Independence and reproducibility across microarray platforms

Jennie E. Larkin[1], Bryan C. Frank[1], Haralambos Gavras[2], John Quackenbush[1,3*]

[1]The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850
[2]Boston University Medical Center, 715 Albany Street, Boston, MA 02118
[3]Department of Biochemistry, The George Washington University, Washington, DC, USA and Department of Statistics, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, MD, USA

[*]Corresponding author:
John Quackenbush, PhD
The Institute for Genomic Research
9712 Medical Center Drive
Rockville, MD 20850
Phone: (301)
Email: johnq@tigr.org
Fax: (301) 838-0208

ABSTRACT

Microarrays have been widely used in the analysis of gene expression, but the issue of reproducibility across platforms is an issue that has yet to be fully resolved. Although all microarray platforms measure RNA abundance and should therefore provide comparable data, direct comparisons between platforms often yield disparate results. This apparent non-reproducibility has caused some researchers to doubt the reliability of the microarray results despite evidence from RT-PCR that suggest they perform well. To address this apparent problem, we compared gene expression between two microarray platforms: the short oligonucleotide Affymetrix Mouse Genome 430 2.0 GeneChip™ and a spotted cDNA array. RNA was extracted from hearts of mice treated with Angiotensin II (Ang II) for 1 day or 14 days to induce hypertension; matched control samples were taken from mice treated with saline. RNA samples were amplified to provide a sufficient RNA for hybridization on both Affymetrix and TIGR cDNA platforms and for subsequent qRT-PCR validation of the expression of specific genes. Links between platforms were established using the TIGR Mouse Gene Index Tentative Consensus (TC) sequences as represented in RESOURCERER, which identified 11,710 genes shared on both arrays. Data were expressed as the $\log_2$ of the experimental (Ang II) expression levels relative to the mean of the appropriate saline control group. We assessed the relative impacts of experimental treatment and platform on gene expression using several analytical techniques, including two-factor ANOVA and Principal Components Analysis. All analyses agreed, showing that the effect of biological treatment had a greater impact on gene expression values than did platform in more than 90% of the genes surveyed, a result validated by quantitative RT-PCR. In the small number of cases where platforms were discrepant, qRT-PCR generally failed to confirm either result, suggesting that sequence-specific effects may make expression estimates difficult for any technique.

INTRODUCTION

DNA microarrays have afforded biological research scientists the opportunity to assay patterns gene expression on a global scale. Although there have been many successful applications of this technology, often with high rates of validation using an alternate technology such as Northern analysis or quantitative RT-PCR, a number of published studies have called into question the validity of the various technologies used for microarray assays, in part because of observed disparities between results obtained by different groups analyzing similar samples [1-8]. As confident practitioners of spotted cDNA microarray technology, we have often been puzzled by the apparent dichotomy of these competing views. In many instances, it seems that the failure to find concordance between various microarray platforms designed to assay biologically relevant patterns of expression is a failure not of the platform or the biological system, but rather a reflection of metrics used to evaluate concordance. Other meta-analyses focus on overlapping lists of significant genes, neglecting the fact that in many instances these represent not only different platforms but also vastly different approaches to data analysis [9-11] – something that we have seen even in looking at a single dataset generated on a single platform.

Based on our experience with hybridization-based approaches to assaying patterns of gene expression, we decided to test platform-dependence in assessing a simple biological system, asking whether platform or treatment were the major factors influencing the patterns of gene expression that are observed. We chose to use a model system we have previously studied using cDNA microarrays, the effects of short and long-term Angiotensin II (Ang II) exposure on cardiac gene expression in a mouse model of hypertension [12] (although this analysis used an independent collection biological replicate RNA samples). We chose to compare treated animals to matched controls using cDNA microarrays and Affymetrix GeneChips™; the former because it is a two-color platform with which we have a great deal of experience and the latter as it is a widely used commercial oligonucleotide-based platform.

One challenging aspect in designing this experiment is the difference in the way that data is collected on these two platforms. While individual Affymetrix arrays are used to assess single biological samples, cDNA arrays typically use co-hybridization of two RNA samples: one representing a treated sample and the second representing a reference or control sample. Consequently, for the cDNA array assays, we chose to use a common reference design in which each experimental RNA sample is co-hybridized with a reference RNA pool as this design most closely mimics the Affymetrix approach.

At every step in the process, from the initial amplification of the RNA to the final steps in the analysis, care was taken to treat the biological samples and the resulting data in an identical fashion as to not introduce artifacts; only in the platform-specific stages involving RNA labeling, hybridization, data extraction, and normalization was there a distinction made between platforms.

The resulting data were transformed to produce a comparison between treated and matched control animals with the goal of addressing a simple biological question: What is the difference in response to elevated levels of Ang II as the length of the exposure increases? In the context of comparing platforms, the question then becomes: Given a biological question evaluated on two different microarray platforms, are there platform-specific differences that mask the underlying biological response? The answer to this second question is no, but an answer qualified by some minor effects we observed that point to challenges inherent in using any hybridization based assay.

EXPERIMENTAL PROCEDURES

Animal handling

Acute or chronic Ang II infusion treatments were conducted using four groups of male 10-week old C57BL/6J mice obtained from the Jackson Laboratories (Bar Harbor, Maine). Animals were housed in the animal quarters with a 12-hour light/dark cycle, in pathogen free, temperature

and humidity controlled room (22°C and 45-55% respectively) with food (Purina Rodent Chow 5002) and distilled water *ad libitum*.

Acute Ang II infusion: Two groups of mice were submitted to intravenous infusion of Ang II (N = 4) or saline (N = 2) for 24 hours while under constant direct (intra-arterial) blood pressure (BP) monitoring. Under anesthesia with intraperitoneal pentobarbital (50 mg/kg), the right iliac artery was catheterized using a modified polyethylene catheter PE-50 tubing flushed with approximately 50µl of 50 IU/ml heparin in 0.9 % saline. The right iliac vein was catheterized with silastic silicon tubing for drug administration. Both lines were exteriorized at the back of the neck and sealed with heat. After surgery, the mice were allowed a recovery period and were housed overnight in separate cages with food and water.

Following a 24 hour recovery period, the two lines were unsealed and attached to a swivel. The arterial line was connected to a BP transducer, and mean BP was recorded with a computerized data-acquisition system (Power Lab/400, AD Instrument Pty Ltd, Castle Hill, Australia). The venous line was connected to a Harvard infusion pump (Harvard Apparatus, Holliston, MA) for drug infusion. Baseline BP was recorded until it became stable. At this point, infusion of Ang II (30 ng/minute) or normal saline started and continued for a period of 24hours.

Chronic Ang II infusion: Two other groups of received Ang II (N = 4) or saline (N = 4) infusion, respectively, for 14 days via an osmotic minipump. The osmotic minipump (model 2002, Alzet Co., Colorado City, CO) was implanted subcutaneously, slightly posterior to the scapula under anesthesia with pentobarbital 50mg/kg ip. Ang II was dissolved in 0.5 mol/l NaCl and 1mmol/l acetic acid, at concentrations sufficient to allow an infusion rate of 40ng/min, known to produce hypertension. Control mice received saline solution via the osmotic minipump. Indirect systolic BP was monitored daily for 14 days by a noninvasive tail-cuff system (BP-2000, Visitech System, Apec, NC).

Tissue harvesting: At the end of treatment (2:00- 4:00 PM), the mice were euthanized with overdose pentobarbital and the hearts were quickly (<3 minutes) removed for mRNA preparation.

The dissected tissues (<0.5 cm in any length) were submerged in approximately five volumes of RNAlater (Ambion, Austin, TX). Samples were kept at 4°C overnight and stored at –20°C individually.

RNA Amplification

Total RNA was extracted using Trizol (Invitrogen, Carlsbad, CA), according to manufacturer specifications. To ensure sufficient RNA for hybridizations to both the TIGR cDNA arrays and the Affymetrix Mouse Genome 430 2.0 GeneChip™, each RNA sample was amplified (N = 14) following a modified Eberwine protocol [13, 14], resulting in antisense cRNA. Additionally, Stratagene Universal mouse reference RNA was also amplified using this protocol, to provide an amplified reference sample for all TIGR microarray hybridizations. Amplification of 2.0 μg of total RNA resulted in 47.0 ± 4.3 μg cRNA. The cRNA was then processed separately for hybridization on TIGR 25K cDNA arrays and Affymetrix Mouse Genome 430 2.0 GeneChip™ (see below). Details of the amplification protocol may be found at <http://pga.tigr.org/protocols.shtml>.

The amplification protocol uses novel priming strategies that allow production of amplified RNA in either the sense or antisense orientation [14], Although cDNA arrays use double-stranded probes allowing use of either sense or antisense, production of antisense cRNA was chosen as it was necessary for hybridization to the Affymetrix GeneChip™ arrays. First-strand cDNA synthesis from total RNA was primed with an Oligo(dT)24 primer (Invitrogen, Carlsbad CA) followed by heat denaturing and then cooling to facilitate primer annealing. Reverse transcription produced cDNA-mRNA hybrids that were subjected to alkali hydrolysis to remove template mRNA. First-strand cDNA (fs-cDNA) was separated from residual enzymes, nucleotides, and mRNA fragments using the MinElute reaction cleanup kit (Qiagen, Valencia CA). Second-strand cDNA synthesis was primed with random hexamers. The sample was heated to denature the fs-cDNA and to eliminate secondary structure. The temperature was then rapidly dropped to the upper limit of the annealing

range and then ramped more slowly to a final temperature of 4ºC. Second-strand cDNA (ss-cDNA) synthesis was carried out using E. coli DNA polymerase and ligase followed by blunt-ending with T4 DNA Polymerase. The double-stranded cDNA was purified using a MinElute reaction cleanup kit (Qiagen, Valencia, CA). In vitro transcription (IVT) from the T7 promoter sequences incorporated into the fs-cDNA produced amplified RNA in the antisense orientation that were subsequently cleaned up using MinElute columns. Antisense cRNA concentrations were measured using the Nanodrop ND-1000 UV-Vis Spectrophotometer (Nanodrop, Rockland DE).

TIGR cDNA Microarray fabrication, RNA labeling and hybridization

TIGR cDNA microarrays were constructed using the NIA 15k and BMAP mouse cDNA clone sets that together contain 27,010 clones representing approximately 22,000 unique transcripts. PCR amplicons were prepared for printing as described previously [15]. Following amplification and purification, amplicons were resuspended at 100-200 nM in 50% DMSO and printed onto aminosaline-coated Nexterion Slide A (Schott Nexterion, Duryea PA) using an Intelligent Automation System (IAS) arrayer (Cambridge, MA). After printing, DNA was cross-linked to the slides by UV irradiation with a Stratalinker UV Crosslinker (Stratagene, La Jolla, CA) and stored in a vacuum chamber until used.

Detailed cDNA target preparation and hybridization protocols are available at <http://pga.tigr.org/protocols.shtml>. Briefly, cDNA was synthesized by random-primed reverse transcription in the presence aminoallyl dUTP, using 2 μg of cRNA. Reaction products were purified and coupled to Cy3 or Cy5 NHS-ester (Amersham, Piscataway, NJ). The labeled cDNAs were purified, combined as appropriate for each hybridization, and lyophilized. All samples were hybridized in duplicate with dye-reversal replicates, against amplified cRNA prepared from the Stratagene Universal Mouse Reference RNA (Stratagene, La Jolla, ,CA).

Slides were prehybridized in 1% Bovine Serum albumin in 5× SSC, 0.1% SDS for 45 minutes at 42°C, after which the slides were washed and dried. Cy3 and Cy5 labeled cDNA was resuspended in 30ml of 50% formamide, 5×SSC, 0.1% SDS containing 0.5 μg mouse COT1-DNA, 1 μg poly-dA and hybridized to the microarray at 42°C for 16 hours under glass coverslips. Following hybridization, slides were washed for 4 minutes at 42°C in solution containing 1×SSC and 0.2% SDS, followed by a 4 minute wash of 0.1×SSC, 0.1% SDS at ambient temperature; then by two 2.5 minute washes of 0.1×SSC; at the ambient temperature. Slides were dried by centrifugation and scanned without delay at 10 μm resolution using an Axon 4000B scanner. Data were saved as 16-bit TIFF files and expression levels were extracted using TIGR Spotfinder [16]. Data were consistent across biological and technical replicates, with 87.7 ± 0.7% "good" spots identified by SpotFinder QC parameters.

TIGR Microarray experimental design and data analysis

To eliminate any possible bias in labeling or detection, paired dye-reversal hybridizations were performed for each comparison made. Prior to data analyses, signals were normalized in the MIDAS software package (http://www.tigr.org/software/tm4; [16]) using a locally weighted scatterplot smoothing regression (LOWESS) algorithm [17] and standard deviation regularization between array subgrids, followed by dye-flip consistency checking set to keep data within the range of ± 2.0SD from the mean.

As all samples were hybridized against the Stratagene control, the resulting $\log_2$ expression values did not directly indicate change of expression levels relative to the control condition. In order to calculate $\log_2$ (AngII-treated Saline control), mean $\log_2$ values for each array element were determined for both the acute saline (N = 2) and chronic saline (N = 4) control treatments. The appropriate mean saline control $\log_2$ value was then subtracted from the associated $\log_2$-transformed acute Ang II treated samples (N = 4) or chronic AngII samples (N = 4), as $\log_2$ (AngII-treated/mean

Saline control) = log2 (AngII-treated) - log2 (mean Saline control). This approach of calculating the mean saline control values for each element had the added benefit of providing robust values for each array element, incorporating both biological and technical replicates to more accurately estimate the value. Consistency of expression across biological replicates within each of the two saline control groups was assessed by comparing the each individual measurement to the mean of the appropriate group using the same procedure (data not shown).

All subsequent analyses were performed on these normalized dataset of biologically relevant measures, using only those 24,759 array elements for which detectable hybridization signals were available for more than 50% of the hybridization assays.

Affymetrix Mouse Genome 430 2.0 GeneChip™ RNA labeling and hybridization

For each sample, 600 ng of amplified, antisense cRNA was used as starting material for the second cycle of the Affymetrix small sample protocol, producing biotinylated antisense cRNA. These samples were then fragmented and labeled, then hybridized on Affymetrix Mouse Genome 430 2.0 GeneChips™, washed and stained according to manufacturer specifications, following the 430 V5 fluidics protocol. The two acute saline biological replicate samples were each subjected to three independent amplification and hybridization cycles; as these technical replicates showed very high reproducibility, all other 12 biological samples (4 chronic saline, 4 acute Ang II, 4 chronic Ang II) had a single technical replicate each. Affymetrix GCOS QC parameters indicated high quality, with consistent hybridizations for all samples; background measurements were nominal (28.1 ± 0.6) and noise was low (1.02 ± 0.06) across all eighteen chips.

CEL files were exported from Affymetrix GCOS software and normalized in dChip [18] to the median intensity using two models, the PM-MM model and the PM-only model. Gene expression values were then $\log_2$ transformed. For comparison with the biological measures on the TIGR cDNA arrays, mean values for each probe set were calculated for both the acute saline (N = 6) and chronic

saline (N = 4) control groups. The normalized, log_2-transformed experimental values from the acute and chronic Ang II treatments were calculated to express the ratio of log2 (AngII-treated/mean Saline control) by subtracting log2 (AngII-treated) - log2 (Saline control). In the PM-MM model, only genes with present calls (51.3 ± 0.6%) were included in subsequent data analysis, resulting in 22,212 probe sets from the PM-MM model. The PM-only model provided data for all probe sets.

Data Availability

Expression data from all assays on all platforms was collected in accordance with the MIAME guidelines has been submitted to the ArrayExpress database and is available with accessions XXXX.

Real-Time Reverse Transcription PCR Analysis

Ten genes identified as consistent across platforms, as determined by non-significant *p*-values in the 2-factor ANOVA for both platform and interaction terms. qRT-PCR was used to validate gene expression for the following genes: TC1210147 amylase 1 salivary, TC1110939 a disintegrin and metalloproteinase with thrombospondin motifs 1(ADAMTS-1), TC1152487 matrix metalloproteinase-2 (MMP), TC1212532 uridine monophosphokinase 2 (UCK2), TC1162269 uridine monophosphokinase 2 (UCK2), TC1091339 Dickkpof related protein 3, TC1227026 insulin-like growth factor binding protein 3 (IGFBP-3), TC1227209 Calcium/calmodulin-dependent protein kinase type I (CAM kinase 1), TC1091182 adipose differentiation related protein, TC1091293 corticosteroid 11-beta dehydrogenase isozyme 1

Eleven genes had significantly different gene expression profiles between the TIGR and Affymetrix microarray platforms, as determined by significant platform and/or interaction terms in the 2-factor ANOVA. qRT-PCR was used to distinguish which platform provided a more accurate measurement of gene expression. These eleven genes were: TC1153012 spastic paraplegia 4

homolog, TC1154682 rho GAP protein, TC1170966 vesicle trafficking protein sec22b, TC1221856 Heat-shock 20 kDa like-protein p20 (HSP 20kDa), TC1196867 unnamed protein product, TC1119502 Unknown (protein for IMAGE:3354845), TC1212535 Ras GTPase-activating protein 2 (Ras GAP2), TC1208706 ORF unknown protein (ID:sll0809), TC1138844 Ectonucleotide pyrophosphatase/phosphodiesterase 2 (E-NPP2), TC1171266 Plasminogen activator inhibitor-1 (PAI-1), TC1207615 unknown.

For each gene, qRT-PCR was run in duplicate assays. Single-stranded cDNA was generated by reverse transcription of 2μg of amplified antisense cRNA in a 40μl reaction volume using the iScript cDNA Synthesis Kit (Bio-Rad Laboratories, Hercules, CA). The cDNA was then diluted to 100μl and subjected to real time PCR using the iTaq SYBR Green Supermix with ROX (Bio-Rad Laboratories, Hercules, CA). Forward and reverse primers were added to the reaction at a final concentration of 200 nM and 1μl of the diluted cDNA was used. The incorporation of the SYBR Green dye into the PCR products was detected in real time with the ABI Prism 7900HT Sequence Detection System. The ROX passive reference dye was used to factor in well and pipetting variability. The incorporation of the SYBR Green resulted in the determination of the threshold cycle (Ct), which identifies the PCR cycle at which exponential growth of the PCR products begins. Standard curves were established for each amplified cRNA sample being analyzed using the 18S ribosomal RNA (Universal 18S rRNA kit; Ambion, Austin, TX). The standard curves were normalized to each other through the 18S rRNA amplification, and quantitation was determined.

Affymetrix and TIGR Data Analysis

Normalized $\log_2$-transformed data expressing the ratio of the experimental value relative to the appropriate mean saline control were imported into TIGR Microarray Experiment Viewer (MeV) analysis software for statistical and functional analysis package (http://www.tigr.org/software/tm4; [16]). Each platform contributed 8 values, four for acute Ang II and four for chronic Ang II treatments.

MeV incorporates the Expression Analysis Systematic Explorer (EASE) software [19], to identify significant Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Microarray Pathway Profiler (GenMAPP) pathways and functional classes of genes based on Gene Ontology (GO) classifications. All statistical tests were calculated in MeV using the F distribution with $\alpha = 0.01$, unless otherwise noted.

To ensure standard treatment of both datasets in analysis, all Affymetrix probe sets and TIGR cDNA clones were mapped to TIGR Mouse Gene Index Tentative Consensus sequences (TC) using RESOURCERER <http://www.tigr.org/tdb/tgi/>. In instances when a TC was represented by two or more probes on the array, which occurred on both TIGR and Affymetrix platforms, the mean of the $\log_2$ ratios of gene expression for that gene was calculated in each experiment. All functional analyses were based on TC assignnments; GO terms were mapped directly to TIGR TCs, whereas KEGG and GenMAPP pathways were mapped to TCs via LocusLink identifiers.

The Affymetrix Mouse Genome 430 2.0 GeneChip™ contains more than 39,000 probe sets, and TIGR platform contains greater than 27,000 array elements. These two platforms overlap by 11,710 TCs, but not all of these elements provided useful hybridization data; although the PM-only model provided data for all probe sets, the Affymetrix PM-MM model had a 49% absent call rate and the TIGR platform had a 12% absent call rate. Of the 11,710 TCs annotated on both platforms 10,177 were present in 50% (8/16) of the experiments, but these generally were missing data on one of the two platforms. However, 5,853 genes had data in 80% of the 16 combined TIGR and AffymetrixPM-MM experiments, and these were used for subsequent analysis.

The expression patterns for the 5,853 “good” genes were subjected to comparative analysis of the platforms, using both 2-factor ANOVA, Principle Components Analysis, and Hierarchical Clustering in MeV. Twenty-one genes were selected for qRT-PCR validation of gene expression, representing two groups of genes: those with consistent expression across platforms (N = 10) and those with divergent expression between platforms (N = 11).

Expression for each gene was determined as the ratio of the $\log_2$ transformed qRT-PCR measurement for each experimental (Ang II) treated sample relative to the mean value of the matched saline-treated controls, as was also determined for both microarray platforms. This equivalency allowed for direct comparisons of microarray and qRT-PCR values. Correlations between data derived from the two Affymetrix models (PM-only and PM-MM), the TIGR cDNA array, and the qRT-PCR measurements were determined in Excel. Separate correlations were run on 10 genes that were consistent between microarray platforms and on 11 genes that differed between the platforms. For disparate genes, correlations were run on each gene individually as well, to identify whether qRT-PCR values correlated more highly for Affymetrix or TIGR microarray values.

RESULTS

Two-factor ANOVA was used to quantify the impact of platform (Affymetrix mouse GeneChip™ 430 2.0 or TIGR mouse cDNA array) and experimental treatment (acute or chronic Ang II treatment) on gene expression values for the 5,853 genes that contained data in > 80% of experiments. For most of the genes shared between the two arrays, the gene expression data was remarkably consistent and was independent of platform (Figure 1), as biological treatment had a greater impact on gene expression values than did microarray platform. 88% of the genes had no significant effect of microarray platform on the expression values (N=5,144). Analysis of these indicated that in most instances, the pattern of expression across samples was similar, independent of platform, but that the relative amplitude of the change was greater on one platform than the other. The interaction term in the ANOVA model identifies genes with divergent genes expression responses between the two platforms (Figure 1). These terms were of particular interest as they defined a small subset of genes for which the two platforms gave strongly divergent measurements, both in amplitude and direction of gene expression. Only 9% of genes (N = 504) of the 5,853 genes

had significant interaction terms in the 2-factor ANOVA; the majority of these showed a strong transcriptional response on the Affymetrix GeneChip™ but not the TIGR cDNA array (Figure 1).

Principal Components Analysis (PCA) is used to reduce the dimensionality of multi-dimensional datasets. PCA was performed on the two Affymetrix data models (PM-MM and PM-only) and on the TIGR cDNA microarray data, to determine whether experiments clustered primarily by platform or by experimental treatment. The primary principle component accounted for 32% of the variation in the data and differentiated between acute and chronic AngII treatments (Figure 2). The second and third principle components accounted for 28% and 11% of the variation in the data and differentiated between biological replicates within each treatment and platform differences. Biological replicates were more tightly clustered in the acute AngII samples than in the chronic AngII samples.

We used qRT-PCR to validate gene expression for 10 genes that shared similar expression profiles across both platforms, representing the group of genes that had non-significant interaction terms in the 2-factor ANOVA. We also performed qRT-PCR on 11 genes for which there was a significant interaction term; these genes had disparate expression profiles across the two microarray platforms. Our goal was to use the qRT-PCR to identify which platform gave the more accurate results when the two platforms differed. As noted previously, all gene expression values, whether derived from microarray or from qRT-PCR, were represented as the $\log_2$-transformed ratios of the experimental (Ang II-treated) gene expression relative to the mean of the time-matched saline-control values. The expression vectors for each gene on each of the four platforms was recorded and pair-wise Pearson correlation coefficients were calculated to assess the degree of concordance between platforms.

For the ten genes that shared similar expression profiles across Affymetrix and TIGR microarrays, there was strong concordance between Affymetrix and TIGR values (0.81 for PM-MM and 0.85 for PM-only), as was expected. Correlations were very tight (0.98) between the two

Affymetrix data models, PM-only and PM-MM, for these ten genes. When the microarray platforms gave consistent results, qRT-PCR also shared a robust correlation between both platforms, with correlations between 0.61 – 0.67 (Figs 3 and 4).

For the eleven genes with disparate profiles between platforms only one gene, plasminogen activator inhibitor 1 (PAI-1), gave robust confirmation of one platform over the other; qRT-PCR values for PAI-1 mirrored Affymetrix PM-only and PM-MM values, with correlation coefficients exceeding 0.89. For the remaining ten genes with disparate gene expression profiles, qRT-PCR validated neither platform. This was not the result of poor quality qRT-PCR runs, as each PCR reaction was run in quadruplicate, with common and disparate genes assayed in the same run. Thus, for the majority of the genes whose profiles disagreed across microarray platforms, qRT-PCR validated neither platform but provided yet a third expression profile.

DISCUSSION

Despite the common perception that gene expression values are not reproducible across platforms [1-8], our analysis of cardiac gene expression showed consistency for greater than 90% of genes shared between the Affymetrix GeneChip™ and TIGR cDNA arrays. qRT-PCR analysis independently verified expression for the genes that had similar expression values in both platforms. There are a variety of factors that may contribute to the reproducibility of the results, and the independence of these results from platform.

The first, and most obvious, reason is that the science and expertise of usage of microarrays as a reliable research tool and repeatability within any one platform has progressed rapidly over the last five years. Whereas earlier microarray experiments sometimes had difficulty in reproducing results between laboratories using the same RNA and same microarray technology [8]. Each platform, both GeneChip™ and cDNA arrays have progressed substantially in recent years, in reliability and reproducibility. Cross-platform comparability can only be achieved once within-platform

consistency issues have been fully addressed; this is one of the signs of a maturing field. It is important that both platforms in the comparison give high quality consistent results. If only one of the two platforms being compared gives consistent, reliable data, then comparisons between the two platforms are meaningless as they cannot give consistent results.

As stated by Jarvinen *et al.* [9], it is essential to have a reliable, consisted method of identifying genes on both platforms. Only if genes are accurately identified on both can their gene expression values be compared effectively. This can be challenging, as oligonucleotide arrays may be generated from very different information than are EST-based cDNA arrays. In this study, we used TIGR Tentative Consensus (TC) sequences for both platforms. The cross-platform comparisons can only be as good as the gene identification method. The results of this study indicate that TCs reliably identify the vast majority of genes correctly, be they based on short oligonucleotide sequences or long cDNAs generated from ESTs.

The methods used for data handling may also influence the repeatability of gene expression values across platforms (REF – 2004 paper). In this study, measurements on both platforms were presented as the log ratio of gene expression in response to Ang II treatment relative to the mean value of the matching saline control. One reason expressing gene expression as a relative ratio may give more consistent results across platforms is that this represents a more biologically meaningful value than intensity measures. Hybridization-based assays rely on a wide range of parameters reflecting the properties of the probe and target molecules, such as length, nucleotide composition, melting temperature, and secondary structure, to produce a detectable signal; the variability in response to small changes in any of these is evident in the range of signals from a collection of Affymetrix probe pairs designed to detect a single transcript. While we hope that hybridization-based assays may give us absolute quatitation of transcript abundance, these factors make it difficult to compare between probe sets on a single platform, much less across platforms. However, one would expect that these factors would be less important when comparing a single RNA species

between different conditions within the same platform – the question which most often is the one we want to address in biological assays. While knowing the absolute level of a transcript is quite useful, knowing how a transcript level changes in response to a particular stimulus often provides greater insight and consequently is the more relevant parameter to use in comparing platforms.

Having optimally pure and consistent starting material may also improve the reliability between platforms. It is possible that in previous studies, some of the difference in results between Affymetrix GeneChip™ results and cDNA arrays may have been due to the discrepancy in RNA handling: a round of amplification is built into the Affymetrix procedure, but not the cDNA protocol. As initial RNA quantities were limited in this study, all RNA was subjected to a round of RNA amplification, producing antisense mRNA. Thus both Affymetrix and TIGR cDNA arrays used amplified RNA as a starting material, minimizing and difference that amplification versus no amplification may have caused, otherwise [20].

A fraction of the genes examined (8-10%) showed divergent results between the two platforms. The common perception upon this type of inconsistency between platforms is to identify one platform as providing superior and more consistent results than does the other platform. However, when gene expression was verified using qRT-PCR, results of only one of the eleven genes tested supported one microarray platform over the other. For the remaining ten genes, qRT-PCR validated neither platform; both microarray platforms gave values inconsistent with the qRT-PCR values as well as with each other. It is reassuring that these non-verifiable genes represented only a minority of genes, but it is necessary to try to identify some factors that may be contributing to these discrepancies. One possibility is that these non-verifiable genes may represent splice variants, with each platform measuring one of the multiple splice variants that may be expressed differently from each other. The qRT-PCR primer is based on the TIGR TC, which is derived from multiple sequence alignment of multiple ESTs. In these cases, the Affymetrix probe may target one gene, the TIGR EST probe may target another section of a related gene, and the TC may represent in

fact an entire gene family rather than any one gene, and therefore will interrogate neither the Affymetrix gene nor the TIGR gene, per se, and thus give results that differ from those two platforms.

This study demonstrates that microarray measurement of gene expression and RNA abundance can be a robust method, providing comparable results from different platforms and validates the findings of a recent, related report[21] that demonstrated consistency between laboratories and platforms provided a consistent analytic approach. However, this requires not only careful attention to the experimental details surrounding data collection and analysis, but consistent gene annotation and reliable means of assessing the quality of each experimental assay. If careful attention is paid to these elements, our data indicate that for the majority of genes, expression is independent of platform in the sense that biological effects are greater than platform effects.

In doing such analysis, researchers should carefully consider the methods used to compare microarray results as these can have a profound effect on the conclusions that are ultimately derived. In this study, we used biological end-points, in our case changes in expression levels in treated animals relative to saline controls, rather than arbitrary measurements that were based on technology. Multiple comparative techniques (two factor ANOVA, principle components analysis, hierarchical clustering) all gave similar results, affirming our conclusion that microarrays can produce reliable, consistent data that are largely independent of platform. As public databases of microarray experiments (GEO and ArrayExpress) continue to rapidly accumulate expression data, the results presented here should provide some level of confidence that high-quality microarray results can provide a valuable resource for meta-analysis directed at uncovering biological phenomena.

References

1. Kuo, W.P., Jenssen, T.K., Butte, A.J., Ohno-Machado, L. & Kohane, I.S. Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics* **18**, 405-412 (2002).
2. Shippy, R. et al. Performance evaluation of commercial short-oligonucleotide microarrays and the impact of noise in making cross-platform correlations. *BMC Genomics* **5**, 61 (2004).
3. Yauk, C.L., Berndt, M.L., Williams, A. & Douglas, G.R. Comprehensive comparison of six microarray technologies. *Nucleic Acids Res* **32**, e124 (2004).
4. Park, P.J. et al. Current issues for DNA microarrays: platform comparison, double linear amplification, and universal RNA reference. *J Biotechnol* **112**, 225-245 (2004).
5. Mah, N. et al. A comparison of oligonucleotide and cDNA-based microarray systems. *Physiol Genomics* **16**, 361-370 (2004).
6. Rogojina, A.T., Orr, W.E., Song, B.K. & Geisert, E.E., Jr. Comparing the use of Affymetrix to spotted oligonucleotide microarrays using two retinal pigment epithelium cell lines. *Mol Vis* **9**, 482-496 (2003).
7. Maitra, A. et al. Multicomponent analysis of the pancreatic adenocarcinoma progression model using a pancreatic intraepithelial neoplasia tissue microarray. *Mod Pathol* **16**, 902-912 (2003).
8. Ulrich, R.G., Rockett, J.C., Gibson, G.G. & Pettit, S.D. Overview of an interlaboratory collaboration on evaluating the effects of model hepatotoxicants on hepatic gene expression. *Environ Health Perspect* **112**, 423-427 (2004).
9. Jarvinen, A.-K. et al. Are data from different gene expression microarray platforms comparable? *Genomics* **83**, 1164-1168 (2004).
10. Mah, N. et al. A comparison of oligonucleotide and cDNA-based microarray systems. *Physiol. Genomics* **16**, 361-370 (2004).
11. Tan, P.K. et al. Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res* **31**, 5676-5684 (2003).
12. Larkin, J.E. et al. Cardiac transcriptional response to acute and chronic angiotensin II treatments. *Physiol Genomics* **18**, 152-166 (2004).
13. Phillips, J. & Eberwine, J.H. Antisense RNA Amplification: A Linear Amplification Method for Analyzing the mRNA Population from Single Living Cells. *Methods* **10**, 283-288 (1996).
14. Marko, N.F., Frank, B.C., Quackenbush, J. & Lee, N.H. A robust method for amplification of RNA in the Sense orientation. (*submitted*).
15. Hegde, P. et al. A concise guide to cDNA microarray analysis. *Biotechniques* **29**, 548-550, 552-544, 556 passim (2000).
16. Saeed, A.I. et al. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**, 374-378 (2003).
17. Cleveland, W.S. Robust locally weighted regression and smoothing scatterplots. *J. Amer. Stat. Assoc* **74**, 829-836 (1979).
18. Cheng, L. & Wong, W.H. in The analysis of gene expression data: methods and software. (eds. G. Parmigiani, E.S. Garrett, R. Irizarry & S.L. Zeger) (Springer, New York; 2003).
19. Hosack, D.A., Dennis, G., Jr., Sherman, B.T., Lane, H.C. & Lempicki, R.A. Identifying biological themes within lists of genes with EASE. *Genome Biol* **4**, R70 (2003).
20. Park, P.J. et al. Current issues for DNA microarrays: platform comparison, double linear amplification, and universal RNA reference. *Journal of Biotechnology* **112**, 225-245 (2004).
21. Irizarry, R.A. et al. Multiple Lab Comparison of Microarray Platforms. *Nat Methods* **submitted** (2004).

Table 1: Forward and reverse primers used for qRT-PCR. Primer design based on TIGR TC sequence.

Primer	TIGR TC ID	Direction	Gene Name
GCTCCTGGTCAACCACCTT	TC1171266_F	forward	PAI1
CACGGTACCTTCTTGTGCAGT	TC1171266_R	reverse	PAI1
ACAACAGCCAGCCAGAAAGT	TC1208706_F	forward	unknown
GACTGCCCAGCAAGTCAAA	TC1208706_R	reverse	unknown
CCAGGTAGATTGTAGGCCATT	TC1153012_F	forward	SPG4
GGAGATGCAGGCATGAAGA	TC1153012_R	reverse	SPG4
GCACTGCCATCCTGAATCTAC	TC1154682_F	forward	RHO GAP
GATCTGTCAGCATTTCCATCA	TC1154682_R	reverse	RHO GAP
GATGGAGCCATCAGTGAGATT	TC1170966_F	forward	Sec22b
CACACTGCTACCAACACAGGA	TC1170966_R	reverse	Sec22b
TTCCACCCCTAAGCCTACCT	TC1221856_F	forward	HSP 20
AAGTGGGGTCAATTTGATGG	TC1221856_R	reverse	HSP 20
TCAGCATCACTCCAATCCTG	TC1196867_F	forward	unnamed
CAGGGCTTGACCTCTCTGTC	TC1196867_R	reverse	unnamed
GTTCCTGTTTGGGCTGTGTT	TC1138844_F	forward	E-NPP2
GAAGGGCGACAAAGAGAATG	TC1138844_R	reverse	E-NPP2
TCCACCCACTATGGAGGAAA	TC1119502_F	forward	unknown
TGCAGTAGGTGAAGGCAAGA	TC1119502_R	reverse	unknown
TACAATCAAACGCCCAACAA	TC1091339_F	forward	Dkk3
TTTGCTTCCTGATCCTCCAC	TC1091339_R	reverse	Dkk3
TCACACGGGTGATGTCAAGT	TC1210147_F	forward	Amylase 1
CATTGCCACAAGTGCTGTCT	TC1210147_R	reverse	Amylase 1
CAATCACTGGAGGCAAGGAG	TC1110939_F	forward	ADAMTS-1
CGGAGATGAGCCTTTCTGTC	TC1110939_R	reverse	ADAMTS-1
AAACGTGCCATTCTTCTTGG	TC1207615_F	forward	unknown
CCATTTTCCTTTCTGCCTTTT	TC1207615_R	reverse	unknown
CCAGATACCTGCACCACCTT	TC1152487_F	forward	MMP2
ATGTCAGACAACCCGAGTCC	TC1152487_R	reverse	MMP2
AAGCTCAGGGTCAAGTTCCA	TC1212532_F	forward	UCK2
CCTGTACATACGCCTCCTCA	TC1212532_R	reverse	UCK2
CAACACAGAAGGGAGCGTTT	TC1162269_F	forward	UCK2
ACTTCAGTGCAGAGGCAGGT	TC1162269_R	reverse	UCK2
AAGGCCGAATTGCATTCTTT	TC1212535_F	forward	Ras GAP2
AGCACTGCGGAGTCCATC	TC1212535_R	reverse	Ras GAP2
TGTTTTCTGGTCCAGCCTCT	TC1227026_F	forward	IGFBP3
GGCTCTCAGACAAGCCACTC	TC1227026_R	reverse	IGFBP3
TCTAGGGCCATGGATTGAAC	TC1227209_F	forward	CAMK1
GAAGAGGTGTGGGGTCAGAA	TC1227209_R	reverse	CAMK1
GCTCAGTAAACCCCTCCTTG	TC1091182_F	forward	ADFP
CATAATGAGGCCCTTGGTTC	TC1091182_R	reverse	ADFP
GCGGACTGGACATGCTTATT	TC1091293_F	forward	HDS11b1
CATGACCACGTAGCTGAGGA	TC1091293_R	reverse	HDS11b1

Figure Captions

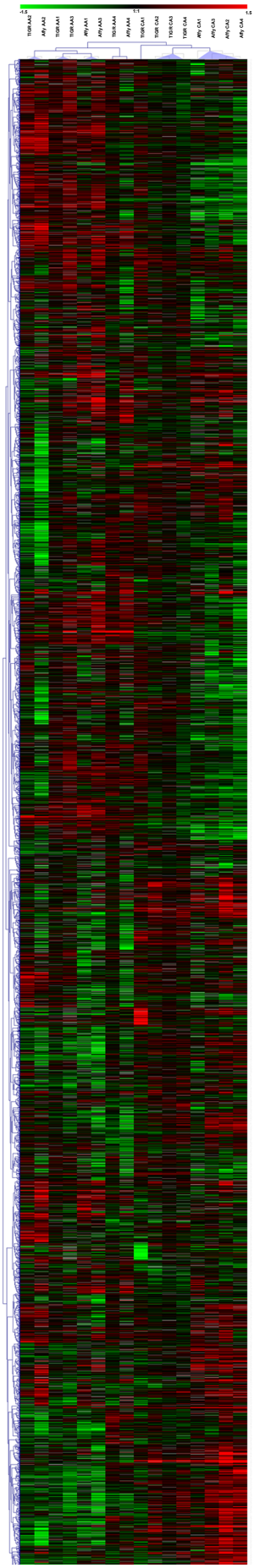
Figure 1: HCL and gene profiles of 2-factor ANOVA results comparing Affymetrix PM-MM to TIGR cDNA microarray data.

Figure 2: Principle Components Analysis of microarray data from TIGR cDNA array and Affymetrix mouse GeneChip™ using both PM-only and PM-MM models. Each experiment is labeled by Platform and experiment ID. Acute AngII experiments fall above the X-Y plane and are colored in shades of pink, whereas chronic AngII experiments fall below the X-Y plan and are colored blue or green.

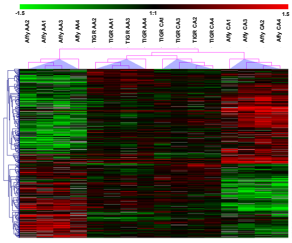
Figure 3: qRT-PCR validation of microarray results. Correlation table for the gene plasminogen activator inhibitor 1 (PAI1, top), which showed high correlations of Affymetrix gene expression measurements to qRT-PCR, but TIGR expression measurements were divergent. Expression of PAI-1, with each platform graphed independently. Correlation tables of microarray measurements to qRT-PCR values for 10 genes that showed no agreement in expression between Affymetrix and TIGR microarray platforms and for 10 genes that were consistent across microarray platforms.

Figure 4: Heirarchical clusters of gene expression as measured by microarrays (Affymetrix using PM-only and PM-MM models), cDNA array (TIGR mouse array), and qRT-PCR. Ten genes that showed no agreement in expression between Affymetrix and TIGR microarray platforms (top) and for 10 genes that were consistent across microarray platforms (bottom).

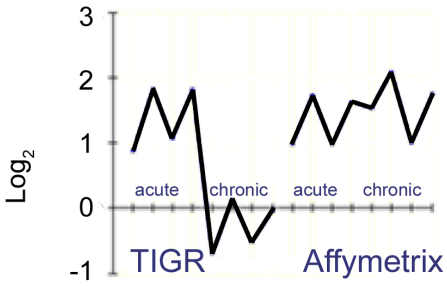
Interaction Non-Significant genes
N = 5439, p < 0.01



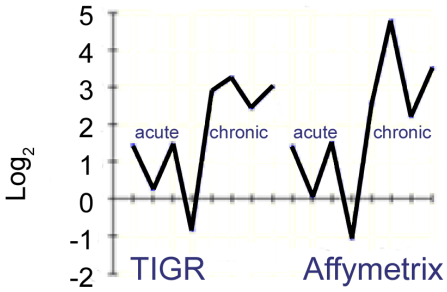
Interaction Significant genes
N = 504, p < 0.01



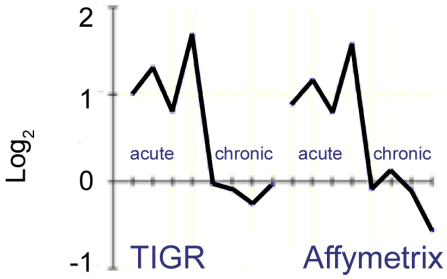
metallothionein-II



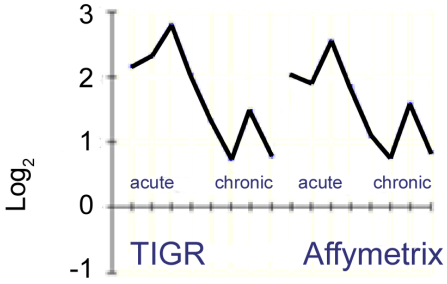
osteoblast specific factor 2

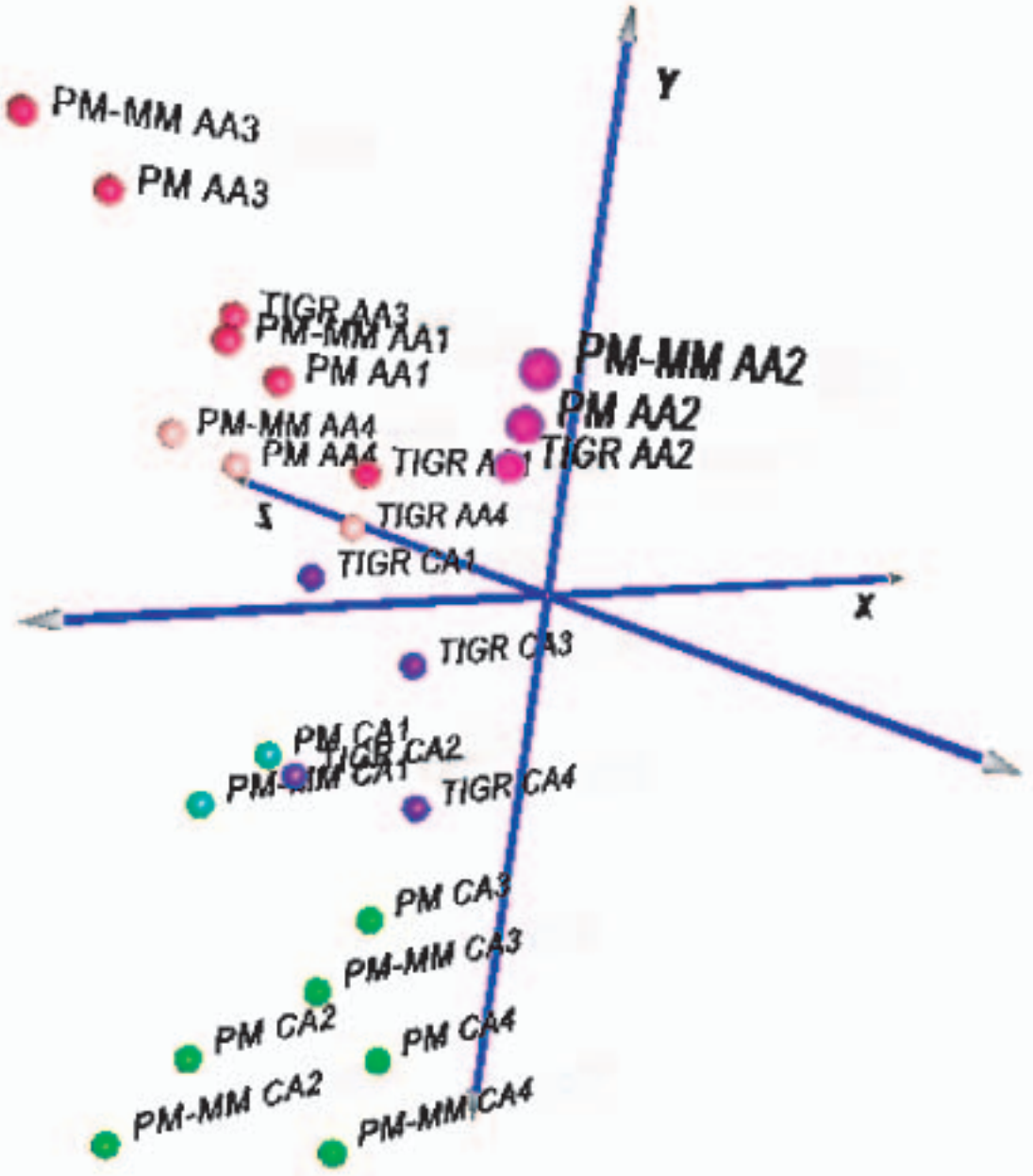


adipose differentiation related protein



ADAMTS-1



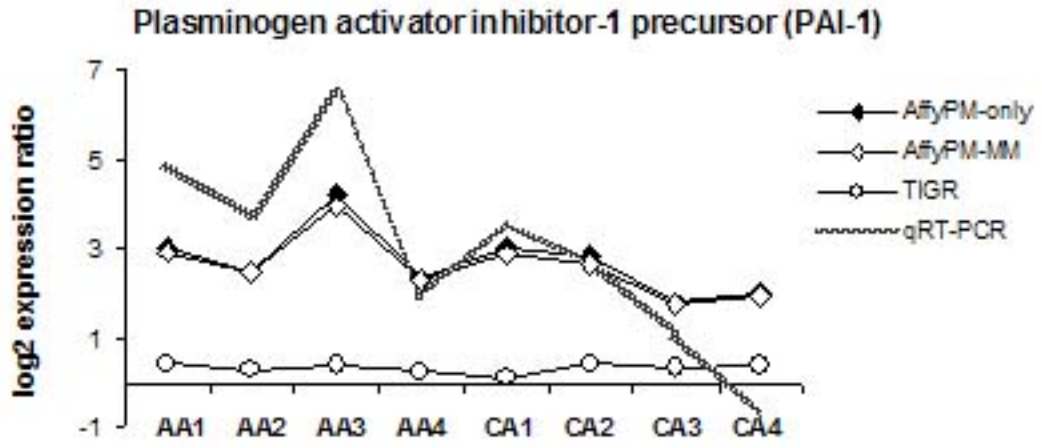


Eigenvalues

Principal Component 1 01.71 32.3 %
Principal Component 2 01.46 27.6 %
Principal Component 3 00.57 10.7 %
Principal Component 4 00.31 05.8 %
Principal Component 5 00.24 04.5 %
Principal Component 6 00.16 03.0 %
Principal Component 7 00.15 02.9 %
Principal Component 8 00.14 02.6 %
Principal Component 9 00.10 02.0 %
Principal Component 10 00.09 01.7 %

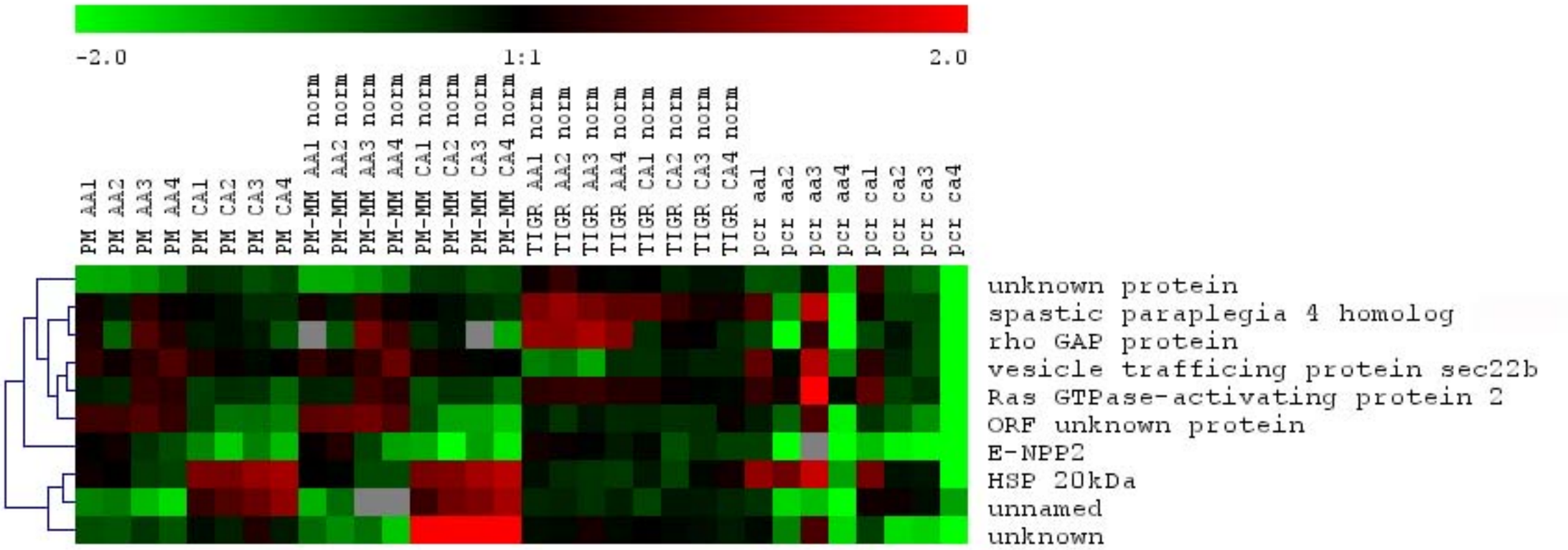
First 2 components: 59.9 %
First 3 components: 70.6 %

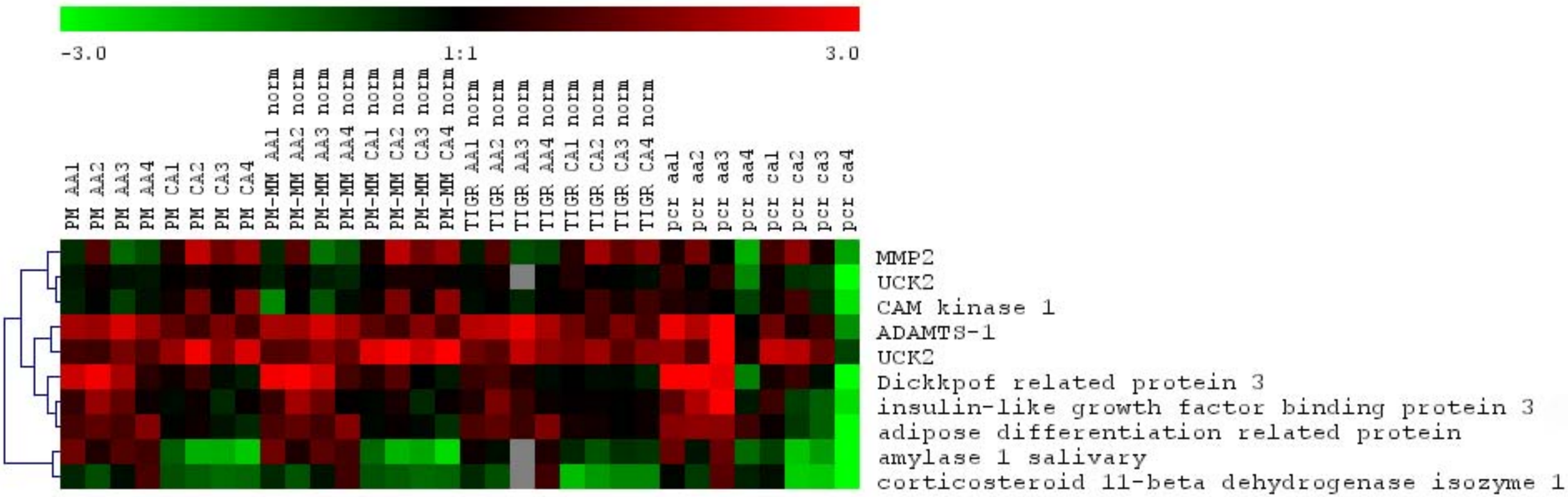
Plasminogen activator inhibitor-1				
	Affy PM-only	*Affy PM-MM*	*TIGR*	*qRT-PCR*
Affy PM-only	1			
Affy PM-MM	0.998	1		
TIGR	0.089	0.087	1	
qRT-PCR	0.893	0.914	0.082	1



Genes that disagree across platforms, without PAI1				
	Affy PM-only	*Affy PM-MM*	*TIGR*	*qRT-PCR*
Affy PM-only	1			
Affy PM-MM	0.756	1		
TIGR	0.081	0.044	1	
qRT-PCR	0.386	0.119	0.024	1

Genes that agree across platforms				
	Affy PM-only	*Affy PM-MM*	*TIGR*	*qRT-PCR*
Affy PM-only	1			
Affy PM-MM	0.984	1		
TIGR	0.847	0.809	1	
qRT-PCR	0.668	0.644	0.614	1


-2.0
1:1
2.0
PM AA1
PM AA2
PM AA3
PM AA4
PM CA1
PM CA2
PM CA3
PM CA4
PM-MM AA1 norm
PM-MM AA2 norm
PM-MM AA3 norm
PM-MM AA4 norm
PM-MM CA1 norm
PM-MM CA2 norm
PM-MM CA3 norm
PM-MM CA4 norm
TIGR AA1 norm
TIGR AA2 norm
TIGR AA3 norm
TIGR AA4 norm
TIGR CA1 norm
TIGR CA2 norm
TIGR CA3 norm
TIGR CA4 norm
pcr aa1
pcr aa2
pcr aa3
pcr aa4
pcr ca1
pcr ca2
pcr ca3
pcr ca4
unknown protein
spastic paraplegia 4 homolog
rho GAP protein
vesicle trafficing protein sec22b
Ras GTPase-activating protein 2
ORF unknown protein
E-NPP2
HSP 20kDa
unnamed
unknown


-3.0
1:1
3.0
PM AA1
PM AA2
PM AA3
PM AA4
PM CA1
PM CA2
PM CA3
PM CA4
PM-MM AA1 norm
PM-MM AA2 norm
PM-MM AA3 norm
PM-MM AA4 norm
PM-MM CA1 norm
PM-MM CA2 norm
PM-MM CA3 norm
PM-MM CA4 norm
TIGR AA1 norm
TIGR AA2 norm
TIGR AA3 norm
TIGR AA4 norm
TIGR CA1 norm
TIGR CA2 norm
TIGR CA3 norm
TIGR CA4 norm
pcr aa1
pcr aa2
pcr aa3
pcr aa4
pcr ca1
pcr ca2
pcr ca3
pcr ca4
MMP2
UCK2
CAM kinase 1
ADAMTS-1
UCK2
Dickkpof related protein 3
insulin-like growth factor binding protein 3
adipose differentiation related protein
amylase 1 salivary
corticosteroid 11-beta dehydrogenase isozyme 1