The MIAME Checklist for Dataset for Resident Arrays

[*All Supplementary information, including materials, methods and primary data can be accessed at http://www.niehs.nih.gov/research/atniehs/core/microarrays/trc/.]

Experiment Design:

- **Type of experiment**: The experiment was designed to determine the reproducibility of gene transcript level measurements between multiple laboratories (n=7) and across multiple microarray platforms (n=10). Each laboratory was provided with large aliquots from two different mouse standard RNA samples: a liver RNA sample and a five-tissue pooled sample (liver, kidney, lung, brain, spleen in equal amounts) that were prepared in one of the consortium laboratories (see URL for Supplementary Methods, Standard Protocols, Tissue Extraction and Standard RNA Sample Preparation). Each laboratory utilized a variety of microarray platforms that were resident (in-house) to their laboratory (refer to Figure 1 of the paper). There were ten total platforms, seven were spotted arrays (manufactured in-house) and three were commercial arrays. The seven spotted resident platforms included: (i) a spotted cDNA array with target cDNAs obtained from The Institute for Genome Research, TIGR (spotted by Laboratory 1); (ii) four unique spotted cDNA arrays with target cDNAs obtained from the National Institute of Aging (NIA) mouse clone sets (spotted by Laboratories 2, 3, 5 and 7 using different releases of the NIA clone sets); (iii) a spotted oligo array with target oligos obtained from Operon (spotted by Laboratory 4); and (iv) a spotted oligo array with target oligos purchased from Compugen (Laboratory 6). The spotted cDNA arrays that were spotted using the NIA clone set and the TIGR cDNA clone set had similar array content; however, we considered them unique platforms because different releases of the clone sets were used and the arrays were spotted at separate laboratories. The three commercial resident platforms included: (i) a long commercial oligo array purchased from Agilent, the Mouse Development (MD) array (Laboratories 2 and 3); (ii) a short commercial oligo array purchased from Affymetrix (Laboratories 2 and 3), and (iii) a long commercial oligo array purchased from Amersham (Laboratory 7).
- **Experimental factors**: Using the standard RNA samples allowed us to focus on variation in the technical and analytical approaches to microarray experimentation, such labeling and hybridization protocols and image analysis, rather than the biological variation.
- **The number of hybridizations performed in the experiment.** Each of the 7 laboratories carried out 4 to 8 hybridizations for each resident microarray platform used by their laboratory. For each of the two-color Resident Arrays, the laboratories conducted 8 hybridizations, 4 that co-hybridized liver RNA versus liver RNA (LvsL) and four that co-hybridized liver RNA versus pooled RNA (LvsP); each set of four hybridizations had two dye-swapped samples (Cy3 versus Cy5). For the single channel arrays (Affymetrix), the laboratories ran duplicate arrays for liver RNA and pooled RNA samples.
- **The type of reference used for the hybridizations**: See answer to “Type of experiment” (above).
- **Hybridization design**: See answer to “Type of experiment” and “The number of hybridizations” (above).
- **Quality control steps taken**: See answer to “The number of hybridizations” (above).
- **URL of any supplemental websites or database accession numbers**: Supplemental information, including materials, methods and primary data can be accessed at http://www.niehs.nih.gov/research/atniehs/core/microarrays/trc/.

Samples used, extract preparation and labeling:

- **The origin of the biological sample:** For tissue extraction, twenty-five C57 black male mice (10-11 weeks) were purchased from Taconic Farms and housed five in a cage in pathogen-free rooms determined by sentinel animal survey. They acclimated for 7-10 days until they reached 12 weeks of age. They were housed in shoebox cages on sawdust and fed NTP 2000 chow; RNA isolation from tissue was performed by one laboratory (see Supplementary Methods, Standard Protocols, Tissue Extraction and Standard RNA Sample Preparation).
- **Manipulation of biological samples and protocols used**: See URL for Supplementary Methods, Standard Protocols, Tissue Extraction and Standard RNA Sample Preparation.
- **Protocol for preparing the hybridization extract**: RNA isolation from tissue was performed by one laboratory (see URL for Supplementary Methods, Standard Protocols, Tissue Extraction and Standard RNA Sample Preparation).
- **Labeling protocol(s)**. The individual laboratories were provided the standard RNA samples and used a variety of resident (in-house) protocols for RNA labeling (see URL for Supplementary Methods, Supplementary Table 4 and Resident Protocols for Laboratory 1-7).
- **External controls (spikes)**. cRNA control *in vitro* Arabidopsis transcripts (70mers) provided by The Institute for Genomics Research (http:pga.tigr.org) were added to each of the standard RNA samples as a means of assessing experimental variability for individual RNA analyses (labeling and hybridizations). A total of 10 Arabidopsis genes were chosen for this spike in design, with fixed concentrations of each transcript in two distinct sets, A3 and A5 (see URL for Supplementary Methods, Standard Protocols, Quality Control Genes).

Hybridization procedures and parameters:

- **The protocol and conditions used during hybridization, blocking and washing**. The individual laboratories were provided the standard RNA samples and used a variety of resident (in-house) protocols for hybridizations to the Resident Arrays (see URL for Supplementary Methods, Supplementary Table 4 and Resident Protocols for Laboratories 1-7).

Measurement data and specifications:

- **The quantitations based on the images**: The individual laboratories used a variety of resident (in-house) protocols for scanning and image analysis (see URL for Supplementary Methods, Supplementary Table 5 and Resident Protocols for Laboratories 1-7). Gene expression florescence intensities (of red and green channels) were measured for all spots represented on the Resident Arrays.
- **The set of quantitations from several arrays upon which the authors base their conclusions.**
  - **Type of scanning hardware and software used**: The individual laboratories used a variety of resident (in-house) equipment for scanning the Resident Arrays (see URL for Supplementary Methods, Supplementary Table 5 and Resident Protocols for Laboratories 1-7).
  - **Type of image analysis software used**: The laboratories used a variety of resident (in-house) image analysis software packages for analyzing the raw images for the Resident Arrays (see URL for Supplementary Methods, Supplementary Table 5 and Resident Protocols for Laboratory 1-7).

- **A description of the measurements produced by the image-analysis software and a description of which measurements were used in the analysis**. The laboratories used a variety of resident (in-house) protocols for scanning and image analysis of the Resident Arrays. The measurements used in the analysis is summarized below for the various image analysis software packages used:

Software	Foreground Intensity Used	Background Intensity Used
GenePix	F532 Mean, F635 Mean	B532 Median, B635Median
MolecularWare	Spot Mean Intensity w595, Spot Mean Intensity w685	Background Median Intensity w595, Background Median Intensity w685
ArraySuite	SR_Mean, SG_Mean	SR_bkMean, SG_bkMean
QuantArray	ch1 intensity, ch2 intensity	Ch1 background, ch2 background
SpotOn	Rmean, Gmean	bgRmed, bgGmed

- **The complete output of the image analysis *before* data selection and transformation**. See URL for Supplementary Information, Download the Primary Data, Raw Data for Resident Arrays.
- **Data selection and transformation procedures**. To compare gene expression measurements across the resident array platforms, a set of common genes present on all platforms (n=502) were identified using a combination of accession and clone based identifier matching, as well as sequence homology searches (see URL for Supplementary Methods, Standard Protocols, Identifying Common Genes Represented on Microarray Platforms). Lowess Normalization without background adjustment was applied to the gene expression measurements generated from all LvsP sample comparisons using the two-color Resident Arrays. For the single color Resident Arrays (Affymetrix), quantile normalization was applied to the gene expression measurements generated from all Liver and Pooled RNA samples. The average gene expression ratios were calculated across the replicate arrays and median Pearson correlation coefficients were calculated for the set of 502 common genes. Pearson correlation coefficients were calculated between log intensity values for all nucleotide sequences represented on the arrays using the ANOVA mixed model (see Methods).
- **Final gene expression data table(s) used by the authors to make their conclusions *after* data selection and transformation**. See URL for Supplementary Information, Download the Primary Data, Raw Data for Resident Arrays.

Array Design:

- **General array design.** Ten different Resident Array platforms were used by the laboratories. Seven of the platforms were resident spotted arrays (cDNA, oligonucleotide) and three were commercial arrays (Affymetrix, Amersham Codelink and Agilent Mouse Development Array). Array design was specific to the platform and can be provided if accepted.
- **For each feature (spot) on the array, its location on the array and the ID of its respective reporter**. Array design was specific to the platform and can be provided if accepted.
- **For each reporter, its type.** The array design including reporters was specific to the platform and can be provided if accepted. In addition, several of the laboratories using cDNA resident microarrays either printed 70-mer

olgonucleotide Arabidopsis probes (Invitrogen) onto their arrays or developed corresponding cDNA probes using a clone set provided by TIGR (http:pga.tigr.org).

- **For commercial arrays**. See vendor for details about commercial arrays.
- **For non-commercial arrays, the following details should be provided**:
- **The source of the reporter molecules.** The seven spotted resident platforms included: (i) a spotted cDNA array with target cDNAs obtained from The Institute for Genome Research, TIGR (spotted by Laboratory 1); (ii) four unique spotted cDNA arrays with target cDNAs obtained from the National Institute of Aging (NIA) mouse clone sets (spotted by Laboratories 2, 3, 5 and 7 using different releases of the NIA clone sets); (iii) a spotted oligo array with target oligos obtained from Operon (spotted by Laboratory 4).
- **The method of reporter preparation**. To be provided, if accepted.
- **The spotting protocols used**. To be provided if accepted.
- **Any additional treatment performed prior to hybridization**.