Short Technical Reports

Analysis of Result Variability from High-Density Oligonucleotide Arrays Comparing Same-Species and Cross-Species Hybridizations

BioTechniques 33:516-524 (September 2002)

ABSTRACT

There exists a significant limitation in the variety of organisms for which microarrays have been developed because of a lack of genomic sequence data. A near-term solution to this limitation is to use microarrays designed for one species to analyze RNA samples from closely related species. The assumption is that conservation of gene sequences between species will be sufficient to generate a reasonable amount of good-quality data. While there have been relatively few published reports describing the use of microarrays for cross-species hybridizations, this technique is potentially a powerful tool for understanding genomics in model organisms such as nonhuman primates. Here we describe the analysis and comparison of hybridization characteristics and data variability from a set of cross-species (rhesus macaque) and same-species (human) hybridization experiments using human high-density Affymetrix oligonucleotide arrays. The data reveal that a large fraction of probe sets are effective at transcript detection in the cross-species hybridization, validating the application of cross-species hybridizations for nonhuman primate genomics research.

INTRODUCTION

The use of microarray technology to measure changes in gene expression has gained widespread acceptance in the research community and is likely to remain a key research tool into the foreseeable future. One important development that has moved this field forward is the availability of standardized, com-

mercially available expression array technology, such as the GeneChip[®] system (Affymetrix, Santa Clara, CA, USA) (10). This system allows researchers worldwide to access the same standardized chips, sample processing methods, and image acquisition instrumentation to study gene expression in humans, mice, yeast, and several other model organisms. The study of gene expression in nonhuman primates represents a critical area of research directly related to the understanding of human health and physiology. Nonhuman primate studies play a crucial role in organ transplantation, vaccine development, viral pathogenesis, gene therapy, and a host of other human health-related technologies.

While the use of gene expression microarrays in nonhuman primate models will clearly advance our understanding of human health and disease, the availability of adequate sequence data keeps this technology out of reach in the near term. One potentially powerful short-term solution to this problem is the use of cross-species hybridizations: human sequence-based microarrays used to analyze nonhuman primate genomics. There are a few publications (1,2,5–7) describing the use of cross-species hybridizations to study nonhuman primates. However, none of these deals with issues of data variability and the integrity of gene expression measurements. Thus, there is a paucity of information on the distribution of genes in such nonhuman primate cross hybridizations and the comparison of hybridization characteristics and data variability obtained from same-species (human RNA hybridized to a human array) and cross-species (nonhuman primate RNA hybridized to a human array) hybridizations. Such comparison data are critical to advancing this strategy of using existing human chips for genomics research with nonhuman primates. Several recent publications (8,11–13) have begun to address some of the sources of non-biological variation found in same-species hybridizations, including both chip-to-chip variation and variation introduced by sample labeling and amplification methods. Here we describe the characterization of microarray variability in the context of same-species and cross-species microarray hybridizations.

MATERIALS AND METHODS

RNA Samples

Total RNA samples derived from human and nonhuman primate (rhesus macaque) brain tissue were used as starting material for RNA amplification and labeling reactions. All human and nonhuman primate studies and sample acquisitions were approved and monitored by the University of California, San Diego IRB, and the IACUC of The Scripps Research Institute. A complete description of the source and collection of brain-tissue samples will be described in detail in Roberts et al. (manuscript in preparation). Total RNA was purified from frontal lobe samples with TRIzol[®] Reagent (Invitrogen, Carlsbad, CA, USA) following the manufacturer's protocol, with an additional centrifugation step to remove unwanted cellular debris. RNA was then further purified using the RNeasy[®] mini kit (Qiagen, Valencia, CA, USA). The quantity of total RNA was assessed by 260-nm UV absorption and qualified by 260/280 ratios. One microgram from each sample was run on 1% agarose gel to assess RNA quality. One sample of the 21 samples processed was excluded from analysis because of RNA degradation. For protocol purposes, 5 µg total RNA were utilized per experiment.

Affymetrix GeneChips

Standard Affymetrix-recommended GeneChip protocols were used for all sample labeling, hybridization, staining, and scanning procedures. All GeneChip data were analyzed using Microarray Suite 5.0 software (MAS v5.0; Affymetrix). The GeneChip expression analysis system uses 11–20 probe pairs to interrogate each gene sequence represented on the chip (8). Each probe pair consists of one perfect-match and one mismatch oligonucleotide. The perfect-match oligonucleotide is a 25-bp sequence complementary to a given portion of the targeted transcript. The mismatch oligonucleotide is identical in sequence to the perfect match, except for a single mismatch at the 13th position. The difference in the hybridization signal between the per-

Table 1. Summary of Samples

Sample	Species	Control	Infected	% Present	
				Replicate 1	Replicate 2
R1	Rhesus		X	33.2	27.4
R2	Rhesus		X	32.0	26.2
R3	Rhesus		X	24.9	23.3
R4	Rhesus		X	30.5	33.7
R5	Rhesus		X	27.0	25.8
R6	Rhesus	X		28.1	26.4
R7	Rhesus		X	34.1	31.5
R8	Rhesus		X	31.1	31.3
R9	Rhesus		X	29.7	29.5
R10	Rhesus		X	30.5	34.5
R11	Rhesus		X	23.2	26.2
R12	Rhesus		X	34.4	30.7
H1	Human	X		49.3	40.7
H2	Human		X	50.2	45.7
H3	Human	X		45.6	49.6
H4	Human	X		46.8	51.1
H5	Human		X	42.5	44.6
H6	Human		X	49.8	48.4
H7	Human	X		44.1	48.1
H8	Human		X	41.4	44.1

fect-match and mismatch probes across all probe pairs used for a given gene generates the hybridization “signal”. Signal is a quantitative metric calculated for each probe set, representing the relative abundance of a transcript. In addition, a “detection” P-value is calculated based on an algorithm using the probe pair intensities compared to a user-definable threshold. A one-sided Wilcoxon’s signed rank test is then applied to this distribution of probe pair intensities. Thus, the user can define cut-offs to describe the confidence that a given transcript is detected (“present”), not detected (“absent”), or “marginal”. The reader is referred to the Affymetrix Web site (http://www.affymetrix.com/products/statistical_algorithms_reference_guide.html) for more explanation of the signal and detection algorithms. For these studies, all user-definable parameters were set at the Affymetrix default values.

We performed 20 pairs of replicate GeneChip hybridizations (8 same species, 12 cross-species). Biotin-labeled cRNA, prepared by in vitro tran-

scription, were obtained using standard Affymetrix protocols. A hybridization cocktail was prepared with 15 μ g cRNA (300 μ L total volume). A 200- μ L aliquot was placed on a GeneChip, and after hybridization this aliquot was returned to the original hybridization cocktail mixture and frozen (-20°C). For replicate chips, a 200- μ L aliquot of the frozen hybridization cocktail mixture was thawed and reapplied to a new GeneChip within 30 days of the initial hybridization (following standard Affymetrix protocols). GeneChips used in these experiments were derived from six different manufacturing lots. All but four of the 20 replicate hybridizations were performed using GeneChips from the same manufacturing lots. The four replicate hybridizations that utilized GeneChips from different lots were for samples R2, R6, R11, and R12. No increased data variability was noted for these samples. The distribution of human and rhesus macaque samples hybridized to replicate GeneChips and the percent present gene calls for each chip are given in Table 1.

Statistical Methodology

To determine if the flux genes were random or nonrandom, genes were binned as either “flux” or “not flux” on each of the 20 pairs of replicate chips, setting up a binomial (Bernoulli) distribution analysis. Each gene on the Hu95Av2 GeneChip (12 625 genes) was defined as a Bernoulli variable. Each of the 20 measurements for each Bernoulli variable was called a Bernoulli trial; successful Bernoulli trials represent flux-positive measurements, and these trials were tallied for each Bernoulli variable. Next, the variables were grouped by number of successful trials, and this relationship was plotted as a histogram of the actual distribution of flux genes. To determine the theoretical random distribution of flux genes, we took the probability, $P(x)$, that any gene is flux on any chip (the total number of flux-positive measurements, 20 285, divided by the total number of measurements, 252 500: 8.0%) and then calculated the probability of any gene being flux in 1–20 experiments based on a binomial distribution, using $P(x)$. This distribution was multiplied by the total number of Bernoulli variables to give a theoretical histogram of successful Bernoulli trials, based on 20 measurements per variable. Based on these calculations, the actual distribution was determined to be nonrandom ($P < 0.001$). To calculate the increased probability of a gene identified as flux in one trial falling into the flux category in a subsequent trial, the variables were binned into either one successful trial or greater than one successful trial out of 20. This probability was compared to the theoretical probability $(0.08)^2$ of a random event.

RESULTS AND DISCUSSION

Normalization and Transcript Detection

The U95Av2 GeneChip contains probe sets for 12 625 human gene sequences. Our data set consisted of RNA from human and nonhuman primate brain samples of the frontal lobes from normal and either human immunodeficiency virus or simian immunodeficiency

Short Technical Reports

ciency virus-infected (rhesus macaque) subjects. Our initial analysis focused on the percent of genes detected as “present” using the Affymetrix MASv5.0 algorithm with default settings. A target intensity value of 250 was used to standardize the GeneChip data from 16 same-species (human RNA on the human U95Av2 GeneChips) and 24 cross-species hybridizations (rhesus macaque RNA on human U95Av2 GeneChips).

In Figure 1, the differences in the number of genes called as present between same- and cross-species hybridizations could represent fundamental differences in monkey and human biology reflected by differences in gene expression patterns. However, it is important to consider the possibility that a subset of nonhuman primate sequences will not hybridize well to the probes designed for the human orthologs leading to apparent differences in gene expression patterns. Since the Affymetrix arrays use 25-bp oligonucleotides and there is approximately 5% sequence difference between the genomes of humans and rhesus monkeys, it is expected that many of the perfect-match oligonucleotides will differ from the rhesus sequence, lowering the number of genes called present by the Affymetrix algorithm. Indeed, this will also be the case for the mismatch oligonucleotide probes, further complicating the patterns. That the percent of genes called present in the nonhuman primate hybridizations ($29.4\% \pm 3.5$) is indeed significantly lower as compared to the percent in the human/human hybridizations ($46.4\% \pm 3.3$; $P < 0.001$) supports the hypothesis that sequence differences do impact gene detection calls. Thus, caution is necessary in interpreting data from genes called as negative or marginal, as important transcript abundance information may be missed if gene calls are arbitrarily excluded based on this criterion. A more detailed analysis of target and probe pair sequences, probe set hybridization data, or additional confirmation obtained by nonhuman primate sequence-specific approaches (i.e., RNA protection assays or PCR) may be helpful in this case.

Conversely, positive gene calls confirmed by replicate chips serve to establish the validity of specific GeneChip

probe sets for use in cross-species hybridizations. In our data set consisting of 12 replicate pairs, 4846 genes (38% of the probe sets on the U95Av2 chip) were called present on both replicates in at least one of the replicate pairs. Thus, as experience with cross-species hybridizations increases and these comparisons of specific gene probe sets on replicate chips in multiple laboratories are shared and refined, confidence in gene calls will be enhanced. A complete set of the genes that were consistently present in our replicate sets are posted at http://www.scripps.edu/services/dna_array. It is important to acknowledge that the threshold values used in the Affymetrix algorithm to define present/absent gene calls has not been rigorously validated to date. Moreover, it is possible that the use of alternative array technologies based on longer probe sequences ranging from 70-mer oligonucleotides to greater than 1000-bp cDNAs might increase the chances of successfully detecting present transcripts in a cross-species hybridization.

GeneChip normalization or “scaling” allows data from different arrays to be compared to each other and is accomplished by adjusting the signals on each chip such that the median signal is equal to a predetermined target intensity value set by the user. In these experiments, total RNA samples collected for analysis (8 human and 12 rhesus) were prepared using Affymetrix protocols (10). Each of these labeled hybridization cocktails were then hybridized to two identical GeneChip arrays (40 total hybridizations).

Figure 1 shows that the value of the scale factor for a given GeneChip is inversely related to the number of genes called as present. In fact, between each of the 20 replicate pairs analyzed, the GeneChip with the higher scale factor consistently had a lower percent present value

(not shown). This result is expected, as a highly sensitive hybridization provides brighter signals, detects more genes, and therefore requires a lower scale factor to bring the median GeneChip signal value to the preset target intensity. That this trend is also observed in the cross-species hybridizations is consistent with the hypothesis that the mean GeneChip hybridization intensities are directly correlated with specific hybridization events and not random experimental noise.

Variability and Flux

Across different replicate GeneChip hybridizations, we examined the number of genes that changed from present to absent or absent to present. In a perfect set of hybridizations, there should be no changes between replicate GeneChips. However, the reality is that there is a subset of genes that change from absent to present and present to absent. We have called this subset defined by replicate chips as flux genes. The results shown in Figure 2 reveal that the flux gene set comprises an average of close to 8% of the total 12 625 genes on the human U95Av2 GeneChip for both same- and cross-species hybridizations. Given the significantly higher percent of genes called present in same- as compared to cross-species hy-

Figure 1. Percent present as a function of scale factor. The percent present (the percent of gene sequences on a given GeneChip called as present by the Affymetrix Microarray Suite 5.0 software) is shown for both human (same-species) and rhesus (cross-species) RNA samples prepared and hybridized to Affymetrix human U95Av2 GeneChips. The overall percent present calls for cross-species hybridizations are lower ($\bar{x} \pm \text{SD}$, $29.4\% \pm 3.5$) than same-species hybridizations ($\bar{x} \pm \text{SD}$, $46.4\% \pm 3.3$). For both hybridization types, higher scale factors are correlated with lower percent present values.

Short Technical Reports

Figure 2. Replicate GeneChip variability. This figure shows the number of genes that change from present to absent, absent to present, and the sum of these two changes (total Flux) for same-species (H1–H8) and cross-species (R1–R12) replicate hybridizations. The total flux remains relatively constant ($\bar{x} \pm \text{SD}$, $7.8\% \pm 0.9$ for same species and $8.2\% \pm 1.4$ for cross-species). Note that the replicate chips were paired arbitrarily so that the first chip has the higher scale factor of the two. Because a higher scale factor gives a lower percent of genes called present (see Figure 1), this pairing scheme results in a greater number of genes changing from absent to present within each pair.

bridizations (46% vs. 29%; $P < 0.001$), it is clear that the size of the flux gene set as a percentage of the number of genes called present is higher for cross-species hybridizations (27% vs. 17%, respectively). It is important to note that the flux gene set cannot be defined unless replicate chips are done and compared. Thus, we are not using this analysis of flux genes to set a statistical confidence level in a positive gene call for a single GeneChip data set.

There are several possible explanations for flux genes. They may represent transcripts present at levels too low to be reliably detected using current Affymetrix gene call algorithms. In this context, the latest version of the Affymetrix data analysis software (version 5.0) used for all the analyses described here reduced the number of flux genes by approximately 35% as compared to the previous version (data not shown). Bioinformatics that refine the param-

Short Technical Reports

Figure 3. Percent of genes called present as a function of signal intensity. Gene expression measurements from the each data set were binned by signal intensity and the percent of genes called present for each bin is shown for same- and cross-species hybridizations.

Figure 4. Percent of genes showing 2-fold or greater change in signal between replicates. Flux genes and genes called present in both replicates (PP) were binned by signal intensity. The percent of genes showing greater than 2-fold change in signal is shown for same- and cross-species hybridizations. Values for flux human stop at the 300–399 bin because there were less than five genes per bin in the remaining bins.

ters for gene call thresholds could reduce the size of the flux gene subset. Theoretically, another source of flux genes would be from relatively ineffective probe set designs, resulting in poor hybridizations for particular probe pairs within a given probe set. Thus, with fewer working probe pairs, increased variability with respect to signals may occur within a given probe set. A logical corollary would be the prediction that species-specific sequence differences can reduce the efficiency of cross-species hybridizations to a subset of GeneChip probes. In fact, our data demonstrate a large number of flux genes for the cross-species hybridizations, a result not surprising given that the probe sets have been designed on the basis of human gene sequences.

If species-specific sequence differences in hybridization efficiencies contribute to specific genes falling into the flux gene set, then it follows that the flux genes should not be a random subset of the GeneChip probe sets. Therefore, we determined the distribution of flux genes across replicate GeneChips using Bernoulli distribution analysis. A gene found in the flux category on one pair of replicate GeneChips has a greater than 11.5-fold higher chance of falling into the flux category in a subsequent pair of replicate GeneChips compared to a randomly selected gene.

Therefore, the flux genes are not a random subset. If fewer probe pairs are working for a specific gene transcript because of marginal hybridization efficiencies, the remaining probe pairs that do work will require higher signal differences between perfect and mismatch probes (resulting in higher signals) to call the gene present. In fact, an analysis of average signal intensities shows that cross-species flux genes have signal intensities 2-fold higher than same-species Flux genes (228 ± 16 vs. 117 ± 16). These results further support the hypothesis that a portion of flux genes in cross-species hybridizations represents genes detected by less-than-optimal probe sets. It is likely that a subset of flux genes represented on the same-species hybridizations also reflect less than optimal probe sets.

Hybridization Signals of Present Genes

Signal intensity is a primary parameter for comparing chip data points and determining the patterns of gene expression. Therefore, we compared same-species and cross-species hybridizations to determine the differences in signal intensity and the relationship to the percent of genes called present (Figure 3). All gene measurements from each data set were binned by signal intensity, and

the percent of genes called as present for each bin was calculated. For example, it is evident that 90% of genes with signal intensities in the 400–499 bin are identified as present for the human/human hybridizations. In contrast, only 60% of genes in that same bin are called as present for the cross-species hybridizations. It is not until the 1000–1099 bin for the cross-species hybridization, over double the intensity found for the human/human hybridizations, that 90% level of genes with a present call is reached using the Affymetrix algorithms. It is also clear that this relationship is consistent for signal intensities up to the 1100–1199 bin, albeit the gap narrows as signal intensity increases. These data further support the hypothesis that genomic sequence differences affect the number of usable probe sets in cross-species hybridizations, resulting in a lower percentage of genes called present.

Hybridization Signal Variability

We next examined the variability in signal intensities between replicate chips comparing same- and cross-species hybridizations. Using our data sets for replicate chips, we first determined the percent of genes with a greater than 2-fold change in signal intensity between the replicates. Ideally,

Short Technical Reports

replicate chips should yield no fold changes in signal intensities, though obviously the imperfection of the technology will produce some variation. Thus, we made the assumption that the percent of genes compared between replicates with greater than 2-fold changes is a measure of signal intensity variability or system noise that is greater than simple technological variance.

In Figure 4, the percent of genes that were called present on both chips (PP) or were Flux genes (A→P plus P→A) were compared by sorting them into bins as a function of signal intensity. Replicate pairs were arbitrarily assigned to signal bins by the highest value of the two. Less than 5% of the genes called present on both replicate chips (PP) show a variability greater than 2-fold. This is true for both same- and cross-species hybridizations and is regardless of signal intensities. Thus, an investigator doing cross-species hybridizations of nonhuman primate RNA to human chips should have as much confidence in the value of genes consistently called as present as with a human/human hybridization. In contrast, flux genes show a greater variability in signal intensities on replicate chips, and this is particularly evident for the cross-species hybridizations (flux rhesus as compared to flux human). Moreover, the majority of flux genes (194/241; 80%) are found in the lower intensity bins (i.e., <300). This also supports our hypothesis that an important determinant of the flux gene set is inefficient hybridization with the human sequence-specific probe sets.

Further work is clearly needed, particularly in the area of comparative primate genomics combined with follow-up validation of microarray data. For example, it is important to acknowledge that our present work is based on doing comparisons of only two replicate chips. Further testing with more replicates will be important to refine this data set, though it should not change our basic premise. It seems reasonable to go forward to identify and document working probe sets for various nonhuman primate species based on replicate cross-species hybridizations. These probe sets can be used now with confidence by investigators. Similarly, we propose that identifying the

subset of flux genes that represent probe sets that are not optimal for transcripts of different nonhuman primate species could direct the design of future GeneChips. However, until the nonhuman primate genome sequences are available, it is clear that there will continue to be a need for doing cross-species hybridizations.

REFERENCES

1. **Bigger, C.B., K.M. Brasky, and R.E. Lanford.** 2001. DNA microarray analysis of chimpanzee liver during acute resolving hepatitis C virus infection. *J. Virol.* 75:7059-7066.
2. **Cavallaro, S., B.G. Schreurs, W. Zhao, V. D'Agata, and D.L. Alkon.** 2001. Gene expression profiles during long-term memory consolidation. *Eur. J. Neurosci.* 9:1809-1815.
3. **Enard, W., P. Khaitovich, J. Klose, S. Zollner, F. Heissig, P. Giavalisco, K. Nieselt-Struwe, E. Muchmore, et al.** 2002. Intra- and interspecific variation in primate gene expression patterns. *Science* 5566:340-343.
4. **Frazer, K.A., J.B. Sheehan, R.P. Stokowski, X. Chen, R. Hosseini, J.F. Cheng, S.P. Fodor, D.R. Cox, and N. Patil.** 2001. Evolutionarily conserved sequences on human chromosome 21. *Genome Res.* 11:1651-1659.
5. **Hacia, J.G., J.B. Fan, O. Ryder, L. Jin, K. Edgemon, G. Ghandour, R.A. Mayer, B. Sun, et al.** 1999. Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. *Nat. Genet.* 2:164-167.
6. **Hacia, J.G., W. Makalowski, K. Edgemon, M.R. Erdos, C.M. Robbins, S.P. Fodor, L.C. Brody, and F.S. Collins.** 1998. Evolutionary sequence comparisons using high-density oligonucleotide arrays. *Nat. Genet.* 18:155-158.
7. **Kayo, T., D.B. Allison, R. Weindruch, T.A. Prolla.** 2001. Influences of aging and caloric restriction on the transcriptional profile of skeletal muscle from rhesus monkeys. *Proc. Natl. Acad. Sci. USA* 9:5093-5098.
8. **Li, C. and W.H. Wong.** 2001. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl. Acad. Sci. USA* 98:31-36.
9. **Lipshutz, R.J., S.P. Fodor, T.R. Gingeras, and D.J. Lockhart.** 1999. High density synthetic oligonucleotide arrays. *Nat. Genet.* 21:20-24.
10. **Lockhart, D.J., H. Dong, M.C. Byrne, M.T. Follettie, M.V. Gallo, M.S. Chee, M. Mittmann, C. Wang, et al.** 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* 13:1675-1680.
11. **Nau, M.E., R.E. Lyndal, R.K. Martin, D.E. Kyle, D.F. Wirth, and M. Vahey.** 2000. Technical assessment of the Affymetrix yeast expression GeneChip YE6100 platform in a heterologous model of genes that confer resistance to antimalarial drugs in yeast. *J. Clin. Microbiol.* 5:1901-1908.
12. **Mills, J.C. and J.I. Gordon.** 2001. A new approach for filtering noise from high-density oligonucleotide microarray datasets. *Nucleic Acids Res.* 15:E72-2.
13. **Schadt, E.E., C. Li, C. Su, and W.H. Wong.** 2000. Analyzing high-density oligonucleotide gene expression array data. *J. Cell Biochem.* 80:192-202.

TSRI manuscript no. 14583-MEM. We wish to thank Michelle Zandonatti for RNA preparation, Thomas Whisenant and David Rocke for assistance with statistical analysis, and Drs. Ernest Beutler, Martin Friedlander, and James Paulson for their support. This work was supported by the following grants: MH62261, MH59468, and MH61224 from the National Institutes of Mental Health to H.S.F. and M01RR00833 (Beutler), P30 EY12598-01 (Friedlander), and 1U54-GM62116 (Paulson) from the National Institutes of Health. Address correspondence to Dr. Jeffrey D. Chismar, Department of Shared Services, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA. e-mail: jchismar@scripps.edu

Received 11 March 2002; accepted 29 April 2002.

J.D. Chismar, T. Mondala, H.S. Fox, E. Roberts, D. Langford¹, E. Masliah¹, D.R. Salomon, and S.R. Head
The Scripps Research Institute
¹*University of California, San Diego*
La Jolla, CA, USA

For reprints of this or any other article, contact
Reprints@BioTechniques.com