



The Evolutionary Landscape of Alternative Splicing in Vertebrate Species

Nuno L. Barbosa-Morais *et al.*
Science **338**, 1587 (2012);
DOI: 10.1126/science.1230612

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of May 3, 2013):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/338/6114/1587.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2012/12/19/338.6114.1587.DC1.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/content/338/6114/1587.full.html#related>

This article **cites 75 articles**, 35 of which can be accessed free:

<http://www.sciencemag.org/content/338/6114/1587.full.html#ref-list-1>

This article has been **cited by** 3 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/338/6114/1587.full.html#related-urls>

This article appears in the following **subject collections**:

Evolution

<http://www.sciencemag.org/cgi/collection/evolution>

octatomic S, dimethyltrisulfide, and dimethyl-tetrasulfide, the latter two for about a combined 1 mmole/g; the fragment contained in addition only naphthalene at 2 to 8 nmol/g, plus methyl-naphthalenes and biphenol in subnanomole amounts, but no alkanes or anthracene/phenanthrene.

The SM meteorite demonstrates that the complexity of C-class asteroid surfaces is greater than previously assumed. Rapid terrestrial alteration probably erases many vestiges of the internal and external processes on the asteroid that remain to be explored in spacecraft sample-return missions.

References and Notes

- Fries, J. Fries, *Meteorit. Planet. Sci.* **45**, 1476 (2010).
- Jenniskens, M. Zolensky, *Meteorit. Bull.* **46**, 1 (2012).
- Materials and methods are available as supplementary materials on Science Online.
- Christie, P. Campus, in *Infrasound Monitoring for Atmospheric Studies*, A. Le Pichon, E. Blanc, A. Hauchecorne, Eds. (Springer, Dordrecht, Netherlands, 2010), pp. 29–75.
- Enns, P. G. Brown, W. N. Edwards, E. A. Silber, *J. Atmos. Sol. Terr. Phys.* **80**, 208 (2012).
- Silber, E. A. Le Pichon, P. Brown, *Geophys. Res. Lett.* **38**, L12201 (2011).
- Shaddad, M. H. *Meteorit. Planet. Sci.* **45**, 1557 (2010).
- Smith, K. D. *Science* **305**, 1277 (2004).
- Haack, H. *Meteorit. Planet. Sci.* **47**, 30 (2012).
- Haack, H. *Workshop on the First Solids in the Solar System*, 7 to 9 November 2011, Koloa, Kauai, Hawai'i (Lunar and Planetary Institute, Houston, TX, 2011), abstr. 9100.
- Nishiizumi, M. W. Caffee, *Meteorit. Planet. Sci.* **44** (suppl.), 5358 (2009).
- Ott, U. *Rev. Mineral. Geochem.* **47**, 71 (2002).
- Heber, V. S. *Geochim. Cosmochim. Acta* **73**, 7414 (2009).
- Leya, H.-J. Lange, S. Neumann, R. Wieler, R. Michel, *Meteorit. Planet. Sci.* **35**, 259 (2000).
- Gounelle, M. *The Solar System Beyond Neptune*, M. A. Barucci, H. Boehnhardt, D. P. Cruikshank, A. Morbidelli, Eds. (Univ. of Arizona Press, Tucson, AZ, 2008), pp. 525–541.
- Nesvorný, D. *Astrophys. J.* **713**, 816 (2010).
- Walsh, M. Delbo, W. F. Bottke, paper presented at the 44th American Astronomical Society/Division of Planetary Sciences Meeting, Reno, NV, 14 to 19 October, 2012, abstr. 305.04.
- Tsuchiyama, A. *Meteorit. Planet. Sci.* **44** (suppl.), 5189 (2009).
- Popova, O. *Meteorit. Planet. Sci.* **46**, 1525 (2011).
- Ebel, M. L. Rivers, *Meteorit. Planet. Sci.* **42**, 1627 (2007).
- Macke, D. T. Britt, G. J. Consolmagno, *Meteorit. Planet. Sci.* **46**, 311 (2011).
- Britt, D. T. *paper presented at the Meteoritical Society Meeting, Cairns, Australia, 12 to 17 August 2012*, abstr. 5250.
- Spergel, R. C. Reedy, O. W. Lazareth, P. W. Levy, L. A. Slates, *J. Geophys. Res.* **91**, 483 (1986).
- Rochette, P. *Meteorit. Planet. Sci.* **43**, 959 (2008).
- Gattacceca, P. Rochette, *Earth Planet. Sci. Lett.* **227**, 377 (2004).
- Bischoff, E. R. D. Scott, K. Metzler, C. A. Goodrich, in *Meteorites and the Early Solar System II*, D. S. Lauretta, H. Y. McSween Jr., Eds. (Univ. of Arizona Press, Tucson, AZ, 2006), pp. 679–712.
- Walker, R. J. *Geochim. Cosmochim. Acta* **66**, 4187 (2002).
- Yin, Q.-Z. Yamashita, A. Yamakawa, R. Tanaka, B. Jacobsen, D. S. Ebel, I. D. Hutcheon, and E. Nakamura, *Lunar Planet. Sci.* **XL**, abstr. 2006 (2009).
- Friedrich, J. M.-S. Wang, M. E. Lipschutz, *Meteorit. Planet. Sci.* **37**, 677 (2002).
- Tonui, M. E. Zolensky, M. E. Lipschutz, *Proc. NIPR Symp. Antarct. Meteorites* **15**, 38 (2002).
- Haq, F. A. Hasan, D. W. G. Sears, C. B. Moore, C. F. Lewis, *Geochim. Cosmochim. Acta* **53**, 1435 (1989).
- Cody, G. D. *Earth Planet. Sci. Lett.* **272**, 446 (2008).
- Young, R. D. Ash, P. England, D. Rumble 3rd, *Science* **286**, 1331 (1999).
- Benedix, G. K. Leshin, J. Farquhar, T. Jackson, M. H. Thiems, *Geochim. Cosmochim. Acta* **67**, 1577 (2003).
- Grady, M. M. *Antarct. Geophys.* **50**, 4.21 (2009).
- Schmitt-Kopplin, A. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2763 (2010).
- Cooper, G. W. M. Onwo, J. R. Cronin, *Geochim. Cosmochim. Acta* **56**, 4109 (1992).
- Monroe, A. S. Pizzarello, *Geochim. Cosmochim. Acta* **75**, 7585 (2011).
- Glavin, D. P. Callahan, J. P. Dworkin, J. E. Elsila, *Meteorit. Planet. Sci.* **45**, 1948 (2010).
- Cronin, J. R. Pizzarello, *Geochim. Cosmochim. Acta* **54**, 2859 (1990).
- Lodders, *Astrophys. J.* **591**, 1220 (2003).
- Jarosewich, R. S. Clarke Jr., J. N. Barrows, Eds., *Smithson. Contrib. Earth Sci.* **27**, 1 (1986).
- Brown, P. G. *Science* **290**, 320 (2000).
- Clayton, R. N. K. Mayeda, *Geochim. Cosmochim. Acta* **63**, 2089 (1999).
- Clayton, R. N. K. Mayeda, J. N. Goswami, E. J. Olsen, *Geochim. Cosmochim. Acta* **55**, 2317 (1991).
- Clayton, R. N. K. Mayeda, A. E. Rubin, *J. Geophys. Res.* **89**, C245 (1984).
- MacPherson, G. J. *Geochim. Cosmochim. Acta* **73**, 5493 (2009).
- Young, E. D. Russell, *Science* **282**, 452 (1998).

Acknowledgments: The rapid recovery of SM was made possible by the local communities of Lotus, Coloma, Pilot Hill, and Grass Valley, due in part to the support of property owners, community leaders, SETI Institute and University of California at Davis (UC Davis) student and volunteer searchers, and officials of the Marshall Gold Discovery State Historic Park and the State of California. We thank J. Howard, M. Carter, D. E. Macon, and L. Hofland of the NASA Ames Research Center; M. Hill of the American Museum of Natural History; and G. Baxter, N. Botto, D. Rowland, O. Kreylos, L. Kellogg, and D. Summer of UC Davis for assistance with reported measurements. The NASA Ames Research Center-coordinated search efforts were supported by the NASA Lunar Science Institute. E.D.Y., M.Z., A.M.D., D.S.E., J.E.E., K.N., K.C.W., M.H.T., M.W.C., Q.-Z.Y., and S.B.S. acknowledge the NASA Cosmochemistry Program for support of work reported here. In addition, the following NASA programs are sincerely acknowledged: Near Earth Object Observation Program (P.J.), Origins of Solar Systems (S.A.S.), Planetary Major Equipment (Q.-Z.Y.), Sample Return Laboratory Instruments and Data Analysis (M.W.C.), the NASA Astrobiology Institute (J.E.E., J.P.D., and D.P.G.), the SPing-8 Grant Program (M.Z.), and Post Doctoral Programs (A.S.B.). P.R.H. acknowledges support from the Tawani Foundation. R.O. acknowledges KAKENHI grants from the Japan Society for the Promotion of Science.

Supplementary Materials

www.sciencemag.org/cgi/content/full/338/6114/1583/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S36
Tables S1 to S22
References (49–105)

9 July 2012; accepted 12 November 2012
10.1126/science.1227163

The Evolutionary Landscape of Alternative Splicing in Vertebrate Species

Nuno L. Barbosa-Morais,^{1,2} Manuel Irimia,^{1,*} Qun Pan,^{1,*} Hui Y. Xiong,^{3,*} Serge Gueroussov,^{1,4,*} Leo J. Lee,³ Valentina Slobodeniuc,¹ Claudia Kutter,⁵ Stephen Watt,⁵ Recep Çolak,^{1,6} TaeHyung Kim,^{1,7} Christine M. Misquitta-Ali,¹ Michael D. Wilson,^{4,5,7} Philip M. Kim,^{1,4,6} Duncan T. Odom,^{5,8} Brendan J. Frey,^{1,3} Benjamin J. Blencowe^{1,4,†}

How species with similar repertoires of protein-coding genes differ so markedly at the phenotypic level is poorly understood. By comparing organ transcriptomes from vertebrate species spanning ~350 million years of evolution, we observed significant differences in alternative splicing complexity between vertebrate lineages, with the highest complexity in primates. Within 6 million years, the splicing profiles of physiologically equivalent organs diverged such that they are more strongly related to the identity of a species than they are to organ type. Most vertebrate species-specific splicing patterns are cis-directed. However, a subset of pronounced splicing changes are predicted to remodel protein interactions involving trans-acting regulators. These events likely further contributed to the diversification of splicing and other transcriptomic changes that underlie phenotypic differences among vertebrate species.

Vertebrate species possess diverse phenotypic characteristics, yet they share similar repertoires of coding genes (*1*).

Evolutionary changes in transcriptomes underlie structural and regulatory differences associated with species-specific characteristics. For

example, species-dependent mRNA and non-coding RNA (ncRNA) expression patterns have been linked to mutational changes in cis- and trans-acting regulatory factors, as well as to phenotypic differences (2–5). However, because organ-dependent mRNA expression levels within individual species have been largely conserved during vertebrate evolution (6, 7), it seems unlikely that changes in gene expression (GE)

¹Banting and Best Department of Medical Research, Donnelly Centre, University of Toronto, Toronto, Ontario M5S 3E1, Canada. ²Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, 1649-028 Lisboa, Portugal. ³Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario M5S 3G4, Canada. ⁴Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. ⁵University of Cambridge, CRUK-Cambridge Institute, Robinson Way, Cambridge CB2 0RE, UK. ⁶Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada. ⁷Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. ⁸Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridge CB10 1SA, UK.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: b.blencowe@utoronto.ca

account for the majority of phenotypic diversity among vertebrates.

Through the variable use of cis-acting RNA elements in exons and flanking introns that are recognized by trans-acting factors, different pairs of splice sites in primary transcripts can be se-

lected in a cell type-, condition-, or species-specific manner (8–15). Changes in alternative splicing (AS) may therefore represent a major source of species-specific differences (16–25). Here, we describe a genome-wide investigation of AS differences among physiologically equi-

valent organs from vertebrate species spanning the major tetrapod lineages.

Evolution of alternative splicing complexity. High-throughput RNA sequencing (RNA-Seq) data were collected from whole brain, forebrain cortex, cerebellum, heart, skeletal muscle, liver,

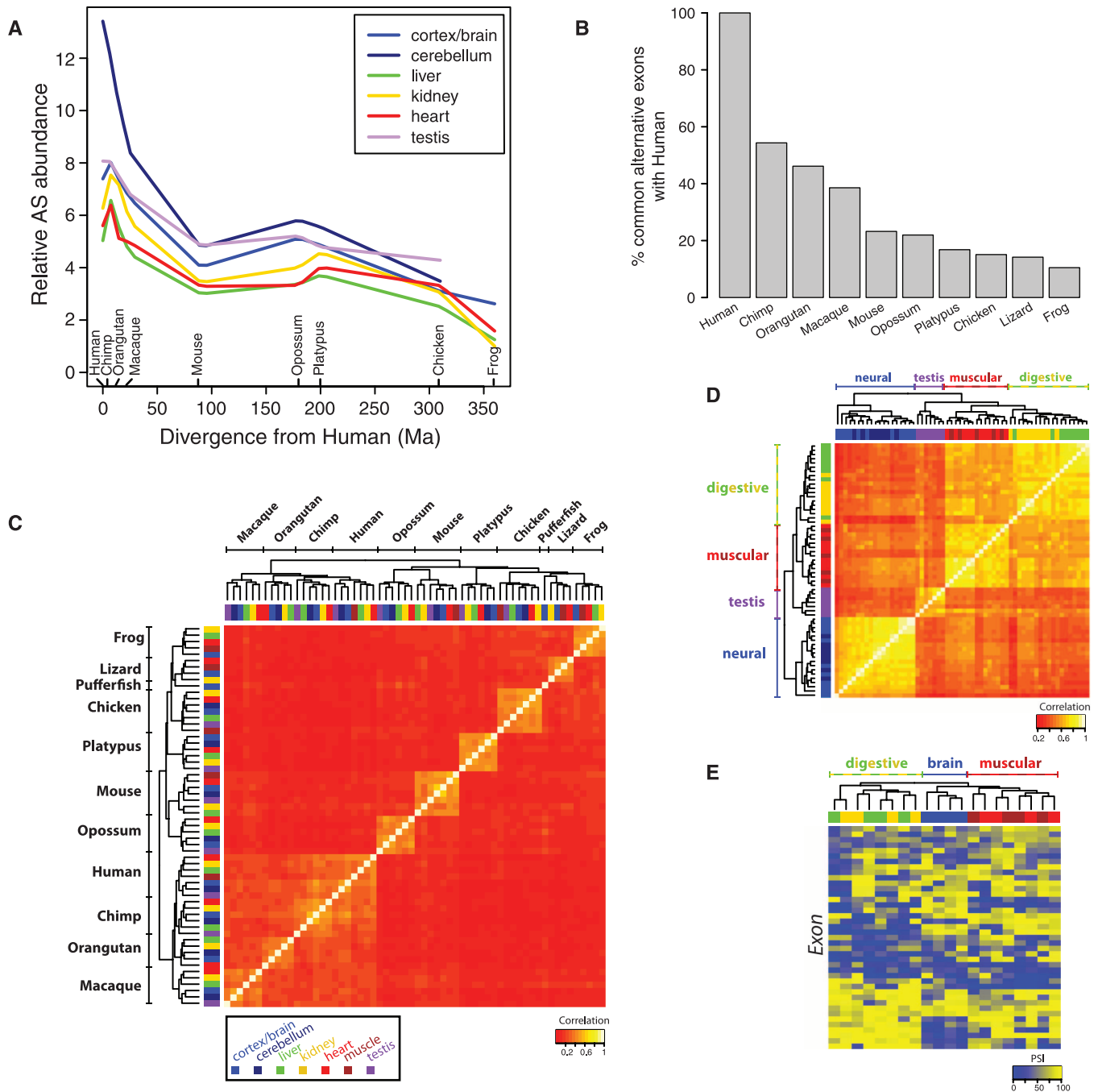


Fig. 1. Profiling of alternative splicing (AS) in vertebrates. **(A)** Relative proportions of exons undergoing AS in each sample, as measured by detection of middle exon skipping in random exon triplets, where the three exons are represented by orthologs in the analyzed species (y-axis units relative to the sample with lowest AS frequency). See fig. S1, A and B, for a more detailed version; see table S5 for details on samples, including replicates, and RNA-Seq data sets. **(B)** Percentage of common AS events between human and other species. **(C)** Symmetrical heat map of Spearman correlations from PSI profiles. For each sample, PSI values for the 1550

orthologous exons in the 11 analyzed species were estimated. See fig. S4A for a more detailed version. **(D)** Symmetrical heat map of Pearson correlations from gene expression (GE) profiles. For each sample, mRNA expression [log cRPKM values (26)] of 1809 analyzed orthologous genes in the 11 analyzed species were estimated. Key as in (C). See fig. S4B for a more detailed version. **(E)** Heat map of PSI values for 41 conserved cassette alternative exons. Rows, exons; columns, samples. Key as in (C). See fig. S11B for a more detailed version. Data are hierarchically clustered (complete method, Euclidean distance) for heat maps in (C) to (E).

kidney, and testis from human, chimpanzee, orangutan, macaque, mouse, opossum, platypus, chicken, lizard, and frog (26). For each species, we considered all internal exons as potential cassette AS events and created nonredundant databases of splice junction sequences formed by inclusion or skipping of each exon. RNA-Seq reads were mapped to the junction databases to determine “percent spliced-in” (PSI) values, and also to representative transcript sequences from each gene to estimate GE levels, represented as “corrected reads per kilobase transcript model per million mapped reads” (cRPKM) values (26). Orthology relationships between genes and exons were established to enable direct cross-species comparisons.

The relative proportions of orthologous exons detected to undergo AS in each sample were determined. Equal numbers of reads were randomly sampled from each RNA-Seq data set to control for coverage differences (26). AS detection is approximately twice as frequent in all analyzed pri-

mate organs as in the equivalent organs from mouse and other species (Fig. 1A and fig. S1). Moreover, there is an overall decline in AS frequency as the evolutionary distance from primates increases. These differences are significant ($P < 10^{-10}$, Mann-Whitney U tests), are robust to different methods of AS frequency detection, and are independent of the variability in AS detection rates between individuals within the same species (Fig. 1A and fig. S1). Genes with the highest AS complexity in human are significantly enriched in cytoskeleton-associated functions ($P < 0.03$) (table S1), which suggests that AS-directed diversification of the cytoskeleton may have been a driving force in the evolution of increased cellular complexity in vertebrate species.

Rapid evolution of organ-specific alternative splicing. We next compared AS profiles across organs and species. Approximately half of alternatively spliced exons among species separated by ~6 million years of evolution are different

(Fig. 1B and figs. S2A and S3A). When clustering organ AS profiles on the basis of how overall PSI values correlate in pairwise comparisons, the samples segregate by individual species (Fig. 1C and fig. S4A). This is in contrast to clustering samples on the basis of how their overall GE levels correlate, where the samples segregate according to tissue type (Fig. 1D and fig. S4B) (6, 7). Principal components analysis confirms that species type and tissue type are the primary sources of variability underlying the overall AS and GE patterns, respectively (figs. S5 and S6). The species-dependent clustering of AS profiles is also observed when analyzing subsets of alternative exons associated with a wide range of splice site strengths, exon length, increased magnitudes of PSI change between tissues, increased read coverage, and also when using independently validated (16) PSI differences (figs. S7 to S10). These results indicate that overall organ-specific AS patterns have evolved at a much more rapid rate than organ-specific GE patterns.

However, when restricting the clustering analysis to all ($n = 41$) orthologous exons that are alternatively spliced in four species (human, mouse, chicken, and frog) representing the main tetrapod lineages, the samples segregate by tissue type, with similar results obtained with the larger data set of tissues and species (Fig. 1E and fig. S11, A to H). The 41 exons, on average, display a wider range of inclusion levels across the samples, indicating that they have a higher degree of regulatory potential (fig. S11I). Consistent with this observation, they are also associated with elevated exonic and flanking intronic sequence conservation, implying that they are under increased selection pressure to maintain binding sites for regulatory trans-acting factors (26). Therefore, although overall AS patterns of multiple organs distinguish vertebrate species, a small subset of exons that undergo AS in multiple species spanning ~350 million years of evolution display conserved patterns of regulation that reflect organ type.

We next investigated the relative rates at which AS and GE have evolved. From pairwise comparisons of PSI values in homologous tissues, we observe an overall increase of PSI divergence from human with evolutionary time (Fig. 2A). In contrast to results from analyzing GE divergence (Fig. 2, C and D) (6, 7), AS levels in testis have not diverged more rapidly than in other tissues (Fig. 2B), and AS events detected in neural tissues display the slowest rate of divergence ($P < 10^{-6}$, Mann-Whitney U tests). A significantly higher proportion (27% more on average) of neural AS events are conserved between vertebrate species than are AS events specific to other organs ($P < 0.002$, Mann-Whitney U test; fig. S2B). These AS events are enriched in genes associated with synaptic transmission, axon guidance, neural development, and actin cytoskeleton reorganization (table S2), indicating that AS regulation of these processes is a highly conserved feature of vertebrate nervous system development.

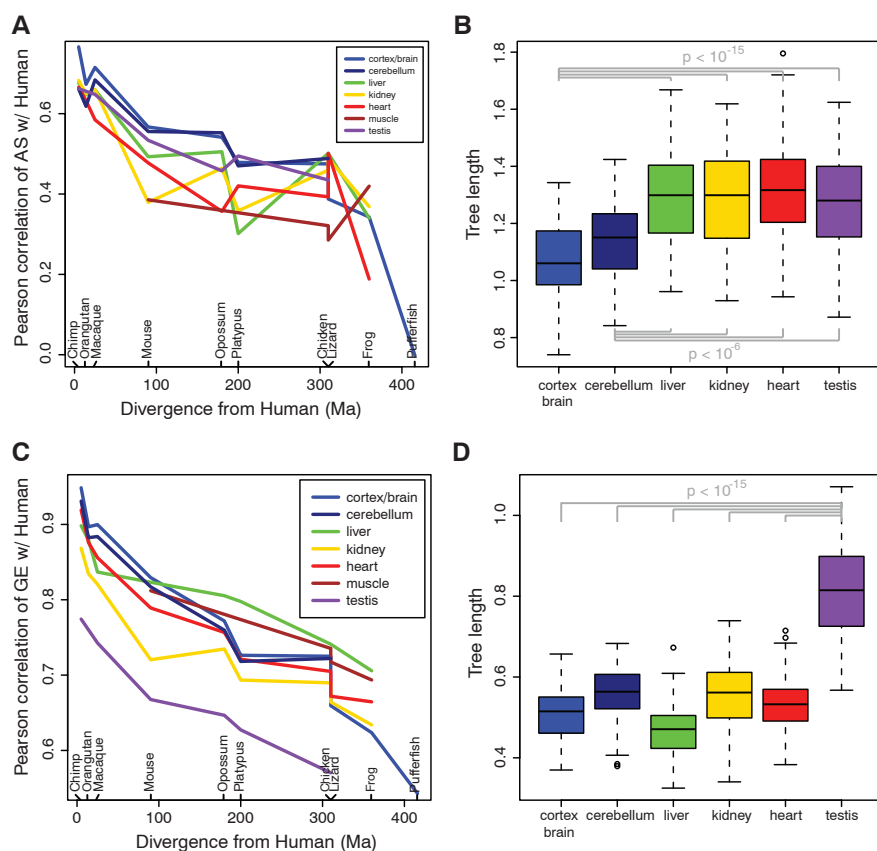


Fig. 2. Different rates of AS and GE divergence. (A) Pearson correlations between human and other species when comparing PSI values pairwise for conserved tissue-specific alternative exons in each tissue. For each pair of samples, correlation is analyzed for PSI values for all exons undergoing AS in both samples. (B) Comparisons of total tree lengths (bootstrapping, 100 replicates) of PSI trees for conserved tissue-specific alternative exons from six tissues in seven species (human, chimp, macaque, mouse, opossum, platypus, chicken). Statistically significant differences between the neural and each of the other tissues are indicated (Mann-Whitney U tests). Full PSI trees for each tissue are shown in fig. S3B. (C) Pearson correlations between human and other species when comparing GE pairwise (cRPKM, log scale; 1809 orthologous genes in the 11 analyzed species) in each organ. (D) Comparisons of total tree lengths (bootstrapping, 100 replicates) of expression trees from analyzing six organs from seven species as in (B). Statistically significant differences between testis and each of the other tissues are indicated (Mann-Whitney U tests). Full expression trees for each tissue are shown in fig. S3C.

Evolution of vertebrate splicing codes. To investigate mechanisms underlying the divergence in organ AS profiles, we used a splicing code derived from mouse data (27, 28) to compare cis-regulatory elements that are predictive of tissue-dependent splicing patterns of five organs (brain, heart, skeletal muscle, liver, and kidney) from the four representative species analyzed above. The splicing code achieved high classification rates when predicting organ-specific AS patterns from sequence alone. The average true positive rate [AUC, area under the receiver operating characteristic (ROC) curve] ranges from 68 to 70% for heart exon skipping to 80 to 88% for brain exon inclusion (Fig. 3, A and B, and fig. S12). An exception was a subset of approximately 200 exon-skipping AS events in human brain,

which were predicted less well (AUC = 62%) (fig. S12) (26).

A comparison of the most strongly predictive cis-elements accounting for each species' organ-dependent splicing patterns revealed that these significantly overlap in most tissues (Fig. 3C and fig. S12). Such cis-elements may represent features comprising an ancestral vertebrate splicing code (Fig. 3D). However, the overlap between sets of splicing code features used in a given pair of species decreases with increased evolutionary distance (Fig. 3C and fig. S12). Thus, organ-dependent AS patterns appear to be generally controlled by significantly overlapping cis-regulatory codes, although progressive divergence in these codes likely also contributes to AS differences.

Species-specific alternative splicing is primarily cis-directed. The extent to which evolutionary change in cis-regulatory codes (versus trans-acting factors) accounts for species-dependent AS differences is not known. To address this, we used a mouse strain, Tc1, carrying the majority of human chromosome 21 (HsChr21) (29). We compared PSI values of exons from HsChr21 transcripts expressed in multiple organs (brain, liver, heart, testis) from the Tc1 strain, with PSI values of the identical exons in the corresponding human organs. We also compared PSI values for the orthologous mouse exons between wild-type and Tc1 mouse strains. For all comparisons, we analyzed a comprehensive set of 13 HsChr21 and orthologous mouse exons that were detected, using RNA-Seq data, to

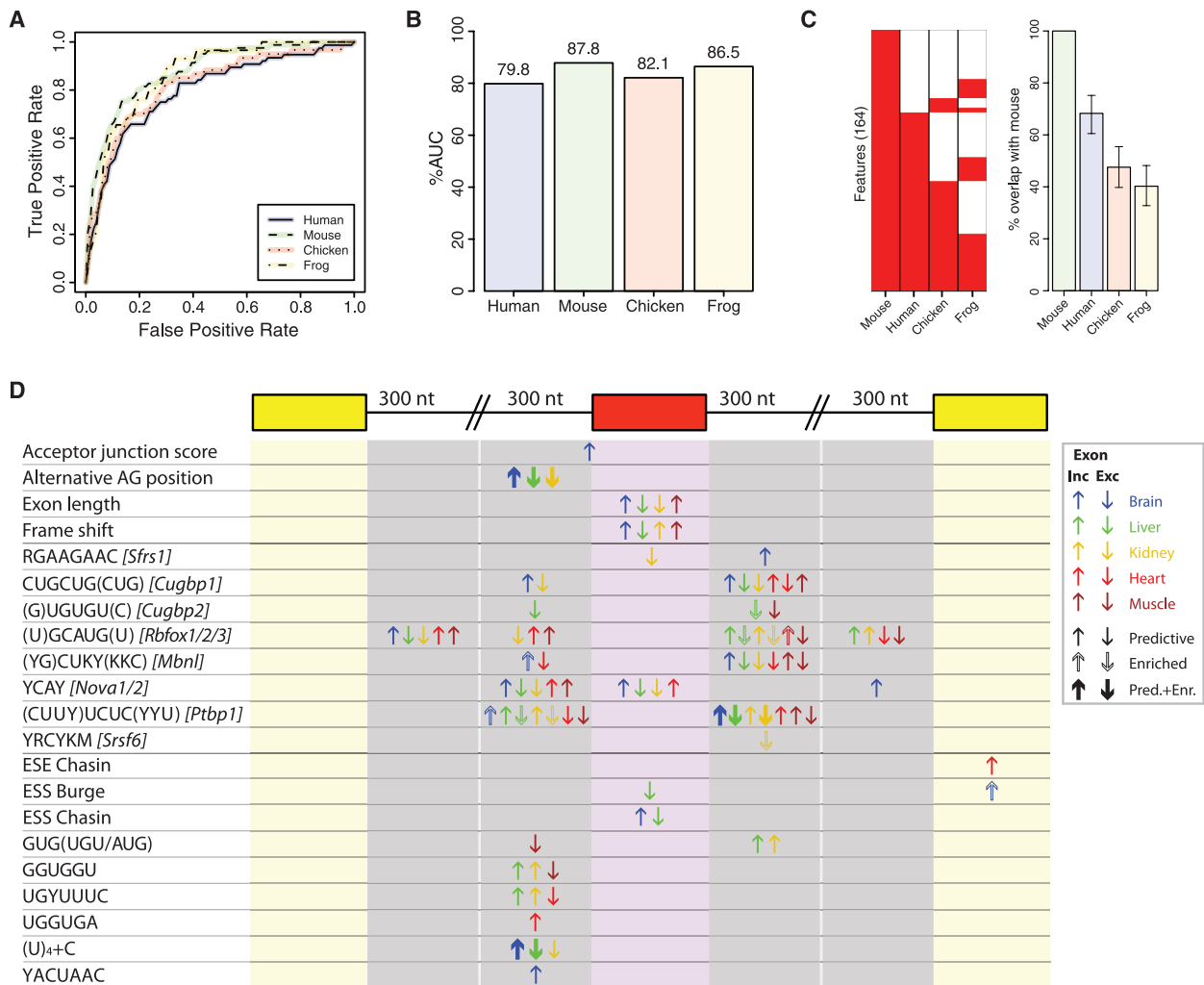


Fig. 3. Inference and comparative analysis of vertebrate splicing codes. **(A)** ROC curves of splicing code predictions for brain-dependent exon inclusion in human, mouse, chicken, and frog. ROC curves for other tissues are shown in fig. S12. **(B)** AUC in percent for each of the ROC curves in (A). **(C)** Left: Heat map of splicing code features with significant prediction scores associated with brain-specific exon inclusion in each species (red for $P < 0.05$, Mann-Whitney U test); statistically significant features in common with all significant mouse features associated with brain-specific exon inclusion are shown. Right: Proportions of features significantly associated with mouse brain-specific exon inclusion also significant for brain-specific exon inclusion in other species.

Error bars: 95% confidence intervals, Pearson's χ^2 proportion tests. Analyses of predictive features for other tissues are shown in fig. S12. **(D)** Region-specific distribution of code features significantly associated with tissue-specific exon inclusion or exclusion in all four species, as determined by their region-specific enrichment or by their predictive power as inferred by the splicing code. Splicing factors associated with code features are in square brackets. Significantly enriched features are indicated by hollow arrows; features both significantly enriched and predictive are indicated by bold arrows. Arrows are colored according to the organ with which the features are associated (refer to key).

have undergone AS in at least one of the two species.

All analyzed HsChr21 exons that are alternatively spliced in human for which the orthologous exons in mouse are constitutively spliced are also alternatively spliced in the Tc1 mouse; likewise, all HsChr21 constitutively spliced exons for which the orthologous exons in mouse are alternatively spliced are also constitutively spliced in the Tc1 mouse ($P < 0.02$, both comparisons; one-sided Fisher's exact tests; Fig. 4, A and B, and fig. S13A). For the orthologous exons that are alternatively spliced in human and mouse, we observe a significantly higher correlation between their inclusion levels in Tc1 mouse and normal human organs relative to the correlation observed when comparing inclusion levels of all orthologous human and mouse exons ($r = 0.89$ versus $r = 0.52$, $P < 0.0008$, one-sided Z-test; fig. S13B) (26).

Collectively, the results indicate that changes among predominantly conserved cis-regulatory elements are sufficient to direct the majority of the species-specific AS patterns, at least in human and mouse. However, because our results also indicate that vertebrate splicing codes diverged with increasing evolutionary distance, and because specific subsets of AS events could not be reliably predicted using the splicing code, changes in trans-acting factors likely also contributed to evolutionary differences in AS.

Species-classifying alternative splicing events in trans-acting regulatory factors.

To investigate which splicing changes during evolution likely had the greatest functional impact, we identified AS events that best discriminate or "classify" species (26). These AS events have relatively large and widely expressed PSI differences between species or lineages, and were validated at a high rate by reverse transcription polymerase chain reaction (RT-PCR) assays ($r = 0.90$, $n = 180$; fig. S14). Gene Ontology enrichment analysis reveals that the corresponding genes are functionally diverse, with "nucleic acid binding" among the most frequently represented categories (Fig. 5A, fig. S14, and tables S3 and S4). Consistent with a major role for conserved cis-regulatory elements governing species-dependent AS profiles (Figs. 3 and 4), the species-classifying AS events are significantly underrepresented in exons that overlap nucleic acid binding domains, relative to other classes of alternative exons in the same genes ($P < 0.04$, Pearson's χ^2 proportion test; Fig. 5C).

An interesting example of a species-classifying AS event is the activity-modulating exon 9 of the splicing regulator PTBP1, located between RNA recognition motifs 2 and 3 in this protein (30, 31). This exon is skipped in mammalian organs but is fully included in chicken and frog organs (Fig. 5A). Consistent with the possibility of a more

variable and extensive regulatory role for PTBP1 in mammalian-specific AS, we observe significant enrichment of putative PTBP1 binding sites in sequences surrounding mammalian-specific AS events relative to sequences surrounding chicken- and frog-specific AS events ($P < 0.002$, one-sided Fisher's exact test; Fig. 5B).

The species-classifying AS events in genes associated with other functions are also significantly underrepresented in exons that overlap folded protein domains ($P < 0.0002$, Pearson's χ^2 proportion test; fig. S15A). However, the species-classifying AS events are significantly enriched in frame-preserving exons relative to nonclassifying species-specific AS events ($P \leq 0.05$, Pearson's χ^2 proportion test; Fig. 5D and fig. S15B), and they are also enriched in protein regions predicted to be disordered. Disordered regions generally reside on protein surfaces and are known to play critical roles in ligand interactions and cell signaling (32, 33), and recent work has shown that tissue-regulated exons enriched in predicted disordered sequences often function in remodeling protein-protein interactions (PPIs) (34, 35). The species-classifying AS events thus possess multiple features of functionally important exons.

Discussion. Our results show that organs of primate species have significantly higher cassette exon AS frequencies than do organs of other vertebrate species. Moreover, overall organ AS

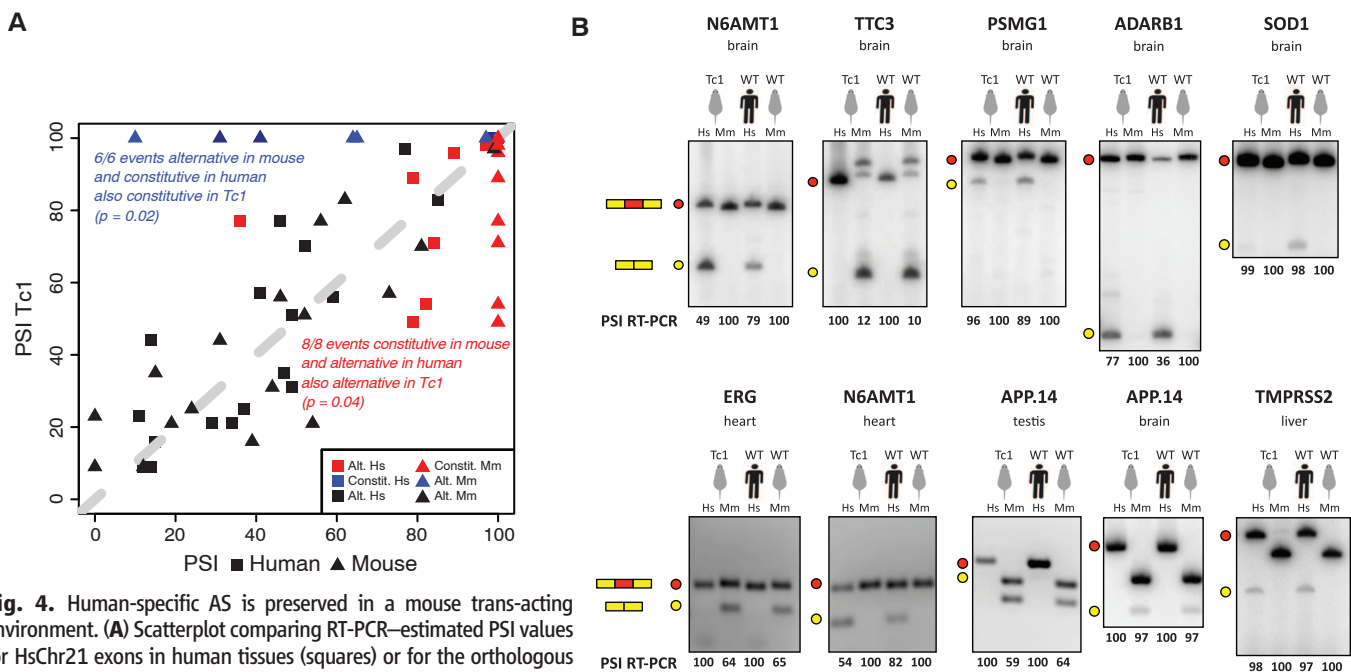


Fig. 4. Human-specific AS is preserved in a mouse trans-acting environment. **(A)** Scatterplot comparing RT-PCR-estimated PSI values for HsChr21 exons in human tissues (squares) or for the orthologous exons in the corresponding wild-type mouse tissues (triangles) (x axis), and the same HsChr21 exons in Tc1 mouse tissues (y axis) [13 different pairs of orthologous exons, 31 total pairs of orthologous AS events (26)]. Data points are colored according to whether the represented splicing events are alternative in human and constitutive in mouse (red), constitutive in human and alternative in mouse (blue), or alternative in both species (black) (26). Dark blue: exons with $PSI > 95\%$ and $< 100\%$ in human and $PSI < 50\%$ in mouse. Identity line is in dashed gray. P values for one-sided Fisher's exact tests are indicated. **(B)** RT-PCR experiments measuring PSI levels for pairs of orthologous human and mouse exons using species-specific primer pairs (26), for exons

alternative in one species and constitutive or near-constitutive ($PSI > 95\%$) in the other. Human Chr21 exons were analyzed in Tc1 mouse tissues and normal human tissues; the orthologous mouse exons were analyzed in corresponding tissues from both the Tc1 and wild-type mouse strains. Quantification of PSI levels, human gene names (followed by the exon number, when more than one exon for the same gene was studied), and tissues are indicated. Red and yellow dots indicate exon-included and exon-skipped isoforms, respectively. Note that each set of species-specific primers amplifying the orthologous splice isoforms generates size-distinct RT-PCR products.

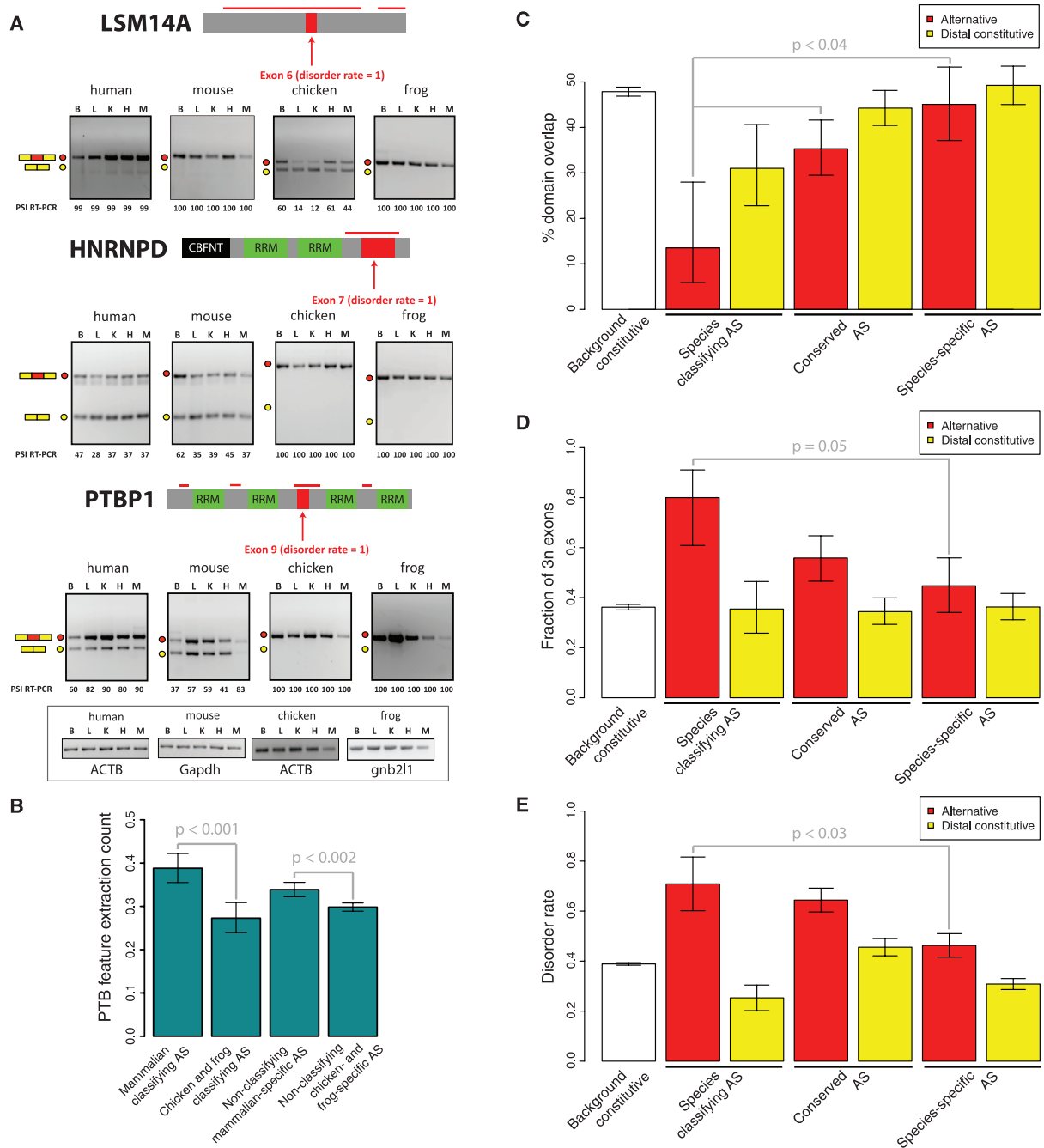


Fig. 5. Characterization of species-classifying AS events. **(A)** Domain diagrams and RT-PCR experiments measuring PSI levels for representative species-classifying AS events. Predicted highly disordered protein regions are depicted by red bars above the domain diagrams. Human gene names are indicated, as are the locations of the classifying alternative exon, species, and organs (B, brain; L, liver; K, kidney; H, heart; M, muscle) in which the validations were performed, as well as quantification of PSI levels. Red and yellow dots indicate exon-included and exon-skipped isoforms, respectively. RT-PCR assays were performed as in Fig. 4. **(B)** Comparisons of the average number of PTBP1-related cis-features associated with mammalian-classifying and nonclassifying species-dependent AS events, and the corresponding classes of species-dependent AS events in chicken and frog. Error bars: 95% confidence intervals, Pearson's χ^2 proportion tests. Statistically significant differences are indicated. **(C)** Comparison of the proportion of residues in different types of coding exons in proteins with nucleic acid-binding domains that overlap folded protein domains. Classes of alternative exons (red bars) analyzed are (i) "species-classifiers," AS events that discriminate species; (ii) "conserved AS," exons detected as AS in at least two of the analyzed species; and

(iii) "species-specific AS," exons that are alternatively spliced in only one of the analyzed species. For each class of alternative exon, distal constitutive exons (separated from the alternative exon by at least two exons) in the same genes are analyzed (yellow bars). "Background constitutive" refers to all exons that are constitutively spliced in all tissues (white bars). Error bars: 95% confidence intervals, Pearson's χ^2 proportion tests. Statistically significant differences between species-classifying AS events and each of the other classes of AS events are indicated. **(D)** Proportion of exons in proteins with nucleic acid binding domains that preserve the reading frame when included/skipped in transcripts (3n exons, exons with nucleotide length multiple of three nucleotides). Types of exons analyzed are as in (C). Error bars: 95% confidence intervals, Pearson's χ^2 proportion tests. Statistically significant differences between species-classifying AS events and the other species-specific AS events are indicated. **(E)** Proportion of predicted intrinsically disordered amino acids in different classes of exons in proteins with nucleic acid binding domains, as described in (C). Error bars denote SE of the associated distributions. Statistically significant differences between species-classifying AS events and the other species-specific AS events are indicated (Student's t test).

profiles more strongly reflect the identity of a species than they do organ type. This contrasts with organ-dependent differences in mRNA expression, which are largely conserved throughout vertebrate evolution [this study and (6, 7)]. We propose that the rapid divergence in AS patterns in vertebrate organs may have played a more widespread role in shaping species-specific differences than did changes in mRNA expression.

Our work offers conclusive evidence that the reassortment of splicing code features can account for the majority of AS differences between vertebrate species. Consistent with this observation, species-classifying exons identified in this study are often found in genes encoding trans-acting regulators but are underrepresented in the nucleic acid binding domains of these proteins. Instead, they are highly enriched in disordered regions, which are known to function in signaling and in mediating PPIs. Because these AS changes affect trans-acting factors involved in gene regulation, they represent an additional mechanistic basis for the remarkable diversification in AS and other transcriptomic changes associated with phenotypic change among vertebrate species.

References and Notes

1. C. P. Ponting, *Nat. Rev. Genet.* **9**, 689 (2008).
2. A. M. Heimberg, L. F. Sempere, V. N. Moy, P. C. Donoghue, K. J. Peterson, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2946 (2008).
3. V. J. Lynch, G. P. Wagner, *Evolution* **62**, 2131 (2008).

4. A. C. Meireles-Filho, A. Stark, *Curr. Opin. Genet. Dev.* **19**, 565 (2009).
5. P. J. Wittkopp, G. Kalay, *Nat. Rev. Genet.* **13**, 59 (2011).
6. D. Brawand *et al.*, *Nature* **478**, 343 (2011).
7. E. T. Chan *et al.*, *J. Biol.* **8**, 33 (2009).
8. A. N. Brooks *et al.*, *Genome Res.* **21**, 193 (2011).
9. J. C. Castle *et al.*, *Nat. Genet.* **40**, 1416 (2008).
10. M. Irimia, B. J. Blencowe, *Curr. Opin. Cell Biol.* **24**, 323 (2012).
11. N. Jelen, J. Ule, M. Živin, R. B. Darnell, *PLoS Genet.* **3**, e173 (2007).
12. A. Kalsotra, T. A. Cooper, *Nat. Rev. Genet.* **12**, 715 (2011).
13. H. Kerai, G. Lev-Maor, G. Ast, *Nat. Rev. Genet.* **11**, 345 (2010).
14. D. D. Licatalosi, R. B. Darnell, *Nat. Rev. Genet.* **11**, 75 (2010).
15. T. W. Nilsen, B. R. Graveley, *Nature* **463**, 457 (2010).
16. J. A. Calarco *et al.*, *Genes Dev.* **21**, 2963 (2007).
17. S. Gelfman *et al.*, *Genome Res.* **22**, 35 (2012).
18. E. O. Gracheva *et al.*, *Nature* **476**, 88 (2011).
19. M. Irimia, J. L. Rukov, S. W. Roy, J. Vinther, J. Garcia-Fernandez, *Bioessays* **31**, 40 (2009).
20. G. Lev-Maor, R. Sorek, N. Shomron, G. Ast, *Science* **300**, 1288 (2003).
21. B. Modrek, C. J. Lee, *Nat. Genet.* **34**, 177 (2003).
22. Q. Pan *et al.*, *Trends Genet.* **21**, 73 (2005).
23. Y. Terai, N. Morikawa, K. Kawakami, N. Okada, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12798 (2003).
24. G. W. Yeo, E. Van Nostrand, D. Holste, T. Poggio, C. B. Burge, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2850 (2005).
25. X. H. Zhang, L. A. Chasin, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 13427 (2006).
26. See supplementary materials on Science Online.
27. Y. Barash *et al.*, *Nature* **465**, 53 (2010).
28. H. Y. Xiong, Y. Barash, B. J. Frey, *Bioinformatics* **27**, 2554 (2011).
29. A. O'Doherty *et al.*, *Science* **309**, 2033 (2005).

30. A. Corriero, J. Valcárcel, *Mol. Cell* **36**, 918 (2009).
31. M. C. Wollerton *et al.*, *RNA* **7**, 819 (2001).
32. J. Bellay *et al.*, *Mol. Biosyst.* **8**, 185 (2012).
33. A. K. Dunker, I. Silman, V. N. Uversky, J. L. Sussman, *Curr. Opin. Struct. Biol.* **18**, 756 (2008).
34. J. D. Ellis *et al.*, *Mol. Cell* **46**, 884 (2012).
35. M. Buljan *et al.*, *Mol. Cell* **46**, 871 (2012).

Acknowledgments: We thank H. Han, D. O'Hanlon, M. Lukk, and the Cambridge Research Institute Biorepository, Genomics, and Bioinformatics Units for expert assistance; V. Tybulewicz, E. Fisher, R. Cohen, J. Wade, D. Simpson, and J. Gurdon for tissues and mouse strains; and T. Hughes, C. Ouzounis, F. Roth, and past and present members of the Blencowe laboratory for helpful discussions and comments on the manuscript. Supported by grants from the Canadian Institutes of Health Research (B.J.B. and B.J.F.); Canadian Cancer Society (B.J.B.); Genome Canada (through the Ontario Genomics Institute) and the Ontario Research Fund (B.J.B., B.J.F. and others); Natural Sciences and Engineering Research Council of Canada (B.J.F.); EMBO YIP and ERC starting Grants (D.T.O.); CIHR postdoctoral and Marie Curie IOF fellowships (N.L.B.-M.); an SNF postdoctoral fellowship (C.K.); an NSERC studentship (S.G.); and a Human Frontiers Science Program Organization long-term fellowship (M.I.). RNA-Seq data sets analyzed in this study can be accessed through GEO (GSE41338, GSE30352) and ArrayExpress (E-MTAB-513).

Supplementary Materials

www.sciencemag.org/cgi/content/full/338/6114/1587/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S15
Tables S1 to S9
References (36–78)

24 September 2012; accepted 14 November 2012
10.1126/science.1230612

Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues

Jason Merkin,¹ Caitlin Russell,¹ Ping Chen,^{1,3} Christopher B. Burge^{1,2*}

Most mammalian genes produce multiple distinct messenger RNAs through alternative splicing, but the extent of splicing conservation is not clear. To assess tissue-specific transcriptome variation across mammals, we sequenced complementary DNA from nine tissues from four mammals and one bird in biological triplicate, at unprecedented depth. We find that while tissue-specific gene expression programs are largely conserved, alternative splicing is well conserved in only a subset of tissues and is frequently lineage-specific. Thousands of previously unknown, lineage-specific, and conserved alternative exons were identified; widely conserved alternative exons had signatures of binding by MBNL, PTB, RBFOX, STAR, and TIA family splicing factors, implicating them as ancestral mammalian splicing regulators. Our data also indicate that alternative splicing often alters protein phosphorylatability, delimiting the scope of kinase signaling.

Alternative pre-mRNA processing can result in mRNA isoforms that encode distinct protein products, or may differ exclusively in untranslated regions, potentially affecting mRNA stability, localization, or translation (1). It can also produce nonfunctional mRNAs that are targets of nonsense-mediated mRNA decay, serving to control gene expression (2). Most human alternative splicing is tissue-regulated (3, 4), but the extent to which tissue-specific splicing patterns are conserved across

mammalian species has not yet been comprehensively studied.

To address outstanding questions about the conservation and functional importance of tissue-specific splicing, we conducted transcriptome sequencing (RNA-Seq) analysis of nine tissues from five vertebrates, consisting of four mammals and one bird. The species, chosen on the basis of the quality of their genomes (all high-coverage finished or draft genomes) and their evolutionary relationships, include the rodents mouse and rat,

the rhesus macaque, a nonrodent/nonprimate boreutherian, cow, and chicken as an outgroup. These relationships allow for the evaluation of transcriptome changes between species with divergence times ranging from <30 million years to >300 million years (Fig. 1A). Our sequencing strategy used paired-end short or long read sequencing of poly(A)-selected RNA. In total, we generated more than 16 billion reads (>8 billion read pairs) totaling over 1 trillion bases (3, 5) (table S1). The data were mapped to the relevant genomes with software that can identify novel splice junctions and isoforms (6).

To assess coverage of genes, we compared these de novo annotations with existing Ensembl annotations. We detected >211,000 (97%) of the ~217,000 annotated exons in mouse, and similarly high fractions in most other species, including more than 99% of exons in chicken (table S1). We estimated that nearly all multi-exonic genes in the species studied are alternatively spliced (fig. S1) (3).

¹Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA. ²Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02142, USA. ³Systems Biology Laboratory, Research Programs Unit, Genome-Scale Biology and Institute of Biomedicine, Biochemistry and Developmental Biology, University of Helsinki, Haartmaninkatu 8, Helsinki, FIN-00014, Finland.

*To whom correspondence should be addressed. E-mail: cburge@mit.edu