OMICS A Journal of Integrative Biology
Volume 8, Number 1, 2004

Incongruent Expression Profiles between Human and Mouse Orthologous Genes Suggest Widespread Neutral Evolution of Transcription Control

ITAI YANAI,[1,2] DAN GRAUR,[3] and RON OPHIR[2]

ABSTRACT

Rapid rates of evolution can signify either a lack of selective constraint and the consequent accumulation of neutral alleles, or positive Darwinian selection driving the fixation of advantageous alleles. Based on a comparison of 1,350 orthologous gene pairs from human and mouse, we show that the evolution of gene expression profiles is so rapid that it is comparable to that of paralogous gene pairs or randomly paired genes. The expression divergence in the entire set of orthologous pairs neither strongly correlates with sequence divergence, nor focuses in any particular tissue. Moreover, comparing tissue expressions across the orthologous gene pairs, we observe that any human tissue is more similar to any other human tissue examined than to its corresponding mouse tissue. Collectively, these results indicate that, while some differences in expression profiles may be due to adaptive evolution, the levels of divergence are mostly compatible with a neutral mode of evolution, in which a mutation for ectopic expression may rise to fixation by random drift without significantly affecting the fitness. A disturbing corollary of these findings is that knowledge of where the gene is expressed may not carry information about its function.

INTRODUCTION

Transcription levels, locations, and timing of expression constitute a good first indication of a gene's activity. Microarray experiments have enabled the exploration of an organism's gene expression program across conditions (Chu et al., 1998; DeRisi et al., 1997), tissues (Su et al., 2002), and pathological and ontogenetical stages (Arbeitman et al., 2002; Golub et al., 1999), thereby providing a high-throughput means for estimating the latitude of a gene's function.

Commonly overlooked, however, is the contribution, if any, of the transcription in each particular context of expression to the fitness of the organism. Drawing from predicted protein-coding sequences from the genomes of human (Lander et al., 2001) and mouse (Waterston et al., 2002), which diverged more than 70 million years ago (Springer et al., 2003), it is possible to identify pairs of genes between the two organisms that are most likely to retain the same function, that is, orthologs. Human and mouse gene pairs

[1] Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel.
[2] Bioinformatics Unit, Department of Biological Services, Weizmann Institute of Science, Rehovot, Israel.
[3] Department of Biology and Biochemistry, University of Houston, Houston, Texas.

are considered orthologous if their divergence is solely due to a speciation event, that is, if they are direct descendants of a single gene that existed in the most recent common ancestor of the two taxa. Such gene pairs are "direct evolutionary counterparts" (Koonin, 2001) and are expected, with very few exceptions, to retain the same function. The degree to which the human–mouse orthologs have retained the same function can be estimated from their pattern of expression. A recently published "gene expression atlas" (Su et al., 2002) of over 200 high-density oligonucleotide arrays of human and mouse normal tissues contains 64 arrays corresponding to 16 homologous tissues from human and mouse, each in duplicate.

A high degree of divergence between the expression profiles of orthologs from closely related taxa indicates either (1) a lack of selective constraint on expression and the consequent accumulation of neutral controlling alleles by random genetic drift or (2) positive Darwinian selection driving the fixation of advantageous alleles controlling tissue specificity of expression (Graur and Li, 2000). Here we present evidence in support of the existence of widespread neutral expression in the two organisms, that is, expression that confers no selective advantage onto the organism. The existence of neutral expression would imply that comparative expression studies—rather than "high-throughput bioinformatics"—are necessary to distinguish between instances in which expression at the RNA level indicates function and those in which expression neither fulfils a function nor lowers the fitness of the organism.

MATERIALS AND METHODS

Sequence analysis

Orthology between human and mouse sequences was determined at the level of protein sequence by using sequences retrieved from NCBI (www.ncbi.nlm.nih.gov). Sequences assigned to the same LocusLink (Pruitt and Maglott, 2001) correspond to alternative splice variants and were interpreted as products of the same gene. The Inparanoid program (Remm et al., 2001) was used to detect orthologs, in which each orthologous cluster has at least 80% sequence coverage. Of the 13,666 families found, we selected the 12,678 families of size one, that is, families containing one human–mouse homologous gene pair. We identified paralogs by searching for pairs of sequences with an alignment of BLAST expectation value of 10^{-10} or less, and a bidirectional alignment coverage of at least 80%. We detected 1,314 and 2,600 human and mouse paralogous pairs, respectively. Of the 12,678 orthologous families of size one, 6,894 have paralogs in at least one of the two genomes. Genetic distances between proteins were calculated according to Kimura's protein distance, $D = -\ln(1 - p - 0.2p^2)$, where p denotes the sequence identity between two proteins (Kimura, 1983).

Microarray analysis

Of the 46 and 45 human and mouse tissues, respectively, studied with Affymetrix chips in the original study (Su et al., 2002), we selected 16 tissues (listed in Fig. 2 below) that were common to both organisms and in duplicate. We pre-processed the primary data of the study using the Bioconductor R-package (www.bioconductor.org). Briefly, we used robust multi-array averaging (RMA) background correction and median polish summary as described by Irizarry et al. (2003) and quantile normalization as described by Bostald et al. (2003) on the $\log_2$ transformed data. Furthermore, we standardized the intensities of each array to a mean of 0 and a standard deviation of 1. The microarray probesets were matched to mRNAs by sequence similarity (sequences were retrieved from the Affymetrix and NCBI websites). We removed from further analysis those probesets whose probes may be assigned to more than one LocusLink. Additionally, we required that each LocusLink gene be linked to only one probeset. Thus, splice variants associated with different probesets were also removed from the analyses. Of the 12,678 orthologous families, 2,268 had both human and mouse probesets on the chips. To control for the quality of the expression data, we excluded tissues in which the expression intensities between replicates differed by more than 20%. Next, we filtered those pairs whose profiles had less than 12 common tissues with consistent intensities. Finally, we removed those pairs where one or both genes did not show at least one significant expression across the tissues. We were left with 1,350 orthologous pairs, of which 762 do not have paralogs and the remaining have paralogs that predate the human–mouse divergence (outparalogs [Remm et al., 2001]). As far as paralogous pairs are concerned, 83 are from human and 82 are from mouse.

Human–mouse Comparisons

To analyze the differences between the expression profiles, we calculated a normalized Euclidean distance (henceforth distance) in the tissue space, which reflects the differences in both expression pattern and intensity. Euclidean distances were calculated for "consistent" tissues, and were normalized by the square root of the number of such tissues. To isolate pattern effects from intensity values, and to detect similar expression patterns, we calculated Pearson's correlations between the two taxa. Pearson's correlations were calculated only for tissues of the human–mouse profile pairs that were judged to be consistent with respect to their replicates (i.e., intensities between replicates differed by less than 20%). As each gene-pair is represented by four profiles (two replicates in mouse and two in human), the average normalized Euclidean distance and Pearson's correlation of all possible combinations (H1-M1, H1-M2, H2-M1, and H2-M2, where H = human, M = mouse, and 1 and 2 are the replicates) was calculated and used in the distributions. Taking all four pairs without averaging yielded similar results. The random and most similar human and mouse pairs were calculated, respectively, from 7,309 and 5,221 human and mouse expression profiles that met the consistency criterion. Whenever normality cannot be assumed, non-parametric statistic tests are used to test for differences between distributions. Accordingly Wilcoxon's rank sum tests replaced *t*-tests in our analyses.

Tissue expression dendrogram

From the set of 1350 orthologous pairs, we selected those 159 pairs whose member genes are differentially expressed (ANOVA, $p < 0.05$, df = 15) in both human and mouse. Hierarchical clustering of this set yielded a dendrogram of relationships (Chu et al., 1998) which was bootstrapped by sampling with replacement 1000 times (Felsenstein, 1985). The consensus dendrogram was constructed using the consense program (Margush and McMorris, 1981).

RESULTS

Similarities between human and mouse orthologous expression profiles

The distribution of distances between the replicated arrays (Fig. 1A) has a median value of 0.025. This result indicates that the experimental error is rather small. We proceeded by matching each human gene expression profile with the most similar mouse gene profile and vice versa. The median distance for this matched set was 0.04, quite similar to the distance obtained for the experimental replicates (Fig. 1A). Surprisingly, none of these most similar profile pairs corresponded to orthologs. Furthermore, even when greatly relaxing the criteria to include not one, but the top one hundred best hits, only eight orthologs were detected (~1%).

The distributions of distances between orthologous pairs has a median of 0.14 and 0.15 for orthologs with or without paralogs, respectively (Fig. 1A). The fact that the experimental error is significantly smaller than the distance between orthologs ($p < 10^{-16}$, Wilcoxon rank test) strongly supports the notion that the distance between orthologous genes is real, rather than due to experimental noise. Furthermore, the correspondence between the distributions of the two groups of orthologs suggests that the paralogs do not appreciably influence the distance between the expression profiles of the orthologs (Fig. 1A). As a control, we calculated the distances between pairs of randomly chosen human and mouse genes, and found a median value of 0.25. Although similar in shape (Fig. 1A), the median of the random set is significantly larger than that for the orthologous pairs ($p < 10^{-16}$, Wilcoxon rank test).

It is also important to compare the orthologous distances with the distances between paralogous pairs, since paralogs are expected to diverge in function from one another quite frequently (Force et al., 1999; Lynch and Conery, 2000; Wagner, 2002). Thus, it is expected that the median distance would be higher for paralogs than for orthologs. Indeed, the median distance were 0.20 and 0.16 for the human and mouse par-

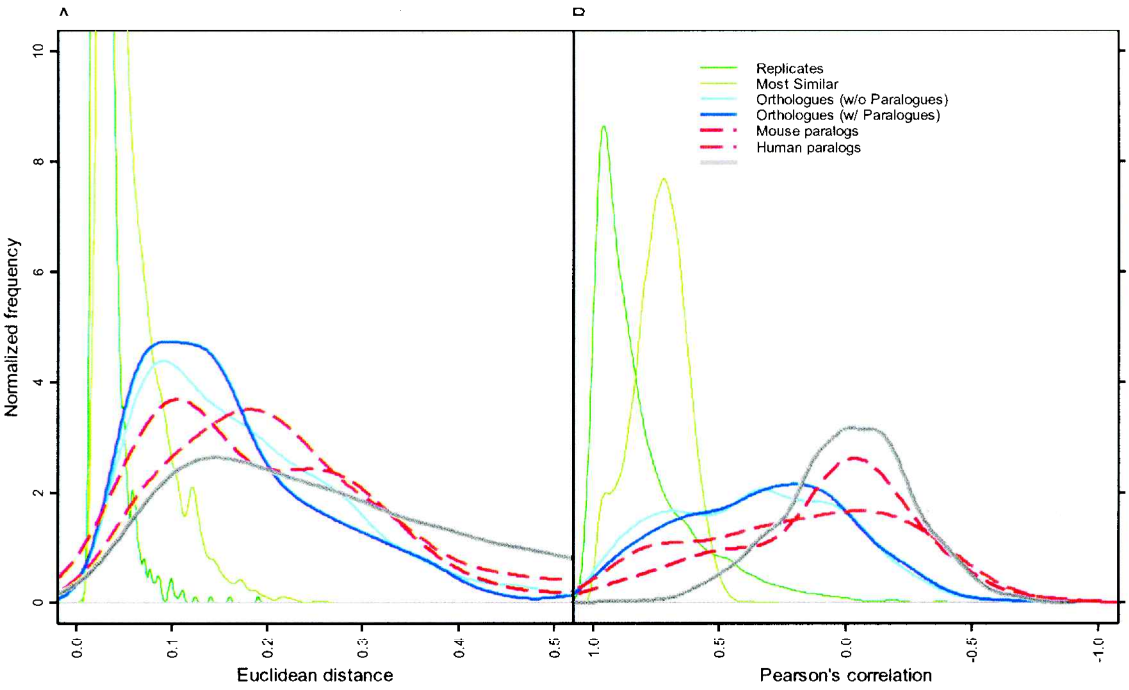

FIG. 1. Distribution of the normalized Euclidean distances (**A**) and Pearson's correlation (**B**) between the expression profiles. The curves were smoothed according to density estimation using R (www.r-project.org).

alogous pairs, respectively. Both values are situated between the distributions of the corresponding orthologs and random pairs.

We found that the patterns of all the distributions in Figure 1A are robust to changes in the distance metric. In addition to the Euclidean distance, we also examined Pearson's correlation and a new "highest-tissue congruence measure." Pearson's correlation produced a similar order of gene pairs (Fig. 1B) with two minor differences: (1) the orthologs with paralogs appear indistinguishable from those without paralogs, whereas the distributions were slightly different with the Euclidean distances (Table 1), and (2) the order of the human and mouse paralogs relative to the other pairs is reversed. As Pearson's correlation identifies linear relationships between the profiles and the Euclidean distances are dependent upon the amplitude of the differences, the two metrics assess different aspects of the data. It is, thus, noteworthy that identical relationships between the distributions are revealed by both metrics.

Our new "highest-tissue congruence measure" is a simple indicator of whether two profiles agree as far as the tissue with the highest expression intensity is concerned. This measure is useful for investigating the possibility that expression divergence occurs only in tissues with low levels of transcription. Again, we are struck by the relative incongruence between orthologous pairs, with less than a third of them being expressed at the highest level in the same tissue in mouse and human (Table 1). Altogether, the order of the replicates, most similar profiles, orthologs, paralogs, and random pairs, is invariant across the three measures which query three different properties of profile-relationships.

Relationships between tissue profiles reveal a dichotomy according to organisms

Our derived human and mouse expression sets are linked by both corresponding tissues and orthologous genes, thus allowing the relationships between tissue profiles to also be examined. Since corresponding human and mouse tissues carry out homologous functions in both organisms, it is expected that the 16 pairs of human–mouse tissues be closely linked in terms of their expression. For example, it is recognized that

TABLE 1. EXPRESSION AND SEQUENCE DIVERGENCE BETWEEN PAIRED GROUPS

	Expression divergence between paired profiles				*Kendal's correlation and p-value between expression and sequence distance*	
	Median of normalized Euclidean distances	*Median of Pearson's correlations*	*Fraction of pairs with the same highest tissue*	*Sequence distance: median of protein sequence evolutionary distances*	*Euclidean distance as expression divergence*	*Pearson's correlation as expression divergence*
Replicates						
Human	0.0258	0.9140	0.7052	0	—	—
Mouse	0.0226	0.8273	0.6185	0	—	—
Human–mouse most similar profiles	0.0416	0.7336	0.7[a], 0.5[b]	∞[c]	—	—
Orthologs						
Without paralogs	0.1548	0.3542	0.3189	0.096	$\tau = 0.04$ $p \leq 0.129$	$\tau = 0.03$ $p \leq 0.171$
With paralogs	0.1422	0.3130	0.2908	0.143	$\tau = 0.04$ $p \leq 0.139$	$\tau = 0.04$ $p \leq 0.109$
Paralogs						
Human	0.2010	0.1331	0.1807	0.697	$\tau = -0.02$ $p \leq 0.837$	$\tau = 0.05$ $p \leq 0.52$
Mouse	0.1620	0.0214	0.1341	0.651	$\tau = -0.02$ $p \leq 0.777$	$\tau = 0.09$ $p \leq 0.23$
Random pairs	0.2461	−0.0186	0.0748	∞[c]	—	—

[a] Best hit detected according to Pearson's correlation.

[b] Best hit detected according the Euclidean distance.

[c] No sequence similarity was detected.

cancerous human tissues are closer in their expression to their corresponding normal tissue than to other tissues (Alon et al., 1999).

Strikingly, the relationships between the 16 human–mouse tissue pairs show a clear and significant dichotomy according to organisms, not corresponding tissues, across the set of orthologous genes (Fig. 2). The dendrogram of tissues clusters into two monophyletic groups the human and mouse tissues, respectively, with a 100% bootstrapping value. Thus, contrary to expectation, any human tissue is more similar in its expression profile to any other human tissue, than to its corresponding tissue in mouse. Furthermore, the close relationships between pairs of tissues is conserved among the human and mouse branches. For example, liver and kidney are sister tissues in both the human and mouse branches, as well as salivary gland and thyroid, and the nervous system tissues (dorsal root ganglion, cerebellum, and amygdala). The overall topologies, while similar, are not identical—probably due to the relatively small set of orthologous genes examined here.

Independent sequence and expression distance between orthologs

We next asked whether or not a correlation exists between protein sequence divergence and expression distance between the orthologs. As Table 1 shows, we found that the degree of divergence in the coding sequence of orthologs is neither a strong nor a significant indicator of divergence at the level of expression

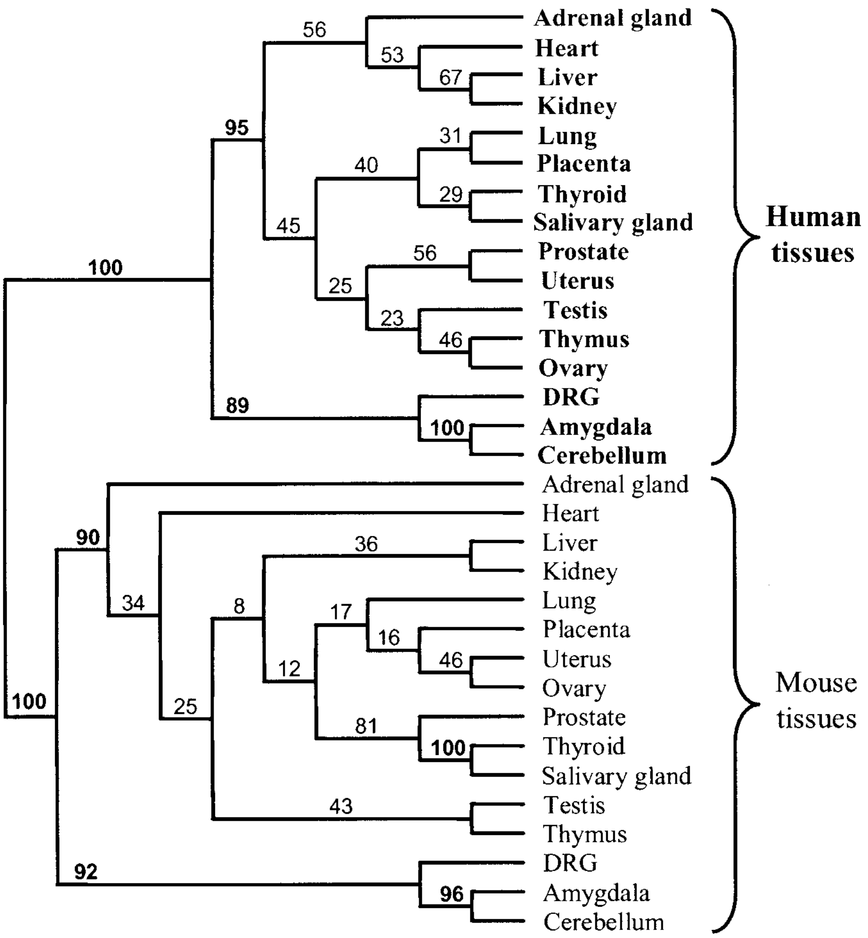

FIG. 2. Expression dendrogram of 32 tissues from human (dark) and mouse (light). For each internal branch, a confidence value was calculated, and robust topologies with bootstrap values of more than 85% are in bold.

profiles. The correlations detected explain less than 2% of the variation in transcription levels regardless of either the metric used (Euclidean distance or Pearson's correlation) or the orthologous set (with or without paralogs) (Table 1). The situation is similar in paralogs. These results are consistent with those found in *S. cerevisiae*, in which sequence similarity between paralogs did not correlate particularly well with expression patterns (Gu et al., 2002; Wagner, 2000). This lack of correlation in conjunction with the distribution of the expression distances suggest a rapid divergence in the orthologous expression profiles between human and mouse, which on *a priori* grounds were expected to behave conservatively. Moreover, the orthologs in our study have a mean nonsynonymous to synonymous substitution ratio of 0.16, which indicates that they have not experienced substantial adaptive evolution in the coding region, and were mostly subject to purifying selection.

Differences are evenly distributed among tissues

In order to gain insight into the nature of the differences between the human and mouse ortholog profiles, we analyzed the variation among the 16 tissues. For each tissue, we computed the number of instances in which expression was detected in both human and mouse ("conservative expression") and the number of instances in which expression was only found in one taxon ("divergent expression"). We found that, for any expression threshold ($0 \leq t \leq 1.8$), the variances among the "divergent-expression" tissues were lower, often significantly lower (Bartlett test, $p < 0.05$), than those in the "conservative" instances. The same result was found for the mouse expression profiles with respect to their human expression conservations. In summary, we found that changes between the human and mouse expression profiles do not correspond to mainly changes in one or a few tissues but are roughly evenly distributed across all tissues.

Overlap between human and mouse profiles tend to uncover gene function

As an example of incongruent expression profiles, we selected four orthologous pairs that are known to have a role in the nervous system (Fig. 3). As expected, high expression levels were conserved between human and mouse tissues pertaining to the nervous system (cerebellum, amygdala, and the dorsal root ganglion). We note, however, that transcription was also detected in other tissues but is not conserved between human and mouse. The ENO2 gene produces a neuronal enolase (Oliva et al., 1991) with unexpected expression found in the human uterus. ENO2 expression in human uterus was further substantiated by EST representation (Diehn et al.. 2003). Expression in the mouse uterus was not detected by either method.

The NSF gene product has a role in the regulation of AMPA receptors which are important determinants of synaptic strength (Braithwaite et al., 2002). Although the highest tissues of expression (amygdala) in the two profiles is conserved, the profiles are very distant in terms of their Euclidean distance and Pearson's correlation due to ectopic expression in the human trachea, uterus, testes, salivary gland, prostate, and lung and in mouse placenta (Fig. 3). The GABRD gene product, an ion channel subunit whose function is thought to be neuron specific (Windpassinger et al., 2002), shows conserved human–mouse expression in the amygdala and cerebellum while expression in human liver, placenta, and salivary gland is not conserved in mouse. Finally, the conserved expression pattern of APBB1, which codes for an amyloid beta A4 protein and plays an important role in the pathogenesis of Alzheimer's disease (Chu et al., 1998), is limited to the brain, however, transcription in mouse is widespread across all tissues.

Finally, let us now consider a protein-coding gene whose function is currently unknown. The expression profiles of FLJ13110 in human and mouse (Fig. 3) overlap only in the dorsal root ganglion tissue, suggesting that despite its significant expression in human uterus and testis, FLJ13110's normal function is associated with the nervous system, and has probably nothing to do with either uterus or testis.

DISCUSSION

Our main and most striking finding is that the expression profiles of human–mouse orthologs are far more different than expected, and are, in fact, quite similar to those exhibited by profiles of paralogs or, even worse, by randomly paired expression profiles. A dendrogram relating tissues according to expression surprisingly does not link corresponding tissues between human and mouse. The weak correlation between the divergence of orthologous coding sequences and that of the corresponding expression profiles explains very little of the variation. Furthermore the divergences between the human and mouse profiles are significantly more evenly distributed across tissues than expressions conserved between the two organisms. Interestingly, we found examples of orthologous profiles where the tissues of conserved expression relate to gene function, whereas divergent expression does not.

Several caveats regarding our results are important to consider. First, high-throughput expression data are known to be noisy. Accordingly, we only used those genes for which replicate experiments yielded congruent results. Second, some noise is expected due to intra-organismal variability (Enard et al., 2002). However, since all human and mouse tissues were derived from normal adult organs, we expect this type of noise to be insignificant. Indeed, the original authors (Su et al., 2002) compared orthologous expression under less stringent criteria, and concluded that intra-tissue variability is negligible. Finally, the effects of intraspecific polymorphisms were offset by pooling together mRNAs from several individuals (Su et al., 2002).

As stated previously, the differences between human and mouse expression profiles might be attributed to positive Darwinian selective pressures on gene expression acting in a species-specific manner in the two lineages since their divergence from a common ancestor. However, since one set of orthologous pairs in our study was limited to those genes that did not experience duplication either before or after the divergence between human and mouse, it is expected that both members of an orthologous pair will maintain

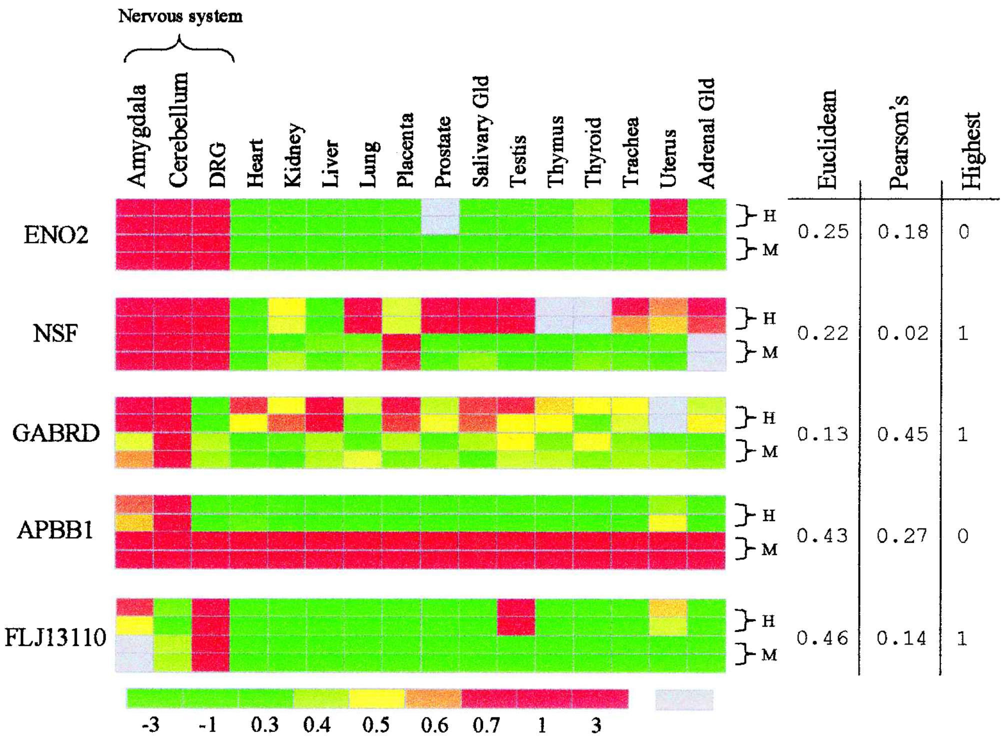

FIG. 3. Divergence in human and mouse orthologous expression profiles. For four ortholog pairs of genes involved in the nervous system and an ortholog pair of genes of unknown function (FLJ13110), the expression in each of the 64 chips is shown according to the 16 tissues (columns), in replicates (rows), for both human (H) and mouse (M). Intensity of expression is shown in a color code whose values are shown in the legend at the bottom. Instances where the tissue replicates are not consistent are indicated in gray. DRG, dorsal root ganglion. To the right of each profile, the Euclidean distance, Pearson's correlation, and highest tissue congruence assessment is given.

the same original function. Considering the number of reproducible differences in the expression patterns between human and mouse, the number of mutations that have been fixed in either promoter sequences or the regulating transcription factors may have been very large. In fact, since we have only analyzed expression in 16 tissues, which is but a small fraction of the complete transcriptional repertoire across conditional, developmental, and temporal axes, the number of fixed changes is likely to be enormous. We have also shown that the differences in expression do not correlate with the sequence divergence for ortholog pairs. Finally, the differences in expression are not particular to any subset of tissues and are fairly evenly distributed across the tissues. These considerations suggest that positive selection cannot explain the divergence between human and mouse expression patterns.

The dichotomy of human and mouse tissues in the dendrogram shown in Figure 2 indicates that expression changes in a particular tissue are correlated with those of other tissues in the same organism. The fact that the dendrogram is so accurately dichotomized is evidence to the large amount of changes that occurred between the expression programs of the two organisms. Such divergence is consistent with a neutral mode of evolution but is not easily explained by adaptation processes.

We are, therefore, left with the second possible explanation, that is, that many of the differences in tissue-expression patterns between human and mouse are selectively neutral. By neutral expression, we refer to a particular spatial, temporal, or conditional pattern of transcription whose genetic contribution (Cheung et al., 2003) has not been selected for, for example, by virtue of its optimal adaptedness to a particular cellular environment, but was fixed in a population through random drift because of its inconsequentiality as far as fitness is concerned. We may even envision a situation in which fixation of an expression variant may occur despite slightly deleterious misregulatory effects. In fact, many examples of neutral expression are known in the literature. The most striking instances concern pseudogenes. For example, human myosin XVBP, which is an unprocessed pseudogene located at 17q25, is highly expressed kidney and stomach tissues (Boger et al., 2001). In a recent study, Rifkin et al. (2003) studied gene expression at the start of metamorphosis in *Drosophila simulans*, *Drosophila yakuba*, and four strains of *Drosophila melanogaster.* They found that, in ~7% of the genes whose expression changes in at least one lineage, the change of expression appears to be neutral.

Our results collectively suggest the existence of widespread neutral expression in the two organisms, that is, expression that confers no selective advantage or disadvantage onto the organism. The proposal that changes in transcription may not affect fitness can be seen as an extension of the neutral mutation theory of molecular evolution (Kimura, 1983). The neutral theory is in essence a theory about the fitness relevance of mutations. Our results can be explained by assuming that a genetic mutation causing ectopic expression of a gene may not be sufficiently deleterious to be eliminated by purifying selection, and may, thus, be fixed in the population by random drift. In other words, we claim that many mutations affecting gene expression may be neutral.

A fundamental assumption of gene-expression studies is that the location and timing of expression can teach us about the function of a gene (Bassett et al., 1999; Chu et al., 1998; DeRisi et al., 1997). Our findings concerning neutral expression imply that each expression profile is an "overestimate" of functionality, and that information of where the gene is expressed cannot be easily translated into knowledge of what the gene does. Given this state of affairs, distinguishing functional expression from neutral expression could be accomplished by comparing the expression profiles of orthologous genes and identifying functional overlaps between the two (as in Fig. 3). Such a method is analogous to detecting functionally related residues in protein sequences by identifying conserved residues in a multiple sequence alignment.

ACKNOWLEDGMENTS

We thank Doron Lancet, Yitzhak Pilpel, Yoav Gilad, Shmuel Pietrokovsky Adnan Derti, Carlos J. Camacho, Eitan Rubin, and Giddy Landan for fruitful discussions and valuable comments. I.Y. and R.O. wish to particularly acknowledge the support and encouragement of Chaim Kahana. We also thank John Hogenesch for access to the primary data. I.Y. is a Koshland Scholar.

REFERENCES

ALON, U., BARKAI, N., NOTTERMAN, D.A., et al. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci USA **96,** 6745–6750.

ARBEITMAN, M.N., FURLONG, E.E., IMAM, F., et al. (2002). Gene expression during the life cycle of *Drosophila melanogaster.* Science **297,** 2270–2275.

BASSETT, D.E., Jr., EISEN, M.B., and BOGUSKI, M.S. (1999). Gene expression informatics—it's all in your mine. Nat Genet **21,** 51–55.

BOGER, E.T., SELLERS, J.R., and FRIEDMAN, T.B. (2001). Human myosin XVBP is a transcribed pseudogene. J Muscle Res Cell Motil **22,** 477–483.

BOLSTAD, B.M., IRIZARRY, R.A., ASTRAND, M., et al. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics **19,** 185–193.

BRAITHWAITE, S.P., XIA, H., and MALENKA, R.C. (2002). Differential roles for NSF and GRIP/ABP in AMPA receptor cycling. Proc Natl Acad Sci USA **99,** 7096–7101.

CHEUNG, V.G., CONLIN, L.K., WEBER, T.M., et al. (2003). Natural variation in human gene expression assessed in lymphoblastoid cells. Nat Genet **33,** 422–425.

CHU, S., DERISI, J., EISEN, M., et al. (1998). The transcriptional program of sporulation in budding yeast. Science **282,** 699–705.

DERISI, J.L., IYER, V.R., and BROWN, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science **278,** 680–686.

DIEHN, M., SHERLOCK, G., BINKLEY, G., et al. (2003). SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data. Nucleic Acids Res **31,** 219–223.

ENARD, W., KHAITOVICH, P., KLOSE, J., et al. (2002). Intra- and interspecific variation in primate gene expression patterns. Science **296,** 340–343.

FELSENSTEIN, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39,** 783–791.

FORCE, A., LYNCH, M., PICKETT, F.B., et al. (1999). Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151,** 1531–1545.

GOLUB, T.R., SLONIM, D.K., TAMAYO, P., et al. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science **286,** 531–537.

GRAUR, D., and LI, W.-H. 2000. *Fundamentals of Molecular Evolution* (Sinauer, Sunderland, MA).

GU, Z., NICOLAE, D., LU, H.H., et al. (2002). Rapid divergence in expression between duplicate genes inferred from microarray data. Trends Genet **18,** 609–613.

IRIZARRY, R.A., HOBBS, B., COLLIN, F., et al. (2003). Exploration, normalization, and summaries of high-density oligonucleotide array probe level data. Biostatistics **4,** 249–264.

KIMURA, M. (1983). *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge).

KOONIN, E.V. (2001). An apology for orthologs—or brave new memes [Comment]. Genome Biol **2,** 1005.

LANDER, E.S. LINTON, L.M. BIRREN, B., et al. (2001). Initial sequencing and analysis of the human genome. Nature **409,** 860–921.

LYNCH, M., and CONERY, J.S. (2000). The evolutionary fate and consequences of duplicate genes. Science **290,** 1151–1155.

MARGUSH, T., and McMORRIS, F.R. (1981). Consensus n-trees. Bull Math Biol **43,** 239–244.

OLIVA, D., CALI, L., FEO, S., et al. (1991). Complete structure of the human gene encoding neuron-specific enolase. Genomics **10,** 157–165.

PRUITT, K.D., and MAGLOTT, D.R. (2001). RefSeq and LocusLink: NCBI gene-centered resources. Nucleic Acids Res **29,** 137–140.

REMM, M., STORM, C.E., and SONNHAMMER, E.L. (2001). Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J Mol Biol **314,** 1041–1052.

RIFKIN, S.A., KIM, J., and WHITE, K.P. (2003). Evolution of gene expression in the Drosophila melanogaster subgroup. Nat Genet **33,** 138–144.

SPRINGER, M.S., MURPHY, W.J., EIZIRIK, E., et al. (2003). Placental mammal diversification and the Cretaceous-Tertiary boundary. Proc Natl Acad Sci USA **100,** 1056–1061.

SU, A.I., COOKE, M.P., CHING, K.A., et al. (2002). Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci USA **99,** 4465–4470.

WAGNER, A. (2000). Decoupled evolution of coding region and mRNA expression patterns after gene duplication: implications for the neutralist-selectionist debate. Proc Natl Acad Sci USA **97,** 6579–6584.

WAGNER, A. (2002). Asymmetric functional divergence of duplicate genes in yeast. Mol Biol Evol **19,** 1760–1768.

WATERSTON, R.H., LINDBLAD-TOH, K., BIRNEY, E., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature **420,** 520–562.

WINDPASSINGER, C., KROISEL, P.M., WAGNER, K., et al. (2002). The human gamma-aminobutyric acid A receptor delta (GABRD) gene: molecular characterisation and tissue-specific expression. Gene **292,** 25–31.

Address reprint requests to:
Dr. Itai Yanai
Department of Molecular Genetics
Weizmann Institute of Science
Rehovot 76100, Israel

E-mail: itai.yanai@weizmann.ac.il