A new method to remove hybridization bias for interspecies comparison of global gene expression profiles uncovers an association between mRNA sequence divergence and differential gene expression in Xenopus

Maureen A. Sartor¹,²,⁴, Aaron M. Zorn⁶, Jennifer A. Schwanekamp¹,³, Danielle Halbleib¹,³, Sai Kumar Karyala¹,³, Michael L. Howell⁸, Gary E. Dean⁸,⁹, Mario Medvedovic¹,²,⁴,⁵,⁷ and Craig R. Tomlinson¹,³,⁴,⁷,*

¹University of Cincinnati, Department of Environmental Health, Cincinnati, OH, 45267-0056, USA, ²Division of Biostatistics and Epidemiology, ³Division of Environmental Genetics and Molecular Toxicology, ⁴Center for Environmental Genetics, ⁵Center for Genome Information and ⁶Division of Developmental Biology, Children's Hospital, Cincinnati, OH 45229-3039, USA, ⁷Hyacinth Genomics, LLC, 3431 Stettinius Avenue, Cincinnati, OH 45208, USA, ⁸Protein Express, Inc., 9940 Reading Road, Cincinnati, OH 45241, USA and ⁹Department of Molecular Genetics, Biochemistry and Microbiology, University of Cincinnati, Cincinnati, OH, 45267-0524, USA

Received September 13, 2005; Revised November 29, 2005; Accepted December 9, 2005

ABSTRACT
The recent sequencing of a large number of Xenopus tropicalis expressed sequences has allowed development of a high-throughput approach to study Xenopus global RNA gene expression. We examined the global gene expression similarities and differences between the historically significant Xenopus laevis model system and the increasingly used X.tropicalis model system and assessed whether an X.tropicalis microarray platform can be used for X.laevis. These closely related species were also used to investigate a more general question: is there an association between mRNA sequence divergence and differences in gene expression levels? We carried out a comprehensive comparison of global gene expression profiles using microarrays of different tissues and developmental stages of X.laevis and X.tropicalis. We (i) show that the X.tropicalis probes provide an efficacious microarray platform for X.laevis, (ii) describe methods to compare interspecies mRNA profiles that correct differences in hybridization efficiency and (iii) show independently of hybridization bias that as mRNA sequence divergence increases between X.laevis and X.tropicalis differences in mRNA expression levels also increase.

INTRODUCTION
Xenopus has played a prominent role in many seminal discoveries in biology (1–3). The eggs and embryos of Xenopus in comparison with most vertebrates are larger, more plentiful, simpler to obtain and easier to manipulate. These virtues led researchers of the last century to use Xenopus laevis as a model system of choice to investigate countless questions in developmental and cellular biology. However, X.laevis has some shortcomings that the closely related species Xenopus tropicalis does not. A major advantage of X.tropicalis is that it has nearly one-half the genome content of X.laevis because the X.tropicalis genome is diploid while the X.laevis genome is allotetraploid, which for the most part precludes genetic and gene expression knockdown manipulations in X.laevis that can be readily carried out in X.tropicalis. Furthermore, X.tropicalis develops in one-third the time and requires one-fifth the housing space, yet is genetically and embryologically very similar to X.laevis.

To take full advantage of the Xenopus model system for studies in basic and medical science, genomic information
regarding the relative RNA expression levels of *X. laevis* and *X. tropicalis* must be made available. With the advent of the *X. tropicalis* system for genetic and genomic studies (4,5), it would be highly valuable to the research community to ascertain the similarities and differences that exist between the historically significant *X. laevis* system and the increasingly used and genetically amenable *X. tropicalis* system. Thus, a primary intent of the work described here was to provide a comprehensive comparison of global gene expression profiles of multiple tissues and developmental stages of the two *Xenopus* species using microarrays and to make the data accessible to the *Xenopus* research community.

Microarray studies with *Xenopus* were first carried out using a cDNA-based platform to examine the temporal regulation of global gene expression during development and neural induction (6,7). More recently, cDNA microarray platforms were used to analyse the global effect on gene expression by VegT mRNA from *X. tropicalis* found that the based on *X. laevis* levels of 96 genes between selty relevant to the work described here, mRNA expression tissues and developmental stages (9). In published work clo-

**Microarray hybridization**

Total RNA from the tissues and developmental stages was amplified two rounds using the Amino Allyl MessageAmp II aRNA Amplification Kit (Ambion, Austin, TX; catalog no. 1753) according to the accompanying protocol. The amplified RNA (aRNA) samples were concentrated to 2 μg/μl by vacuum drying. The aRNA was labeled with reactive monofunctional cyanine-3 (Cy3) and cyanine-5 (Cy5) dyes (Amersham, Piscataway, NJ; catalog no. RPN5661) by an indirect amino allyl labeling method as described in Guo et al. (14) with the following exception. The labeling reaction was initiated by adding 5 μl (10 μg of aRNA) to 5 μl of coupling buffer, which in turn was used to suspend the Cy3 and Cy5 dyes.

The *X. tropicalis* 70mer oligonucleotide library (Operon Biotechnologies, Inc., Huntsville, AL), representing 10,989 mRNA transcripts was suspended in 3× SSC at 30 μM and printed at 22°C and 65% relative humidity on aminosilane-coated slides (Cel Associates, Inc., Pearland, TX; VSA-25C) using a high-speed robotic OmniGrid machine (GeneMachines, San Carlos, CA) with Stealth SMP3 pins (Telechem, Sunnyvale, CA) (14,15).

A pre-hybridization step was carried out by placing slides in slide rack and immersing them in a staining dish containing 5× SSC, 0.1% SDS and 1% BSA, stirring at 48°C for 1 h. Following pre-hybridization, the slides were dipped ~10 times in two dishes containing denized water at room temperature and excess water was gently shaken off. The slides were dipped 10 times in isopropanol at room temperature and spun dry. The vacuum dried Cy3 and Cy5 labeled aRNAs were suspended in 9 μl water, and the mixture heated at 95°C for 3 min and centrifuged at 10,000×g for 1 min. A 2× hybridization buffer was prepared containing 50% formamide, 10× SSC and 0.2% SDS and preheated to 48°C. To the denatured Cy3/Cy5 target mixture, 21 μl of the 2× hybridization buffer preheated to 48°C was added. To block non-specific hybridization, 8 μl of calf thymus DNA (1 μg/μl) (Sigma, St Louis, MO; catalog no. D6861), 2 μl poly(A)-DNA (10 μg/μl) (Sigma, catalog no. P9403), 2 μl yeast tRNA (4 μg/μl) (Sigma, catalog no. R8759) were added to a total volume of 42 μl. The hybridization mixture was applied to the pre-hybridized microarray slide and covered with a 22 × 60 coverslip (Fisher, Pittsburgh, PA; catalog no. 12-545-J). The slide was placed in a CMT Hybridization Chamber (Corning, Acton, MA; catalog no. 2551) and 12 μl water was added to the small reservoirs at each end of the chamber. The sealed hybridization chambers were placed in a water bath at 48°C for 66 h (16). After hybridization, the slides were placed in a slide rack, set in a staining dish containing 1× SSC with 0.1% SDS preheated to 48°C, the coverslips were removed, and the slides were washed for 15 min at 48°C with agitation. The slides were washed further with agitation for 5 min at 48°C three times in 0.1× SSC and 0.1% SDS. The slides were transferred to a staining jar containing 0.1× SSC and washed twice at room temperature for 5 min with agitation. The slides were spun dried immediately after washing, and imaging and data analyses were carried out as described (16).

**Data normalization and analysis**

The data representing raw spot intensities generated by GenePix® Pro version 5.0 was analysed to identify

**MATERIALS AND METHODS**

**Xenopus cultures and RNA isolation**

*X. laevis* and *X. tropicalis* embryos were generated by in vitro fertilization as previously described (12), and the embryos were staged (13). Three separate matings were performed. Each biological replicate was from a separate mating, and 50 sibling embryos from the same mating were pooled to generate the RNA. Embryos from the different matings were always kept separate and not pooled. Total RNA was extracted from pooled embryos or 200 mg of ovary or liver tissue using Trizol Reagent (Invitrogen, Carlsbad, CA) according to the manufacturer’s protocol. The total RNA was further purified by phenol–chloroform extraction and ethanol precipitation.

*X. laevis* and *X. tropicalis* embryos were generated by in vitro fertilization as previously described (12), and the embryos were staged (13). Three separate matings were performed. Each biological replicate was from a separate mating, and 50 sibling embryos from the same mating were pooled to generate the RNA. Embryos from the different matings were always kept separate and not pooled. Total RNA was extracted from pooled embryos or 200 mg of ovary or liver tissue using Trizol Reagent (Invitrogen, Carlsbad, CA) according to the manufacturer’s protocol. The total RNA was further purified by phenol–chloroform extraction and ethanol precipitation.
differentially expressed genes. Data normalization was performed in three steps for each microarray separately (16). Channel specific local background intensities were subtracted from the median intensity of each channel (Cy3 and Cy5). Second, background adjusted intensities were log-transformed and the differences ($R$) and averages ($A$) of log-transformed values were calculated as $R = \log_2(X1) - \log_2(X2)$ and $A = (\log_2(X1) + \log_2(X2))/2$, where $X1$ and $X2$ denote the Cy5 and Cy3 intensities after subtracting local backgrounds, respectively. Third, data centering was performed by fitting the array-specific local regression model of $R$ as a function of $A$. The difference between the observed log-ratio and the corresponding fitted value represented the normalized log-transformed gene expression ratio. Normalized log-intensities for the two channels were then calculated by adding half of the normalized ratio to $A$ for the Cy5 channel and subtracting half of the normalized ratio from $A$ for the Cy3 channel. A statistical analysis was performed for each gene and for each Xenopus species separately by fitting the following mixed effects linear model (9). $Y_{ijk} = \mu + S_j + C_k + \mu_{ijk}$, where $Y_{ijk}$ corresponds to the normalized log-intensity on the $i$-th array ($i = 1, \ldots, 15$), with the $j$-th tissue/sample type and labeled with the $k$-th dye ($k = 1$ for Cy5 and 2 for Cy3), $\mu$ is the overall mean log-intensity, $A_i$ is the effect of the $i$-th array, $S_j$ is the effect of the $j$-th tissue/sample type and $C_k$ is the effect of the $k$-th dye. Assumptions about model parameters were the same as described by Wolfinger et al. (17), with array effects assumed to be random and treatment and dye effects assumed to be fixed. Additionally, a similar statistical analysis was performed for interspecies comparisons using the data from both Xenopus species. Statistical significance of differential expression among RNA samples, after adjusting for array and dye effects, was assessed by calculating $P$-values and estimates of fold-change were calculated. Multiple hypotheses testing adjustment was performed for the full analysis by calculating false discovery rates (FDR) (18,19) and Bonferroni adjusted $P$-values. Data normalization and statistical analyses were performed using SAS statistical software package (SAS Institute Inc., Cary, NC).

Cluster analysis

Clustering was performed using Bayesian infinite mixture (BIM) model based clustering (20) using normalized expression values for each comparison. BIM model based clustering allowed for the fitting of the statistical mixture model without knowing the number of clusters in the data (21). The statistical model was fitted using the Gibbs sampler, and hierarchical clustering was produced by treating pairwise posterior probabilities as the similarity measure and applying the traditional complete-linkage principle. The clustering results were displayed using the TreeView program (http://www.treeview.net/) (22).

Sequence Analysis

Measures of sequence similarity were obtained from Basic Local Alignment Search Tool (BLAST) searches of each of the 70mer oligonucleotides against an X.laevis EST database performed by Operon Biotechnologies, Inc. Measures of whole gene sequence similarity were obtained by similarly blasting the X.tropicalis sequences from which the 70mer oligonucleotides were derived against all X.laevis complete coding sequences available from the NCBI database. In both cases, the value used as the measure of sequence similarity was the highest ‘bit’ score among the significant matches for each BLAST result. We chose to use bits rather than $E$-scores because we were searching against only one database. Bits is a commonly used BLAST outcome score and is defined as follows: Score (bits) = $\lambda^* \text{Score (raw)} - \ln K / \ln 2$, where $\lambda$ and $K$ are Karlin-Altschul parameters (23).

Real-time quantitative polymerase chain reaction

Quantitative polymerase chain reaction (QPCR) analysis using SYBR green and designed primers was carried out following a described protocol (14,15) to confirm the microarray results. Ribosomal protein 60S L4 (RPL4) was selected for use as the reference RNA because there was little difference in RPL4 mRNA levels among the three developmental stages and two tissues of X.tropicalis to X.laevis. The forward and reverse primer sequences for RPL4 were 5‘CCAGAATTCTGAAAAGCAGGAG3’ and 5‘TCTCAAGCTGCTGAGATAGCG3’, respectively (product size 167 bp). The average cycle threshold ($C_T$) value for the reference RNA was used to normalize the tested gene hypoxia inducible factor $\alpha$ (HIF1$\alpha$). The forward and reverse primer sequences for HIF1$\alpha$ were 5‘GTAGTTCAAGGC-TTTGATGC3’ and 5‘GCAATGAAATCAAATACCAAGC3’, respectively (product size ~150 bp). The primer sequences for RPL4 and HIF1$\alpha$ were 100% complementary to the corresponding gene regions in both Xenopus species. All the PCR products produced single bands of the predicted sizes. For the negative control, no template was added to the reference RNA primers from which no PCR products were detected after 40 cycles. Approximately 20–30 µg of aRNA used in the microarray analysis (see earlier description) was used as template for cDNA synthesis using random primers. The QPCR was performed two times using 125 ng (25 ng/µl) and 60 ng (20 ng/µl) of starting cDNA template. The average $C_T$ values for the PCR amplifications and the reference RNAs were determined by carrying out a QPCR measurements from three biological replicates for each gene in each experimental condition (two tissues and three developmental stages). The results from the two QPCR runs were combined for statistical analysis.

Differential gene expression ratios were calculated based on $C_T$ values of the reference RNAs using the following calculations. ANOVA was used to calculate the $\Delta C_{T\text{array}}$ for each tissue/developmental stage and $P$-values. The $\Delta C_{T\text{array}}$ values were then transformed back to the original scale, where normalized RNA expression levels were $2^{\Delta C_{T\text{array}}}$.

RESULTS

We used a high-throughput microarray approach to investigate genomics of the Xenopus model system. The microarray experiments were carried out using a printed library of 70mer oligonucleotides representing 10 898 mRNA transcripts. The sequences were derived from the X.tropicalis expressed sequence tag (EST) project carried out at the Wellcome Trust/Cancer Research Gurdon Institute in Cambridge, UK. The entire experimental design for the microarray experiments carried out for the studies discussed in this paper is shown in Figure 1.
Hybridization efficiencies of *X.laevis* transcript sequences to *X.tropicalis* probes

Before additional questions could be investigated, we first needed to determine whether *X.laevis* transcript sequences could efficiently hybridize to *X.tropicalis* probes. The studies were carried out to ascertain a quantitative measure of how well *X.laevis* transcripts bound to the *X.tropicalis* microarray platform and to determine transcript-specific hybridization efficiency correction factors for direct interspecies comparisons. Hybridization efficiencies were determined by directly comparing the transcript levels of the corresponding tissues and developmental stages for the two *Xenopus* species, and the experimental design used for this portion of the study is shown in Figure 2A. Hybridization conditions for the microarray studies were relatively stringent (48°C, 25% formamide) to minimize as much as possible any experimental variability owing to non-specific binding.

The calculation of hybridization efficiencies was carried out using two methods. The first method is the method of choice for single channel arrays, and the other method is the choice for dual channel arrays. The methods included a local regression of *X.laevis* log spot intensity (single channel arrays, Figure 2B) and a local regression of the log of the ratios of the expression levels for each gene (dual channel arrays, Figure 2C) (plotted on the Y-axis) versus values that are a measure of sequence similarity in the corresponding 70mer probe (X-axis). In other words, the X-axis represents the degree to which the 70mer oligonucleotide probe contained significant sequence matches to the corresponding mRNA sequence (see Materials and Methods for a more detailed description). By assuming that for any sequence similarity level, the mRNA levels for an approximately equal number of genes increased or decreased, we could define the predicted values form local regression as the correction factors to use for hybridization efficiency. Local regression was chosen for two reasons: it is readily available in microarray analysis software because it is commonly used for the normalization of microarray data, and it avoids defining a specific parametric function for the dependence of hybridization efficiencies on sequence similarity. For the 9346 probes tested, i.e. those probes that produced signal over background levels, the Spearman rank correlation coefficient for spot intensity versus sequence similarity was $R = 0.4091$ ($P$-value < 0.0001). These results indicated that overall there was a reasonable increase in spot intensity as sequence similarity increased (Figure 2B).

A primary intent of our studies was to determine whether *X.laevis* gene expression profiles can be determined on an *X.tropicalis* microarray platform. Thus, we asked whether hybridization efficiency seemed to be a factor in differential gene expression measurements (Figure 2C). A measurement of the log ratios versus sequence similarity showed that for the ~4000 probes tested from the direct comparison of the different tissues and developmental stages, as sequence similarity decreased, differential gene expression increased between the two species (Figure 2C). The negative correlation was greatest for transcripts from stage 40 ($R = -0.5328$) and least from liver ($R = -0.3479$) with $P$-values at <0.0001 for all the RNA samples. The predicted values obtained from the local regression analysis were used as the correction factors for hybridization analysis. The magnitude of correction needed showed the likely need of using a correction factor in microarray interspecies comparisons, and although our correction factors are experiment-specific and therefore not listed here, this method for determining correction factors can be used for any interspecies microarray comparisons. In conclusion, using methods that directly compared *X.laevis* and *X.tropicalis* transcript sequences with the *X.tropicalis* platform, we showed that (i) the greater the divergence of RNA transcripts, the greater the difference in spot intensities and differential mRNA levels and (ii) correction factors were calculated for the *X.laevis* sequences.

Differential mRNA expression levels increase as gene sequence divergence increases

A reasonable argument to explain the results in the previous section is that the observed association between sequence divergence versus spot intensities and differential gene expression is due to the inability of divergent *X.laevis* sequences to efficiently hybridize to the *X.tropicalis* probes on the microarray. That is, of course spot intensities will decrease and differences in mRNA levels will increase if one of the sets of transcripts has greater sequence divergence to the arrayed probes. Therefore, we developed a new method to pursue the intriguing question of whether there is an association between RNA transcript divergence and differential gene expression in a way that was free of bias owing to differences in interspecies hybridization efficiencies.

We hypothesized that on the whole, as mRNA sequence divergence increased, differences in gene expression between *X.laevis* and *X.tropicalis* would also increase. The hypothesis was based on the premise that the more similar an mRNA is
expressed in time and space, the more probable the gene sequence and (therefore the encoded protein) will be similar. To test our hypothesis, we designed an experimental approach that was independent of differences in hybridization efficiencies (Figure 3A), in which mRNA levels from liver, ovary, egg, stage 10 (St. 10) and stage 40 (St. 40) were compared with a corresponding RNA samples from three biological replicates of the tissues and developmental stages for a given *Xenopus* species were directly compared to each other by microarray analysis. (B) A plot of log average spot intensities from *X.laevis* (Y-axis) versus *X.tropicalis–X.laevis* probe sequence similarity (X-axis). (C) A plot of the log-ratio hybridization bias (Y-axis) versus *X.tropicalis–X.laevis* probe sequence similarity (X-axis). The Y-axis represents the expected offset from actual relative mRNA expression levels (X.tropicalis/X.laevis).

![Figure 2](image_url)

**Figure 2.** A measurement of microarray hybridization efficiencies for *X.laevis* versus *X.tropicalis*. (A) The experimental design to determine the hybridization efficiencies of *X.laevis* transcripts from ovary, liver, egg; stage 10 (St. 10) and stage 40 (St. 40) to the *X.tropicalis* DNA microarray probes. Corresponding RNA samples from three biological replicates of the tissues and developmental stages for a given *Xenopus* species were directly compared to each other by microarray analysis. (B) A plot of log average spot intensities from *X.laevis* (Y-axis) versus *X.tropicalis–X.laevis* probe sequence similarity (X-axis). (C) A plot of the log-ratio hybridization bias (Y-axis) versus *X.tropicalis–X.laevis* probe sequence similarity (X-axis). The Y-axis represents the expected offset from actual relative mRNA expression levels (X.tropicalis/X.laevis).

Of the 10,898 RNA transcripts screened on the microarray slide, 1681 were identified as significantly different from the corresponding reference RNA for at least one tissue or developmental stage, using the relatively strict criteria of ≥2-fold difference in mRNA levels and FDR (18) of ≤0.05.
Three approaches were used to examine mRNA expression levels versus sequence divergence (Figure 3B and C, Table 1). The first approach was an investigation into the relationship between the difference in mRNA expression levels for each tissue and stage normalized to their respective reference RNAs (Y-axis) versus sequence similarity (X-axis) (Figure 3B). Of the five tissues and developmental stages tested, all but liver (P-value 0.203) showed a significant negative correlation (the remaining: P-value < 0.0001) between mRNA expression levels normalized to reference and sequence similarity (Figure 3B). The range of the correlation values were $R = -0.0785$ for egg to $R = -0.1260$ for ovary. These results showed that as sequence divergence increased so did the divergence in relative mRNA expression levels.

The second approach (Figure 3C) was an examination of the relationship between the square of the correlation coefficients (data from Figure 4B) for all the genes that were significantly changed relative to the corresponding reference RNA versus probe sequence similarity (X-axis). Only the 1681 genes described in the text that were significantly changed relative to the corresponding reference RNA were plotted. (C) A plot of the squared correlation coefficient values (Y-axis) versus probe sequence similarity (X-axis). The Y-axis represents the square of the correlation coefficients for the 1681 genes.

Three approaches were used to examine mRNA expression levels versus sequence divergence (Figure 3B and C, Table 1). The first approach was an investigation into the relationship between the difference in mRNA expression levels for each tissue and stage normalized to their respective reference RNAs (Y-axis) versus sequence similarity (X-axis) (Figure 3B). Of the five tissues and developmental stages tested, all but liver (P-value 0.203) showed a significant negative correlation (the remaining: P-value < 0.0001) between mRNA expression levels normalized to reference and sequence similarity (Figure 3B). The range of the correlation values were $R = -0.0785$ for egg to $R = -0.1260$ for ovary. These results showed that as sequence divergence increased so did the divergence in relative mRNA expression levels.

The second approach (Figure 3C) was an examination of the relationship between the square of the correlation coefficients (data from Figure 4B) for all the genes that were significantly different from the reference RNA (Y-axis) versus the sequence similarity measure (X-axis). For each gene, the square of the correlation coefficient is a measure of mRNA expression similarity of each gene across tissues in *X.laevis* versus *X.tropicalis*. The spearman rank correlation coefficient was $R = 0.1234$ (P-value < 0.0001) showing that as sequence similarity decreased, the correlation between *X.laevis* and
X. tropicalis tissues and developmental stages also decreased. Again, these results are consistent with the contention that the more similar a gene is between X. laevis and X. tropicalis the more likely the genes are expressed similarly.

The above two approaches provided global quantitative measures in the relationship between sequence divergence and gene expression. To totally rule out hybridization efficiency as a confounding factor, the third method we used was an examination of the sequence similarities of specific RNA transcripts from homologous genes of X. laevis and X. tropicalis (Table 1) and was performed to eliminate the possibility that the correlation seen above was due to variability in physical binding properties. In this case, RNA transcript sequence divergence was determined for those genes in which the 70mer probe sequence on the microarray matched at least 69 of 70 nt within the corresponding X. laevis mRNA sequence. This approach allowed a direct measure of sequence divergence independent of hybridization efficiencies because both the X. laevis and X. tropicalis sequences shared nearly identical regions for hybridization to the microarray probes. A total of 154 genes were found that met this condition as well as being present in all tissues and in developmental stages. The sequence similarities of the 154 genes were then compared with the corresponding differential gene expression levels (Table 1).

The results in Table 1 confirmed the results from the global approaches (Figure 3B and C). The 154 genes were evenly divided into two groups in which Group 1 contained those genes that were more similar in sequence outside the complementary 70mer region and Group 2 contained those genes that were less similar outside the 70mer region. The two gene groups were then compared for the five tissues and developmental stages, in which the normalized log difference of the differential mRNA levels were calculated by the formula log [(Xt / Xt Ref) / (Xl / Xl Ref)] described earlier. Differential mRNA expression levels between the two gene groups were significantly different when the data for all the stages and tissues were combined (P-value < 0.007). Additionally, each developmental stage and tissue showed a trend in which differential gene expression widened as sequence divergence increased, and for egg and ovary, the difference was significant (P-value < 0.05). We note that our measure of sequence similarity is dependent on the database used for BLAST, and as the relevant sequence databases are updated, the analysis will be more precise. Nevertheless, we were able to detect significance even in the midst of the noise created by incomplete sequence data. In conclusion, as shown by different approaches, as sequence divergence in mRNA transcripts increased between X. laevis and X. tropicalis, differential gene expression also increased.

**Table 1. Differential mRNA expression levels between X. laevis and X. tropicalis as mRNA sequence divergence increases**

| Gene Group | Medians of absolute normalized log differences | Bit score range (median) |
|------------|-----------------------------------------------|--------------------------|
|            | Egg Stage 10 Stage 40 Liver Ovary            |                          |
| Group 1 (Higher similarity scores) | 0.46 0.51 0.61 0.87 0.32 735–4230 (1201) |                          |
| Group 2 (Lower similarity scores) | 0.67 0.62 0.64 1.08 0.47 34–722 (348)    |                          |

*a* There were 77 genes in each group.

*b* Overall P-value = 0.007 from Wilcoxon non-parametric test.

*c* Absolute normalized log difference calculation: |log [(Xt / Xt Ref) / (Xl / Xl Ref)]|.

*d* Significant difference: P-value < 0.05.

**Figure 4.** The gene expression profiles for X. laevis (Xl) and X. tropicalis (Xt) are similar. (A) The experimental design to determine how X. laevis and X. tropicalis compare in their global gene expression profiles for selected tissues (ovary and liver) and developmental stages (egg, stage 10 and stage 40). mRNA levels from three biological replicates from the two tissues and three developmental stages were compared to a reference RNA (Ref.) for a given Xenopus species. The estimates of log [(Xt/Xt Ref)/(Xl/Xl Ref)] for each gene were compared between X. laevis and X. tropicalis to determine correlation coefficients. (B) Histogram representing the correlation coefficients between X. laevis and X. tropicalis using 1681 transcript levels that changed significantly among the different tissues and stages.
The global gene expression profiles for *X.laevis* and *X.tropicalis* are generally similar

The third question we asked was in regard to how *X.laevis* and *X.tropicalis* compare in their global gene expression profiles for selected tissues and developmental stages. For this part of the study, we used those data derived from the experimental design shown in Figure 4A, in which mRNA levels from liver, ovary, egg, stage 10 and stage 40 were compared with the corresponding *X.laevis* or *X.tropicalis* reference RNA described earlier. Again, separate comparisons with a commonly composed reference RNA allowed the determination of relative gene expression changes for each *Xenopus* species independent of differences in hybridization efficiencies.

As described earlier, 1681 genes were identified as significantly different in mRNA expression levels from the corresponding reference RNA for at least one tissue or developmental stage. Use of the significantly changed genes allowed us to compare ratio-based correlation coefficients for like genes between *X.laevis* and *X.tropicalis* among tissues (Figure 4B). Of the 10 898 total transcripts, 2510 transcripts (23%) were present in all tissues and met the signal to noise criterion for analysis of *X.laevis* versus the corresponding reference RNA, and 8317 transcripts (76%) were similarly analysed for *X.tropicalis* versus its reference RNA. Of the 1681 significantly changed genes ~24% had a correlation of 0.90 or greater, and 68% of the genes for the two *Xenopus* species had correlation values of 0.5 or greater. Thus, the global gene expression levels for *X.laevis* and *X.tropicalis* are generally similar for the examined tissues and developmental stages.

The EASE program (24), was used to functionally group the most positively (>0.90) and negatively (<−0.35) correlated genes between *X.laevis* and *X.tropicalis* based on the Gene Ontology (GO, http://www.geneontology.org/) database, in which gene products are described in the context of biological processes, cellular components and molecular functions in a species-independent manner (24,25) (Table 2). Using Fisher’s Z-score in EASE, the two gene lists were used to test for GO categories significantly enriched with either the most highly, or negatively, correlated genes. The results revealed that the most highly correlated expressed genes were involved primarily in protein synthesis. The lowly correlated genes did not yield any significant results. The 20 most positively and negatively correlated genes are shown in Table 3. The merging of all 1681 significantly changed genes into the GO program (Table 4) showed that the categories whose genes correlated most highly between *X.laevis* and *X.tropicalis*, on average were related to development, growth, reproduction, cell death, biosynthesis and response to abiotic/external stimulus.

QPCR assays were carried out to confirm the microarray results. The HIFIα gene was selected for the QPCR assays based on the criteria that it (i) was expressed at relatively high intensities; (ii) was shown relatively large fold changes for at least some of the different tissues and developmental stages and (iii) was a gene that encoded a product of known function. The fold-change increase in HIFIα RNA expression levels in *X.tropicalis* relative to *X.laevis* from the microarray studies were: egg, 5.2; stage 10, 9.4; stage 40, 8.5; liver, 27.3 and ovary, 6.3 (P-values ≤ 0.001). Using RPL4 as a reference, the fold-change increase in HIFIα RNA expression levels in *X.tropicalis* relative to *X.laevis* from the QPCR results were: egg, 7.4; stage 10, 1.2; stage 40, 3.3; liver, 2.8 and ovary, 1.5 (egg, stage 40, and liver P-value ≤ 0.001; ovary P-value ≤ 0.05). Thus, because the QPCR results for the five tissues/developmental stages all followed the same trend as the microarray results and at statistically significant levels (except for stage 10), the QPCR results validated the microarray results.

**Table 2.** The biological process, molecular function, or cellular component involving those genes with the most highly correlated (>0.90) gene expression levels between *X.laevis* and *X.tropicalis*.

| GO Gene Category | List hits | List total | P-value | Bonferroni P-value | Benjamini FDR |
|------------------|-----------|------------|---------|--------------------|---------------|
| Structural molecule activity | 52 | 211 | 0.000 | 0.000 | 0.000 |
| Cytosolic ribosome | 29 | 208 | 0.000 | 0.000 | 0.000 |
| Protein biosynthesis | 46 | 211 | 0.000 | 0.135 | 0.012 |
| Nucleic acid binding | 88 | 211 | 0.000 | 0.142 | 0.012 |
| Ribonucleoprotein complex | 46 | 208 | 0.000 | 0.149 | 0.012 |
| Protein metabolism | 85 | 211 | 0.000 | 0.263 | 0.020 |
| Cytosol | 56 | 208 | 0.001 | 0.725 | 0.051 |
| Binding | 141 | 211 | 0.001 | 0.772 | 0.051 |
| Biosynthesis | 53 | 211 | 0.009 | 1.000 | 0.446 |

*Gene Ontology (GO) program (http://www.geneontology.org).*  
*bFalse Discovery Rate (18,19).*

**The global gene expression profiles for *X.laevis* and *X.tropicalis* follow parallel temporal developmental programs**

The last question posed was how the two *Xenopus* species compare in their gene expression profiles during development. That is, do the temporal gene expression profiles correspond similarly to the observed embryological stages for the two species? Identifying the genes that behave similarly and differently at the different developmental stages may lead to the identification of determinants. To carry out this portion of the study, mRNA levels from egg, stage 10 and stage 40 from a given *Xenopus* species was compared with the same corresponding reference RNAs described earlier (Figure 5A). In addition, for each *Xenopus* species, mRNA from egg was compared with stage 10 mRNA and with stage 40 mRNA relative to the reference RNA. By comparing gene expression levels in relation with the corresponding reference RNA, the experimental design again allowed for the detection of differentially expressed genes during the course of development without the skewing effects from differences in hybridization efficiencies for the two *Xenopus* species on the *X.tropicalis*-based microarray platform.

The heat map in Figure 5B shows a gene cluster diagram for stage 10 and stage 40 mRNA levels versus corresponding *X.laevis* and *X.tropicalis* egg mRNA levels and includes the top ranked most significantly changed 200 genes in each comparison that changed at least 50%. There were overlapping genes that were top ranked in more than one comparison, thus, a total of 547 different genes were included. The heat map shows that for the vast majority of genes, a given gene that either increased (red), decreased (green) or did not change (black) relative to the corresponding egg mRNA, also increased, decreased or did not change accordingly in the other *Xenopus* species. Thus, the heat map is generally
Table 3. The 20 genes of the highest (upper tier) and lowest (lower tier) correlation values

| NCBI clone ID | Gene product description                                                                 | Pearson correlation |
|---------------|------------------------------------------------------------------------------------------|---------------------|
| NP_000970.1   | Ribosomal protein L18; 60S ribosomal protein L18 [Homo sapiens]                         | 0.997               |
| NP_009092.2   | NRAS-related gene; upstream of NRAS [H.sapiens]                                        | 0.997               |
| NP_003398.1   | Zinc finger protein 36                                                                   | 0.997               |
| NP_115684.1   | Hypothetical protein MGC4189 [H.sapiens]                                                | 0.995               |
| NP_497827.1   | Cytosolic juvenile hormone binding protein subunit like (32.1 kD) (3F243)               | 0.995               |
| NP_001001.2   | Ribosomal protein S6; 40S ribosomal protein S6; phosphoprotein NP53                      | 0.995               |
| NP_004692.1   | Cyclin B2 [H.sapiens]                                                                   | 0.994               |
| AAH61315.5    | Unknown (protein for MGC:75795) [Silurana tropicalis]                                   | 0.994               |
| BAC98186.1    | mKIAA1504 protein [Mus musculus]                                                        | 0.994               |
| NP_114172.1   | Cyclin B1; G2/mitotic-specific cyclin B1 [H.sapiens]                                    | 0.994               |
| P30985        | Transcription factor 12 (Class A helix–loop–helix transcription factor GE1)             | 0.993               |
| NP_001924.2   | Dihydrolipoamide S-succinyltransferase                                                   | 0.992               |
| NP_057735.2   | DAPPER1 [H.sapiens]                                                                     | 0.992               |
| NP_057018.1   | Nucleolar protein NOP5/NOP58 [H.sapiens]                                                | 0.992               |
| NP_014936.1   | Homology to rat S10; Rps10ap [Saccharomyces cerevisiae]                                 | 0.992               |
| P98199        | Potential phospholipid-transporting ATPase (ATPase class I type 8B member 2)             | 0.992               |
| NP_000995.1   | Ribosomal protein P2; 60S acidic ribosomal protein P2                                    | 0.991               |
| NP_001025.1   | Ribonucleotide reductase M2 polypeptide [H.sapiens]                                      | 0.991               |
| NP_002257.1   | Karyopherin alpha 2; RAG cohort 1; importin alpha 1 [H.sapiens]                         | 0.991               |
| XP_252671.2   | Similar to Probable chromodomain-helicase-DNA-binding protein KIA1A416                   | 0.991               |
| NP_060695.1   | Homolog of Caenorhabditis elegans smu-1; ortholog of rat brain-enriched WD-repeat protein| –0.774              |
| AAA70336.1    | LATS                                                                                   | –0.784              |
| AAH60352.1    | Unknown (protein for MGC:68448) [Xenopus laevis]                                        | –0.787              |
| NP_735366.1   | CG2139-PB [Drosophila melanogaster]                                                      | –0.790              |
| NP_07287.1    | Hypothetical protein ET [H.sapiens]                                                     | –0.796              |
| NP_15097.1    | Mitochondrial ribosomal protein S36 [H.sapiens]                                         | –0.802              |
| NP_006295.1   | Deleted in split-hand/split-foot 1 region [H.sapiens]                                   | –0.829              |
| NP_057124.2   | CGI-100 protein [H.sapiens]                                                             | –0.831              |
| BAC04638.1    | Unnamed protein product [H.sapiens]                                                     | –0.837              |
| NP_005889.2   | Membrane component, chromosome 11, surface marker 1 [H.sapiens]                         | –0.841              |
| XP_358556.1   | Hypothetical protein XP_358556 [Mus musculus]                                          | –0.859              |
| NP_001780.1   | Cell division cycle 25A; Cdc25A; protein-tyrosine-phosphatase [H.sapiens]              | –0.865              |
| NP_006590.1   | Armadillo repeat-containing protein [H.sapiens]                                         | –0.869              |
| NP_733778.1   | Muscle-specific beta 1 integrin binding protein [H.sapiens]                             | –0.890              |
| NP_056299.1   | GCIP-interacting protein p29 [H.sapiens]                                                | –0.896              |
| NP_780399.1   | RIKEN CDNA 290001023 [Mus musculus]                                                     | –0.905              |
| XP_215053.2   | Similar to D7Wsu128e protein [Rattus norvegicus]                                       | –0.908              |
| NP_068681.1   | Quaking protein [Mus musculus]                                                         | –0.912              |
| NP_001800.1   | Centromere protein A; centromere protein A (17kD) [H.sapiens]                         | –0.935              |

a block of genes with decreased mRNA levels (green upper half) over a block of genes with increased mRNA expression levels (lower red half) and indicates that the gene expression profiles during development for X. laevis and X. tropicalis are very similar. Nonetheless, there were several small clusters of genes that were contrary to the general trend and are designated to the right of the heat map by the numbers 1–3 (Figure 5B). For example, the cluster of genes designated number one shows a group of genes that decreased at stages 10 and 40 in X. tropicalis, whereas, in X. laevis, the same genes remained relatively unchanged. However, in each of the groups 1–3, no functional relationship among the genes could be discerned and suggests that there are no major differences in any specific developmental program. In conclusion, for the great majority of the genes, mRNA levels for the same genes rose and declined similarly during development in both Xenopus species.

Additional evidence that the developmental programs for X. laevis and X. tropicalis are similarly regulated is shown in Table 5. Table 5 lists the correlation coefficients between the two Xenopus species for the 116 genes identified as significantly changed among tissues or developmental stages, i.e. those genes involved in development according to the GO database (first row in Table 4, GO term GO:0007275). Of the 116 genes listed, 88 (76%) had a correlation coefficient of 0.5 or greater (above the black dividing line) suggesting that many of the key genes that direct or participate in development are regulated similarly in the two Xenopus species.

**DISCUSSION**

**Sequence divergence and gene expression**

We introduced four questions to investigate. The first question was whether the association between gene sequence divergence and hybridization efficiency can be effectively measured and used to generate a correction factor when mRNA levels are directly compared (Figure 2), and the second question was whether there is an association between sequence divergence and differences in gene expression levels for X. laevis and X. tropicalis when hybridization bias is removed (Figure 3). The prediction for the first question is obvious, i.e. that there would be a direct relationship between sequence similarity and hybridization efficiency, and our intent was to quantify sequence divergence and hybridization efficiency between the two Xenopus species. By plotting spot intensity...
and the relative mRNA expression levels for each gene (Figure 2C) as measures of hybridization efficiency versus sequence matches in the 70mer probe, we showed, not surprisingly, that overall, as sequence similarity for the genes from the two *Xenopus* species increased, so did hybridization efficiency. Moreover, we obtained a hybridization efficiency measure for each *Xenopus* gene based on its sequence similarity measure. Such information may be useful as a correction factor when directly comparing *X.laevis* and *X.tropicalis* mRNA levels or the mRNA levels of any species.

For the second question, our hypothesis was that as sequence divergence increased so would the difference in gene expression levels (Figure 3, Table 1). The hypothesis was based on the premise that as sequence divergence increased for a given transcript between two species, the greater the likelihood that the transcript encoded a different protein and that the encoded protein carried out a different function, and therefore, the greater the likelihood the transcript would be expressed at a different time and/or level for the encoded protein to carry out the different function. That is,

| Level | GO term | GO description | Correlation Genes | Mean | Median | SD |
|-------|---------|----------------|-------------------|------|--------|----|
| 1     | GO:0007275 | Development    | 116               | 0.61 | 0.80   | 0.45 |
| 1     | GO:0050793 | Regulation of biological process | 203               | 0.60 | 0.74   | 0.41 |
| 1     | GO:0009987 | Cellular process | 797               | 0.58 | 0.74   | 0.43 |
| 1     | GO:0007582 | Physiological process | 793              | 0.57 | 0.73   | 0.43 |
| 2     | GO:0040007 | Growth         | 18                | 0.70 | 0.85   | 0.32 |
| 2     | GO:0050793 | Regulation of development | 13               | 0.69 | 0.84   | 0.29 |
| 2     | GO:0000003 | Reproduction   | 17                | 0.54 | 0.85   | 0.57 |
| 2     | GO:0016265 | Death          | 36                | 0.67 | 0.82   | 0.35 |
| 2     | GO:0009605 | Response to external stimulus | 51              | 0.65 | 0.81   | 0.40 |
| 2     | GO:0042592 | Homeostasis    | 11                | 0.59 | 0.80   | 0.45 |
| 2     | GO:0050794 | Regulation of cellular process | 174             | 0.61 | 0.77   | 0.41 |
| 2     | GO:0009653 | Morphogenesis  | 71                | 0.59 | 0.75   | 0.45 |
| 2     | GO:0050791 | Regulation of physiological process | 193          | 0.60 | 0.75   | 0.42 |
| 2     | GO:0030154 | Cell differentiation | 22             | 0.64 | 0.74   | 0.42 |
| 2     | GO:0008152 | Metabolism     | 622               | 0.58 | 0.74   | 0.42 |
| 2     | GO:0007154 | Cell communication | 111              | 0.57 | 0.73   | 0.41 |
| 2     | GO:0006950 | Response to stress | 59              | 0.59 | 0.70   | 0.38 |
| 2     | GO:0050790 | Regulation of enzyme activity | 13           | 0.53 | 0.70   | 0.53 |
| 2     | GO:0006928 | Cell motility   | 20                | 0.54 | 0.67   | 0.41 |
| 2     | GO:0009719 | Response to endogenous stimulus | 15          | 0.65 | 0.60   | 0.25 |
| 2     | GO:0016049 | Cell growth    | 12                | 0.72 | 0.89   | 0.34 |
| 2     | GO:0009628 | Response to abiotic stimulus | 17          | 0.63 | 0.85   | 0.49 |
| 2     | GO:0040008 | Regulation of growth | 11             | 0.68 | 0.84   | 0.31 |
| 2     | GO:0009581 | Detection of external stimulus | 16          | 0.57 | 0.83   | 0.53 |
| 2     | GO:0008219 | Cell death     | 36                | 0.67 | 0.82   | 0.35 |
| 2     | GO:0009058 | Biosynthesis   | 172               | 0.63 | 0.81   | 0.42 |
| 2     | GO:019953  | Sexual reproduction | 16             | 0.52 | 0.81   | 0.58 |
| 2     | GO:019538  | Protein metabolism | 298             | 0.62 | 0.80   | 0.42 |
| 2     | GO:016043  | Cell organization and biogenesis | 78          | 0.60 | 0.79   | 0.44 |
| 2     | GO:0009607 | Response to biotic stimulus | 45            | 0.64 | 0.75   | 0.34 |
| 2     | GO:007267  | Cell-cell signaling | 11             | 0.55 | 0.74   | 0.46 |
| 2     | GO:008283  | Cell proliferation | 47             | 0.56 | 0.74   | 0.46 |
| 2     | GO:006793  | Phosphorus metabolism | 64            | 0.62 | 0.73   | 0.37 |
| 2     | GO:006118  | Electron transport | 46             | 0.55 | 0.73   | 0.38 |
| 2     | GO:007155  | Cell adhesion   | 21                | 0.65 | 0.72   | 0.29 |
| 2     | GO:009887  | Organogenesis   | 53                | 0.60 | 0.72   | 0.41 |
| 2     | GO:0019222 | Regulation of metabolism | 123          | 0.57 | 0.70   | 0.42 |
| 2     | GO:006139  | Nucleobase, nucleoside, nucleotide and nucleic acid synthesis | 238          | 0.55 | 0.69   | 0.44 |
| 2     | GO:007165  | Signal transduction | 89             | 0.56 | 0.69   | 0.41 |
| 2     | GO:009308  | Amines metabolism | 30             | 0.52 | 0.69   | 0.42 |
| 2     | GO:006519  | Amino acid and derivative metabolism | 26          | 0.51 | 0.68   | 0.44 |
| 2     | GO:005975  | Carbohydrate metabolism | 39            | 0.57 | 0.67   | 0.40 |
| 2     | GO:004011  | Locomotion      | 20                | 0.54 | 0.67   | 0.41 |
| 2     | GO:006810  | Transport       | 162               | 0.53 | 0.66   | 0.45 |
| 2     | GO:006119  | Oxidative phosphorylation | 23           | 0.63 | 0.66   | 0.27 |
| 2     | GO:009056  | Catabolism      | 74                | 0.56 | 0.65   | 0.42 |
| 2     | GO:009611  | Response to wounding | 17             | 0.56 | 0.63   | 0.41 |
| 2     | GO:006091  | Generation of precursor metabolites and energy | 83           | 0.53 | 0.61   | 0.39 |
| 2     | GO:006086  | Alcohol metabolism | 26             | 0.56 | 0.61   | 0.36 |
| 2     | GO:006082  | Organic acid metabolism | 37           | 0.46 | 0.60   | 0.44 |
| 2     | GO:006974  | Response to DNA damage stimulus | 14           | 0.63 | 0.60   | 0.24 |

aGene Ontology (GO) program (http://www.geneontology.org).
bOverall mean = 0.56.
cOverall median = 0.73.
Figure 5. The global gene expression profiles for *X.laevis* (XI) and *X.tropicalis* (Xt) follow parallel temporally-regulated developmental programs. (A) The part of the experimental design used to determine how *X.laevis* and *X.tropicalis* compare in their global gene expression profiles for selected developmental stages (egg; stage 10, St. 10 and stage 40, St. 40). mRNA levels from three biological replicates from the three developmental stages for a given Xenopus species were compared with a reference RNA (Ref.) and mRNA from egg was compared with stage 10 and to stage 40 via the reference RNA. (B) Hierarchical tree of genes and heat map of the developmental stages, in which corresponding stage 10 and stage 40 mRNA levels were compared to egg mRNA levels for each Xenopus species. The top 200 ranked genes in each comparison that were at least 50% changed were included. The heat map columns left to right are: *X.laevis* stage 10 versus *X.laevis* egg, *X.tropicalis* stage 10 versus *X.tropicalis* egg, *X.laevis* stage 40 versus *X.laevis* egg, and *X.tropicalis* stage 40 versus *X.tropicalis* egg. The brackets to the right of the heat map numbered 1–3, designate groups of genes that are contrary to the overall clustering trend and are described in the text.
## Table 5. Correlation coefficients of genes known to be involved in Xenopus development

| Gene ID | Description of development genes | Pearson correlation |
|---------|----------------------------------|--------------------|
| 7812    | NRAS-related gene; upstream of NRAS | 1.00               |
| 51602   | Nucleolar protein NOP5/NOP58      | 0.99               |
| 4678    | Nuclear autoantigenic sperm protein isofrom 1 | 0.99              |
| 21974   | Topoisomerase (DNA) II beta       | 0.98               |
| 1786    | DNA (cytosine-5')-methyltransferase 1; DNA methyltransferase 1; DNAmethyltransferase | 0.98 |
| 1466    | Cysteine and glycine-rich protein 2; SmLIM; LIM domain only 5, smooth muscle | 0.98 |
| 175621  | EMB-5, abnormal EMBryogenesis EMB-5 (175.8 kD) (emb-5) | 0.98 |
| 4624    | Myosin heavy chain 6; myosin heavy chain alpha isoform | 0.97 |
| 851389  | Required for Start B in mitosis and for meiosis I spindle pole body separation; Cdk36p | 0.97 |
| 1277    | Alpha 1 type I collagen prepropeptide | 0.97 |
| 1459    | Casein kinase 2, alpha prime polypeptide | 0.97 |
| 7273    | Titin isofrom N2-B; connectin; CMH9, included; cardiomyopathy, dilated 1G | 0.97 |
| 13822   | Unnamed protein product | 0.96 |
| 30096   | Zic       | 0.95   |
| 5351    | Lysyl hydroxylase precursor; lysine hydroxylase | 0.95 |
| 3149    | High-mobility group box 3; high-mobility group (nonhistone chromosomal) protein4 | 0.95 |
| 172244  | Cytoplasmic Polyadenylation Element-Binding protein (cpb-3) | 0.95 |
| 665     | Bone morphogenetic protein 7 precursor; osteogenic protein 1 | 0.95 |
| 3399    | Inhibitor of DNA binding 3 | 0.94 |
| 326340  | Zygote arrest 1 | 0.94 |
| 70      | Actin, alpha, cardiac muscle precursor | 0.94 |
| 3622    | Inhibitor of growth 1-like | 0.94 |
| 71320   | Membrane-bound casein kinase I homolog; Yck2p | 0.93 |
| 54766   | B-cell translocation gene 4; putative transcriptional regulator | 0.93 |
| 1301    | Alpha 1 type XI collagen isofrom B prepropeptide; collagen XI, alpha-1-polypeptide | 0.93 |
| 8651    | Suppressor of cytokine signaling 1; STAT induced SH3 protein 1 | 0.93 |
| 4116    | Mago-nashi homolog | 0.92 |
| 23411   | Siruin 1; sir2-like 1; siruin type 1 | 0.92 |
| 51654   | CDK5 regulatory subunit associated protein 1 isofrom a | 0.92 |
| 7784    | Zona pellucida glycoprotein 3 prepropeptide | 0.91 |
| 6520    | Solute carrier family 3 (activators of dibasic and neutral amino acid transport), member 2 | 0.91 |
| 86      | BAFA53a; hArpN beta; actin-related protein; BAF complex 53 kDa subunit; BRG1-associated factor | 0.91 |
| 57829   | Zona pellucida glycoprotein 4 prepropeptide; zona pellucida B protein | 0.90 |
| 3475    | Interferon-related developmental regulator 1 | 0.90 |
| 854549  | Homolog of chicken calponin, thus the name S.cerevisiae CalPonin; Scp1p | 0.89 |
| 70      | Actin, alpha, cardiac muscle precursor | 0.89 |
| 5757    | Prothymosin, alpha (gene sequence 2B) | 0.89 |
| 8857    | Fc fragment of IgG binding protein; IgG Fc binding protein | 0.89 |
| 12505   | CD44 antigen precursor (Phucygotic glycoprotein I) (PGP-1) (HUTCH-1) | 0.88 |
| 323630  | Similar to dishevelled 2, dsh homolog | 0.87 |
| 64063   | T-box transcription factor comesosdermin | 0.87 |
| 6678    | Secreted protein, acidic, cysteine-rich (osteocentin) | 0.87 |
| 4729    | NADH dehydrogenase (ubiquinone) flavoprotein 224kDa | 0.87 |
| 984     | Cell division cycle 2-like 1 (PITSLRE proteins); Cell division cycle 2-like 1 | 0.86 |
| 3491    | Cysteine-rich, angiogenic inducer, 61 | 0.85 |
| 2266    | Fibrinogen, gamma chain isofrom gamma-A precursor | 0.85 |
| 20687   | trans-acting transcription factor 3 | 0.85 |
| 23481   | Pescadillo homolog 1, containing BRCT domain | 0.85 |
| 174044  | SMAI body size SMA-6, Serine-threonine kinase, transforming growth factor betatype I receptor | 0.84 |
| 10361   | Nucleoplasmin 2 | 0.83 |
| 180357  | ForkHead transcription factor family member, defective PHArynx development | 0.82 |
| 6159    | Ribosomal protein L29; 60S ribosomal protein L29; heparin/heparansulfate-interacting protein | 0.82 |
| 176688  | Serine/arginine rich splicing factor SF2, substrate of the SR protein kinasesPK-1 (28.7 kDa) | 0.82 |
| 1278    | Alpha 2 type I collagen; Collagen I, alpha-2 polypeptide; Collagen of skin,tendon and bone, alpha-2 chain | 0.81 |
| 5292    | Pim-1 oncogene; Oncogene PIM1 | 0.80 |
| 54993   | Zinc finger protein 29 | 0.80 |
| 51399   | Synbindin; TRS23 homolog; hematopoietic stem/progenitor cell protein 172 | 0.80 |
| 70      | Actin, alpha, cardiac muscle precursor | 0.79 |
| 6658    | SRY (sex determining region Y)-box 3; transcription factor SOX-3 | 0.78 |
| 3398    | Inhibitor of DNA binding 2; inhibitor of differentiation 2; DNA-binding protein inhibitor ID2 | 0.77 |
| 54514   | DEAD (Asp-Glu-Ala-Asp) box polypeptide 4; VASA protein | 0.76 |
| 26578   | Osteocalcin stimulating factor 1 | 0.75 |
| 10643   | IGF-II mRNA-binding protein 3; KH domain containing protein overexpressed in carcinoma | 0.75 |
| 5743    | Prostaglandin-endoperoxide synthase 2 precursor; prostaglandin G/H synthase andcyloxygenase | 0.75 |
| 6227    | Ribosomal protein S21; 40S ribosomal protein S21 | 0.74 |
| 8943    | Adaptor-related protein complex 3, delta 1 subunit; adaptin, delta | 0.73 |
| 5515    | Protein phosphatase 2 (formerly 2A), catalytic subunit, alpha isoform | 0.72 |
| 6223    | Ribosomal protein S19; 40S ribosomal protein S19 | 0.72 |
| 5052    | Peroxiredoxin 1; natural killer-enhancing factor A; proliferation-associated gene A | 0.71 |
across tissues and developmental stages, the more similar orthologous genes are expressed in time and space, then the more likely the gene sequence and therefore the encoded protein sequence will be similar. To test that hypothesis, we designed a method to compare interspecies transcriptomes free of hybridization bias using a same species reference RNA. It should be noted that the method in theory can be used to compare any two transcriptomes free of binding bias on any given array of probes, although to increase sensitivity it may positively select for a mutation in the coding region of the gene.

Indeed, we observed that for the transcripts on the whole, as mRNA sequence divergence increased, so did mRNA expression levels, which raises a cause and effect question. In the case where sequence divergence in the mRNA sequence could be a cause for differential gene expression, a gene with a non-lethal, non-synonymous mutation would encode a different protein that would either be better or worse at carrying out its function than would the formerly encoded protein. Because no mutation would have yet occurred in the regulatory regions for the gene in question, the transcript for the new protein would probably be expressed at the same time and level of the previous transcript. A mutation in the regulatory region of the gene would be required to alter transcriptional timing or transcript levels such that the new protein could function better, serving as a positive selective mechanism for the gene.

It could also be the case that a change in gene expression could be the selective force for mRNA sequence divergence. A mutation that first occurred in the regulatory region that caused a change in transcript levels and/or timing could be a means for selecting those mutations in the corresponding coding region that encoded a better functioning protein under the new expression conditions. For example, a mutation in the promoter that decreased the transcription rate of the gene would probably be expressed at the same time and level of the previous transcript. A mutation in the regulatory region of the gene would be required to alter transcriptional timing or transcript levels such that the new protein could function better, serving as a positive selective mechanism for the gene.

Table 5

| Gene ID    | Description of development genes                                                                 | Pearson correlation |
|------------|---------------------------------------------------------------------------------------------------|---------------------|
| 2010       | Emerin                                                                                           | 0.71                |
| 2147       | Coagulation factor II precursor; prothrombin                                                    | 0.70                |
| 1209       | Cleft lip and palate associated transmembrane protein 1                                         | 0.68                |
| 1281       | Alpha 1 type III collagen; Collagen III, alpha-1 polypeptide; collagen, fetal                    | 0.67                |
| 27289      | GTP-binding protein RHO6                                                                        | 0.67                |
| 80781      | Alpha 1 type XVIII collagen isofrom 2 precursor; endostatin                                      | 0.65                |
| 652        | Bone morphogenetic protein 4 preproprotein; bone morphogenetic protein 2B                      | 0.63                |
| 856856     | Suppressor of Choline Synthesis Likely to be involved in regulating INO1expression                | 0.62                |
| 859        | Caveolin 3; M-caveolin; caveolin-3                                                               | 0.61                |
| 851676     | Brain Modulosignalin Homolog; Bmh2p [S.cerevisiae]                                              | 0.60                |
| 176702     | Human Mortality factor-Related Gene related (38.3 kDa) (mrg-1)                                   | 0.59                |
| 35070      | Cadherin-N CG7100-PH                                                                           | 0.58                |
| 43510      | Kayak CG15509-PB                                                                               | 0.54                |
| 2195       | FAT gene product                                                                                | 0.53                |
| 6665       | SRY (sex determining region Y)-box 15; SRY (sex determining region Y)-box 20                    | 0.51                |
| 1994       | ELAV-like 1; embryonic lethal, abnormal vision, Drosophila, homolog-like 1; Huantien R          | 0.50                |
| 5274       | Serine (or cysteine) proteinase inhibitor, clade I (neuroserpin), member 1; protease inhibitor 12 (neuroserpin) | 0.50                |
| 4733       | Developmentally regulated GTP binding protein 11                                                 | 0.50                |
| 6997       | Teratocarcinoma-derived growth factor 1                                                          | 0.45                |
| 11146      | FKBP-associated protein isofamp; FK506-binding protein-associated protein; glomulin              | 0.44                |
| 1634       | Decorin isofrom b precursor; dermatan sulphate proteoglycans II                                 | 0.41                |
| 694        | B-cell translocation protein 1                                                                  | 0.30                |
| 10856      | RuvB-like 2; erythrocite cystosol protein, 51-KD; TBP-interacting protein, 48-KD; Reptin52       | 0.24                |
| 3852       | Keratin 5; Keratin-5; 58 kda cytokeratin; keratin, type II cytoskeletal 5;cystokeratin 5         | 0.23                |
| 7125       | Tropomin C2, fast                                                                               | 0.19                |
| 173233     | UNCoordinated locomotion UNC-59, septin (52.9 kDa) (unc-59)                                     | 0.12                |
| 7092       | Tolloid-like 1                                                                                  | 0.09                |
| 2879       | Glutathione peroxidase 4; phospholipid hydroperoxidase; sperm nucleglutathione peroxidase         | 0.03                |
| 1655       | DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 5                                                  | 0.00                |
| 224        | Aldehyde dehydrogenase 3A2; aldehyde dehydrogenase 10; fatty aldehyde dehydrogenase           | 0.00                |
| 928        | CD9 antigen; motility related protein; leukocyte antigen MAC3                                   | 0.01                |
| 43383      | Fork head CG10002-PA [Drosophila melanogaster]                                                   | 0.01                |
| 8861       | LIM domain binding 1; carboxy terminal LIM domain protein 2; LIM domain-bindingfactor-1         | 0.03                |
| 41062      | Polychaetoid CG31434-PA                                                                         | 0.05                |
| 5931       | Retinoblastoma binding protein 7                                                                | 0.07                |
| 4637       | Smooth muscle and non-muscle myosin alkali light chain isoform 1                                | 0.08                |
| 7171       | Tropomyosin 4                                                                                   | 0.10                |
| 7168       | Tropomyosin 1 (alpha)                                                                           | 0.30                |
| 8324       | Frizzled 7; frizzled (Drosophila) homolog 7                                                     | 0.51                |
| 4869       | Nucleophosmin (nucleolar phosphoprotein B23, numatrin); Nucleophosmin 1                         | 0.51                |
| 4738       | Neural precursor cell expressed, developmentally down-regulated 8                               | 0.52                |
| 197988     | Cadherin protein                                                                                | 0.53                |
| 51588      | Protein inhibitor of activated STAT protein PIASy                                               | 0.64                |
| 7979       | Deleted in split-hand/split-foot 1 region                                                        | -0.83               |
| 19317      | Quaking protein                                                                                 | -0.91               |
that increased the activity of the translated protein. When more sequence data for both Xenopus species become available, we would predict that the genes in Xenopus with the greatest sequence divergence in the coding region would be a means for compensation, also have the greatest sequence divergence in the regulatory regions. Furthermore, it will prove interesting to tease out whether certain groups of genes diverged more rapidly than others and whether these genes play roles in developmental programs. Our new method for interspecies transcriptome comparisons may also prove useful in the testing of the neutral theory of evolution (26, 27) to determine whether the mRNA sequence divergence between species is primarily stochastic and neutral or the result of natural selection. Our results would support the latter because the mRNA sequence divergence we observed did have an effect on mRNA expression levels, i.e. if the transcript sequence divergence we observed was neutral, little or no change in gene expression would be predicted.

An apparent complication that would not be faced in most interspecies comparisons is that X. laevis is an allotetraploid (28), in which ~50% of the genes in X. laevis are duplicated (29), which arose from the fusion of two nuclei from different species (30). These facts raise the question of (i) which orthologous transcript sequence (which may vary considerably for a given duplicated gene) may have hybridized to a given X. tropicalis probe and (ii) the larger question of whether a given X. laevis transcript sequence may have hybridized specifically to the corresponding orthologous probe. In regard to which X. laevis orthologous transcript may have hybridized to a given X. tropicalis probe, it is impossible with the current microarray probes to determine to what degree dissimilar RNA sequences of a duplicated X. laevis gene pair would have hybridized. The binding of two different labeled targets to a probe would have provided a sum amount of fluorescence signal. In our analysis when applicable (Table 1), we used only the more similar sequence of an X. laevis gene pair because it could not be ascertained how much signal the less similar sequence provided. Thus, for Table 1 the degree of sequence divergence may be underestimated.

In regard to whether a given X. laevis sequence hybridized specifically to the corresponding orthologous probe, because hybridization efficiency is a function of sequence similarity, owing to the stringent hybridization conditions employed for these studies and the divergence of the X. laevis sequences, it is probable that only the orthologous transcripts would have hybridized to the corresponding probe. Furthermore, non-specific hybridization would be expected to be less in closely related interspecies comparisons than with an intraspecies comparison, i.e. if somewhat divergent sequences are expected to hybridize less well to the orthologous probes then they would also be expected to bind less well to the non-orthologous probes. These observations may account for our results described earlier in which only 23% of the genes met the signal to noise criterion for analysis of X. laevis versus its reference RNA while 76% of the X. tropicalis genes were analysed versus its reference RNA.

Other studies have addressed the problem of sequence mismatches for multi-species comparisons on microarrays by using multi-probe arrays representing the different target species (31), by disregarding all data except those representing identical target/probe sequences (27), or by incorporating a non-specific, general normalization procedure (32). Interspecies hybridization approaches in which mRNA levels were directly compared include canine sequences on human probes (33), bovine sequences on human probes (34), porcine sequences on human probes (35) and non-human primate sequences on human probes (36). In each study, orthologous genes were identified with similar and dissimilar gene expression levels. Our results showed that the orthologous genes between X. laevis and X. tropicalis produced generally similar expression profiles. However, to our knowledge, no other published work has attempted to quantify interspecies comparisons in which the bias owing to differences in hybridization efficiencies was removed and in which differential gene expression levels and transcript sequence divergence were correlated.

**Interspecies comparisons**

Third, we asked how differential gene expression patterns compared between corresponding tissues and developmental stages of X. laevis and X. tropicalis. The expressed RNA from the tissues and stages were each compared with a same-species reference RNA, a method that allowed the comparison of interspecies expression patterns free of bias owing to sequence differences (Figure 4A). The homologous Xenopus genes that were significantly differentially expressed relative to their reference RNA in at least one of the tested tissues and stages (1681 genes of ±2-fold or greater, \( P \)-value \( \leq 0.001 \), FDR \( \leq 0.05 \)) were then compared across species. We assumed that the differentially expressed genes were representative of all genes present in all tissues and there was no evidence to suggest otherwise. We suggest that this approach will be useful in comparing gene expression levels of other related species on a given microarray platform.

We found that among the differentially expressed genes, gene expression levels were quite similar between the two species, and our conclusion was that the genes between the two species were generally expressed similarly (Figure 4B). However, that conclusion was based on our findings that 23\% of the 10 898 total transcripts present in the examined tissues and developmental stages were successfully analysed for X. laevis, yet, 76\% of the transcripts were similarly analysed for X. tropicalis. These results suggest that (i) the overall RNA transcript sequence divergence between the Xenopus species may be such that only one-third of the X. laevis transcript sequences relative to X. tropicalis can bind sufficiently to the X. tropicalis platform and/or (ii) the very unlikely prospect that nearly two-thirds of the transcripts expressed in the examined tissues and developmental stages in X. tropicalis are not expressed in X. laevis. Therefore, our conclusion that the two Xenopus species generally have similar global gene expression patterns is somewhat guarded because it is based on approximately one-quarter of the total available Xenopus transcripts.

**Global gene expression during X. laevis and X. tropicalis development**

The fourth question dealt with how well temporal gene expression changes across the observed embryological stages correlated for the two Xenopus species. We found that gene expression profiles between two given developmental stages
were generally similar for each Xenopus species (Figure 5). The results for this part of the study were also free of any bias due to sequence differences in that changes in gene expression over time (egg to stage 10 to stage 40) were determined relative to the like species reference RNA. In this case, all the homologous Xenopus genes that were differentially expressed in any of the three developmental stages relative to the reference RNA were compared with each other (+2-fold or greater, P-value ≤ 0.001, FDR ≤ 0.05). These results led to a total of 547 genes included in the clustering analysis. The data presented in Figure 5B show that X.laevis and X.tropicalis express the majority of genes at the same developmental stage. Although it was beyond the scope of this paper to speculate what implications the differences in temporal expression may mean in the two developmental programs, it will undoubtedly bear out that the genes that are not expressed similarly will prove more interesting.

ACKNOWLEDGEMENTS

We thank Drs Ken Petren and Yolanda Sanchez for critical reading of the manuscript and Subramanian Venkatesan and Srinivasan Raghuraman for technical assistance. This work was supported in part by National Institutes of Health NCRR grant R43 RR019815-01 to MLH and HD42572 to A.M.Z. Funding to pay the Open Access publication charges for this article was provided by the GML at the University of Cincinnati.

Conflict of interest statement. None declared.

REFERENCES

1. Gurdon,J.B. (2002) Perspective on the Xenopus field. Dev. Dyn., 225, 379.
2. Slack,J. (2000) Essential Developmental Biology. Blackwell Science, Oxford.
3. Wolpert,L., Beddington,R., Brockes,J., Jessell,T., Lawrence,P. and Meyerowitz,E. (1998) Principles of Development. Current Biology, London.
4. Amaya,E., Offield,M.F. and Grainger,R.M. (1998) Frog genetics: Xenopus tropicalis jumps into the future. Trends Genet., 14, 253–255.
5. Hirsch,N., Zimmerman,L.B. and Grainger,R.M. (2002) Xenopus, the next generation: X.tropicalis genetics and genomics. Dev. Dyn., 225, 422–433.
6. Altman,C.R., Bell,E., Sczyrba,A., Pun,J., Bekiranov,S., Gaasterland,T. and Brivanlou,A.H. (2001) Microarray-based analysis of early development in Xenopus laevis. Dev. Biol., 236, 64–75.
7. Munoz-Sanjuan,J., Bell,E., Altman,C.R., Viona,A. and Brivanlou,A.H. (2002) Gene profiling during neural induction in Xenopus laevis: regulation of BMP signaling by post-transcriptional mechanisms and TAB3, a novel TAK1-binding protein. Development, 129, 5529–5540.
8. Tavener,N.V., Kofron,M., Shin,Y., Kabitschke,C., Gilchrist,M.J., Wylie,C., Cho,K.W., Heasman,J. and Smith,J.C. (2005) Microarray-based identification of VegT targets in Xenopus. Mech. Dev., 122, 333–347.
9. Baldessari,D., Shin,Y., Krebs,O., Konig,R., Koide,T., Vinayagam,A., Fenger,U., Mochii,M., Terasaka,C., Kitayama,A. et al. (2005) Global gene expression profiling and cluster analysis in Xenopus laevis. Mech. Dev., 122, 441–475.
10. Chalmers,A.D., Goldstone,K., Smith,J.C., Gilchrist,M., Amaya,E. and Papalopulu,N. (2005) A Xenopus tropicalis oligonucleotide microarray works across species using RNA from Xenopus laevis. Mech. Dev., 122, 355–363.
11. Gilchrist,M.J., Zorn,A.M., Voigt,J., Smith,J.C., Papalopulu,N. and Amaya,E. (2004) Defining a large set of full-length clones from a Xenopus tropicalis EST project. Dev. Biol., 271, 498–516.
12. Khojka,M.K., Chung,C., Bustamante,E.L., Gaw,L.W., Trott,K.A., Yeh,J., Lim,N., Lim,J.C., Taverner,N., Amaya,E. et al. (2002) Techniques and probes for the study of Xenopus tropicalis development. Dev. Dyn., 225, 499–510.
13. Nieuwenhuizen,P.D. and Faber,J. (1994) Normal Table of Xenopus laevis (Daudin): A Systematcal and Chronological Survey of the Development from the Fertilized Egg till the End of Metamorphosis. Garland Publishing, Inc., New York.
14. Guo,J., Sartor,M., Karyala,S., Medvedovic,M., Kann,S., Puga,A., Ryan,P. and Tomlinson,C.R. (2004) Expression of genes in the TGF-beta signaling pathway is significantly deregulated in smooth muscle cells from aorta of aryl hydrocarbon receptor knockout mice. Toxicol. Appl. Pharmacol., 194, 79–89.
15. Karyala,S., Guo,J., Sartor,M., Medvedovic,M., Kann,S., Puga,A., Ryan,P. and Tomlinson,C.R. (2004) Different global gene expression profiles in benzo[a]pyrene- and dioxin-treated vascular smooth muscle cells of AHR-knockout and wild-type mice. Cardiovasc Toxicol., 4, 47–73.
16. Sartor,M., Schwaneckamp,J., Halbleib,D., Mohamed,I., Karyala,S., Medvedovic,M. and Tomlinson,C.R. (2004) Microarray results improve significantly as hybridization approaches equilibrium. Biotechniques, 36, 790–796.
17. Wolfinger,R.D., Gibson,G., Wolfinger,E.D., Bennett,L., Hamadeh,H., Bushel,P., Afshari,C. and Paules,R.S. (2001) Assessing gene significance from cDNA microarray expression data via mixed models. J. Comput. Biol., 8, 625–637.
18. Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc., B 57, 289–309.
19. Reiner,A., Yekutieli,D. and Benjamini,Y. (2003) Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics, 19, 388–397.
20. Medvedovic,M. and Sivaganesan,S. (2002) Bayesian infinite mixture model based clustering of gene expression profiles. Bioinformatics, 18, 1194–1206.
21. Medvedovic,M., Yeung,K.Y. and Bumgarner,R.E. (2004) Bayesian mixture model based clustering of replicated microarray data. Bioinformatics, 20, 1222–1232.
22. Eisen,M.B., Spellman,P.T., Brown,P.O. and Botstein,D. (1998) Cluster analysis and display of genome-wide expression patterns. Proc. Natl Acad. Sci. USA, 95, 14863–14868.
23. Karlin,S. and Altschul,S.F. (1990) Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. Proc. Natl Acad. Sci. USA, 87, 2262–2266.
24. Hosack,D.A., Dennis,G., Jr., Sherman,B.T., Lane,H.C. and Lempicki,R.A. (2003) Identifying biological themes within lists of genes with EASE. Genome Biol., 4, R70.
25. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genet., 25, 25–29.
26. Kimura,M. (1983) The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge, UK.
27. Kaitiothov,P., Weiss,G., Lachmann,M., Hellmann,I., Enard,W., Muzezel,B., Winkler,U., Angor,W. and Paabo,S. (2004) A neutral model of transcriptome evolution. PLoS Biol., 2, E132.
28. Bisbec,C.A., Baker,M.A., Wilson,A.C., Haji-Azimi,J. and Fischberg,M. (1977) Albumin phylogeny for clawed frogs (Xenopus). Science, 195, 785–787.
29. Hughes,M.K. and Hughes,A.L. (1993) Evolution of duplicate genes in a tetraploid animal, Xenopus laevis. Mol. Biol. Evol., 10, 1360–1369.
30. Prince,V.E. and Pickett,F.B. (2002) Splitting pairs: the diverging fates of duplicated genes. Nature Rev. Genet., 3, 827–837.
31. Gilad,Y., Rikitin,S.A., Bertone,P., Gerstein,M. and White,K.P. (2005) Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles. Genome Res., 15, 674–680.
32. Ranz,J.M., Castillo-Davis,C.I., Meiklejohn,C.D. and Hart, R.D.L. (2003) Sex-dependent gene expression and evolution of the Drosophila transcriptome. Science, 300, 1742–1745.
33. Grigoryev,D.N., Ma,S.F., Simon,B.A., Irizarry,R.A., Ye,S.Q. and Garcia,J.G. (2005) In vitro identification and in silico utilization of interspecies sequence similarities using GeneChip technology. BMC Genomics, 6, 62.
34. Adjaye, J., Herwig, R., Herrmann, D., Wruck, W., Benkahla, A., Brink, T.C., Nowak, M., Carnwath, J.W., Hultschig, C., Niemann, H., et al. (2004) Cross-species hybridisation of human and bovine orthologous genes on high density cDNA microarrays. *BMC Genomics*, 5, 83.

35. Shah, G., Azizian, M., Bruch, D., Mehta, R. and Kittur, D. (2004) Cross-species comparison of gene expression between human and porcine tissue, using single microarray platform—preliminary results. *Clin. Transplant*, 18 (Suppl 12), 76–80.

36. Wang, Z., Lewis, M.G., Nau, M.E., Arnold, A. and Vahey, M.T. (2004) Identification and utilization of inter-species conserved (ISC) probesets on Affymetrix human GeneChip platforms for the optimization of the assessment of expression patterns in non-human primate (NHP) samples. *BMC Bioinformatics*, 5, 165.