DOI: 10.1534/genetics.108.099960

Note

Segmental Duplications Contribute to Gene Expression Differences Between Humans and Chimpanzees

Ran Blekhman,* Alicia Oshlack† and Yoav Gilad*,[1]

**Department of Human Genetics, University of Chicago, Chicago, Illinois 60637 and †Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia 3052*

Manuscript received December 19, 2008
Accepted for publication March 26, 2009

ABSTRACT

In addition to specific changes in *cis*- and *trans*-regulatory elements, structural changes in the genome are hypothesized to underlie a large number of differences in gene expression between species. Accordingly, we show that species-specific segmental duplications are enriched with genes that are differentially expressed between humans and chimpanzees.

CHANGES in gene regulation have likely played an important role in evolution, including in primates (Britten and Davidson 1971; King and Wilson 1975; Carroll *et al.* 2001; Cresko *et al.* 2004; Gilad *et al.* 2006b; Carroll 2008). In addition to changes in *cis*- and *trans*-regulatory elements, a possible mechanism that might explain differences in gene regulation between species may be structural changes in the genome, such as chromosomal rearrangements, segmental duplications, and copy number variation (*e.g.*, Haberer *et al.* 2004; Huminiecki and Wolfe 2004; Teichmann and Babu 2004; Force *et al.* 2005). In primates, some measure of support for this idea was found in the observation that human-specific large-scale chromosomal rearrangements are slightly, but significantly, enriched with genes that are differentially expressed between humans and chimpanzees (Khaitovich *et al.* 2004; Blekhman *et al.* 2008).

In this context, it is interesting to investigate the contribution of smaller-scale structural genomic differences, such as segmental duplications (Bailey *et al.* 2002; She *et al.* 2006), to differences in gene expression between humans and chimpanzees. Previous microarray studies reported that duplicated genes, in either human or chimpanzee, tend to be highly expressed in the species in which the duplication has occurred (Khaitovich *et al.* 2004; Cheng *et al.* 2005). This observation probably reflects the fact that, unless specific measures are taken, duplicated genes are expected to cross-hybridize to the same probes and therefore their expression level may appear elevated. In that sense, previous observations cannot exclude a technical, rather than a biological, explanation for the enrichment of differentially expressed genes in segmental duplications. Moreover, previous studies used multispecies expression data that were collected using a single-species array. As a result, previous estimates of gene expression differences between species may be confounded by the effect of sequence mismatches on hybridization intensity (Gilad *et al.* 2005, 2006a; Sartor *et al.* 2006).

To study the effect of segmental duplications on the evolution of gene regulation in human and chimpanzee, we used previously published gene expression data from a genomewide multispecies array (Blekhman *et al.* 2008). Gene expression data were collected for 18,109 genes, from three tissues (liver, kidney, and heart), using 18 samples from each species. Genes that are differentially expressed between species were identified using likelihood-ratio tests in the framework of nested mixed linear models (Blekhman *et al.* 2008).

Using a data set of human and chimpanzee segmental duplications, we identified genes located within segmental duplications in one or both species (see supporting information, File S1 and Figure S1). Our approach was to compare estimates of interspecies gene expression differences between genes that are not associated with any duplication and genes within a segmental duplication in one or both species. Using this approach, we found that species-specific segmental duplications are enriched with genes that are differentially expressed between species, regardless of the tissue ($P = 2.4 \times 10^{-3}$,

Supporting information is available online at http://www.genetics.org/cgi/content/full/genetics.108.099960/DC1.

[1]*Corresponding author:* Department of Human Genetics, University of Chicago, Chicago, IL 60637. E-mail: gilad@uchicago.edu

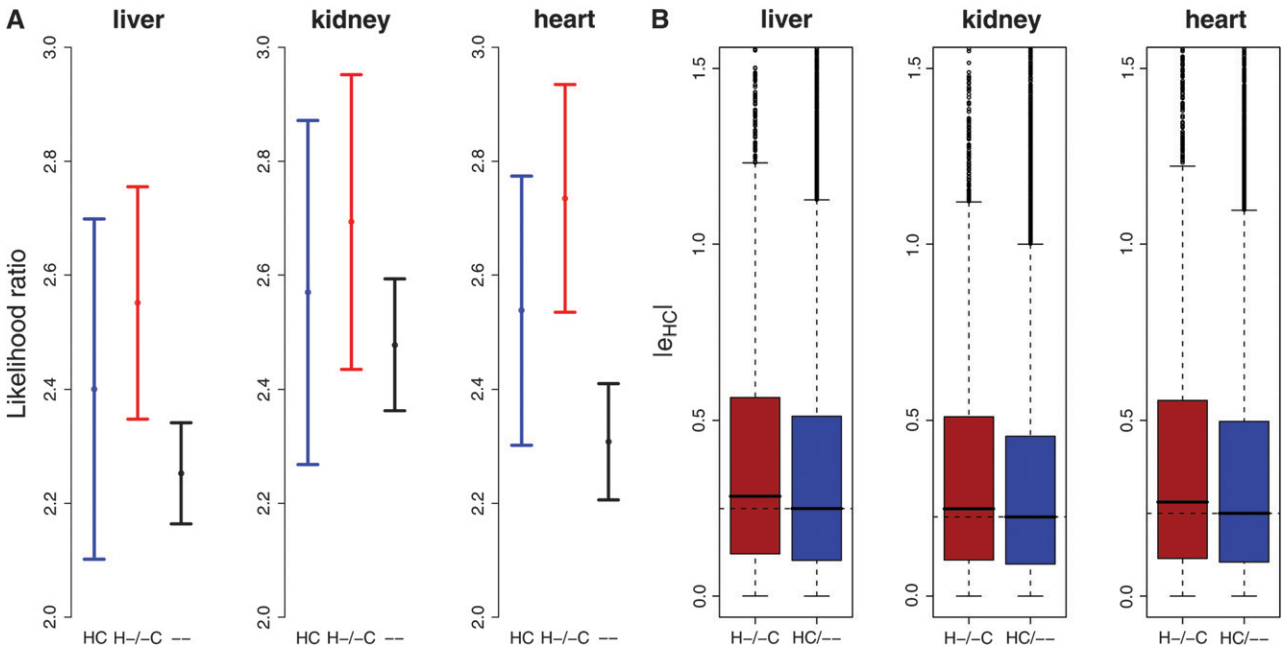

FIGURE 1.—Expression divergence is associated with segmental duplications. (A) Medians of the likelihood-ratio values for testing differential expression between human and chimpanzee. Black, 11,822 genes not associated with duplications (–); blue, 1715 genes associated with duplications in both species (HC); red, 3084 genes associated with either human-specific (H-) or chimpanzee-specific (-C) duplications. The error bars are 95% confidence intervals calculated using bootstrapping (1000 repetitions). See File S1 for more information on the statistical analyses. (B) Box plots of estimates of absolute log fold change in gene expression between the species for liver, kidney, and heart. These estimates were generated using linear models for each species (see File S1); the difference in expression level was estimated as $|e_{HC}| = |\mu_h - \mu_c|$, which is the absolute value of the difference between the log expression level estimates of humans and chimpanzees. The difference between the two distributions is significant in all tissues ($P < 10^{-3}$, using a permutation test on the difference in medians).

$P = 0.057$, and $P = 5 \times 10^{-4}$ in liver, kidney, and heart, respectively, by a permutation test on the medians; see File S1 and Figure 1A). Moreover, genes that are within species-specific segmental duplications (*i.e.*, the duplicated genes) show significantly higher absolute fold difference in expression level between human and chimpanzee compared with genes that are not associated with duplications ($P < 10^{-3}$ in all tissues; Figure 1B and Figure S5).

A possible explanation for the observation that species-specific segmental duplications are enriched with genes that are differentially expressed between humans and chimpanzees is cross-hybridization. For example, if there are more copies of gene *A* in the human genome compared to the chimpanzee genome (*i.e.*, gene *A* is within a human-specific duplication), one might expect mRNA transcribed from all copies of gene *A* to cross-hybridize to the same probe set on the array, resulting in an apparent elevated expression level of gene *A* in humans compared with chimpanzees.

While increased dosage is an intuitive mechanism by which duplications affect gene regulation, we also wanted to address the possibility that duplications may affect gene regulation independently of simple dosage effects—perhaps due to changes in the proximal regulatory elements that affect the expression of duplicated genes. To do so, we looked for evidence of cross-hybridization by plotting the difference between the values of $e_{HC} = \mu_h - \mu_c$ (*i.e.*, the difference in log expression level between humans and chimpanzees) across the three categories of genes mentioned above. If cross-hybridization underlies most interspecies differences in expression for genes within species-specific segmental duplication, one would expect genes within human-specific duplications to have $e_{HC} > 0$ and genes within chimpanzee-specific duplications to have $e_{HC} < 0$.

Importantly, we do not find a trend toward elevated expression levels for genes within species-specific duplications. The proportions of genes with elevated expression level in the species with the duplication are 0.49, 0.48, and 0.49, for genes in liver, kidney, and heart, respectively (Figure S2). We note that some genes within duplications are mapped to regions that are also variable in copy number between individuals. Thus, such genes may not in fact be duplicated in the individuals considered in this study and therefore would not be expected to show elevated expression level in the species with the annotated duplication. However, our observations are virtually unchanged when we exclude genes within segmental duplication that are known to overlap copy number variable regions in humans and chimpanzees (PERRY *et al.* 2008) (Figure S6, Figure S7, and Figure S8).

Thus, cross-hybridization is unlikely to explain the observed association between species-specific segmental

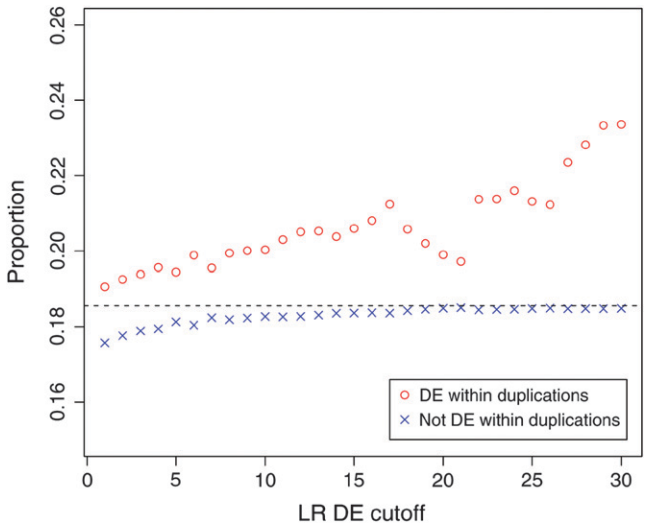

Figure 2.—The proportion of genes within species-specific segmental duplications (*y*-axis) is plotted at different likelihood-ratio cutoffs (*x*-axis) for classifying genes as differentially expressed (DE) between species, using the liver expression data (red circles). The proportion of genes within species-specific segmental duplications among genes that are not classified as differentially expressed (based on the different cutoffs) is plotted with blue x's. The overall proportion of genes within species-specific duplications is shown by the dashed horizontal line. Similar plots using the kidney and heart expression data are available as Figure S3 and Figure S4, respectively.

duplications and interspecies gene expression differences. In other words, our observations cannot be explained by a simple dosage effect as a result of gene duplications. Instead, it is reasonable to assume that orthologous genes within species-specific duplications are regulated by a different set of elements, as their proximal genomic environment has changed (Haberer *et al.* 2004; Force *et al.* 2005; Conrad and Antonarakis 2007). In such cases, duplications may have resulted in the introduction of proximal enhancers, repressors, or boundary regulatory elements, which can result in a shift of expression level in both directions.

Next we wanted to assess the contribution of segmental duplications to the overall differences in gene regulation between humans and chimpanzees. To do so, we calculated the proportion of genes within species-specific duplications among genes that are differentially expressed between the species. Because such a comparison depends on the statistical cutoff chosen to classify genes as differentially expressed, we examined a wide range of possible cutoffs. Interestingly, regardless of the cutoff chosen (in all tissues), the proportion of genes in segmental duplications is always higher for genes that are classified as differentially expressed between humans and chimpanzees compared with genes that are classified as not differentially expressed between the species (Figure 2 and Table S3). Moreover, the proportion of genes within species-specific duplications is higher when more stringent statistical cutoffs are used to classify genes as differentially expressed between the species. Thus, our analysis suggests that segmental duplications might explain as least 2%, but perhaps as much as 8%, of differences in gene expression between humans and chimpanzees (in the three adult tissues studied here; Figure 2, Figure S3, and Figure S4).

Finally, we examined the known functions of genes within species-specific segmental duplication (see File S1). We found that human-specific duplications are somewhat enriched with transcription factors and genes in metabolic pathways compared with chimpanzee-specific duplications (Table S1). This enrichment becomes much more pronounced when we also condition on observing a difference in gene expression levels between the species. Indeed, transcription factors and genes in metabolic pathways are the top gene ontology categories that are overrepresented among genes that are differentially expressed between the species and are within human-specific duplications (Table S2). This result is consistent with our previous observations of overrepresentation of transcription factors and metabolic genes among genes whose regulation likely evolves under directional selection exclusively in humans (Gilad *et al.* 2006b; Blekhman *et al.* 2008), although, importantly, the genes that underlie the two observations are not the same.

In summary, our results provide support for a role of segmental duplications in shaping the evolution of gene regulation. Further, our observations suggest that genes within species-specific duplications are more likely to have either reduced or elevated expression levels compared with genes not associated with duplications. A possible explanation may be that the expression levels of genes within species-specific segmental duplications are affected by different proximal *cis*-regulatory elements compared with those of orthologous genes in their original genomic location.

We thank G. Perry, L. Barreiro, R. Bainer, and C. Cain for helpful discussions and J. Marioni, Z. Gauhar, and N. Zeus for comments on the manuscript. This work was supported by the Sloan Foundation and National Institutes of Health grant GM077959 to Y.G.

LITERATURE CITED

Bailey, J. A., Z. Gu, R. A. Clark, K. Reinert, R. V. Samonte *et al.*, 2002 Recent segmental duplications in the human genome. Science **297:** 1003–1007.

Blekhman, R., A. Oshlack, A. E. Chabot, G. K. Smyth and Y. Gilad, 2008 Gene regulation in primates evolves under tissue-specific selection pressures. PLoS Genet. **4:** e1000271.

Britten, R. J., and E. H. Davidson, 1971 Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Q. Rev. Biol. **46:** 111–138.

Carroll, S. B., 2008 Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell **134:** 25–36.

Carroll, S. B., J. K. Grenier and S. D. Weatherbee, 2001 *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Blackwell Scientific, Malden, MA.

Cheng, Z., M. Ventura, X. She, P. Khaitovich, T. Graves *et al.*, 2005 A genome-wide comparison of recent chimpanzee and human segmental duplications. Nature **437:** 88–93.

Conrad, B., and S. E. Antonarakis, 2007 Gene duplication: a drive for phenotypic diversity and cause of human disease. Annu. Rev. Genomics Hum. Genet. **8:** 17–35.

Cresko, W. A., A. Amores, C. Wilson, J. Murphy, M. Currey *et al.*, 2004 Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. Proc. Natl. Acad. Sci. USA **101:** 6050–6055.

Force, A., W. A. Cresko, F. B. Pickett, S. R. Proulx, C. Amemiya *et al.*, 2005 The origin of subfunctions and modular gene regulation. Genetics **170:** 433–446.

Gilad, Y., S. A. Rifkin, P. Bertone, M. Gerstein and K. P. White, 2005 Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles. Genome Res. **15:** 674–680.

Gilad, Y., A. Oshlack and S. A. Rifkin, 2006a Natural selection on gene expression. Trends Genet. **22:** 456–461.

Gilad, Y., A. Oshlack, G. K. Smyth, T. P. Speed and K. P. White, 2006b Expression profiling in primates reveals a rapid evolution of human transcription factors. Nature **440:** 242–245.

Haberer, G., T. Hindemitt, B. C. Meyers and K. F. Mayer, 2004 Transcriptional similarities, dissimilarities, and conservation of cis-elements in duplicated genes of Arabidopsis. Plant Physiol. **136:** 3009–3022.

Huminiecki, L., and K. H. Wolfe, 2004 Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse. Genome Res. **14:** 1870–1879.

Khaitovich, P., B. Muetzel, X. She, M. Lachmann, I. Hellmann *et al.*, 2004 Regional patterns of gene expression in human and chimpanzee brains. Genome Res. **14:** 1462–1473.

King, M. C., and A. C. Wilson, 1975 Evolution at two levels in humans and chimpanzees. Science **188:** 107–116.

Perry, G. H., F. Yang, T. Marques-Bonet, C. Murphy, T. Fitzgerald *et al.*, 2008 Copy number variation and evolution in humans and chimpanzees. Genome Res. **18:** 1698–1710.

Sartor, M. A., A. M. Zorn, J. A. Schwanekamp, D. Halbleib, S. Karyala *et al.*, 2006 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. Nucleic Acids Res. **34:** 185–200.

She, X., G. Liu, M. Ventura, S. Zhao, D. Misceo *et al.*, 2006 A preliminary comparative analysis of primate segmental duplications shows elevated substitution rates and a great-ape expansion of intrachromosomal duplications. Genome Res. **16:** 576–583.

Teichmann, S. A., and M. M. Babu, 2004 Gene regulatory network growth by duplication. Nat. Genet. **36:** 492–496.

Communicating editor: M. W. Nachman

GENETICS

Supporting Information
http://www.genetics.org/cgi/content/full/genetics.108.099960/DC1

Note

Segmental Duplications Contribute to Gene Expression Differences Between Humans and Chimpanzees

Ran Blekhman, Alicia Oshlack and Yoav Gilad

DOI: 10.1534/genetics.108.099960

FILE S1

Segmental duplications contribute to gene expression differences between humans and chimpanzees

Ran Blekhman[1], Alicia Oshlack[2] and Yoav Gilad[1]

[1]Department of Human Genetics, University of Chicago, Chicago, IL 60637

[2]Walter and Eliza Hall Institute of Medical Research, Parkville, Vic, Australia 3052

Expression data

We used expression data that we previously collected and analyzed (Blekhman et al. 2008). Briefly, the data were collected using a multi-species microarray, containing orthologous probes from three primate species: human, chimpanzee, and rhesus macaque. The array contains probes for 18,109 genes (368,678 probes in total). The data include gene expression estimates from six individuals from three tissues (liver, kidney cortex and heart muscle), from each of the three species. Complete information on sample collection, study design, array hybridizations, low-level analysis, and quality control is available in (Blekhman et al. 2008).

Identifying differentially expressed genes

For all subsequent analyses we excluded probes that did not have corresponding orthologs in all three species (i.e., we only consider probes that have the human, chimpanzee, and rhesus macaque species-specific versions on the array – we refer to these as the "corresponding orthologous probes"). Following this step, we excluded genes that were represented by fewer than three corresponding orthologous probes across all species. Thus, the total number of genes included in all subsequent analyses was 17,231 (95% of genes originally included on the array). Expression estimates were obtained from Blekhman et al. (2008).

To identify genes that are differentially expressed (DE) between human and chimpanzee within a tissue, we used likelihood ratio (LR) tests within the frame work of nested mixed linear models, as previously described (Blekhman et al. 2008). Briefly, we estimated the maximum likelihood of the full model as well as that of a reduced (null) model, in which we assume that the expression level in human and chimpanzee is similar. We then calculated $-2\cdot$(log-likelihood ratio) between the fits of the reduced and full models. We expect genes that deviate from the null (i.e., genes that are truly differentially expressed between human and chimpanzee) to have higher values of this statistic.

Segmental duplications data

Intra-specific segmental duplications are defined as low-copy repeats 1 kb or longer, with at least 90% similarity to another genomic region within the genome (Bailey et al. 2002). Data for human (hg18, March 2006 build) were downloaded from the UCSC genome browser ('genomicSuperDups' table in the 'Segmental Dups' track). Data for chimpanzee (panTro2, March

2006 build) are not yet available on the UCSC database, and so were downloaded from the Segmental Duplications Database at Washington University (http://humanparalogy.gs.washington.edu/pantro2wgac/pantro2wgac.html).

We used the genomic coordinates of the 17,321 genes in human and chimpanzee, for which we had expression data, to find which genes are located within a known segmental duplication. We failed to unambiguously identify the physical location of 700 genes in either the human or (more often) the chimpanzee genome, hence we considered in subsequent analyses data from 16,621 genes.

We classified genes as located within a segmental duplication when any part of the gene sequence overlaps a duplication. Using this approach, we found 2524 genes within segmental duplications in human and 3992 genes within segmental duplications in chimpanzee. Of these, 808 genes are associated with a segmental duplication in humans but not in chimpanzee and 2276 genes are associated with a segmental duplication in chimpanzee but not in human (Figure S1). Thus, 1716 genes are located within segmental duplications in both species, and 3084 genes are located within species-specific segmental duplications.

Importantly, in order to confirm that our results do not depend on the particular cutoff used to identify segmental duplications, we repeated the entire analyses described below with a more stringent cutoff for identifying duplication. In the second analysis we used a cutoff of 94% similarity, which should theoretically enrich our dataset with "younger" duplication events. Our qualitative results remained the same (Figure S5).

Inter-species gene expression differences and segmental duplications

To study the effect of segmental duplications on the evolution of gene regulation in human and chimpanzee we classified genes into three categories: (*i*) genes that are located within segmental duplications in both species (HC), (*ii*) genes that are located within species-specific segmental duplications (either in human (H-) or in chimpanzee (-C)), and (*iii*) genes that are not located within a segmental duplication in either species (--).

As a first step, we compiled a list of likelihood ratio (LR) values, testing the null hypothesis that genes are not differentially expressed between humans and chimpanzees (Figure 1A in the main text). This analysis was done using data from each tissue separately. To test whether the medians of the LR values differ between categories we used a permutation test on D, the difference between pairs of medians. Specifically, we randomly divided all the LR values into two categories while maintaining the original size of each category, and then calculated the medians of the random groups. This permutation was repeated 10,000 times, and each time the difference between the medians of the two randomly selected groups (D_i) was recorded. The test p-value was defined as the number of times where $D_i \geq D$, divided by 10,000.

We also asked about the magnitude of expression difference between the species. To do so, we used the estimated expression levels generated from the linear model for each species (Blekhman et al. 2008). Based on these estimates, we inferred the change in expression level between humans and chimpanzees for each gene as $|e_{HC}| = |\mu_h - \mu_c|$, which is the absolute value

of the inter-species difference in log expression level. We then compared the distribution of these values between the three categories of genes (see Figure 1B in the main text).

Enrichment of GO functional categories

In order to learn more about the possible functions of genes within species-specific duplications, we performed a global "GO analysis". We included all GO categories under 'biological processes' and 'molecular function' using DAVID (http://david.abcc.ncifcrf.gov/). Results for this analysis are shown in Table S1. We then repeated this analysis, considering only genes that are differentially expressed between human and chimpanzee in at least one of the three tissues (Table S2), using a likelihood ratio cutoff of 10 (Blekhman et al. 2008).

We excluded from both tables categories associated with fewer than 5 genes, as well as categories with an uncorrected *P*-value higher than 0.05.

References

Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE (2002) Recent segmental duplications in the human genome. Science 297:1003-1007

Blekhman R, Oshlack A, Chabot AE, Smyth GK, Gilad Y (2008) Gene Regulation in Primates Evolves under Tissue-Specific Selection Pressures. PLoS Genetics 4:e1000271

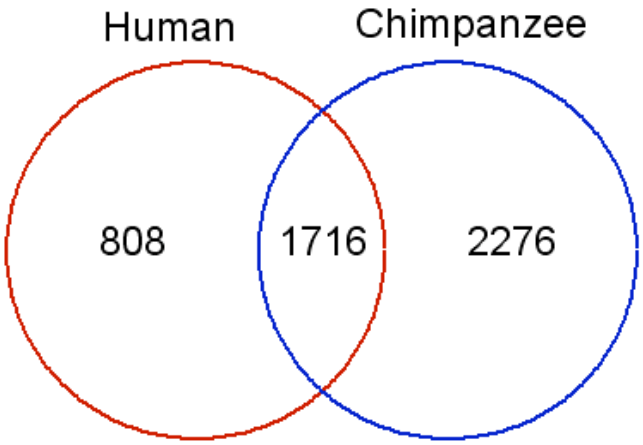

FIGURE S1.—A Venn diagram of the numbers of genes within segmental duplications in human and chimpanzee.

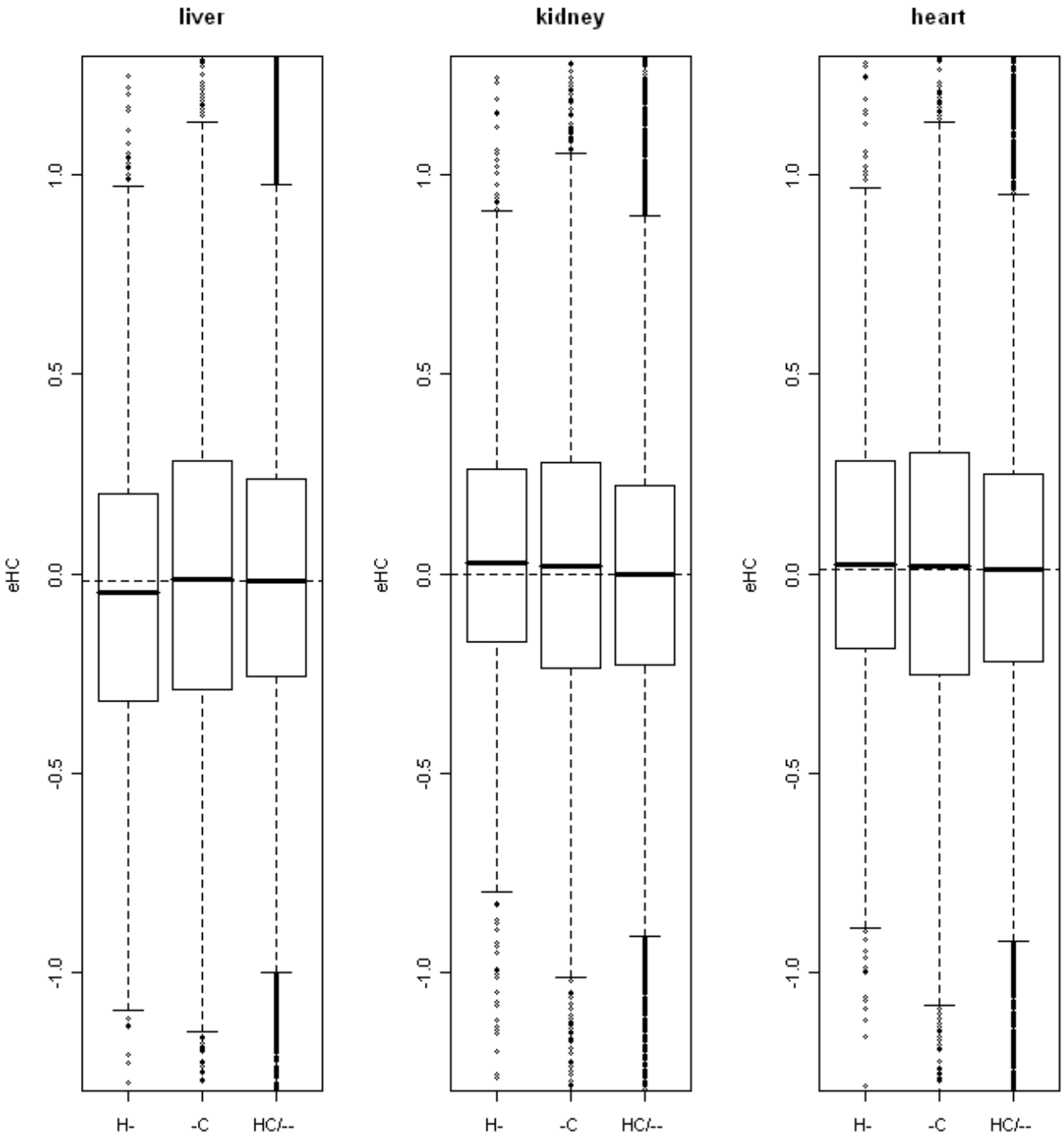

FIGURE S2.—Boxplots showing the distributions of e_{HC} values for genes within human-specific duplications (H-), chimpanzee-specific duplications (-C), and all other genes (HC or --), in liver, kidney, and heart. The horizontal dotted line denotes the median of e_{HC} values in the third group (HC or --). Bars represent the 95 percentiles of each distribution and less than 3% of the extreme data in each distribution are not shown. The complete distributions range from -4.42 to 6.04 in liver; from -3.83 to 4.48 in kidney; and from -4.42 to 5.36 in heart.

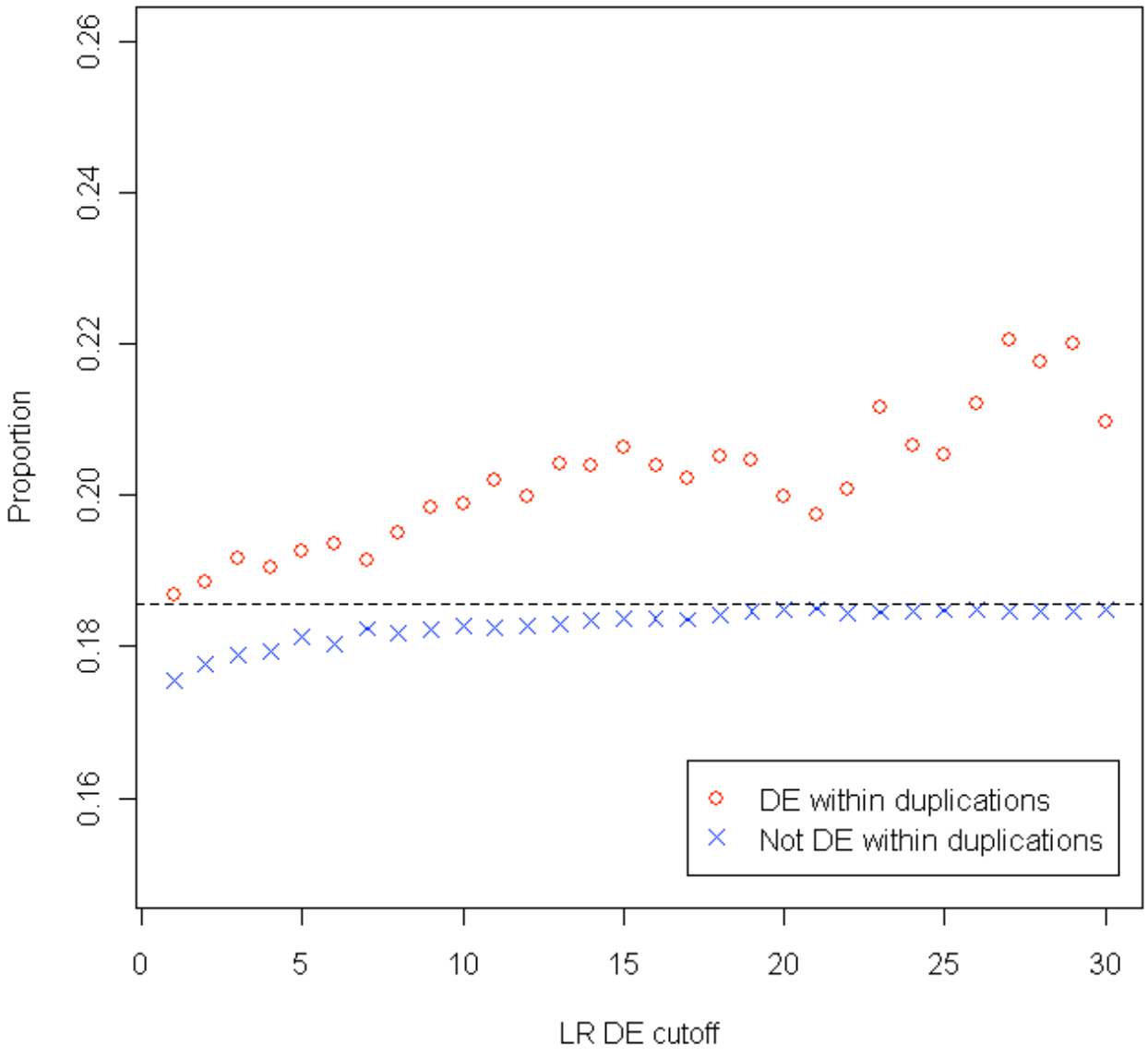

FIGURE S3.—The proportion of genes within species-specific segmental duplications (y-axis), is plotted at different LR cutoffs (x-axis) for classifying genes as differentially expressed between species using the kidney expression data (red circles). The proportion of genes within species-specific segmental duplications among genes that are not classified as differentially expressed (based on the different cutoffs) is plotted in blue crosses. The overall proportion of genes within species-specific duplications is shown by the dashed horizontal line.

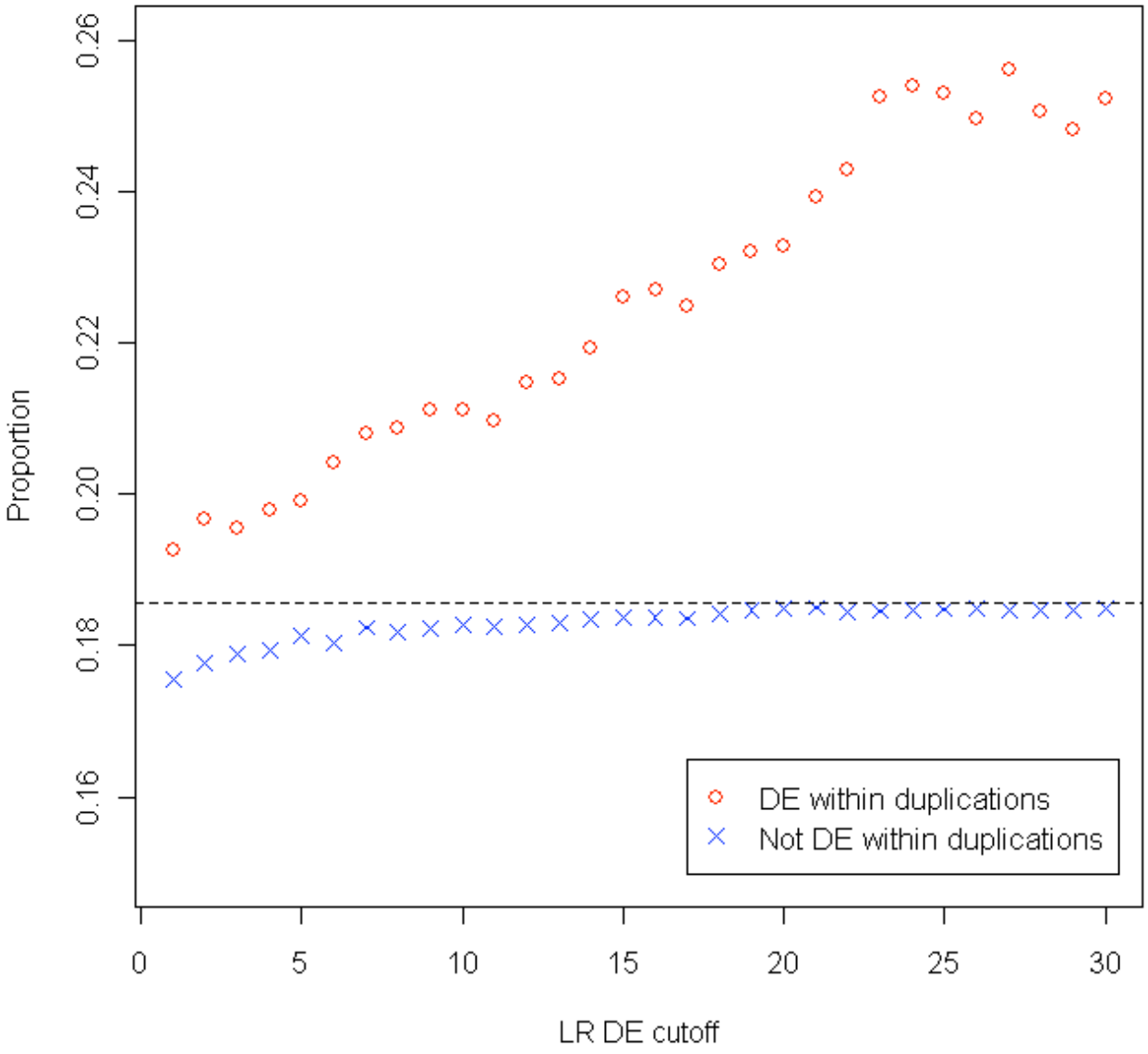

FIGURE S4.—The proportion of genes within species-specific segmental duplications (y-axis), is plotted at different LR cutoffs (x-axis) for classifying genes as differentially expressed between species using the heart expression data (red circles). The proportion of genes within species-specific segmental duplications among genes that are not classified as differentially expressed (based on the different cutoffs) is plotted in blue crosses. The overall proportion of genes within species-specific duplications is shown by the dashed horizontal line.

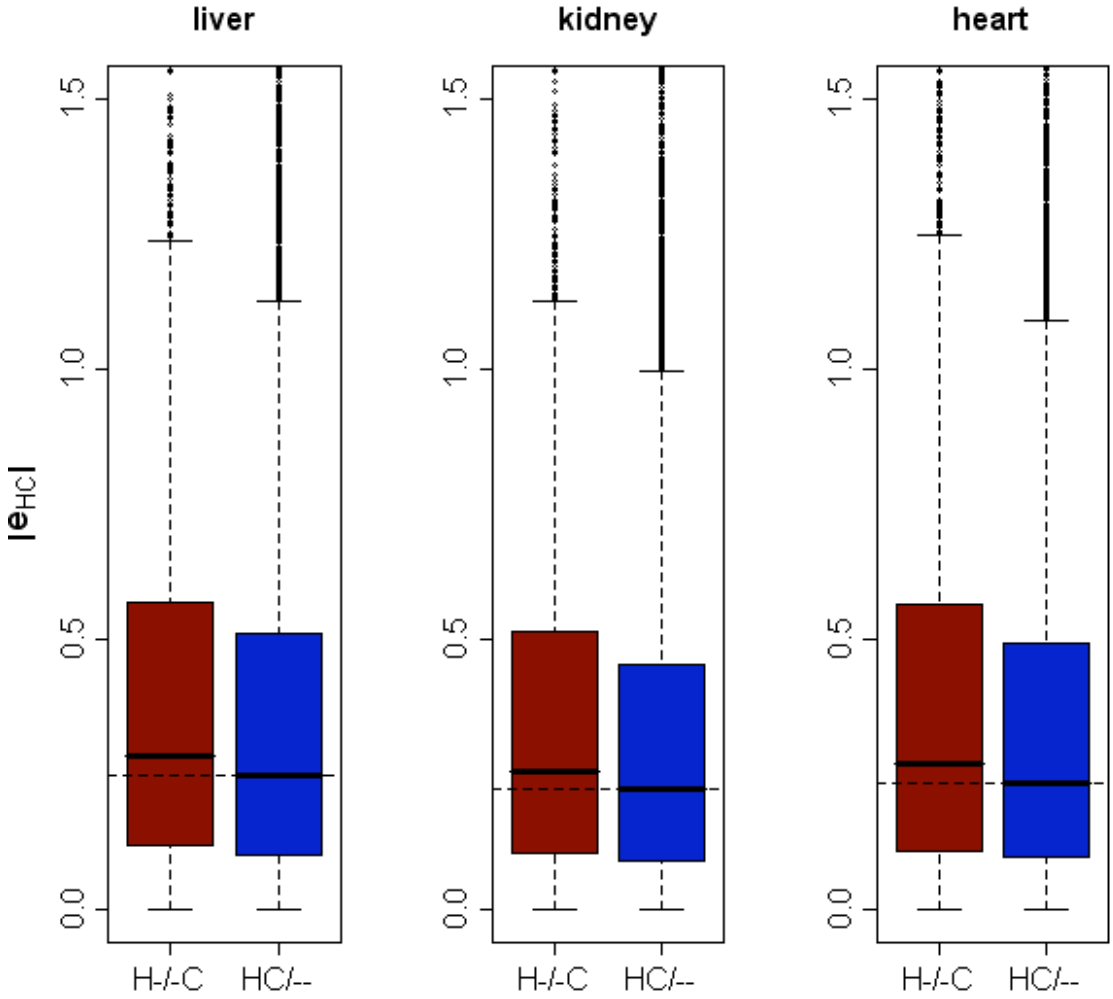

FIGURE S5.—Box plots of estimates of absolute fold change in gene expression between the species for liver, kidney, and heart. These estimates were generated using linear models for each species (see File S1); the difference in log expression level was estimated as $|e_{HC}| = |\mu_h - \mu_c|$, which is the absolute value of the difference between the log expression level estimates of human and chimpanzee. In this analysis, duplications were identified using an alternative identity cutoff of **94%.** Data for genes within species-specific duplications is in red, and for all other genes in blue. The difference between the two distributions is significant in all tissues ($P < 10^{-4}$; using a permutation test on the difference in medians).

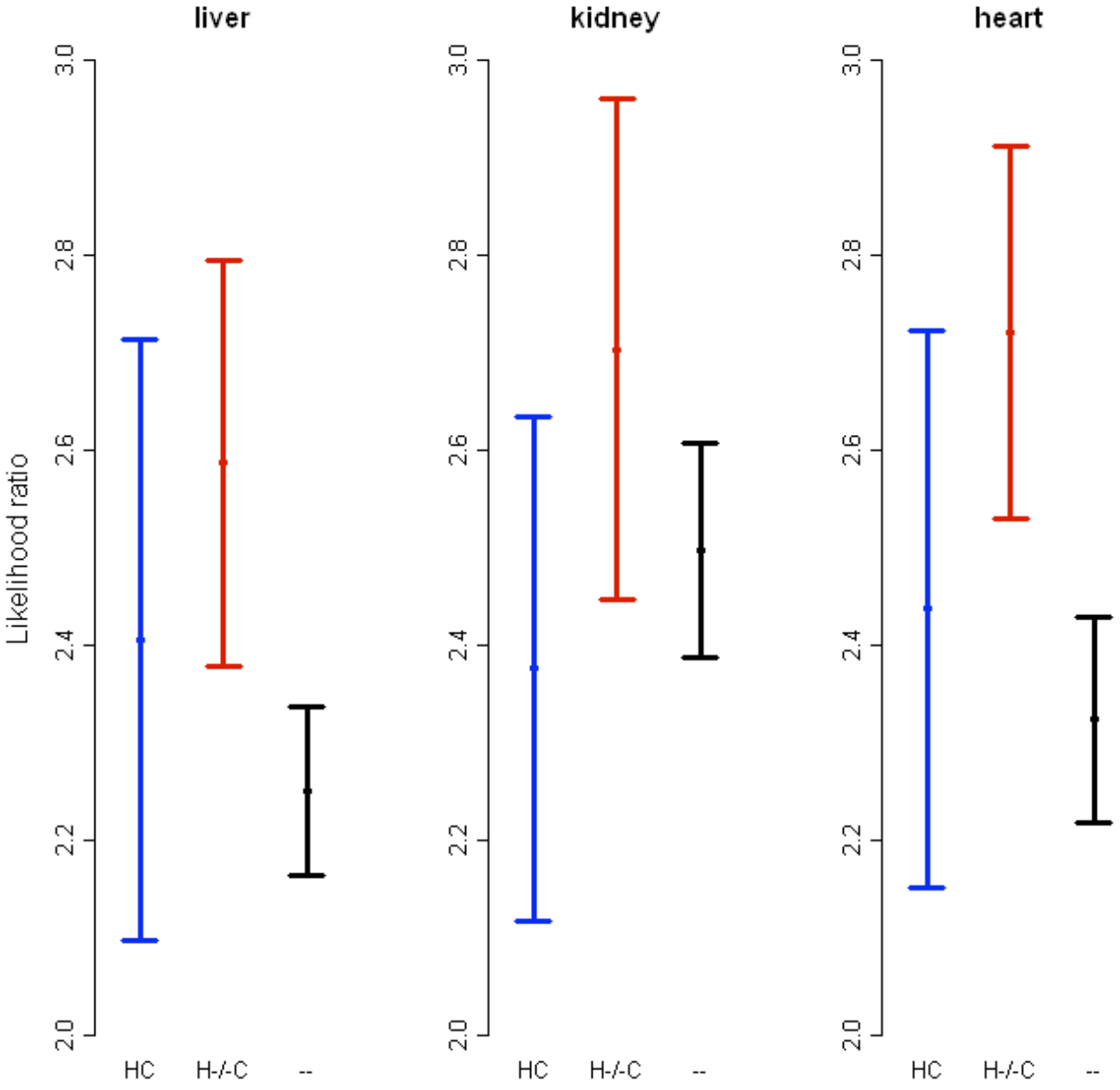

FIGURE S6.—Medians of the likelihood ratio values for testing differential expression between human and chimpanzee. Black: genes not associated with duplications (--). Blue: genes associated with duplications that do not overlap CNVs in both species (HC). Red: genes associated with either human-specific (H-) or chimpanzee-specific (-C) duplications, which do not overlap CNVs in either species. The error bars are 95% confidence intervals calculated using bootstrapping (1000 repetitions).

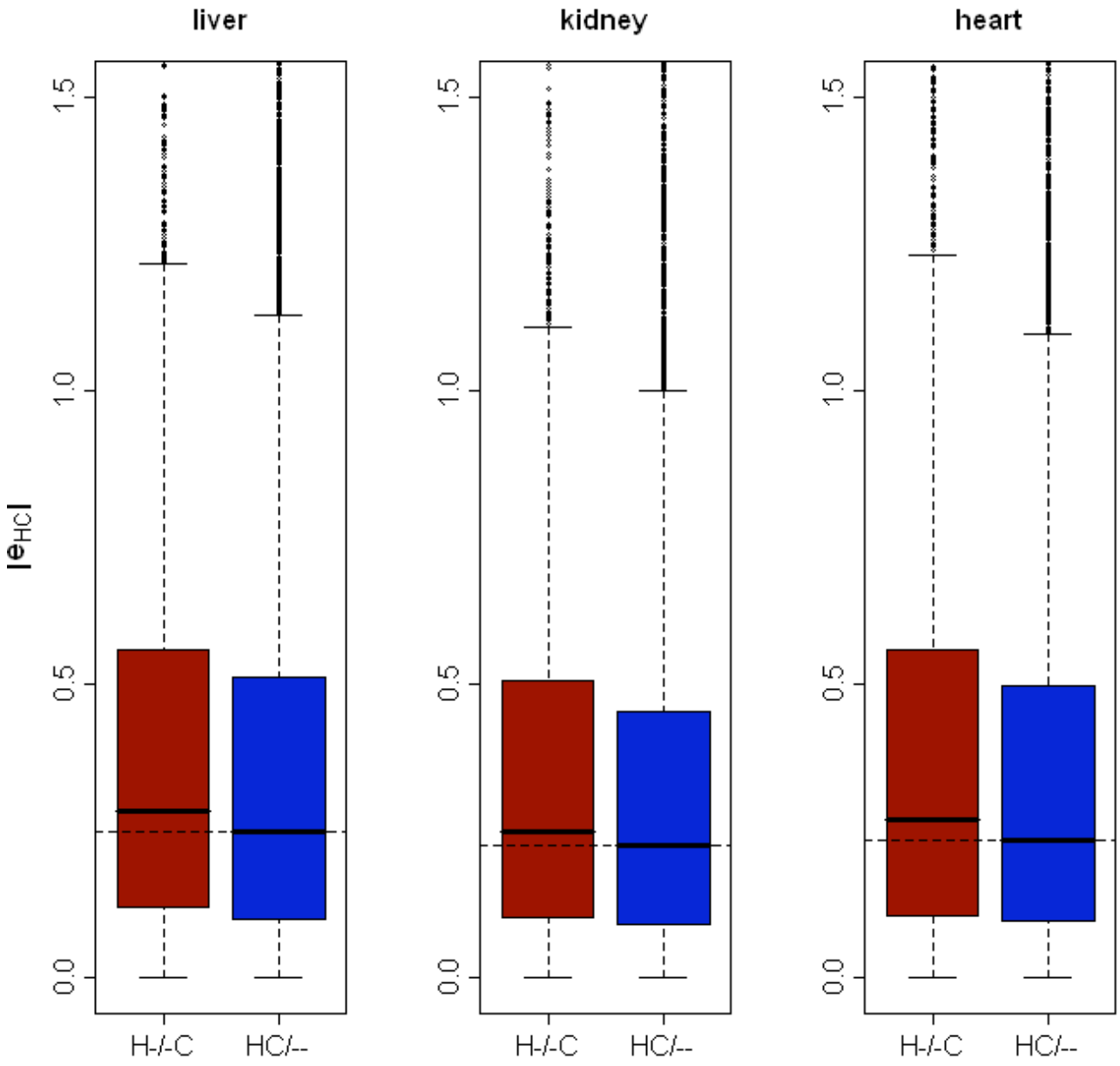

FIGURE S7.—Box plots of estimates of absolute log fold change in gene expression between the species for liver, kidney, and heart. These estimates were generated using linear models for each species (see File S1); the difference in expression level was estimated as $|e_{HC}| = |\mu_h - \mu_c|$, which is the absolute value of the difference between the log expression level estimates of humans and chimpanzees. Blue: genes not associated with duplications (--) as well as genes associated with duplications that do not overlap CNVs in both species (HC). Red: genes associated with either human-specific (H-) or chimpanzee-specific (-C) duplications, which do not overlap CNVs in either species. The difference between the two distributions is significant in all tissues ($P < 10^{-3}$; using a permutation test on the difference in medians).

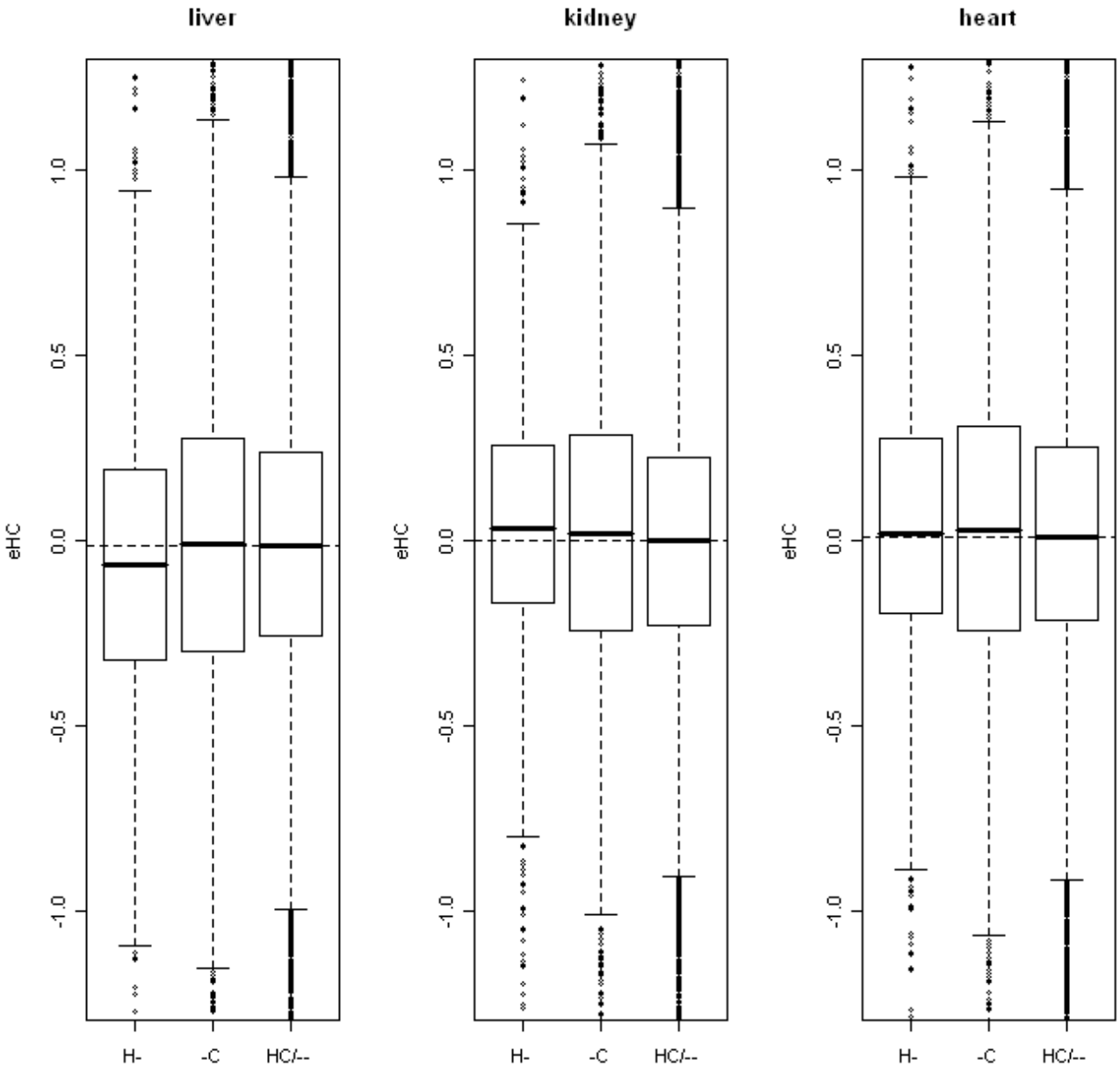

FIGURE S8.—Boxplots showing the distributions of e_{HC} values for genes within human-specific duplications that do not overlap CNVs in humans (H-), chimpanzee-specific duplications that do not overlap CNVs in chimpanzee (-C), and all other genes (HC or --), in liver, kidney, and heart. The horizontal dotted line denotes the median of e_{HC} values in the third group (HC or --). Bars represent the 95 percentiles of each distribution and less than 3% of the extreme data in each distribution are not shown.

TABLE S1

GO terms enriched among 'genes within human-specific duplications' *compared with* 'genes within chimpanzee-specific duplications'.

GO term	Number of human-specific SD genes associated with GO term	Percentage of human-specific SD genes associated with GO term	P-value
GO:0003723~RNA binding	46	5.88%	1.92E-05
GO:0003676~nucleic acid binding	160	20.46%	2.13E-05
GO:0004984~olfactory receptor activity	17	2.17%	5.42E-05
GO:0007608~sensory perception of smell	19	2.43%	6.88E-05
GO:0065004~protein-DNA complex assembly	12	1.53%	3.63E-04
GO:0007606~sensory perception of chemical stimulus	20	2.56%	6.93E-04
GO:0006139~nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	170	21.74%	0.001243
GO:0006333~chromatin assembly or disassembly	11	1.41%	0.002039
GO:0010467~gene expression	144	18.41%	0.002889
GO:0003677~DNA binding	104	13.30%	0.003958
GO:0006334~nucleosome assembly	8	1.02%	0.004148
GO:0031497~chromatin assembly	8	1.02%	0.008892
GO:0022607~cellular component assembly	39	4.99%	0.010151
GO:0006397~mRNA processing	18	2.30%	0.011181
GO:0016070~RNA metabolic process	124	15.86%	0.012323
GO:0006354~RNA elongation	5	0.64%	0.017407
GO:0022618~protein-RNA complex assembly	12	1.53%	0.018163
GO:0019953~sexual reproduction	21	2.69%	0.018248
GO:0006350~transcription	110	14.07%	0.018826
GO:0031323~regulation of cellular metabolic process	118	15.09%	0.020256
GO:0019222~regulation of metabolic process	124	15.86%	0.022046
GO:0010468~regulation of gene expression	112	14.32%	0.023666
GO:0005125~cytokine activity	12	1.53%	0.024868
GO:0016071~mRNA metabolic process	21	2.69%	0.028635
GO:0019219~regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	109	13.94%	0.030038
GO:0044237~cellular metabolic process	323	41.30%	0.030681
GO:0065003~macromolecular complex assembly	35	4.48%	0.032934
GO:0007276~gamete generation	17	2.17%	0.033087
GO:0000003~reproduction	29	3.71%	0.033718
GO:0006351~transcription, DNA-dependent	99	12.66%	0.039702
GO:0032774~RNA biosynthetic process	99	12.66%	0.039702

GO:0003725~double-stranded RNA binding	5	0.64%	0.040758
GO:0001584~rhodopsin-like receptor activity	24	3.07%	0.041592
GO:0006325~establishment and/or maintenance of chromatin architecture	17	2.17%	0.041749
GO:0006323~DNA packaging	17	2.17%	0.041749
GO:0016779~nucleotidyltransferase activity	8	1.02%	0.04287
GO:0045449~regulation of transcription	105	13.43%	0.043015
GO:0040008~regulation of growth	16	2.05%	0.043324
GO:0050896~response to stimulus	114	14.58%	0.049319

Results are ordered by *P*-value, and include only categories with $P<0.05$. GO categories with fewer than 5 genes were not included.

TABLE S2

GO terms enriched among genes that are differentially expressed between humans and chimpanzees, which are located within human-specific duplications, *compared with* genes that are differentially expressed between the species, which are located within chimpanzee-specific duplications.

GO term	Among genes that are DE between humans and chimpanzees		*P*-value
	Number of human-specific SD genes associated with GO term	Percentage human-specific SD genes associated with GO term	
GO:0010467~gene expression	60	18.93%	4.81E-04
GO:0003676~nucleic acid binding	65	20.50%	4.94E-04
GO:0019222~regulation of metabolic process	53	16.72%	6.60E-04
GO:0006139~nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	69	21.77%	0.001230182
GO:0006350~transcription	45	14.20%	0.002495423
GO:0031323~regulation of cellular metabolic process	48	15.14%	0.002712589
GO:0019219~regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	45	14.20%	0.003000729
GO:0042157~lipoprotein metabolic process	6	1.89%	0.004053974
GO:0010468~regulation of gene expression	45	14.20%	0.004282773
GO:0045449~regulation of transcription	43	13.56%	0.005151646
GO:0033036~macromolecule localization	25	7.89%	0.005518004
GO:0016070~RNA metabolic process	50	15.77%	0.007078467
GO:0006351~transcription, DNA-dependent	40	12.62%	0.008041289
GO:0032774~RNA biosynthetic process	40	12.62%	0.008041289
GO:0015931~nucleobase, nucleoside, nucleotide and nucleic acid transport	7	2.21%	0.008592803
GO:0006913~nucleocytoplasmic transport	7	2.21%	0.008592803
GO:0051169~nuclear transport	7	2.21%	0.008592803
GO:0046983~protein dimerization activity	15	4.73%	0.010217501
GO:0003677~DNA binding	40	12.62%	0.010344436
GO:0065004~protein-DNA complex assembly	6	1.89%	0.01136352
GO:0006355~regulation of transcription, DNA-dependent	39	12.30%	0.012250914
GO:0016043~cellular component organization and biogenesis	64	20.19%	0.012510613
GO:0042158~lipoprotein biosynthetic process	5	1.58%	0.014131316
GO:0006497~protein amino acid lipidation	5	1.58%	0.014131316
GO:0000003~reproduction	14	4.42%	0.016448369
GO:0006403~RNA localization	7	2.21%	0.017132283
GO:0019953~sexual reproduction	11	3.47%	0.024002753
GO:0050658~RNA transport	6	1.89%	0.024309125
GO:0050657~nucleic acid transport	6	1.89%	0.024309125
GO:0051236~establishment of RNA localization	6	1.89%	0.024309125
GO:0044238~primary metabolic process	134	42.27%	0.024981521
GO:0030528~transcription regulator activity	26	8.20%	0.027515231
GO:0065003~macromolecular complex assembly	20	6.31%	0.030895767

GO:0004984~olfactory receptor activity	5	1.58%	0.03244891
GO:0006397~mRNA processing	8	2.52%	0.033040608
GO:0045934~negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	10	3.15%	0.034227463
GO:0022607~cellular component assembly	20	6.31%	0.037740318
GO:0044237~cellular metabolic process	133	41.96%	0.038220758
GO:0031324~negative regulation of cellular metabolic process	10	3.15%	0.047461118
GO:0003700~transcription factor activity	18	5.68%	0.047495834
GO:0050794~regulation of cellular process	77	24.29%	0.047965855
GO:0016481~negative regulation of transcription	9	2.84%	0.04889326
GO:0008152~metabolic process	146	46.06%	0.04933884

Results are ordered by *P*-value, and include only categories with $P<0.05$. GO categories with fewer than 5 genes were not included.

TABLE S3

The numbers of genes in each category and statistical cutoff combination that were used to generate Figure 2 and Figures S4 and S5

	Liver				Kidney				Heart			
DE Cutoff	DE	SD & DE	!DE	SD & !DE	DE	SD & DE	not DE	SD & !DE	DE	SD & DE	!DE	SD & !DE
1	10998	2097	5623	987	11077	2070	5544	1014	10897	2099	5724	985
2	8845	1703	7776	1381	9082	1713	7539	1371	8893	1749	7728	1335
3	7402	1435	9219	1649	7724	1481	8897	1603	7536	1474	9085	1610
4	6276	1228	10345	1856	6699	1275	9922	1809	6540	1294	10081	1790
5	5374	1045	11247	2039	5824	1122	10797	1962	5701	1135	10920	1949
6	4616	918	12005	2166	5159	998	11462	2086	5046	1030	11575	2054
7	3994	781	12627	2303	4576	876	12045	2208	4471	930	12150	2154
8	3485	695	13136	2389	4049	790	12572	2294	3939	822	12682	2262
9	3044	609	13577	2475	3608	716	13013	2368	3495	738	13126	2346
10	2691	539	13930	2545	3238	644	13383	2440	3089	652	13532	2432
11	2389	485	14232	2599	2907	587	13714	2497	2763	579	13858	2505
12	2097	430	14524	2654	2649	529	13972	2555	2464	529	14157	2555
13	1856	381	14765	2703	2361	482	14260	2602	2202	474	14419	2610
14	1629	332	14992	2752	2110	430	14511	2654	2003	439	14618	2645
15	1442	297	15179	2787	1885	389	14736	2695	1779	402	14842	2682
16	1264	263	15357	2821	1706	348	14915	2736	1578	358	15043	2726
17	1130	240	15491	2844	1547	313	15074	2771	1410	317	15211	2767
18	1006	207	15615	2877	1405	288	15216	2796	1251	288	15370	2796
19	891	180	15730	2904	1275	261	15346	2823	1112	258	15509	2826
20	794	158	15827	2926	1146	229	15475	2855	988	230	15633	2854
21	715	141	15906	2943	1034	204	15587	2880	890	213	15731	2871
22	627	134	15994	2950	921	185	15700	2899	803	195	15818	2889
23	566	121	16055	2963	818	173	15803	2911	729	184	15892	2900
24	500	108	16121	2976	726	150	15895	2934	650	165	15971	2919
25	441	94	16180	2990	662	136	15959	2948	581	147	16040	2937
26	391	83	16230	3001	599	127	16022	2957	513	128	16108	2956
27	340	76	16281	3008	526	116	16095	2968	457	117	16164	2967
28	298	68	16323	3016	487	106	16134	2978	411	103	16210	2981
29	270	63	16351	3021	432	95	16189	2989	375	93	16246	2991
30	244	57	16377	3027	391	82	16230	3002	333	84	16288	3000

Data for 30 different LR cutoffs are given in each tissue. *DE* is number of inter-species differentially expressed genes; *SD & DE* is the number of inter-species differentially expressed genes that are within species-specific segmental duplications; *!DE* is the number of genes that are not differentially expressed between the species; and *SD & !DE* is the number of genes that are not differentially expressed between the species and are within species-specific segmental duplications.