM: **Prognostic meta-signature of breast cancer developed by two-stage mixture modeling of microarray data.** *BMC Genomics* 2004, **5**:94.
38. Hu P, Greenwood CM, Beyene J: **Integrative analysis of multiple gene expression profiles with quality-adjusted effect size models.** *BMC Bioinformatics* 2005, **6**:128.
39. Marshall E: **Getting the noise out of gene arrays.** *Science* 2004, **306**:630-631.
40. Järvinen A, Hautaniemi S, Edgren H, Auvinen P, Saarela J, Kallioniemi O, Monni O: **Are data from different gene expression microarray platforms comparable?** *Genomics* 2004, **83**:1164-1168.
41. Dumur C, Nasim S, Best A, Archer K, Ladd A, Mas V, Wilkinson D, Garrett C, Ferreira-Gonzalez A: **Evaluation of quality-control criteria for microarray gene expression analysis.** *Clinical Chemistry* 2004, **50**:1994-2002.
42. **Full set of 16 GeneChips from MDX** [http://www.ccrf-cagenomics.vcu.edu/QC_for_MicroarrayGeneExpressionAnalysis.html]
43. Grant G, Fortney A, Gorreta F, Estep M, Giacco LD, Meter AV, Christensen A, Appalla L, Naouar C, Jamison C, Al-Timimi A, Donovan J, Cooper J, Garrett C, Chandhoke V: **Microarrays in cancer research.** *Anticancer Research* 2004, **24**:441-448.
44. Allison D, Page G, Beasley T, Edwards J, Eds: *DNA Microarrays and Related Genomics Techniques: Design, Analysis, and Interpretation of Experiments* Chapman Hall/CRC Press chap. Normalization of microarray data; 2006:9-28.
45. Dudoit S, Yang Y, Callow M, Speed T: **Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments.** *Statistica Sinica* 2002, **12**:111-139.
46. Yang Y, Dudoit S, Luu P, Lin D, Peng V, Ngai J, Speed T: **Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation.** *Nucleic Acids Research* 2002, **30**:e15.
47. Kooperberg C, Fazio T, Delrow J, Tsukiyama T: **Improved background correction for spotted DNA microarrays.** *Journal of Computational Biology* 2002, **9**:55-66.
48. Hubbell E, Lui W, Mei R: **Robust estimators for expression analysis.** *Bioinformatics* 2002, **18**:1585-1592.
49. Irizarry R, Hobbs B, Collin F, Beazer-Barclay Y, Antonellis K, Scherf U, Speed T: **Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data.** *Biostatistics* 2003, **4**:249-264.
50. Wu Z, Irizarry R, Gentleman R, Martinez-Murillo F, Spencer F: **A model-based background adjustment for oligonucleotide expression arrays.** *Journal of the American Statistical Association* 2004, **99**:909-917.
51. R Development Core Team: *R: A language and environment for statistical computing* 2005 [<http://www.R-project.org>]. R Foundation for Statistical Computing, Vienna, Austria [ISBN 3-900051-07-0]
52. Gentleman R, Carey V, Bates D, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini A, Sawitzki G, Smith C, Smyth G, Tierney L, Yang J, Zhang J: **Bioconductor: open software development for computational biology and bioinformatics.** *Genome Biology* 2004.
53. Tsai J, Sultana R, Lee Y, Perte G, Karamycheva S, Antonescu V, Cho J, Parvizi B, Cheung F, Quackenbush J: **RESOURCERER: a database for annotating and linking microarray resources within and across species.** *Genome Biology* 2001, **2**:1-4.
54. Cope L, Zhong X, Garrell E, Parmigiani G: **MergeMaid: R Tools for Merging and Cross-Study Validation of Gene Expression Data.** *Statistical Applications in Genetics and Molecular Biology* 2004, **3**:Article 29.
55. Svensson BAT, Kreeft AJ, van Ommen GJ, den Dunnen JT, Boer J: **GeneHopper: a web-based search engine to link gene expression platforms through GenBank accession numbers.** *Genome Biology* 2003, **4**:R35.
56. Bussey KJ, Kane D, Sunshine M, Narasimhan S, Nishizuka S, Reinhold W, Zeeberg B, Weinstein A, Weinstein JN: **MatchMiner: a tool for batch navigation among gene and gene product identifiers.** *Genome Biology* 2003, **4**:R27.
57. Wang P, Ding F, Chiang H, Thompson RC, Watson SJ, Meng F: **ProbeMatchDB-a web database for finding equivalent probes across microarray platforms and species.** *Bioinformatics* 2002, **18**:488-489.
58. Mecham B, Wetmore D, Szallasi Z, Sadovsky Y, Kohane I, Mariani T: **Increased measurement accuracy for sequence-verified microarray probes.** *Physiological Genomics* 2004, **18**:308-315.
59. Mecham B, Klus G, Strovel J, Augustus M, Byrne D, Bozso P, Wetmore D, Mariani T, Kohane I, Szallasi Z: **Sequence-matched probes produce increased cross-platform consistency and more reproducible biological results in microarray-based gene expression measurements.** *Nucleic Acids Research* 2004, **32**:1-8.
60. Carter S, Eklund A, Mecham B, Kohane I, Szallasi Z: **Redefinition of Affymetrix probe sets by sequence overlap with cDNA microarray probes reduces cross-platform inconsistencies in cancer-associated gene expression measurements.** *BMC Bioinformatics* 2005, **6**:107.
61. **Raw data from three laboratories** [<http://www.people.vcu.edu/~kjarcher/Research/Data.htm>]
62. Neter J, Wasserman W, Kutner M: *Applied Linear Regression Models* Boston, MA: Irwin; 1989.
63. Hardin J, Schmidediche H, Carroll R: **The regression-calibration method for fitting generalized linear models with additive measurement error.** *The Stata Journal* 2003, **3**:361-372.