

Standardizing global gene expression analysis between laboratories and across platforms

Members of the Toxicogenomics Research Consortium¹

To facilitate collaborative research efforts between multi-investigator teams using DNA microarrays, we identified sources of error and data variability between laboratories and across microarray platforms, and methods to accommodate this variability. RNA expression data were generated in seven laboratories, which compared two standard RNA samples using 12 microarray platforms. At least two standard microarray types (one spotted, one commercial) were used by all laboratories. Reproducibility for most platforms within any laboratory was typically good, but reproducibility between platforms and across laboratories was generally poor. Reproducibility between laboratories increased markedly when standardized protocols were implemented for RNA labeling, hybridization, microarray processing, data acquisition and data normalization. Reproducibility was highest when analysis was based on biological themes defined by enriched Gene Ontology (GO) categories. These findings indicate that microarray results can be comparable across multiple laboratories, especially when a common platform and set of procedures are used.

Transcriptional profiling using DNA microarrays is one of many genomic tools that is now being used to characterize biological systems. Despite the increasing reliance on this technology by the scientific community, the reproducibility of microarray data between laboratories and across platforms has not been adequately addressed. Now there is a range of DNA microarray platforms including one- and two-channel formats, cDNA and oligonucleotide microarrays, in-house spotted microarrays, and commercially developed microarrays. There is also great diversity in the protocols used by different laboratories for RNA preparation and labeling, as well as in the instrumentation and software used for these procedures. Moreover, there are many computational and statistical tools for analyzing microarray images, quantitating spot intensities, normalizing and background-correcting these data, and for determining which transcripts are differentially expressed^{1–3}. The impact of these multifaceted approaches toward assessing global gene expression remains inadequately characterized^{4–7}. The issue of data reproducibility and reliability is crucial to the generation of, and ultimately to the utility of, large databases of microarray results^{8,9}. Although the Microarray Gene Expression Data (MGED) Society has coordinated an impressive effort to develop

guidelines for publishing microarray data through the minimal information about microarray experiments (MIAME) standards^{10,11}, these efforts have focused on documentation of experimental details and results, and therefore do not directly address issues of reproducibility between laboratories or across platforms. It is thus critical to determine the effect of methodological variables on the reproducibility, validity and generalizability of the results.

The Toxicogenomics Research Consortium was established in November 2001, with advancing the application of gene expression technologies in toxicology as one of its goals. The first Consortium study systemically assessed microarray data and reproducibility of the results within and between laboratories, as well as within and between microarray platforms. In doing so, potential sources of inter- and intralaboratory error and variability in the microarray experimental results were identified. Two standard microarrays were used by the Consortium laboratories: a spotted long oligonucleotide microarray, produced by one of the Consortium laboratories (designated the standard spotted array), and a commercially produced long oligonucleotide microarray (designated the standard commercial array). Consortium members also used a variety of other microarray platforms that were 'resident' at each laboratory (Fig. 1). The resident arrays included both commercial microarrays and in-house spotted microarrays, in long oligonucleotide, short oligonucleotide and cDNA formats.

Each laboratory was provided with aliquots of two different RNA samples that were prepared in one of the Consortium laboratories—a sample prepared from mouse livers (liver RNA; L), and a sample prepared from equal amounts of RNA isolated from five mouse tissues, liver, kidney, lung, brain and spleen (pooled RNA; P). The microarray hybridizations were designed to determine the reproducibility of gene expression measurements, the reproducibility of measuring differential transcript representation when comparing liver RNA to pooled RNA, and the feasibility of deriving comparable results across disparate laboratories and platforms¹². These common RNA samples allowed us to focus on variation in the technical and analytical approaches to microarray experimentation without biological variation^{1,13,14}. The results demonstrate that the highest level of reproducibility between laboratories was observed when a commercial microarray was used together with standardized protocols for RNA labeling, hybridization, microarray processing, data acquisition and data normalization. Whereas this

¹A list of authors and their affiliations appears in the **Supplementary Note** online. Correspondence should be addressed to B.K. Weis (weis@niehs.nih.gov).

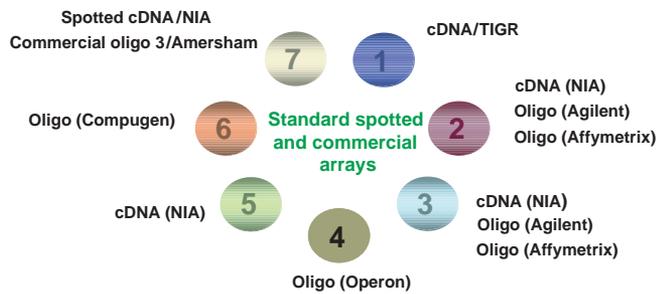


Figure 1 | DNA microarray platforms used across laboratories. Seven laboratories used a total of 12 microarray platforms: seven spotted resident cDNA and oligonucleotide platforms, three commercial resident platforms and two standard microarray platforms. Each colored circle represents a laboratory. Resident platforms are represented around the outer periphery of the ring of circles; the standard array platforms are represented in the middle of ring of circles (see Methods for description of platforms). An eighth laboratory (not shown), the provider of the standard commercial array, contributed to dataset D.

may be expected, the extent of the improvement in reproducibility that was obtained by such standardization was surprising. Notably, even with low levels of data correlation, the biological themes that emerge from these results are remarkably consistent.

RESULTS

Reproducibility of expression intensity with standard arrays

The standard spotted arrays and common RNA samples were used to generate dataset A (Fig. 1 and Supplementary Table 1 online). Researchers in laboratories 1–7 each carried out eight hybridizations: four that cohybridized liver RNA labeled with both Cy3 and Cy5 (LvsL), and four that cohybridized liver RNA and pooled RNA (LvsP). Each set comprised two dye-swapped samples^{15,16}. In each of these seven laboratories researchers used their own protocols for mRNA labeling, microarray hybridization, image acquisition and data analysis (Supplementary Methods online). These data were combined based on Unigene IDs. The reproducibility of raw intensity values was fairly high within each laboratory for LvsL, with median correlation coefficients ranging from 0.73 to 0.90 (Fig. 2a). When the data from each laboratory were compared to the collective data from the other laboratories, however, the correlations were significantly lower, between 0.21 and 0.41 across laboratories (Fig. 2b). The intensity values for the pooled RNA

samples were extracted from LvsP data and used to make PvsP comparisons *in silico*. The same trends were revealed (Figs. 2c,d, green symbols): correlation coefficients for PvsP ranged from 0.68 to 0.91 within laboratories and from 0.23 to 0.44 across laboratories. Thus, inter-laboratory differences negatively impact data reproducibility as measured by raw intensity values.

The first step toward evaluating reproducibility across laboratories was to standardize methods for the entry, storage and retrieval of the microarray data generated from different platforms and analysis software packages. Although this step did not significantly affect data correlation within each laboratory, the correlations across laboratories were improved dramatically (Dataset B; Fig. 2 and Supplementary Table 1 online). Specifically, the median LvsL correlation across laboratories improved from 0.33 to 0.56, and the PvsP correlation improved from 0.32 to 0.59. This indicates that an important source of variability for Dataset A was in the data handling methods; using standardized file formats and gene nomenclature improved the ability to detect correlations that are inherent in the data.

To evaluate the impact of image analysis methods on data reproducibility, we reanalyzed each microarray image from all laboratories (for Dataset B) using the same software package (GenePix Pro v4.1.1.28, Axon) and a common set of feature extraction parameters (dataset C; Supplementary Table 1). Standardizing the method for image analysis did not significantly affect the within-laboratory correlations, but did result in a modest increase in the correlation of intensity data across the seven laboratories (Fig. 2). The median correlation coefficient across laboratories for LvsL increased from 0.56 to 0.59, and for LvsP from 0.59 to 0.64.

Notwithstanding these standardization efforts, the best intensity correlation coefficient values across laboratories (0.59–0.64) were relatively poor (Supplementary Table 1). This was likely due to the diverse RNA labeling and hybridization methods used by the laboratories. To address this issue, the LvsL and LvsP experiments described above were repeated in each of eight laboratories using common protocols for RNA labeling and hybridization, and the standard commercial array (Dataset D). We applied a standard file format, nomenclature and image analysis protocol to these data. Intensity correlation coefficients were markedly improved for the within-laboratory comparisons and marked improvements in across-laboratory correlations were realized (Fig. 2). This was observed for both LvsL (Figs. 2a,b) and PvsP values (Figs. 2c,d), with median correlation coefficients improving to 0.87–0.92

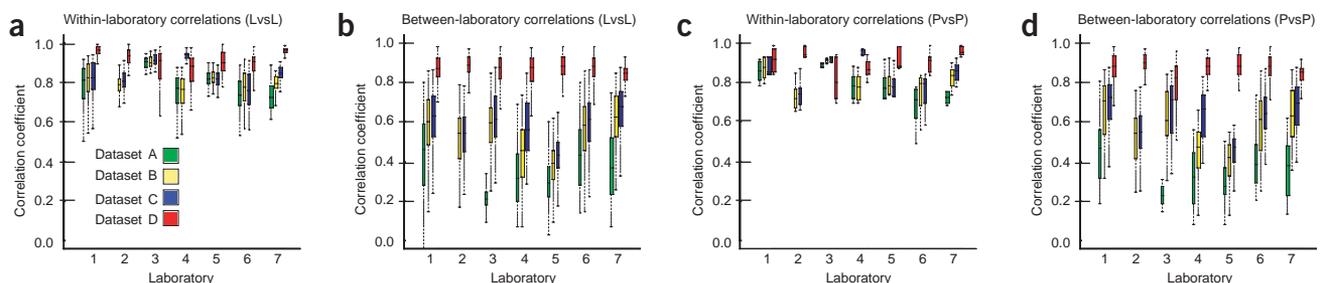
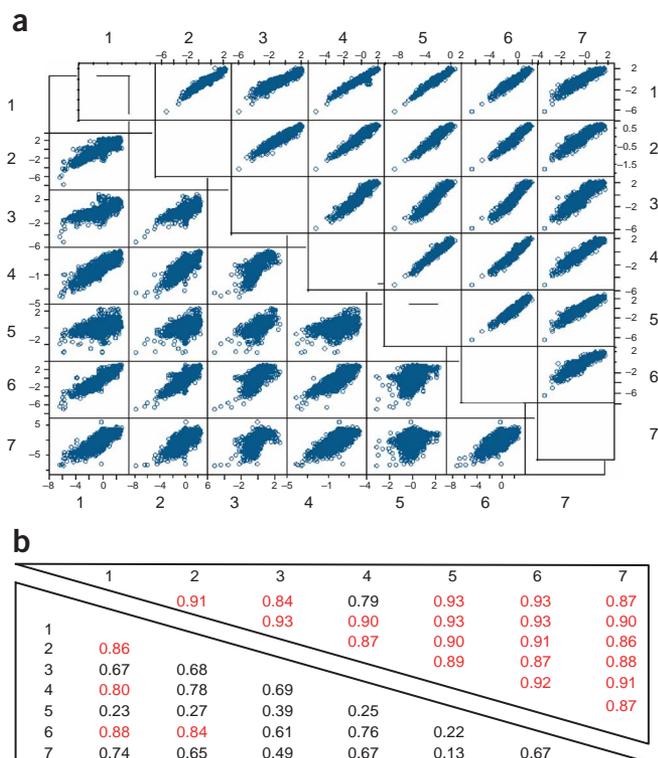


Figure 2 | Within- and between-center Pearson correlation coefficients for gene expression intensity using standard arrays. (a–d) Liver and pooled RNA samples were hybridized to two common platforms. Pearson correlations of raw intensity measurements were calculated for all possible pairwise combinations either within a laboratory (a,c) or between laboratories (b,d) on either the standard spotted arrays (datasets A–C) or the standard commercial array (dataset D). The box plots represent median values with upper and lower quantiles; the dotted lines represent maximum and minimum values.



(Supplementary Table 1). Thus, standardization of RNA labeling and hybridization protocols is an important contributor to signal intensity correlations across laboratories.

Reproducibility of expression ratios with standard arrays

It is arguably important to evaluate reproducibility in gene expression ratio measurements between laboratories and across platforms. Indeed, in most transcriptional profiling studies, it is ultimately the relative changes in gene expression ratios that are used to infer biological mechanisms and state changes. The first task was to establish how the transcript level ratios between liver and pooled RNA (LvsP) varied depending on the method used for data

Figure 4 | Resident array Pearson correlation.

Liver and pooled RNA samples were hybridized to seven different resident microarray platforms at eight laboratories (1–8). The average gene expression ratios were calculated across replicate microarrays within each laboratory and Pearson correlation coefficients were calculated for the set of 502 common genes (white boxes). In addition, average pair-wise correlations between different array replicates for each laboratory/platform combination were calculated (grey boxes). Labels comprise the laboratory number, type of array used and the source of probes for the array.

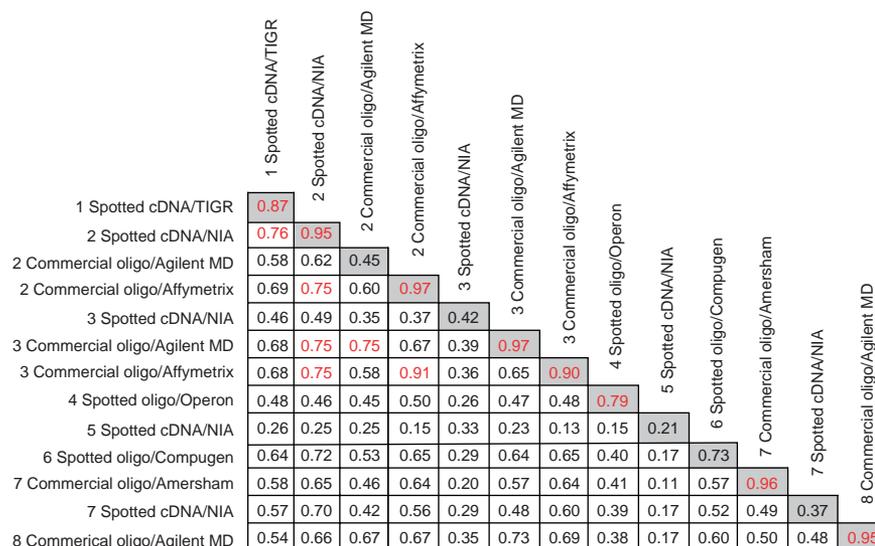


Figure 3 | Within and between laboratory Pearson correlation coefficients for \log_2 gene expression ratios using standard arrays. (a,b) Liver and pooled RNA samples were hybridized to two common array platforms in seven laboratories (1–7). Average \log_2 gene expression ratios were calculated across laboratories and were used to calculate correlation coefficients. Graphic display of pair-wise comparisons of average \log_2 gene expression ratios for liver versus pooled RNA samples plotted for all genes on the standard spotted array (a, lower panel) and the standard commercial array (a, upper panel). Pearson correlation coefficients for pair-wise comparisons across laboratories for the standard spotted array (b, lower panel) and the standard commercial array (b, upper panel). Correlation coefficients greater than 0.80 are highlighted (red).

normalization and background subtraction. We applied four different approaches to Dataset C, wherein file format, nomenclature and image analysis were standardized, but different labeling and hybridization protocols were used in each laboratory. We found the highest median correlation (0.69) of the LvsP \log_2 ratios between laboratories using Lowess normalization without background subtraction (Supplementary Table 2 online). Thus, we applied Lowess normalization without background adjustment to the gene expression measurements generated from all LvsP sample comparisons.

We calculated Pearson correlation coefficients comparing the average expression ratios across laboratories for each transcript feature on each platform, approximately 18,000 for datasets B and C and 20,000 for dataset D (Fig. 3). Similar to the raw intensity correlations (Fig. 2), the highest reproducibility was observed for dataset D, in which essentially all procedures were standardized. Once again, there was a marked increase in reproducibility in dataset D relative to dataset C.

Reproducibility of expression ratios with resident arrays

To compare expression ratios across noncommercial resident arrays (Figs. 1 and 4), we identified a set of common transcripts present on all 12 platforms (Supplementary Table 3 online and Supplementary Methods). Using stringent criteria, only 502 transcripts were matched across the 12 microarray platforms. We limited our analysis to these 502 genes to minimize the possibility that poor correlations between platforms could be due to gene misidentification. We applied Lowess normalization without

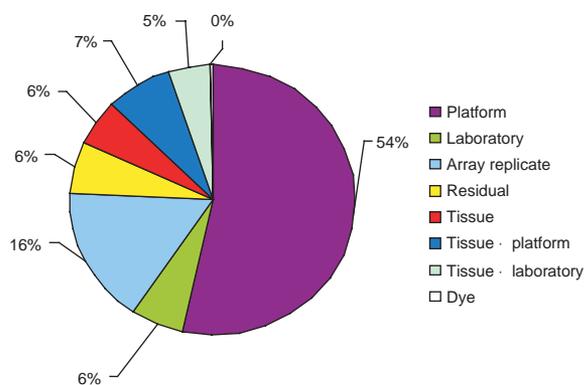


Figure 5 | Sources of variation in gene expression measurements across microarray platforms and laboratories for resident arrays. Contributions of different sources of variability were estimated with an ANOVA mixed model. Microarray platform was the largest source of variability, followed by laboratory and array-to-array replication (array replicate).

background subtraction as described above. For the single-color resident arrays (Affymetrix), we applied quantile normalization¹⁷.

We ran the two-color resident arrays in quadruplicate and the one-color arrays in duplicate. We calculated average expression ratios, as described above, across the replicate microarrays. We calculated median Pearson correlation coefficients for the set of 502 common transcripts (Fig. 4). As before, the reproducibility for each platform within its resident laboratory was generally very good, in particular for the commercial platforms. Overall, the cross-platform correlations were extremely poor both within and between laboratories, although we noted a few acceptable correlations (>0.75). We performed hierarchical clustering of the \log_2 -ratio values for the 502 common genes (Supplementary Fig. 1 online). Overall, we obtained similar ratios for a considerable percentage of the common genes for a majority of laboratory and platform combinations. For example, 69% of all laboratory and platform combinations had correlations greater than 0.70; the highest correlation was observed within dataset D (0.93). The remaining laboratory and platform combinations had lower overall \log_2 ratios and did not correlate as well.

Microarray platform contributes most to reproducibility

To assess the relative contribution of the different sources of technical variability in our gene expression measurements, we fitted an ANOVA random effects model to the LvsL and LvsP normalized data from the resident array platforms. For each of the 502 common transcripts, the model was used to partition the observed variability in the data into variability owing to platform, laboratory, microarray replicate, residual, tissue, tissue \times platform, tissue \times laboratory and dye¹⁸ (Fig. 5 and Supplementary Methods). These results indicate that more than half of the variability observed in these data is attributable to the microarray platform; differences between replicate microarrays and between different laboratories contributed substantially less.

Emergence of biological themes using Gene Ontology

We found considerable variability in gene expression using gene-by-gene comparisons. Subsequently, we determined whether consistent biological themes could nonetheless be identified among

different microarray platforms and laboratories. We identified the differentially expressed genes in the LvsP data for each laboratory and platform combination, and used the lists to identify enriched GO categories using EASE¹⁹ (Supplementary Methods). During the generation of the lists of differentially expressed genes, we noted a marked improvement in concordance (percent overlap in significantly induced or repressed genes based on pair-wise comparisons of gene lists across laboratories) of gene lists for datasets with increased standardization of technical methods (Supplementary Table 4 online). For example, we observed good overall concordance for dataset D, up to 80%. In addition, we found 277 transcripts that were significantly regulated, as defined by fold change plus an error term (described in Supplementary Methods) in all eight laboratories that contributed to Dataset D. For Dataset C, concordance was considerably lower, only as high as 52.4 percent, and only 13 genes were found to be significantly regulated in the six laboratories contributing to Dataset C.

We identified a list of 106 significant GO nodes that clustered into three main branches across 24 laboratory and platform combinations (Fig. 6 and Supplementary Table 5 online). Concordance was highest for branch 2, which primarily represented dataset D. More than 50% of the functionally enriched GO nodes had 70% concordance within branch 2, whereas less than 50%

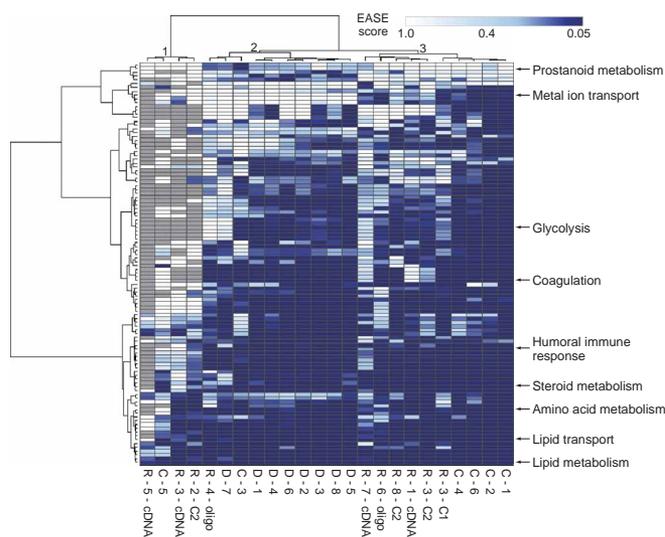


Figure 6 | Clustering of 24 laboratory and platform combinations based on common GO nodes. Common GO nodes were selected by two criteria: an EASE score was calculated for at least 20 of the 24 laboratory and platform combinations, and the EASE score was significant ($P < 0.05$) for at least one of the laboratory and platform combinations. This resulted in a list of 106 common GO nodes. Hierarchical clustering of both the laboratory and platforms, and the common GO nodes was performed using the calculated EASE scores. The relationship between the color intensity and EASE score is illustrated by the color bar. Gray indicates that an EASE score was not calculated for that GO node. The laboratory and platform is denoted by the letter and number combination at the bottom of every column. C = dataset C (standard spotted array with data extracted from a common image analysis software package), D = dataset D (standard commercial array), R = resident array. The number details which of the eight laboratories performed the hybridizations (see Figure 1). Resident arrays are also described by the type of array they are cDNA = spotted cDNA, oligo = spotted oligonucleotide, C#1 = commercial oligonucleotide arrays from Affymetrix and C#2 = commercial oligonucleotide arrays from Agilent. Numbers on the upper x-axis refer to branches of the dendrogram.

concordance was observed across all three branches. The decline in concordance across datasets is likely due to the impact of branch 1, which had relatively little GO node enrichment.

We found many similar biological themes across most laboratory and platform combinations (Fig. 6). For example, three GO nodes demonstrated enrichment across multiple laboratory and platform combinations: steroid metabolism, humoral immune response and coagulation. Enrichment of these nodes is readily explained by the samples used in this study. The liver is a principal site of steroid metabolism; it was thus expected to be an enriched node in liver RNA when compared to pooled RNA. Likewise, the spleen is an initiating organ in the humoral immune response; the presence of spleen RNA in the pooled sample resulted in an enrichment of this node relative to the liver sample. Finally, the liver has a role in coagulation through the synthesis of coagulation factors (for example, coagulation factor IX) and hepatocyte nuclear factors, thus explaining the enrichment of this GO node.

Notably, the EASE score for the coagulation nodes on three resident platforms (R7-cDNA, R1-cDNA and R3-C#2) was not as significant (EASE score > 0.05) as for other laboratory and platform combinations. This observation can be explained by the different transcript representation across the arrays. For both the standard spotted and standard commercial arrays, approximately 60 genes map to the coagulation GO node. In contrast, this node is represented by far fewer genes on the three resident arrays: 19 genes (R7-cDNA), 22 genes (R1-cDNA) and 25 genes (R3-C#2). The EASE score is a function of both the number of genes for a given GO node present on the array and the number of genes present in the list of differentially expressed genes for that node. Evaluating this further, several genes within the coagulation GO node (for example, fibrinogen, coagulation factor X and serine (or cysteine) proteinase inhibitor; *Serpind1*) were identified as differentially expressed in the majority of the laboratory and platform combinations; however, none of these genes were represented on the arrays that were used for R7-cDNA, R1-cDNA, and R3-C#2 (all of which represent distinct platforms). Therefore, the less significant EASE scores (> 0.05) for these resident arrays were likely due to a decreased representation of this GO node on these arrays.

DISCUSSION

Our results indicate that technical variables such as the microarray platform, the labeling and hybridization protocols, and the approaches to data analysis can profoundly affect the comparability of gene expression experiments between laboratories. Comparability is highest when these technical variables are standardized. We found that comparable biological themes emerge from data across disparate platforms and laboratories when GO nodes are used to analyze collections of genes representative of biological themes in lieu of direct gene-by-gene comparisons. This method of analysis may therefore prove useful for mitigating potentially confounding factors inherent in multisite and multiplatform data. The relationships between GO categories, however, take the form of directed acyclic graphs, meaning that 'child' categories can have multiple 'parents'. Thus, differences (and similarities) between datasets can be exaggerated because related nodes (for example, regulation of body fluids, hemostasis and coagulation) can contain some of the same genes. Therefore, the identity of nodes and their interrelatedness should be considered when attempting to assess reproducibility and concordance of disparate datasets.

Our findings have important lessons for the field of genomics. First, as one begins to use genomics to identify biological responses or states, one must carefully assess the platform and experimental (analytical) protocols used by the investigators. Our results demonstrate that the microarray platform can be a source of substantial gene expression variability and that commercial microarrays, for a variety of reasons (such as uniform labeling and hybridization techniques, consistent quality of the microarray itself), yield results that are more comparable between laboratories. Second, using genomics to identify environmental-response genes and biological pathways will require external validation, preferably through focused, independent hypothesis-testing experiments. Thus, gene expression results from microarray studies originating from one laboratory should be considered to be a foundation for developing testable hypotheses that can be addressed in subsequent experiments. Third, our findings demonstrate that the generalizability of gene expression studies can be limited between independent laboratories and across platforms. Although independent laboratories can clearly achieve similar results, this can be greatly facilitated by a substantial commitment to using harmonized experimental protocols, similar approaches to image and data analysis, and similar or identical microarray platforms. Fourth, similar biological themes can emerge from results obtained in the absence of harmonization, although the findings should be treated with caution. Although we found common GO categories across laboratories and platforms, there were also several distinct differences, indicating that it is easy to overinterpret microarray results. Finally, our findings indicate that creation of gene expression databases that incorporate results from multiple laboratories will be most useful if experimental standards are developed, and data filters are applied before consolidation of the individual experimental results. Without these steps, the results from comparisons across laboratories can be misleading and should be considered with appropriate caution.

METHODS

Microarrays. Twelve total microarray platforms, including seven spotted resident microarray platforms, three commercial resident platforms and two standard platforms were used (Figs. 1 and 4). Seven laboratories used the following four types of spotted resident platforms: (i) a spotted cDNA microarray with target cDNAs obtained from TIGR (laboratory 1); (ii) four different spotted cDNA microarrays with target cDNAs obtained from the National Institute of Aging (NIA) mouse clone sets (laboratories 2, 3, 5 and 7); (iii) a spotted oligonucleotide microarray with target oligonucleotides obtained from Operon (laboratory 4); and (iv) a spotted oligonucleotide microarray with target oligonucleotides purchased from Compugen (laboratory 6). Three laboratories used the following commercial resident platforms: (i) a long oligonucleotide Agilent Mouse Development (MD) microarray, (laboratories 2 and 3); (ii) a short oligonucleotide Affymetrix microarray (laboratories 2 and 3); and (iii) a long oligonucleotide Amersham microarray (laboratory 7). Two laboratories used the following standard platforms: (i) the standard spotted array, made in laboratory 1, by depositing 70-mer oligonucleotides (Operon) representing 18,000 unique mouse transcripts onto poly(L-lysine)-coated slides using a GeneMachines OmniGrid Arrayer. Control spots corresponding to the 10-gene *Arabidopsis thaliana* set (<http://pga.tigr.org>) were randomly spotted throughout the microarray; and (ii) the standard commercial array

that comprised *in situ* synthesized 60-mer oligonucleotides representing ~20,000 mouse transcripts, designed through collaboration between the Toxicogenomics Research Consortium and Agilent Technologies.

RNA labeling and hybridization. RNA labeling and hybridization procedures used with the standard spotted array and non-commercial resident arrays (datasets A, B and C) are described in **Supplementary Methods**. Standard protocols used with the standard commercial array are described in **Supplementary Methods**. For the commercial resident arrays, individual laboratories performed labeling and hybridization according to the manufacturer's recommendations.

Microarray scanning and image analysis. Scanning and image analysis methods used by the individual laboratories for the standard spotted array (datasets A and B) and resident arrays are described in **Supplementary Methods**. For dataset C, raw image files for the standard spotted array were reanalyzed using Axon GenePix Pro v4.1.1.28 using uniform image extraction parameters (**Supplementary Methods**). Standardized protocols for microarray scanning and image analysis used by all laboratories for the standard commercial arrays (dataset D) are described in **Supplementary Methods**.

Data preprocessing. Non-normalized intensity measurements obtained from the standardized image processing protocol were used to generate four normalized datasets by applying: (i) global intensity normalization, (ii) global intensity normalization with background adjustment, (iii) Lowess normalization with background adjustment applied to a \log_2 -ratio versus \log_2 -geometric-mean intensity (R-I or M-A plot), and (iv) Lowess normalization without background subtraction applied to an R-I plot^{20,21} (**Supplementary Methods**).

Statistical methods. To examine data reproducibility, Pearson correlation coefficients were calculated between background-corrected \log_2 intensity values for all nucleotide sequences represented on all microarrays. Transcripts represented across all microarray platforms were identified by mapping to the NIA mouse gene index (**Supplementary Methods**). When there was more than one correlation coefficient in a comparison, a median of the relevant correlations was presented. To assess the contributions of different potential sources of variability, an ANOVA mixed model¹⁶ was fitted for each of the 502 genes represented on all the platforms (**Supplementary Methods**). For each laboratory, a list of statistically significant up- or downregulated genes was generated based on a prespecified false discovery rate of 0.05. This rate was calculated in a step-up fashion for the mixed model²².

File format. For dataset A, microarray images were analyzed at individual laboratories using different (in-house) image analysis software packages (**Supplementary Methods**). Raw image extracted data files were stored in a shared database by extracting data from columns of interest from the image analysis output files. For datasets B, C and D, the output files from extracted images were directly stored on an ftp server in a common file location and flat file format without parsing. The datasets were combined based on Unigene IDs.

Scoring and evaluation of Gene Ontology categories. Methods for scoring and evaluating GO categories are described in **Supplementary Methods**.

Additional methods and information. Tissue extraction and RNA isolation procedures are described in **Supplementary Methods**. Additional material and primary (raw) data are available online (<http://dir.niehs.nih.gov/microarray/trc/>) and via GEO database (accession number GSE2458).

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank J. Quackenbush from The Institute for Genomic Research, L. Hartwell from Fred Hutchinson Cancer Research Center and R. Wolfinger from the SAS Institute for their scientific contributions. We thank K.J. Yost (Science Applications International) and P. Cozart (NIEHS ITSS) for their information technology support. Research support was provided by National Institutes of Environmental Health Sciences grants ES11375, ES11384, ES11387, ES11391 and ES11399, and Contract # N01-ES-25497.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 14 December 2004; accepted 25 March 2005

Published online at <http://www.nature.com/naturemethods/>

1. Quackenbush, J. Computational analysis of microarray data. *Nat. Rev. Genet.* **2**, 418–427 (2001).
2. Salter, A.H. & Nilsson, K.C. Informatics and multivariate analysis of toxicogenomics data. *Curr. Opin. Drug Discov. Devel.* **6**, 117–122 (2003).
3. Nadon, R. & Shoemaker, J. Statistical issues with microarrays: processing and analysis. *Trends Genet.* **18**, 265–271 (2002).
4. Spruill, S.E., Lu, J., Hardy, S. & Weir, B. Assessing sources of variability in microarray gene expression data. *Biotechniques* **33**, 916–923 (2002).
5. Tan, P.K. *et al.* Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res.* **31**, 5676–5684 (2003).
6. Yang, Y.H. & Speed, T. Design issues for cDNA microarray experiments. *Nat. Rev. Genet.* **3**, 579–588 (2002).
7. Marshall, E. Getting the noise out of gene arrays. *Science* **306**, 630–631 (2004).
8. Becker, K.G. The sharing of cDNA microarray data. *Nat. Rev. Neurosci.* **2**, 438–440 (2001).
9. Miles, M.F. Microarrays: lost in a storm of data. *Nat. Rev. Neurosci.* **2**, 440–443 (2001).
10. Ball, C.A. *et al.* Standards for microarray data. *Science* **298**, 539 (2002).
11. Campbell, P. Microarray standards at last. *Nature* **418**, 323 (2002).
12. Kim, H. *et al.* Use of RNA and genomic DNA references for inferred comparisons in DNA microarray analyses. *Biotechniques* **33**, 924–930 (2002).
13. Eisen, M.B. & Brown, P.O. DNA arrays for analysis of gene expression. *Methods Enzymol.* **303**, 179–205 (1999).
14. Cronin, M. *et al.* Universal RNA reference material for gene expression. *Clin. Chem.* **50**, 1464–1471 (2004).
15. Kerr, M.K. & Churchill, G.A. Experimental design for gene expression microarrays. *Biostatistics* **2**, 183–201 (2001).
16. Kerr, M.K. Experimental design to make the most of microarray results. *Methods Mol. Biol.* **224**, 137–147 (2003).
17. Irizarry, R.A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15 (2003).
18. Wolfinger, R. *et al.* Assessing gene significance from cDNA microarray expression data via mixed models. *J. Comput. Biol.* **8**, 625–637 (2001).
19. Hosack, D.A., Dennis, G., Sherman, B.T., Lane, H.C. & Lempicki, R.A. Identifying biological themes within lists of genes with EASE. *Genome Biol.* **4**, R60 (2003).
20. Hyduke, D.R., Rohlin, L., Kao, K.C. & Liao, J.C. A software package for cDNA microarray normalization and assessing confidence intervals. *OMICS* **7**, 227–234 (2003).
21. Tseng, G.C., Oh, M.K., Rohlin, L., Liao, J.C. & Wong, W.H. Issues in cDNA microarray Analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res.* **29**, 2549–2557 (2001).
22. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a powerful approach to multiple testing. *J. R. Stat. Soc. (Ser A)* **57**, 289 (1995).