*Nucleic Acids Research, 2005, Vol. 33, No. 17* ***5659–5666***
*doi:10.1093/nar/gki834*

# Evolutionarily conserved and diverged alternative splicing events show different expression and functional profiles

**Zhengyan Kan, Philip W. Garrett-Engele, Jason M. Johnson and John C. Castle***

Rosetta Inpharmatics LLC, a wholly owned subsidiary of Merck & Co., Inc., 401 Terry Ave N, Seattle, WA 98109, USA

Received April 21, 2005; Revised July 1, 2005; Accepted August 29, 2005

## ABSTRACT

**To better decipher the functional impact of alternative splicing, we classified alternative splicing events in 10 818 pairs of human and mouse genes based on conservation at genome and transcript levels. Expression levels of conserved alternative splices in human and mouse expressed sequence tag databases show strong correlation, indicating that alternative splicing is similarly regulated in both species. A total of 43% (8921) of mouse alternative splices could be found in the human genome but not in human transcripts. Five of eleven tested mouse predictions were observed in human tissues, demonstrating that mouse transcripts provide a valuable resource for identifying alternative splicing events in human genes. Combining gene-specific measures of conserved and diverged alternative splicing with both gene classification based on Gene Ontology (GO) and microarray-determined gene expression in 52 diverse human tissues and cell lines, we found conserved alternative splicing most enriched in brain-expressed signaling pathways. Diverged alternative splicing is more prevalent in testis and cancerous cell line up-regulated processes, including protein biosynthesis, responses to stress and responses to endogenous stimuli. Using conservation as a surrogate for functional significance, these results suggest that alternative splicing plays an important role in enhancing the functional capacity of central nervous systems, while non-functional splicing more frequently occurs in testis and cell lines, possibly as a result of cellular stress and rapid proliferation.**

## INTRODUCTION

Alternative splicing of pre-mRNA is a regulatory mechanism that plays important roles in human physiology and disease (1–3) and is a major contributor to the functional complexity of the human genome (4). Microarray and large-scale computational analysis of expressed sequence tag (EST)/cDNA sequences have shown that alternative splicing affects at least 50% of human and mouse genes (5–7). A number of bioinformatics analyses have been performed to elucidate the functional impact of alternative splicing by identifying the enrichment of alternative splicing events in certain classes of genes or by classifying the functional regions affected (8,9) and, while providing valuable insights, the conclusions are not entirely consistent [see Lareau *et al*. (10)]. Recent evidence emerging from comparative genomics studies indicates that the majority of alternatively spliced cassette exons are not evolutionarily conserved, may result from aberrant splicing and likely do not create functional proteins. Conversely, conserved alternative exons more frequently create transcripts capable of producing functional proteins as they frequently occur outside protein domains and frequently preserve the protein reading frame (11–14).

Hence, classifying alternative splicing forms based on evolutionary conservation helps delineate the functionality of alternative splice forms. Previous cross-species splicing studies have largely focused on identifying characteristics to predict individual exons as alternative and conserved (11–14). In this study, we examine not only cassette exons but also additional types of alternative splicing events in 10 818 pairs of human and mouse genes. We use mouse and human genomic and transcript sequences to classify alternative splices into one of the three classes: conserved, novel and diverged. Our experimental validation of novel alternative splicing events in human genes predicted from mouse transcripts suggests that the extent of conserved alternative splicing is greater than previously estimated. Combining

*To whom correspondence should be addressed. Tel: +1 206 802 6337; Fax: +1 206 802 6411; Email: John_castle@merck.com

gene expression and functional annotation, we identified important distinctions between evolutionarily conserved and diverged alternative splicing events.

## MATERIALS AND METHODS

### Delineating alternative splice patterns

A splice, or a junction, is a unique pair of adjacent splice sites identified through genomic transcript alignment. As previously described (15), we used mouse and human transcripts and genomes to identify alternative splices in human and mouse genes. Genomic contig sequences and the LocusLink database were downloaded from NCBI as of April, 2004 (16) and transcript sequences were taken from UniGene releases Hs.168 and Mm.135 (16). Orthologous gene pairing required reciprocal best matches in the Homologene database (17). For each gene, genomic sequence with flanking 10 kb extensions was extracted based on LocusLink coordinate annotation. Exon and splice coordinates were identified by aligning transcript sequences to the genome of the same species using Sim4 (18). A program called TAP was used to identify consensus splice patterns by clustering genomic alignments (19). Cross-species genomic alignments were generated using EST_GENOME (20), enabling mapping of nucleotide splice patterns between human and mouse (21).

### Classification of splices

Based on a comparison of the resulting nucleotide alignment patterns, all splices were classified as 'conserved', reliably mapped in both genomes and identified in transcripts from both species, 'diverged', no matching splice junction can be found within 10 bases in the transcript sequence based on the cross-species alignment or 'novel', mapped in both genomes but identified in transcripts of only one species. Additionally, if the location of an exon–exon boundary within a transcript, as defined by genomic alignment, differs between human and mouse genomic alignments, the splice is classified as diverged. In Figure 1, splices A and B are mapped to splices a and b from the cross-species alignment. A is conserved as its counterpart, splice a, is found in mouse transcripts whereas B is novel as splice b is not found in mouse transcripts but is found in the mouse genome. Splices C and D are diverged because exon 4 has no homologous match in the mouse genome and therefore no matching splices.

For each splice, the occurrence frequency in transcripts, relative to mutually exclusive splices, measures the relative abundance of the underlying splice form, to the extent that EST/cDNA sequencing constitutes an unbiased sampling of the transcriptome (22). We define 'major' splices as those observed more frequently than mutually exclusive splices or those not having any mutually exclusive splice patterns whereas 'minor' splices are those observed less frequently. We define 'alternative splices' as minor splices from non-RefSeq transcripts and 'reference splices' as major splices from RefSeq transcripts. A total of 1268 gene pairs for which over 40% of the reference exons could not be reliably mapped between species were excluded.

### *Z*-score

The binomial *Z*-score, also known as *Z*-ratio, measures the difference, in units of standard deviation, between the observed frequency and the expected frequency of an event based on binomial probability distribution. The *Z*-score approximating the expression level of a splice using transcript occurrences is:

$$Z = \frac{(k/n) - f \pm (0.5/n)}{\sqrt{[f \times (1-f)]/n}}, \qquad \mathbf{1}$$

where $k$ is the observed number of sequences containing a splice, $n$ is the total number of overlapping sequences, and $f$ is the expectation of 50% (22).

### Quantifying gene-specific alternative splicing

To quantify the amount of alternative splicing for a gene, we counted the number of distinct splices found in the associated transcripts, yielding a 'splice count' for each gene. Gene-to-gene comparisons are confounded by factors such as differences in transcript coverage and in the number of exons. We found that the level of alternative splicing can be approximated by the natural logarithm of spliced transcript coverage (Supplementary Figure S1). Thus, to mitigate differences in

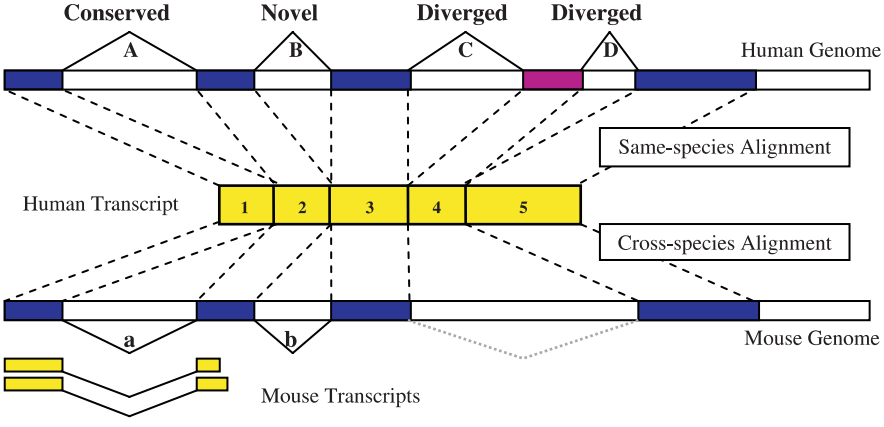


**Figure 1.** Classification of splice patterns based on conservation. Shown above is a human transcript aligned to human and mouse genomic sequences, defining two sets of exons (shaded) and splices (dashed). Each splice corresponds to a splice junction in the transcript sequence and can be matched to a cross-species splice if the positions of two splice junctions on the transcript coincide.

transcript coverage, we normalized each gene splice count by the log of the transcript coverage, resulting in a normalized overall splice score and normalized conserved, diverged, and novel splice scores for each gene.

### Expression and functional classification

We classified tissues and gene functional categories based on conserved and diverged alternative splicing. For each Gene Ontology (GO) Biological Process in the 'generic slim' division, the rate of normalized conserved alternative splicing of the underlying gene sets was used to generate a t-score. For each gene set the t-score is calculated as:

$$t = \frac{k-\mu}{s/\sqrt{n}}, \quad \mathbf{2}$$

where $s$ is the deviation in units of standard error, $k$ is the observed sample mean, $\mu$ is the global population mean, and $n$ is the gene-set size. Associations between individual genes and GO terms were derived from NCBI's LocusLink.

Microarray measurements of 10 000 human genes across 52 human tissues (7) were used to determine gene expression. In this experiment, each gene was monitored by multiple junction-specific probes and we obtained the expression level for the entire gene by taking the mean intensity. Logarithm based 10 ratios for each gene in each tissue were determined by comparison with the average intensity value across all 52 tissues, for the gene. A gene is classified as over-expressed in a tissue if it is expressed above noise levels and the log10 ratio is >2.

### Tissue origin classification of alternative splices

Tissue origins of human ESTs, and thus individual splices, were identified based on gene expression vocabulary (eVOC) annotations of cDNA library sources, as provided by UniGene Library Browser (17) and eVOC (23). Tissue names were taken from level three of the anatomical system hierarchy of eVOC except for four large and heterogeneous categories, 'liver and biliary system', 'genital', 'central nervous system' and 'peripheral nervous system', which were subdivided. This resulted in a total of 30 normal tissue types after excluding diseased and fetal tissues and tissues with very few ESTs. For each splice, we count the transcript sequences containing the splice in each tissue ($k$), in all tissues ($n$), and the percentage of all sequences in each tissue ($f$). The distribution bias of alternative splices in each tissue is also quantified by $Z$-score representing the deviation between the observed frequency of a splice sequence count and the expected frequency, using Equation 1.

### RT–PCR validation

The Qiagen OneStep RT–PCR kit (Qiagen Catalog no. 210212) was used with gene-specific primers. The PCR component involves 35 cycles at 30°C for 94 h, 40°C for 63.5 h and 50–120°C for 72 h depending on the predicted size of the product. All products are resolved on a 2% agarose gel run at 100 V in TAE buffer. A diverse set of 44 human mRNA tissues and cell lines were used: adipose; adrenal gland; bone marrow; brain, cerebellum; brain, frontal lobe; brain, hippocampus; brain, medulla oblongata; brain, pons; brain, putamen; brain, thalamus; colon, descending; colon tumor tissue; fetal brain; fetal kidney; fetal liver; fetal lung; fetal vertebra; heart; HeLa S3; ileum; jejunum; kidney; leukemia promyelocytic (HL-60); liver; lung; lung carcinoma (A549); lymphoma burkitt's (Raji); melanoma (G361); osteosarcoma (MG-63); ovary; pancreas; peripheral leukocytes; pituitary; placenta; prostate; retina; skeletal muscle; skin; spinal cord; stomach; testis; thymus; thyroid; uterus.

## RESULTS

### Genome-wide survey of alternative splices

In our dataset of 10 818 genes and 140 103 distinct human splices, 64% are reference splices whereas 26% are alternative splices (Table 1). When compared with mouse transcript and genome sequences, the vast majority of human alternative splices are either diverged (49%) or novel (44%). Only 7% of alternative splices are conserved in mouse transcripts versus 85% of reference splices, corroborating previous estimates of lower conservation of alternatively spliced exons (11,12,24,25).

### Prediction and experimental validation of novel alternative splices

Almost 43% of mouse alternative splices (8921) can be mapped to the human genome but are not currently found in human transcripts. A significant proportion of these are likely species-specific variants (13). Nevertheless, we were able to validate 5 of 11 predictions using RT–PCR across a panel of 44 human tissues (Supplementary Table S1). These predictions were randomly selected from a list of human genes enriched for genes with therapeutic interest using the following criteria: (i) the predicted alternative splices must have multiple mouse transcripts as support evidence and (ii) the predicted alternative splices are not found in human transcript databases at the time of analysis. Although the sample size is small, the 45% validation rate suggests that a significant fraction of conserved alternative splices identified by one species' transcripts are currently not present in the other species' transcripts.

We can conceptually divide all conserved alternative splicing events in human and mouse genes into four categories—those already found in both human and mouse transcripts (transcript conserved), those found in mouse transcripts but not found in human transcripts (mouse only), those found in human transcripts but not found in mouse transcripts (human

**Table 1.** Alternative splicing statistics

| Alternative splices | Human Number | Genes | Mouse Number | Genes |
|---|---|---|---|---|
| Conserved | 2526 (7%) | 1635 | 2526 (12%) | 1635 |
| Diverged | 17 743 (49%) | 6239 | 9451 (45%) | 4577 |
| Novel | 16 024 (44%) | 6189 | 8921 (43%) | 4691 |
| Total[a] | 35 721 | 7968 | 20 898 | 6697 |
| Estimations | | | | |
| Conserved currently novel | 6997 | | 5206 | |
| Total conserved[b] | 14 729 | | 14 729 | |

[a]From a total of 10 818 genes examined.
[b]Total conserved = 2526 (already identified) + 6997 (estimated conserved from human 'novels') + 5206 (estimated conserved from mouse 'novels').

only) and those not found in either human or mouse transcripts. We performed the following exercise to estimate the extent of 'human only' and 'mouse only' events. Using genomic sequence conservation of exons based on cross-species transcript alignment, we found that 27% of minor cassette exons in human genes and 37% in mouse genes are conserved in the opposing genome at >75% identity (Supplementary Figure S2). This estimate of 27% conservation rate is close to the 25% and 27% rates previously reported (11,12). Assuming that alternative splices are conserved at a similar level as cassette exons, we estimate that human transcripts contain a total of 9523 conserved alternative splices (25% of the 35 721 human splices identified) and mouse transcripts contain 7732 conserved alternative splices (27% of the 20 898 mouse splices identified) (Table 1). We have identified 2526 transcript conserved alternative splices, thus mouse transcripts currently contain 5206 undetected human conserved alternative splices (7732–2526), more than doubling the number of currently identified conserved alternative splices. Moreover, our estimate indicates at least 14 729 conserved alternative splices ($9523 + 7732 - 2526$) exist in these 10 818 genes, yielding an average of 1.4 conserved events per gene.

### Expression of conserved and diverged alternative splices

Using *Z*-score to represent the transcript coverage of a splice relative to mutually exclusive splicing events, we saw a strong correlation ($R = 0.75$) for 51 164 conserved splice pairs between mouse and human (Figure 2). In addition, conserved alternative splices exhibited greater transcript coverage than diverged or novel alternative splices (Figure 3A). Conserved alternative splices also exhibit weaker dependency on transcript coverage than diverged or novel alternative splices, suggesting that they arise from a distinct process and, perhaps obviously, that the total number of conserved alternative splices is limited (Figure 3B). The observation that the level of diverged alternative splicing continues to increase without reaching saturation even at high coverage indicates that diverged alternative splices may be generated by a stochastic process at a rate proportional to gene expression. Furthermore, the observed lower expression of diverged splices is consistent with the hypothesis that diverged alternative splices are frequently aberrant splice forms generated at a low background rate due to somatic mutations or spliceosomal errors (12,22). When examined in terms of expression, exon size, exon size as a factor of three, and associated intron size (Figures 4B, S3-5), the class of novel alternative splices shows characteristics between diverged and conserved splices, suggesting that it is a mixture of the two. Hence, we focused on comparing conserved and diverged alternative splices in subsequent analysis.

We classified genes into tissue-specific sets using microarray measured expression levels across a panel of 52 human

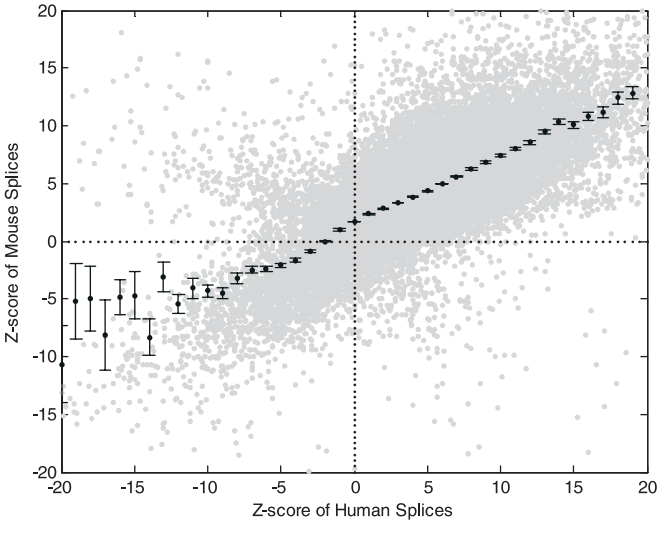


**Figure 2.** Cross-species correlation of alternative splicing expression levels. For each conserved alternative splice, the expression *Z*-score in mouse and human transcripts is plotted. Black dots and error bars show the average and standard deviation of binned *Z*-scores.

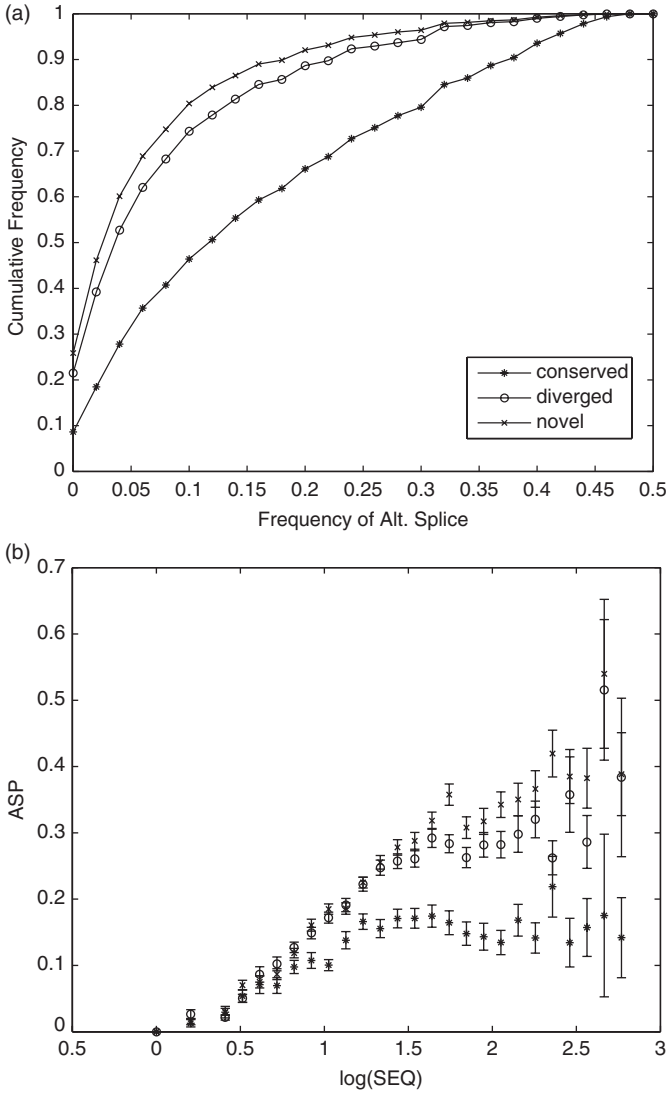


**Figure 3.** Transcript-based frequency analysis of alternative splices. (**A**) Cumulative frequency plot for conserved and diverged alternative splices. Frequency of a splice denotes the fraction of overlapping transcripts containing the splice. By definition, the frequency of minor alternative splices ranges from 0 to 0.5. (**B**) Average number of distinct alternative splices per gene, binned by the log10 of gene transcript coverage. Error bars represent 1 SD.

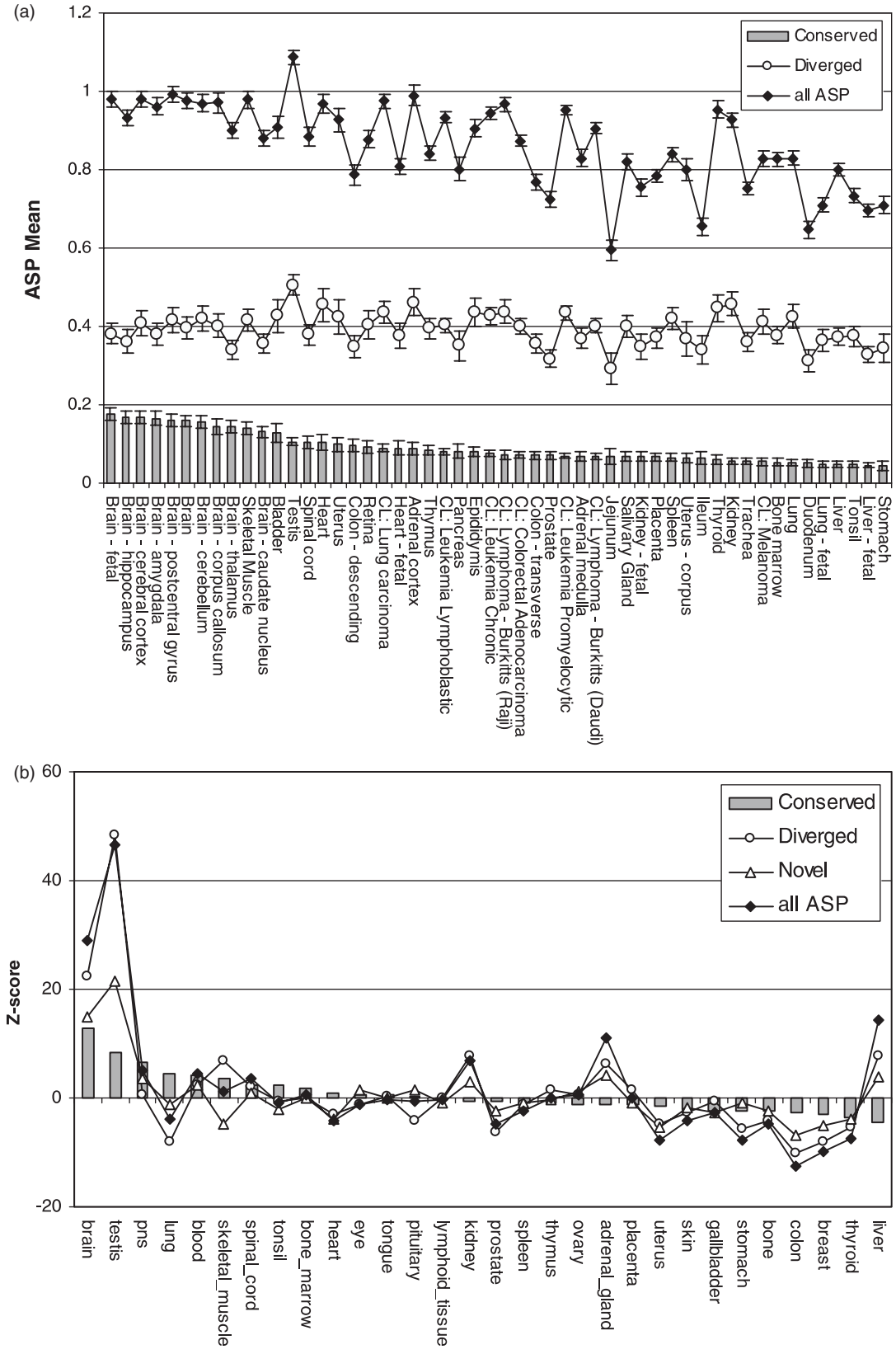


**Figure 4.** Expression and alternative splices. (**A**) Microarray expression of genes with alternative splices. Mean splice scores, normalized splices per gene, for diverged, conserved, and all alternative splices of genes over-expressed in each of 52 human tissues. Tissues are ranked by their average conserved splice scores. 'CL:' denotes cell lines. Novel splices plot between diverged and conserved. (**B**) Tissue distribution of alternative splicing across human tissues from EST eVOC library annotation. The *Z*-score measures the over-representation of ESTs containing alternative splices in a specific tissue relative to the background transcript levels.

tissues (7). Based on average splice scores (see methods) in tissue-expressed genes, including diverged and conserved alternative splicing, we found genes over-expressed in testis, brain and muscle tissues show the highest levels of alternative splicing whereas stomach and intestines show the lowest levels (Figure 4A). However, incorporating conservation classification status shows that the level of conserved alternative splicing is most elevated in brain sub-regions and muscle, while diverged alternative splicing is the most enriched in testis. Examination of the tissue origin of each splice from ESTs further shows brain (the most enriched in conserved alternative splices) and testis (the most enriched in diverged alternative splices) (Figure 4B).

### Integrated expression and function profiles of alternative splices

For each GO Biological Process and tissue, we intersected the associated GO gene set with the tissue over-expressed genes. From this list of genes, we calculated the *t*-score representing the enrichment of alternative splices. Very different patterns of distribution are observed for conserved and diverged alternative splices (Figure 5). The most significant enrichment of conserved alternative splicing is found in brain-expressed genes involved in functional processes, such as signal transduction, protein modification and ion transport (Figure 5A). Diverged alternative splicing is enriched in genes over-expressed in testis as well as cancerous cell lines and involved in processes, such as protein biosynthesis, cellular responses to stress and DNA damages (Figure 5B). Enrichment is also observed in processes, such as protein modification, cell death and cell cycles that are expressed in brain tissues.

## DISCUSSION

We found that the expression levels of conserved splices are strongly correlated in mouse and human genes, suggesting that the 'major' versus 'minor' roles of conserved splice forms were largely established before the divergence of human and rodents and were maintained to the present. Furthermore, as the relative expression levels of these splice forms are similar, the splicing regulatory mechanisms appear conserved between human and mouse. These results complement recent findings of similar splicing regulatory sequences and factors in mouse and human (26).

Species-specific alternative splicing events are not necessarily void of function; however, evolutionary conservation at the sequence level is an established indicator of functional importance. Although the functional roles of individual events need to be experimentally investigated on a case-by-case basis, strong evidence exists that non-conserved alternative splicing events are less likely to generate functional proteins than conserved ones (11–14). For example, non-conserved minor cassette exons are more likely to cause frame-shifts than conserved ones (95 versus 23%) (12). Alternative splicing events that introduce frame-shifts tend to not yield a functional protein product since the resulting mRNA is likely to be degraded by non-sense mediated decay (27). Thus we suggest conservation status as one proxy for the functional status of alternative splices.

The observed distinctions between conserved and diverged alternative splices in terms of relative abundance, functional process and tissue expression indeed suggest a fundamental functional difference between these classes. Previous studies have reported that brain and testis exhibit high levels of alternative splicing (28,29), and microarray studies using exon–exon junction probes revealed higher amounts of alternative splicing in cell lines (7). However, when viewed in terms of conservation, our results markedly demonstrate that the greatest amount of conserved alternative splicing occurs in the central nervous system while the high level of alternative splicing observed in testis and cell lines is largely diverged. Indeed, a recent study predicting conserved exon skipping events also mentioned enrichment in brain samples (14).

Alternative splicing in human and mouse brain may play important roles in enriching the functional diversity of signaling systems. Complexity generated by alternative splicing may confer an advantage in signaling pathways where functional capacity is enhanced by diverse ligand–receptor interactions, similar to expanding a complex communication network through adding nodes and connections. In cancerous cell lines and testis, the prevalence of diverged splicing may result from either more splicing errors or faulty surveillance mechanisms. For example, the increased levels of diverged splicing in these proliferating cells could be the product of stressed splicing machineries rapidly carrying out RNA splicing and protein synthesis, or result from accumulated somatic mutations that disrupt splicing regulation. Furthermore, in cancerous tumor or immortalized cell lines, positive selection for aberrant splicing events that disrupt regulation and promote proliferation may occur, similar to evolutionarily selected advantageous mutations.

In conclusion, we applied novel analysis techniques to 10 818 human and mouse genes, the most comprehensive gene set assembled to date in functional investigation of alternative splicing, and identified important expressional and functional distinctions between evolutionarily conserved and diverged alternative splicing events. Conserved alternative splicing is more highly expressed and less dependent on transcript coverage than diverged alternative splicing, indicating a difference in the biological processes from which these two classes of events originate. Strong correlation between alternative splicing expression levels in human and mouse suggests conservation of the regulatory mechanisms and functional roles for conserved alternative splicing events. Furthermore, conserved alternative splicing is more enriched in genes expressed in neuronal tissues and signaling pathways, perhaps allowing for more complexity and diversity. On the other hand, diverged alternative splicing is enriched in genes expressed in testis and cancerous cell lines where increased rate of aberrant splicing may result from abnormal cellular conditions, rapid cell proliferation or faulty surveillance mechanisms. Finally, computational prediction and experimental validation of novel alternative splicing events in human genes based on mouse transcripts suggest that the extent of conserved alternative splicing, hence the functional impact of alternative splicing, is far from fully revealed.

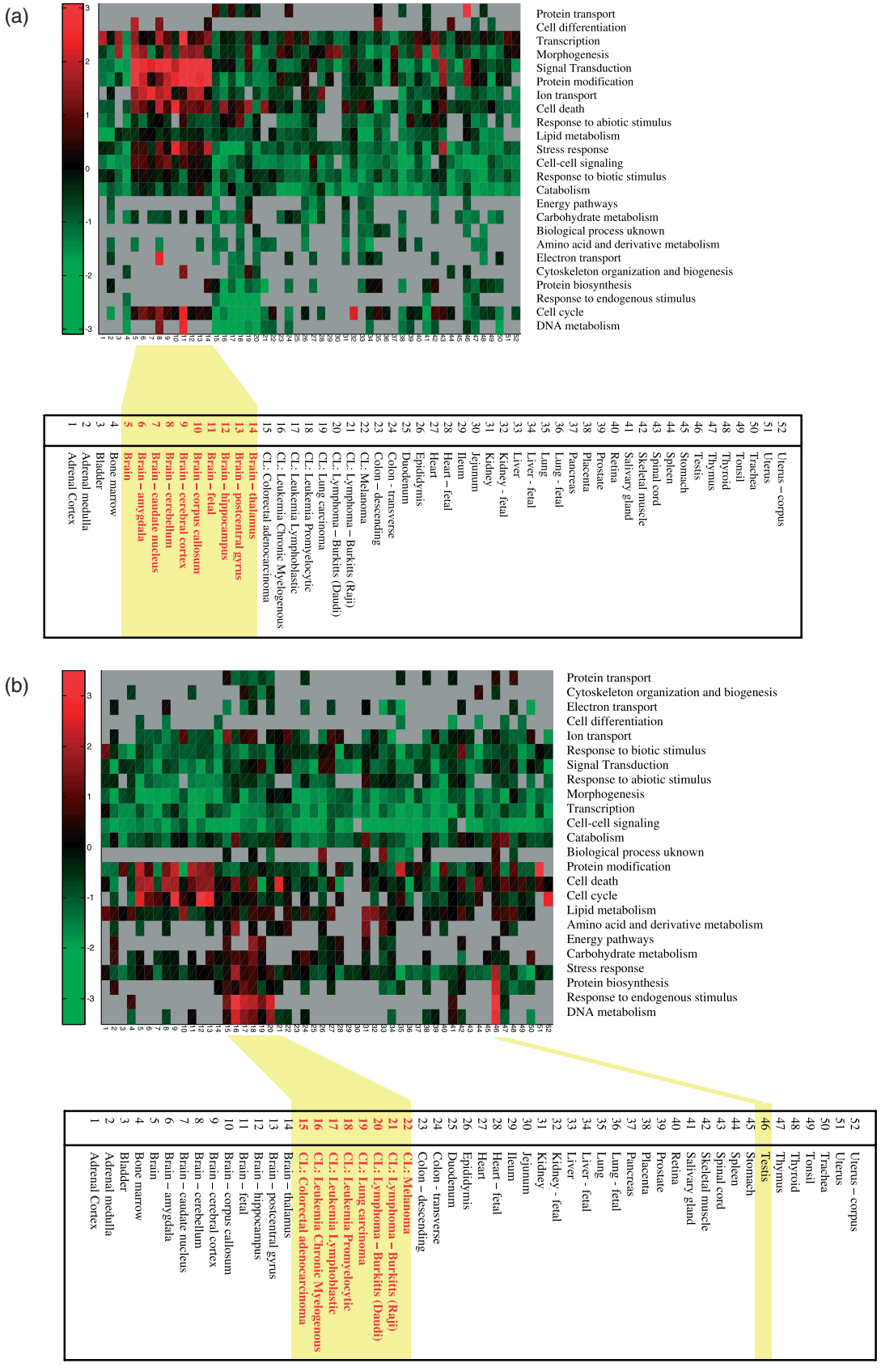


**Figure 5.** Integrated profiles across expression and GO categories. (**A**) *T*-scores measuring the enrichment of conserved alternative splices across functional and expressional categories. Rows correspond to 24 functional categories (GO-biological process, slim) and columns correspond to gene sets identified as over-expressed in each of 52 human tissues. Tissue-defined and GO-defined gene sets with no intersecting genes are shaded gray. (**B**) *T*-scores for diverged alternative splices.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors thank Ben Blencowe for sharing a manuscript prior to publication, Nick Tsinoremas for helpful conservations and Rosetta's Gene Expression Laboratory for microarray data. Funding to pay the Open Access publication charges for this article was provided by Merck & Co., Inc.

*Conflict of interest statement*. None declared.

## REFERENCES

1. Lopez,A.J. (1998) Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu. Rev. Genet.*, **32**, 279–305.
2. Caceres,J.F. and Kornblihtt,A.R. (2002) Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet.*, **18**, 186–193.
3. Black,D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.*, **72**, 291–336.
4. Graveley,B.R. (2001) Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.*, **17**, 100–107.
5. Modrek,B. and Lee,C. (2002) A genomic view of alternative splicing. *Nature Genet.*, **30**, 13–19.
6. Zavolan,M., Kondo,S., Schonbach,C., Adachi,J., Hume,D.A., Hayashizaki,Y. and Gaasterland,T. (2003) Impact of alternative initiation, splicing, and termination on the diversity of the mRNA transcripts encoded by the mouse transcriptome. *Genome Res.*, **13**, 1290–1300.
7. Johnson,J.M., Castle,J., Garrett-Engele,P., Kan,Z., Loerch,P.M., Armour,C.D., Santos,R., Schadt,E.E., Stoughton,R. and Shoemaker,D.D (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science*, **302**, 2141–2144.
8. Modrek,B., Resch,A., Grasso,C. and Lee,C. (2001) Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res.*, **29**, 2850–2859.
9. Liu,S. and Altman,R.B. (2003) Large scale study of protein domain distribution in the context of alternative splicing. *Nucleic Acids Res.*, **31**, 4828–4835.
10. Lareau,L.F., Green,R.E., Bhatnagar,R.S. and Brenner,S.E. (2004) The evolving roles of alternative splicing. *Curr. Opin. Struct. Biol.*, **14**, 273–282.
11. Modrek,B. and Lee,C.J. (2003) Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nature Genet.*, **34**, 177–180.
12. Sorek,R., Shamir,R. and Ast,G. (2004) How prevalent is functional alternative splicing in the human genome? *Trends Genet.*, **20**, 68–71.
13. Pan,Q., Bakowski,M.A., Morris,Q., Zhang,W., Frey,B.J., Hughes,T.R. and Blencowe,J.B. (2005) Alternative splicing of conserved exons is frequently species-specific in human and mouse. *Trends Genet*, **21**, 73–77.
14. Yeo,G.W., Van Nostrand,E., Holste,D., Poggio,T. and Burge,C.B. (2005) Identification and analysis of alternative splicing events conserved in human and mouse. *Proc. Natl Acad. Sci. USA*, **102**, 2850–2855.
15. Kan,Z., Castle,J., Johnson,J.M. and Tsinoremas,N.F. (2004) Detection of novel splice forms in human and mouse using cross-species approach. *Pac. Symp. Biocomput.*, 42–53.
16. Pruitt,K.D. and Maglott,D.R. (2001) RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res.*, **29**, 137–140.
17. Wheeler,D.L., Church,D.M., Federhen,S., Lash,A.E., Madden,T.L., Pontius,J.U., Schuler,G.D., Schriml,L.M., Sequeira,E., Tatusova,T.A. *et al.* (2003) Database resources of the National Center for Biotechnology. *Nucleic Acids Res.*, **31**, 28–33.
18. Florea,L., Hartzell,G., Zhang,Z., Rubin,G.M. and Miller,W. (1998) A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res.*, **8**, 967–974.
19. Kan,Z., Rouchka,E.C., Gish,W.R. and States,D.J. (2001) Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. *Genome Res.*, **11**, 889–900.
20. Mott,R. (1997) EST_GENOME: a program to align spliced DNA sequences to unspliced genomic DNA. *Comput. Appl. Biosci.*, **13**, 477–478.
21. Chuang,T.J., Chen,F.C. and Chou,M.Y. (2004) A comparative method for identification of gene structures and alternatively spliced variants. *Bioinformatics*, **20**, 3064–3079.
22. Kan,Z., States,D. and Gish,W. (2002) Selecting for functional alternative splices in ESTs. *Genome Res.*, **12**, 1837–1845.
23. Kelso,J., Visagie,J., Theiler,G., Christoffels,A., Bardien,S., Smedley,D., Otgaar,D., Greyling,G., Jongeneel,C.V., McCarthy,M.I. *et al.* (2003) eVOC: a controlled vocabulary for unifying gene expression data. *Genome Res.*, **13**, 1222–1230.
24. Thanaraj,T.A., Clark,F. and Muilu,J. (2003) Conservation of human alternative splice events in mouse. *Nucleic Acids Res.*, **31**, 2544–2552.
25. Sugnet,C.W., Kent,W.J., Ares,M.,Jr and Haussler,D. (2004) Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac. Symp. Biocomput.*, 66–77.
26. Yeo,G., Hoon,S., Venkatesh,B. and Burge,C.B. (2004) Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proc. Natl Acad. Sci. USA*, **101**, 15700–15705.
27. Ruiz-Echevarria,M.J., Gonzalez,C.I. and Peltz,S.W. (1998) Identifying the right stop: determining how the surveillance complex recognizes and degrades an aberrant mRNA. *EMBO J.*, **17**, 575–589.
28. Xu,Q., Modrek,B. and Lee,C. (2002) Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res.*, **30**, 3754–3766.
29. Yeo,G., Holste,D., Kreiman,G. and Burge,C.B. (2004) Variation in alternative splicing across human tissues. *Genome Biol.*, **5**, R74.