Evolutionary dynamics of gene and isoform regulation in mammalian tissues*

Jason Merkin[1], **Caitlin Russell**[1], **Ping Chen**[1,3], and **Christopher B. Burge**[1,2,4]

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

Abstract

Most mammalian genes produce multiple distinct mRNAs through alternative splicing, but the extent of splicing conservation is not clear. To assess tissue-specific transcriptome variation across mammals, we sequenced cDNA from 9 tissues from 4 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 novel, 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 (NMD), 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 significance of tissue-specific splicing, we conducted transcriptome sequencing (RNA-Seq) analysis of 9 tissues from 5 vertebrates, consisting of 4 mammals and one bird. The species, chosen based on 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 non-rodent/non-primate boroeutherian, 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 polyA-selected RNA. In total, we generated over 16 billion reads (>8 billion read pairs) totaling over 1 trillion bases (3, 5) (table S1). The data were mapped to the relevant genomes using software that can identify novel splice junctions and isoforms (6).

To assess coverage of genes, these de novo annotations were compared with existing Ensembl annotations. We detected over 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).

All tissues have conserved expression signatures

To explore the expression relationships between the samples, we used hierarchical clustering based on Jensen-Shannon divergence (JSD) distances between the expression of orthologous genes. A clear pattern emerged in the resulting dendrogram (Fig. 1A, table S2). Samples of the same tissue from different individuals of the same species were invariably the most similar, followed by samples from the same tissue from other species, with few exceptions. This "tissue-dominated clustering" pattern indicates that most tissues possess a conserved gene expression signature and suggests that conserved gene expression differences underlie tissue identity in mammals (5, 7). Since gene expression varies by cell type, some observed differences could reflect changes in cell type composition. The most notable exceptions to tissue dominance were that some chicken muscle samples clustered with chicken heart rather than mammalian muscle, and that chicken lung, colon and spleen samples clustered with each other rather than with their mammalian counterparts. These exceptions suggest that species-specific divergence in expression begins to exceed conserved tissue-specific differences at a phylogenetic distance of ~300 MY, corresponding to the split between birds and mammals.

Some tissues have conserved splicing signatures

To understand the splicing relationships between the samples, we performed an analogous clustering analysis using the "percent spliced in" (PSI or Ψ) values of the exons that were alternatively spliced in all species containing them. PSI values, the fraction of a gene's mRNAs that contain the exon, were calculated from transcript abundance measurements (8) (fig. S2), and were clustered using the same metric (Fig. 1B). Samples of the same tissue from individuals of the same species almost invariably clustered together. However, at larger distances a more complex pattern emerged. Tissue-dominated clustering was observed for brain and for the combination of heart and muscle, indicating that these tissues have strong splicing signatures conserved between mammals and chicken, and the rodent testis samples also clustered together. By contrast, samples from the remaining tissues (colon, kidney, liver, lung, spleen) exhibited "species-dominated clustering", forming distinct clusters by species rather than by tissue. This trend suggests that alternative splicing patterns specific to this latter group of tissues are less pronounced or less conserved than those of brain, testis, heart and muscle (fig. S3). The greater prominence of species-dominated

clustering of PSI values suggests that exon splicing is more often affected by lineage-specific changes in *cis*-regulatory elements (9) and/or *trans*-acting factors than is gene expression (6). Lineage-specific changes in splicing factor expression may have contributed to the tendency of splicing patterns to cluster by species more often than by tissue (table S3, fig. S4).

Features of conserved, tissue-specific alternative exons

A subset of several hundred alternative exons exhibited highly conserved tissue-specific splicing patterns. The gene for eukaryotic translation elongation factor 1 delta (*EEF1δ*) (Fig. 2A) and many other examples in our data demonstrate that highly tissue-specific patterns of splicing can be conserved for hundreds of millions of years (9).

To assess the phylogenetic distribution of alternative splicing events across mammals, we grouped exons by the inferred age of alternative splicing, defined as PSI < 97%. Out of ~48,000 internal exons with clear orthologs in chicken and at least two mammals, we identified exons alternatively spliced in a species- or rodent-specific manner as well as ~500 "broadly alternative" exons with alternative splicing observed in all mammals studied (Fig. 2B, table S4). Conversely, we identified exons that were constitutively spliced in a lineage-specific manner (and alternatively spliced elsewhere), representing losses of alternative splicing. Using data from the Illumina human Body Map 2.0 dataset, rhesus-specific alternative exons were twice as likely to be detected as alternatively spliced in human as were exons with exon skipping detected only in a single rodent (fig. S5), consistent with the closer phylogenetic relationship of human to rhesus than to mouse or rat. In addition, more than 500 exons were identified whose phylogenetic splicing patterns imply multiple changes between constitutive and alternative splicing during mammalian evolution, suggesting frequent inter-conversion between constitutive and alternative splicing (10).

We observed a monotonic increase in tissue specificity within mouse as the phylogenetic breadth of alternative splicing increased from 1 to 4 mammals (Fig. 2B,C). The fraction of exons that preserved reading frame in both inclusion and exclusion isoforms also increased from ~40% to ~70% with increasing phylogenetic breadth of alternative splicing. These patterns suggest that more broadly occurring (ancient) alternative splicing events function primarily to generate distinct protein isoforms, which are often tissue-specific (11). On the other hand, while more lineage-restricted (recently evolved) alternative splicing events contribute more often to regulation involving reading frame disruption, which may yield truncated or nonfunctional mRNAs or proteins or serve to down-regulate expression, usually in a less tissue-specific manner.

Splice site changes may convert alternative to constitutive splicing

Exons that recently converted from constitutive to alternative splicing had significantly weaker 3' and 5' splice sites in the alternative splicing species than in their constitutively spliced orthologs (Fig. 2C) (10). However, recent conversion from alternative to constitutive splicing was not associated with significant changes in 5' or 3' splice site strength, suggesting involvement of other events such as loss of negative-acting *cis*-regulatory elements. Constitutive exons that converted to alternative splicing in other species tended to have weaker splice sites than maintained constitutive exons (P < 0.01, rank-sum test), suggesting that exons with weaker splice sites may be predisposed to convert to alternative splicing. We found that exons with nearby G-runs (often bound by hnRNP H family proteins) were 25–60% more likely to have converted from constitutive to alternative splicing (fig. S6) (12).

Exons alternatively spliced in all mammals tended to have the weakest 5' and 3' splice sites, approximately 1 bit weaker than maintained constitutively spliced exons (Fig. 2C) (13). These exons had mean PSI values that were closer to 50% than other exon groups (Fig. 2C), suggesting that weaker splice sites may have evolved in these exons to enable a broader range of exon inclusion levels.

Splicing *cis*-regulatory elements located adjacent to (or within) alternative exons often confer regulation through interaction with cell type- or condition-specific protein factors (14). Using a large set of intronic splicing regulatory elements (ISRE) motifs recognized by both tissue-specific and broadly expressed splicing factors derived from (15, 16), reduced motif turnover was observed in exons alternatively spliced in multiple species relative to constitutive or recently converted alternative exons (Fig. 2D) (11, 17). Exons that converted from alternative to constitutive splicing in one or both rodents showed substantially increased turnover of ISREs than mammalian-wide alternative exons (Fig. 2D), suggesting that mutations affecting ISREs may contribute to these conversions.

Tissue-specific regulatory motifs accumulate in broadly alternative exons

Using vertebrate whole-genome alignments, strong sequence conservation of only the exon and core splice site motifs was observed in broadly constitutive exons and exons that recently acquired alternative splicing. However, increased sequence conservation both within the exon and extending at least 70 bases into the intron on either side was observed with increasing phylogenetic breadth of alternative splicing (Fig. 3A), suggesting the occurrence of purifying selection on adjacent ISREs and providing support for the reliability of these exon classifications.

To assess the nature of potential regulatory elements present in introns adjacent to alternative exons, we ranked pentanucleotides (5mers) based on their relative frequency of occurrence in introns downstream of broadly alternative exons relative to constitutive exons using an information criterion (6). Among the top ten 5mers in this ranking were perfect or near-perfect matches to consensus motifs for tissue-specific splicing regulatory factors, including those of the MBNL, PTB, RBFOX, STAR and TIA families of splicing factors (18) (Fig. 3B, table S5). Presence of motifs associated with almost all of these splicing factor families was conserved downstream of broadly alternative exons more than two standard deviations more often than control motifs (Fig. 3C), implying strong selection to maintain their presence. Pronounced enrichment of these motifs was restricted primarily to exons with broad alternative splicing (4 species), with only modest enrichment downstream of rodent-specific alternative exons and little to no enrichment near mouse-specific alternative exons (Fig. 3D, fig. S7). These observations suggested that exons with more ancient alternative splicing – which are more often tissue-specific (Fig. 2B) – are more reliant for their regulation on a distinct subset of ISRE motifs corresponding to the tissue-specific factors listed above (MBNL, RBFOX, etc.).

To explore this hypothesis, we analyzed cross-linking / immunoprecipitation-sequencing (CLIP-Seq) data to assess the transcriptome-wide binding of the mouse splicing factor muscleblind-like 1 (MBNL1) (19). Greater phylogenetic breadth of alternative splicing was associated with ~3-fold increased frequency of *in vivo* MBNL1 binding (Fig. 3E). Presence of a MBNL motif was associated with increased binding near alternative but not constitutive exons (Fig. 3F), suggesting that motif presence is necessary but not sufficient for strong binding *in vivo*. As a group, broadly alternative exons have somewhat higher density of MBNL motifs (Fig. 3B), but increased frequency of MBNL binding was observed even when comparing to subsets of constitutive or more narrowly alternative exons with identical MBNL motif counts (Fig. 3G). These observations suggest that broadly alternative exons

have evolved features beyond motif abundance (such as favorable RNA structural features) to enhance binding of MBNL family splicing regulators. This phenomenon may extend to other factors (fig. S8).

Alternative splicing alters phosphorylation potential

Exons whose presence was widely conserved (at least 4 out of 5 species) were classified based on the tissue specificity and evolutionary conservation of their splicing patterns using JSD-based metrics (6) into constitutive exons and four groups of alternative exons grouped by the degree of tissue-specificity and evolutionary conservation of their splicing patterns. Functional analysis of species-specific alternative exons yielded few significant biases (table S6). However, analysis of the tissue-specific conserved group using DAVID (20, 21) identified a number of significantly enriched keywords and Gene Ontology categories, including several related to cell-cell junctions and cytoskeleton (Fig. 4A), suggesting that these splicing events may contribute to differences in cell structure, cell motility and tissue architecture (22). The most enriched keyword "alternative splicing", reflects simply the abundant alternative splicing of this set of genes; the next most significantly enriched keyword was phosphoprotein.

To explore this connection to phosphorylation, we used Scansite (23) to predict phosphorylation sites in peptides encoded by different subsets of exons. Tissue-specific alternatively spliced exons, including both the conserved and non-conserved subsets, contained about 40% more predicted phosphorylation sites than other classes of exons (Fig. 4B and fig. S9). A comparable degree of enrichment for phosphorylation sites was observed in these exons using the curated Phosphosite database (24) of experimentally determined phosphorylation sites (Fig. 4B). Phosphorylation site density in exons was correlated with phylogenetic breadth of alternative splicing, (Fig. 4C, fig. S10). These observations suggest that tissue-specific alternative splicing is often used to alter the potential for protein phosphorylation, which can alter protein stability, enzymatic activity, subcellular localization and other properties.

Exon 20 of the mouse tight junction protein 1 (TJP1) gene exhibits strongly tissue-specific alternative splicing and encodes a peptide containing an established phosphorylation site (25) that is predicted to be phosphorylated by ERK1 (aka MAPK3). In rhesus, ERK1 expression was above its median value in colon, lung, testis and brain (therefore referred to as "kinase high" tissues) relative to liver, heart, muscle and spleen ("kinase low" tissues). The PSI value of TJP1 exon 20 was much more variable in the kinase high tissues, ranging from 9% in testis to 90% in in colon – a switch score of 81% – than in the kinase low tissues, where it had a switch score of 38%. Similar trends were observed in cow (Fig. 4D), and in rat and mouse (not shown).

To explore this phenomenon, we analyzed the splicing patterns of exons that contained predicted phosphorylation sites in relation to the expression of the associated kinases. To characterize the relationship between each exon-kinase pair, the "kinase switch index" (KSI) was defined as the switch score in "kinase high" tissues minus the switch score in "kinase low" tissues (see example in Fig. 4D). We used RNA-Seq estimates of kinase expression, which were reasonably well correlated with *in vivo* kinase activity patterns (r = 0.71, fig. S11). We observed that phosphorylation of sites within conserved, tissue-regulated exons is more tissue-specific than in other sets of exons (fig. S12) and that these exons exhibit substantially elevated KSI values relative to shuffled controls (Fig. 4D, table S7).

The above observations suggest a model in which tissue-regulated alternative splicing delimits the scope of tissues in which a kinase can phosphorylate a target. For example, the TJP1 protein mentioned above is a cytoplasmic constituent of the tight junction complex

implicated in the timing of tight junction formation, and its phosphorylation is involved in tight junction dynamics (26, 27). The testes display unique tight junction biology in that tight junctions regularly dissolve and reform to permit passage of preleptotene spermatocytes (28). The specific exclusion of exon 20 in the mammalian testis may allow TJP1 to escape phosphorylation that would otherwise alter the tight junction association kinetics required for normal testis function (29). Another example, with KSI value closer to the mean value, is an alternative exon in Drosha, a protein required for processing of microRNA primary transcripts (fig. S13). Phosphorylation of Drosha confers nuclear localization, which is required for its normal function in microRNA biogenesis (30). Therefore, splicing of Drosha may be used to alter the level of phosphorylatable Drosha protein, potentially influencing microRNA abundance in different cells or tissues.

We identified a large number of other exon-kinase pairs with elevated KSIs, including prominent kinases involved in development, cell cycle and cancer (e.g., Akt1, Clk2, PKC, Src) and targets with important roles in tissue biology such as Atf2, Enah, and Vegfa (fig. 14, table S7). Their elevated KSIs suggest that splicing is often used to focus the scope of signaling networks, connecting specific kinases to specific targets in a more cell- or tissue-restricted fashion than would occur from expression alone.

Taken together, the analyses described here reveal two disparate facets of mammalian alternative splicing. We identify a core of ~500 exons with conserved alternative splicing in mammals and high sequence conservation. These exons often encode phosphorylation sites and their tissue-specific splicing is likely to have substantial impacts on the outputs of signaling networks. Conversely, we observe extensive variation in the splicing of these exons between species, often exceeding intra-species differences between tissues, suggesting that changes in splicing patterns often contribute to evolutionary rewiring of signaling networks.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References and Notes

1. Stamm S, et al. Gene. 2005 Jan 3.344:1. [PubMed: 15656968]
2. Lareau LF, Brooks AN, Soergel DA, Meng Q, Brenner SE. Advances in experimental medicine and biology. 2007; 623:190. [PubMed: 18380348]
3. Wang ET, et al. Nature. 2008 Nov 27.456:470. [PubMed: 18978772]
4. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Nature genetics. 2008 Dec.40:1413. [PubMed: 18978789]
5. Brawand D, et al. Nature. 2011 Oct 20.478:343. [PubMed: 22012392]
6. Materials and methods are available as supplementary material on Science Online.
7. Chan ET, et al. Journal of biology. 2009; 8:33. [PubMed: 19371447]
8. Trapnell C, et al. Nature biotechnology. 2010 May.28:511.
9. Jelen N, Ule J, Zivin M, Darnell RB. PLoS genetics. 2007 Oct.3:1838. [PubMed: 17937501]

10. Lev-Maor G, et al. PLoS genetics. 2007 Nov.3:e203. [PubMed: 18020709]

11. Yeo GW, Van Nostrand E, Holste D, Poggio T, Burge CB. Proc Natl Acad Sci U S A. 2005 Feb 22.102:2850. [PubMed: 15708978]

12. Xiao X, et al. Nature structural & molecular biology. 2009 Oct.16:1094.

13. Baek D, Green P. Proc Natl Acad Sci U S A. 2005 Sep 6.102:12813. [PubMed: 16123126]

14. Matlin AJ, Clark F, Smith CW. Nat Rev Mol Cell Biol. 2005 May.6:386. [PubMed: 15956978]

15. Huelga SC, et al. Cell reports. 2012 Feb 23.1:167. [PubMed: 22574288]

16. Cook KB, Kazan H, Zuberi K, Morris Q, Hughes TR. Nucleic acids research. 2011 Jan.39:D301. [PubMed: 21036867]

17. Sorek R, Ast G. Genome research. 2003 Jul.13:1631. [PubMed: 12840041]

18. Ladd AN, Cooper TA. Genome biology. 2002 Oct 23.3 reviews0008.

19. Wang ET, et al. Cell. 2012 Aug 17.150:710. [PubMed: 22901804]

20. Huang da W, Sherman BT, Lempicki RA. Nature protocols. 2009; 4:44.

21. Huang da W, Sherman BT, Lempicki RA. Nucleic acids research. 2009 Jan.37:1. [PubMed: 19033363]

22. Shapiro IM, et al. PLoS genetics. 2011 Aug.7:e1002218. [PubMed: 21876675]

23. Obenauer JC, Cantley LC, Yaffe MB. Nucleic Acids Res. 2003 Jul 1.31:3635. [PubMed: 12824383]

24. Hornbeck PV, Chabra I, Kornhauser JM, Skrzypek E, Zhang B. Proteomics. 2004 Jun.4:1551. [PubMed: 15174125]

25. Dephoure N, et al. Proceedings of the National Academy of Sciences of the United States of America. 2008 Aug 5.105:10762. [PubMed: 18669648]

26. Samak G, Aggarwal S, Rao RK. American journal of physiology. Gastrointestinal and liver physiology. 2011 Jul.301:G50. [PubMed: 21474650]

27. Sabath E, et al. Journal of cell science. 2008 Mar 15.121:814. [PubMed: 18285450]

28. Mruk DD, Cheng CY. Philosophical transactions of the Royal Society of London. Series B, Biological sciences. 2010 May 27.365:1621.

29. Aggarwal S, Suzuki T, Taylor WL, Bhargava A, Rao RK. The Biochemical journal. 2011 Jan 1.433:51. [PubMed: 20961289]

30. Tang X, Zhang Y, Tucker L, Ramratnam B. Nucleic acids research. 2010 Oct.38:6610. [PubMed: 20554852]

31. Yeo G, Burge CB. J Comput Biol. 2004; 11:377. [PubMed: 15285897]

32. Somel M, et al. Genome research. 2010 Sep.20:1207. [PubMed: 20647238]

33. Yeo G, Holste D, Kreiman G, Burge CB. Genome Biol. 2004; 5:R74. [PubMed: 15461793]

34. Parkhomchuk D, et al. Nucleic Acids Res. 2009 Oct.37:e123. [PubMed: 19620212]

35. Trapnell C, Pachter L, Salzberg SL. Bioinformatics. 2009 May 1.25:1105. [PubMed: 19289445]

36. Roberts A, Pimentel H, Trapnell C, Pachter L. Bioinformatics. 2011 Sep 1.27:2325. [PubMed: 21697122]

37. Paten B, Herrero J, Beal K, Fitzgerald S, Birney E. Genome research. 2008 Nov.18:1814. [PubMed: 18849524]

38. Katz Y, Wang ET, Airoldi EM, Burge CB. Nat Methods. 2010 Dec.7:1009. [PubMed: 21057496]

39. Berretta R, Moscato P. PloS one. 2010; 5:e12262. [PubMed: 20805891]

40. Martinez O, Reyes-Valdes MH. Proceedings of the National Academy of Sciences of the United States of America. 2008 Jul 15.105:9709. [PubMed: 18606989]

41. Lindblad-Toh K, et al. Nature. 2011 Oct 27.478:476. [PubMed: 21993624]

42. Wang Z, et al. PLoS biology. 2010; 8:e1000530. [PubMed: 21048981]

43. Yeo GW, et al. Nature structural & molecular biology. 2009 Feb.16:130.

44. Huttlin EL, et al. Cell. 2010 Dec 23.143:1174. [PubMed: 21183079]

45. Buljan M, et al. Molecular cell. 2012 Jun 29.46:871. [PubMed: 22749400]

46. Knight R, et al. Genome biology. 2007; 8:R171. [PubMed: 17708774]
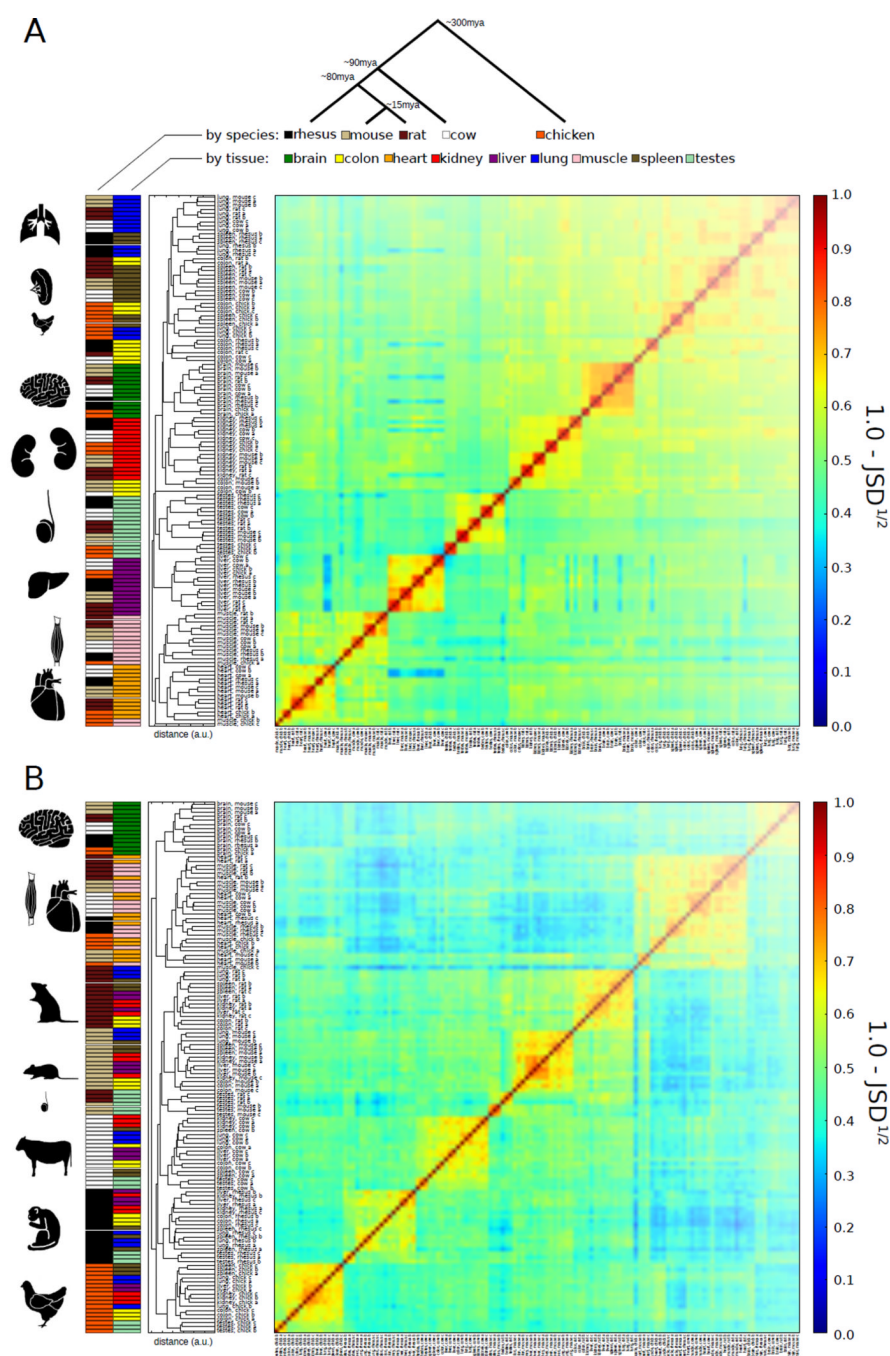
Figure 1. Conservation of expression signatures in all tissues and of alternative splicing signatures in some tissues

A) Clustering of samples based on expression values (FPKM) of singleton orthologous genes present in all 5 species (*n*=7713). Average linkage hierarchical clustering was used with distance between samples measured by the square root of the Jensen-Shannon Divergence (JSD) between the vectors of expression values.

B) Clustering of samples based on PSI values of exons in singleton orthologous genes conserved to chicken, with alternative splicing detected in all individuals analyzed (*n*=489). Clustered as in (A). When the set of genes used in this analysis was clustered by gene

expression rather than PSI values, tissue-dominated clustering was observed, as in (A) (fig. S15).
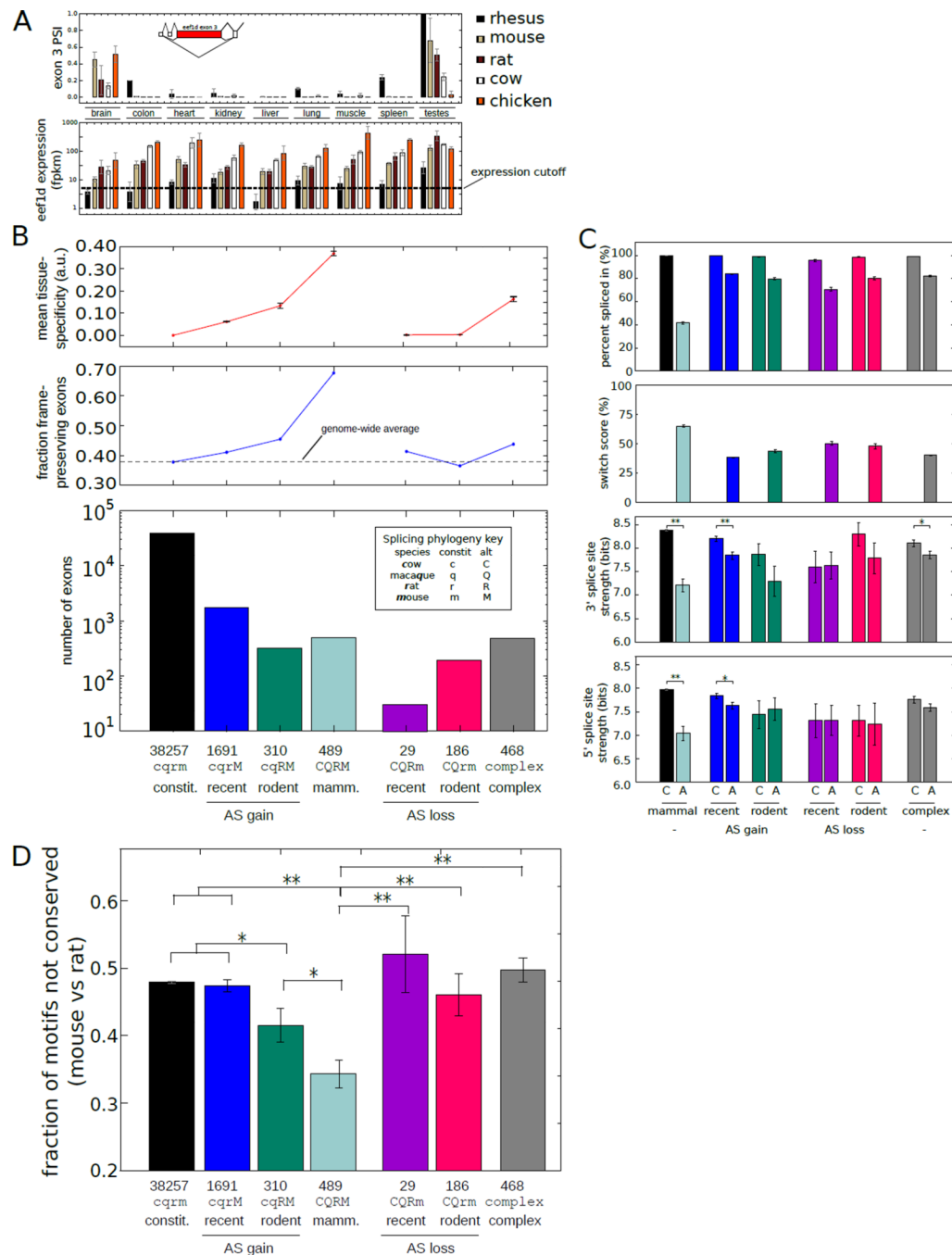
Figure 2. Exonic features associated with evolutionary change in alternative splicing

A) Above: PSI values for eef1d exon 3 across tissues and species analyzed. Below: Eef1d gene expression values. (Mean ± SD of 3 biological replicates). PSI values were calculated only for tissues with FPKM 5. Inset: exon structure of 5' end of eef1d gene (ENSMUSG00000055762).

B) Below: Number of internal exons binned by the age of the inferred alternative splicing based on occurrence in 2 individuals. Above: The fraction of exons with length divisible by 3 and the mean and SEM of the tissue-specificity.

C) Top: mean ± SEM of PSI values of exons binned by the phylogenetic extent of alternative splicing as in (B). Middle: mean ± SEM of 3' splice site scores of exons in each

bin. Bottom: mean ± SEM of 5' splice site scores. Splice sites were scored using the MaxEnt model (31). *Indicates t-test p-value < 0.05. **Indicates P < 0.001.

D) Fraction of regulatory 5mers in the downstream intron that differed between mouse and rat in exons binned by the phylogenetic extent of alternative splicing as in (B) (Mean ± SEM). *Indicates t-test P < 0.05 (t-test). **P < 0.001.
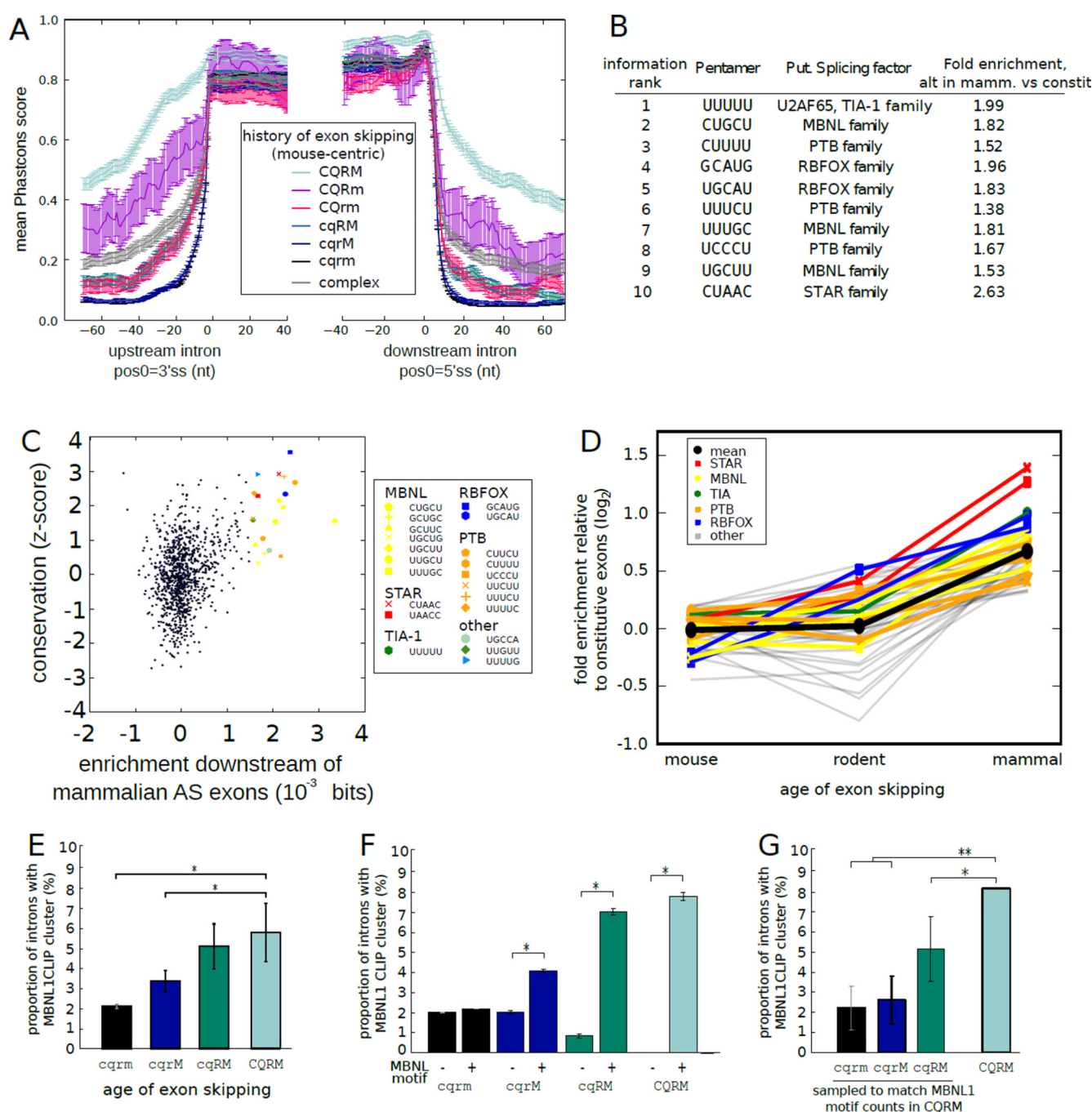
Figure 3. Tissue-specific regulatory motifs accumulate and are preferentially bound in broadly alternative exons

A) Mean ± SEM of Phastcons scores (using the placental mammals alignment, with mouse coordinates) in exons and flanking introns grouped by phylogenetic pattern of alternative splicing. Splicing pattern indicated by letters adjacent to colored bars, as in Fig. 2B.

B) Top ten 5mers in broadly alternative exons relative to constitutive exons ranked by discrimination information (6).

C) The conservation of all 5mers (6) compared with their discrimination information. All 5mers with discrimination information   0.001 bits are highlighted. UUUUU was an outlier in enrichment (0.011 bits) and is not shown.

D) Fold enrichment (log$_2$) relative to constitutive exons in downstream region was plotted for 5mers with discrimination information 0.001 bits for exons grouped by phylogenetic breadth of alternative splicing. 5mers associated with known splicing regulators are shown in color, with mean of all 5mers in black.

E) The fraction of introns containing MBNL1 CLIP-Seq clusters (19) was assessed in introns adjacent to exons with different phylogenetic patterns splicing, as in (A).

F) As in (E), but grouped by presence/absence of a MBNL1 motif. The mean fraction ± SEM of 1000 bootstrap samples is shown. *$P < 0.01$ (binomial test).

G) As in (E), but with exons sampled from each set to match the MBNL motif counts in the CQRM set. Mean ± SD of 1000 samplings is shown for the first 3 groups; observed mean is shown for the CQRM set. *$P < 0.05$, **$P < 0.001$.
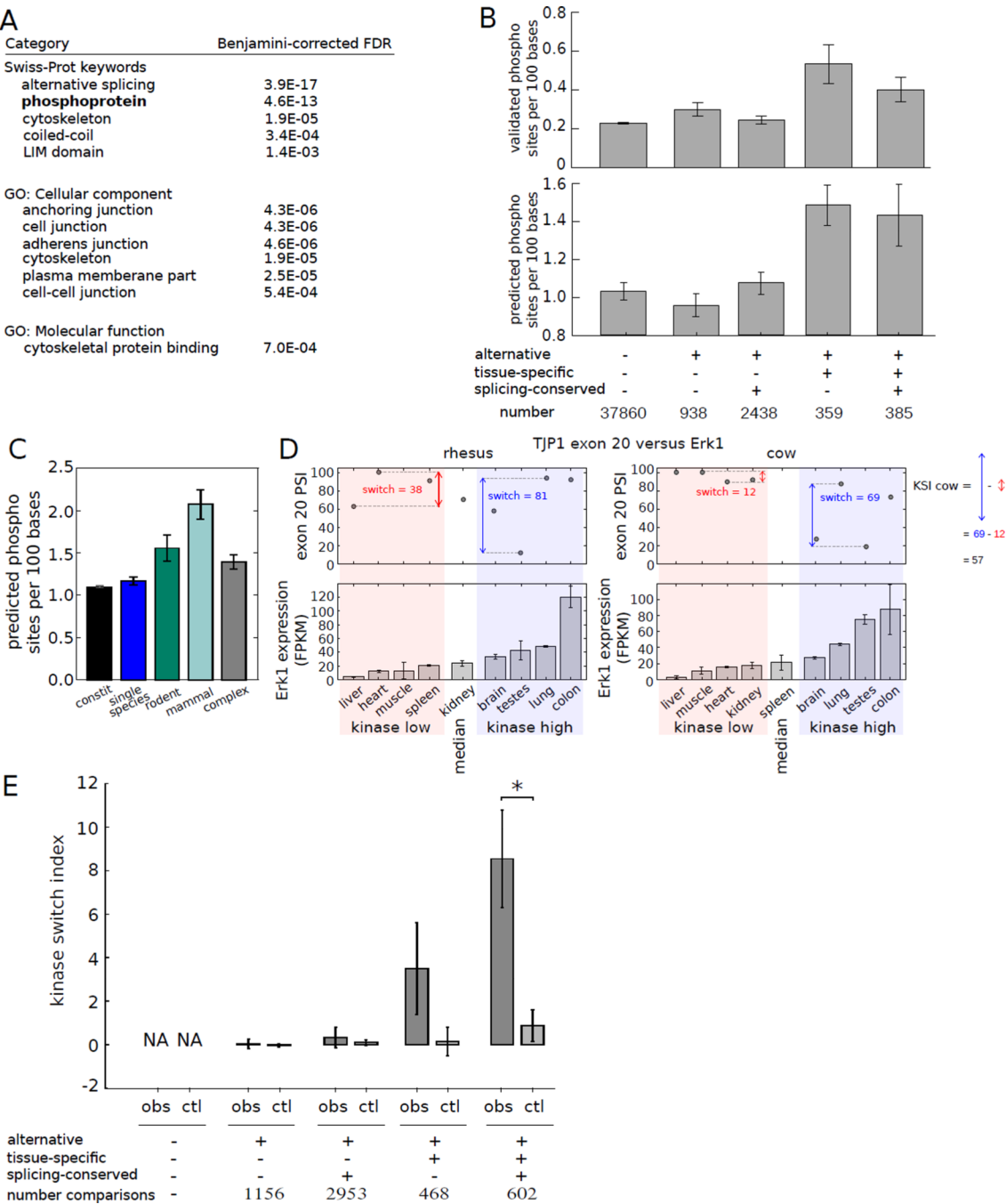
Figure 4. Alternative splicing delimits the scope of protein phosphorylation
A) GO analysis of genes containing tissue-regulated exons whose splicing is conserved.
B) Density of Phosphosite phosphorylation sites (top) or Scansite predicted phosphorylation sites (bottom) in exons grouped by alternative splicing status, tissue-specificity and splicing pattern conservation (6). Mean ± SEM is shown.
C) Mean Scansite predicted phosphorylation site density in exons grouped by phylogenetic breadth of splicing.
D) TJP1 exon 20 splicing has a higher switch score in tissues where Erk1 is expressed above median levels (shaded blue) than where it is expressed below median (shaded pink); KSI is

defined as the difference between these switch scores. PSI value not calculated if TJP1 expression fell below a cutoff (e.g., cow spleen).

E) Mean KSI values for kinase-exon pairs involving the sets of exons as in (B). Mean ± SEM is shown. Observed values (obs) were compared with controls in which PSI values in different tissues were randomly permuted (ctl). Comparisons marked (*) were significant by Mann-Whitney U test (P < 0.005).