Proteomic Evidence for In-Frame and Out-of-Frame Alternatively Spliced Isoforms in Human and Mouse

Rodrigo F. Ramalho, Sujun Li, Predrag Radivojac, and Matthew W. Hahn

Abstract—In order to find evidence for translation of alternatively spliced transcripts, especially those that result in a change in reading frame, we collected exon-skipping cases previously found by RNA-Seq and applied a computational approach to screen millions of mass spectra. These spectra came from seven human and six mouse tissues, five of which are the same between the two organisms: liver, kidney, lung, heart, and brain. Overall, we detected 4 percent of all exon-skipping events found in RNA-seq data, regardless of their effect on reading frame. The fraction of alternative isoforms detected did not differ between out-of-frame and in-frame events. Moreover, the fraction of identified alternative exon-exon junctions and constitutive junctions were similar. Together, our results suggest that both in-frame and out-of-frame translation may be actively used to regulate protein activity or localization.

Index Terms—Mass spectrometry, dual-coding genes, exon-skipping, protein isoform

1 Introduction

The cellular splicing machinery excises introns and joins adjacent exons to make mRNAs. Although the correct ordering and inclusion of exons is critical to the creation of functional transcripts, it is subject to variation by so-called alternative splicing (AS). Alternative splicing results in the exclusion or inclusion of exons and introns into different transcript isoforms from the same gene [1]. There are many different types of alternative splicing, but in mammals the most abundant event is exon-skipping [2]. This type of AS is characterized by the complete presence/absence of an exon in mature transcripts.

Interestingly, half of human and mouse exon-skipping events cause a downstream shift in the codon reading frame because the excluded exon is not a multiple of three in length [3]. These AS events commonly lead to premature termination codons (PTCs) and these transcripts are often degraded by a mechanism known as nonsense mediated decay (NMD; [4], [5], [6]). It has been argued that these frame-disrupting AS events have a major role in the control of gene expression via NMD. This model, known as Regulated Unproductive Splicing and Translation (RUST), proposes a coupling between unproductive AS events (those that result in frame-shifts and premature stop codons in the mRNA) and mRNA degradation through the NMD pathway [4], [7], [8]. However, it is also possible that the function of translated proteins is being regulated by frame-disrupting AS, with alternative protein isoforms having alternative activity levels or alternative localization (e.g., [9]). The possibly functional genes encoding transcripts with multiple reading frames have been called dual-coding genes [10], [11]. These studies have shown evidence for translation of human dual-coding genes by searching translated cDNA sequences against protein sequence databases [10].

Evolutionary conservation has been used extensively to find functionally relevant cases of exon-skipping [12], with many species-specific AS events thought to result from cellular errors that have no function [13]. While skipped exons that are multiples of three in length are over-represented overall relative to random expectations [3], the existence of many evolutionarily conserved AS events that lead to frame disruption [3], [11], [14], [15], [16] suggests that many of them may be functionally relevant. Indeed, the presence of conserved dual-coding genes has also been used to argue for their biological functionality [10], [11].

Despite the widespread occurrence of AS in mammals (e.g., [15], [17]) the evidence for translation of these alternative transcripts is still sparse on a genomic scale. Clearly, this task is complicated by the bias introduced by NMD. Out-of-frame transcripts degraded by NMD will never be detected by sequencing and therefore will be absent in alternative splicing databases. Exploration of human mass spectrometry data has detected tens [18] to hundreds [19] of alternative proteins, while there are at least forty thousand AS events detected in the human transcriptome [20]. This tremendous gap between transcriptomic and proteomic evidence for alternative splicing raises questions about the importance of

- *R.F. Ramalho is with the Department of Biology, Indiana University, 1001 E. 3rd Street, Bloomington, IN 47405. E-mail: rfrusp@gmail.com.*
- *S. Li and P. Radivojac are with the Department of Computer Science and Informatics, Indiana University, 150 S. Woodlawn Avenue, Bloomington, IN 47405. E-mail: {sujli, predrag}@indiana.edu.*
- *M.W. Hahn is with the Department of Biology, Indiana University, 1001 E. 3rd Street, Bloomington, IN 47405 and the Department of Computer Science and Informatics, Indiana University, 150 S. Woodlawn Avenue, Bloomington, IN 47405. E-mail: mwh@indiana.edu.*

Manuscript received 27 Apr. 2015; revised 26 Aug. 2015; accepted 8 Sept. 2015. Date of publication 18 Sept. 2015; date of current version 6 Aug. 2018.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TCBB.2015.2480068

AS in the creation of new proteins, as well as about the power of mass spectrometry platforms to detect such events.

In this study, a large number of mass spectrometry (MS) datasets from several human and mouse tissues were screened to estimate the frequency of exon-exon junctions involved in AS events that maintain or disrupt the reference reading frame. We focused on evolutionarily conserved events recently described in the literature [15], assuming that these events are unlikely to be splicing errors. In order to contextualize our findings, we also estimate the detection frequency of exon-exon junctions involved in non-conserved AS events [17] and the junctions between constitutive exons.

2 Materials and Methods

2.1 Database of Exon-Skipping Events

2.1.1 Reference Protein Sequence Database

We performed MS/MS searches against the reference proteomes of human and mouse. The reference protein sequences were obtained from the files CCDS_protein.20121025.faa and CCDS_protein.20071128.faa available at the NCBI-CCDS database. These files contained 27,523 and 17,704 coding genes from the human and mouse genome, respectively. Fig. S1a, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TCBB.2015.2480068, gives an overview of the bioinformatic workflow for this section. For the analysis of constitutive junctions we obtained the coordinates of exon-exon junctions in the files CCDS.20121025.txt (human) and CCDS.20071128.txt (mouse) and the amino acid sequence encoded by each exon in the files CCDS_protein_exons.20121025.faa (human) CCDS_protein_exons.current.faa (mouse, assembly Mm37.1). By combining information from these files, we found the coordinates of amino acids at the exon boundaries, i.e., exon-exon junctions, of the reference protein sequences. By randomly choosing one exon-exon junction coordinate per gene, we created a database of 18,320 and 16,889 exon-exon junctions for human and mouse, respectively. These coordinates were then searched in the MS/MS search result files, obtained with the reference protein sequences, to find significant peptide-spectrum matches (PSMs) spanning each coordinate (Fig. S1b, available in the online supplemental material).

2.1.2 Alternative Protein Sequence Database

We obtained the coordinates for exon-skipping events that are conserved among mammals (rhesus macaque, mouse, rat, and cow) from supplementary Table S4 in [15]. We obtained the human-specific and mouse-specific exon-skipping events from supplementary Table S3 in [17]. For the exons collected in Merkin et al., 2012 [15], we used Biomart-Ensembl web browser to find the genomic coordinates for the human exons. Each set of exon coordinates were mapped to genomic annotations from the NCBI-CCDS database (file CCDS.20121025.txt). Supplementary Datasets 1 and 2, available online, describe all exon-skipping events analyzed herein. We obtained the exonic nucleotide sequence from the reference human (Hg19) and mouse (NCBI37) genomes via the UCSC genome browser. These data were used to reconstruct the alternative transcripts, i.e. create a coding sequence without the skipped exon. We then used the program transeq from the EMBOSS package (http://emboss.sourceforge.net) to translate all reconstructed alternative transcripts, keeping the reference reading-frame of each gene. For the protein sequences obtained from transcripts with frame-disrupting AS events, only amino acids upstream of the first termination codon were considered in the MS/MS searches (Fig. S1c, available in the online supplemental material).

2.2 Mass Spectrometry Data Analysis

In total, 5,990,793 spectra from seven distinct human tissues and 8,303,412 spectra from six distinct mouse tissues were downloaded from PRIDE [21] and PeptideAtlas [22] (Supplementary Dataset 3, available online). The tissues were chosen to match the tissues analyzed in the RNA-seq studies that define AS events [15], [17].

MS/MS database searches were carried out using Mascot v2.4 (available from Matrix Science under license). Each experimental dataset was analyzed separately against the protein sequence database with the following parameters: tryptic cleavage specificity, up to two missed cleavages; carbamidomethylation of cysteine as fixed modifications and oxidation of methionine as variable modifications were common for all searches. When available, the precursor mass and ion mass tolerance were obtained from the articles describing the experiments. Otherwise, we used the following instrument-specific parameters: for most searches the precursor mass tolerance was set to 2 Da in LCQ or 1.5 Da in LTQ instruments and the ion mass tolerance was set to 1 Da in LCQ or 0.8 Da in LTQ instruments. No peptides outside of the instrument's detection range were included in the analyses.

We obtained the false discovery rate (FDR) using the target/decoy strategy [23] as

$$\mathrm{FDR} = \frac{2 \cdot \mathrm{FP}}{\mathrm{TP} + \mathrm{FP}},$$

where TP is the number of peptide-spectrum matches in the target database and FP is the number of peptide-spectrum matches in the decoy database above a particular score threshold. Each mass spectrometry data set was searched against the target human or mouse proteome (consisted of 27,523 and 17,704 sequences respectively) combined with the same number of shuffled decoy sequences created by the Perl script decoy.pl (http://www.matrixscience.com). The FDR varies in these different human and mouse experiments depending upon the instruments, samples, size of the data, etc. After inspection of the FDR distributions for each human and mouse MS experiment, we observed that peptide-spectrum matches with ion scores $\geq$29 had FDRs ranging from 0.16 to 6.98 percent in various data sets with an average of 1.81 percent.

Each spectrum was searched against our alternative protein sequence database. In the case of frame-preserving exon-skipping events, to report the alternative splicing event the peptide was required to match the junction between the two exons flanking the alternative exon. Peptides that do not span exon-exon junction site cannot distinguish the reference and the alternative proteins in these cases and were therefore not considered. In the case of frame-disrupting exon-skipping events, peptide matches at the junction between the two exons flanking the alternative exon or at sequences downstream of the skipped exon were used as evidence of alternative

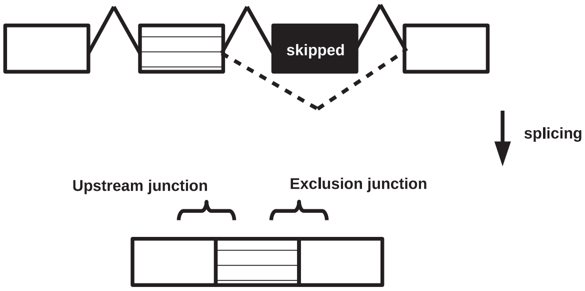


Fig. 1. Types of peptide spectrum matches under comparison. Boxes represent exons, connecting lines represent spliced out introns. The black box represents the skipped exon. Brackets represent PSMs. Upstream junction: Exon-exon junction upstream of the alternative exon. Exclusion junction: Junction reporting exon-skipping.

splicing. For these frame-disrupting cases, all amino acids downstream of the skipped exon are completely different from the reference sequence due to the shift in reading-frame. All significant spectra reporting alternative exon-exon junctions were manually inspected and annotated.

2.3 Gene Expression Data

Gene expression for 21,399 genes from eleven healthy human tissues was obtained from the RNA-Seq Atlas [24].

3 Results

3.1 Conserved Frame-Disrupting Events Drastically Shorten the Reference Protein Length

Evolutionary conservation is used as a proxy for several functionally relevant biological processes, and alternative splicing is not an exception. As previously described [3], [15], [17], [25], we found that at the transcriptional level there is a significant statistical association between the frame-preservation and the evolutionary conservation of the exon-skipping event ($P < 0.0001$, χ^2-test; Table S1a, available in the online supplemental material). The same pattern was observed for mouse; however, the significance was only marginal ($P = 0.05$; Table S1b, available in the online supplemental material). This depletion of frame-disrupting events among all conserved AS events likely reflects the action of natural selection removing splice variants that cause shifts in the reading-frame.

Nevertheless, at the transcript level, there are dozens of frame-disrupting events that are conserved among species diverged for more than 100 million years (see also [11]). Many of these conserved frame-disrupting events cause a drastic shortening of alternative isoforms relative to reference isoforms. We find that 20 percent of the conserved frame-disrupting events shorten the alternative protein to less than 100 amino acids. Conversely, frame-preserving events result in alternative proteins only slightly shorter than the reference in length (Fig. S2, available in the online supplemental material).

The evolutionary conservation of such frame-disrupting AS events suggests that they are unlikely to be products of splicing errors. However, these results do not tell us whether frame-disrupting events have a function at the transcriptomic or proteomic levels. Therefore, we searched major public repositories of mass spectrometry data for spectra reporting the exon-exon junction expected in case of exon skipping.

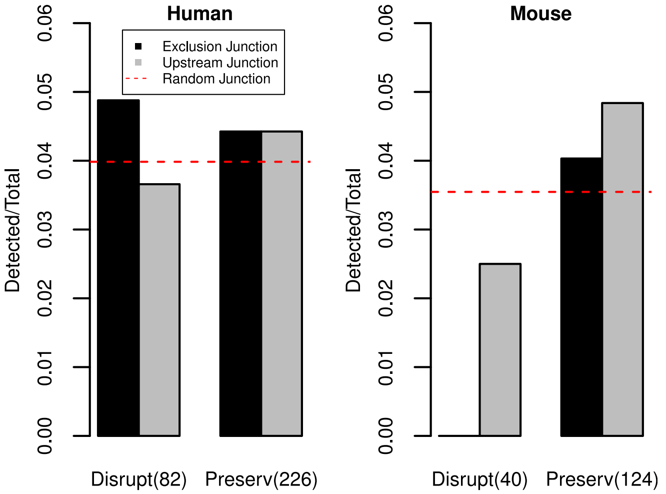


Fig. 2. Detection frequency of distinct exon-exon junctions as described in Fig. 1. a) Human. b) Mouse. Random Junctions: Exon-exon junctions randomly chosen from human or mouse reference proteomes (only one random junction per reference gene was drawn).

3.2 Estimating the Detection Frequency of Frame-Disrupting and Constitutive Events

We used the Mascot program to screen six million human and 8.3 million mouse spectra (Methods, Supplementary Dataset 3, available online). To detect exon-skipping events in MS data we searched for identified PSMs located at the junction between exons that flank the alternative exon (Exclusion junctions; Fig. 1). These PSMs allow us to find exon exclusion of frame-preserving or frame-disrupting exons. Importantly, all significant spectra reporting exclusion junctions were manually inspected. For comparison, we considered PSMs located in junctions of the same gene but immediately upstream (i.e., 5′) of the exon-skipping junctions (Upstream junctions; Fig. 1). The upstream junction set is made up almost entirely of constitutively spliced junctions.

In humans, we detected the exclusion of frame-disrupting exons in a total of four distinct genes from an initial database of 82 genes with detectable junctions (detection frequency: $4/82 = 4.9$ percent; Fig. 2a). Of these, two are evolutionary conserved. In mouse, no frame-disrupting events were detected among a total of 40 genes searched (detection frequency: 0 percent; Fig. 2b). The low detection frequency of frame-disrupting events could reflect a biological mechanism of degradation of transcripts with PTCs, or it could reveal a limitation of analytical platforms in detecting any kind of exon-exon junction. Favoring the latter hypothesis, the results show that, for both species, there is no significant difference in detection frequency between frame-disrupting and constitutive exon-exon junctions (3/82 and 1/40 upstream junctions were detected in human and mouse, respectively; $P = 1$, Fisher's exact test).

On a more technical note, it is important to further comment on the process of FDR estimation used here. It has been previously observed that the accuracy of estimated false discovery rates depends on several factors, including the number of identified spectra, mass tolerance, construction of the decoy database, and the choice of FDR estimators [26]. Consistent with these observations, over the 54 data sets searched in this study, we report that the same ion score threshold of 29 resulted in a relatively high negative correlation between the number of identified spectra and the FDR (Pearson's correlation of -0.47; $P = 4.0 \times 10^{-4}$). Moreover, while the choice of the FDR estimator is not always clear, the "original" formula $\mathrm{FDR} = \mathrm{FP}/\mathrm{TP}$ typically gives a less conservative estimate than the "alternative" formula $\mathrm{FDR} = 2 \cdot \mathrm{FP}/(\mathrm{TP} + \mathrm{FP})$ [26], [27]. Here we again observed similar trends; i.e. the ion score threshold of 29 resulted in $\mathrm{FDR} = 0.93\%$ using the original estimator, whereas the same score threshold resulted in $\mathrm{FDR} = 1.81\%$ using the alternative estimator. Overall, although the final identification results may be somewhat impacted by the decisions made in database search and FDR estimation, we believe that the spectral identification carried out in this work is of good quality, most likely with a true FDR around 1 percent.

3.3 Similar Detection Frequency among Frame-Disrupting, Frame-Preserving, and Random Exon-Exon Junctions

Given that the small size of the frame-disrupting database could lead to weak estimation of the detection frequency, we next focus on the detection frequency of a larger exon-exon junction database, consisting of frame-preserving and random junctions. The latter dataset consists of a set of junctions randomly selected from the reference proteome and therefore is enriched in constitutive junctions.

In humans, we detected the exclusion of frame-preserving exons in a total of 10 distinct genes from an initial database of 226 genes with detectable junctions (detection frequency: $10/226 = 4.4\%$; Fig. 2a). Again, no significant difference in detection frequency between frame-preserving and upstream junctions was observed. In mouse, the exclusion of frame-preserving exons was detected in 4.0 percent of genes (5/124; Fig. 2b). Interestingly, all 10 frame-preserving events detected in humans are evolutionarily conserved. Of the 226 genes containing frame-preserving events, 148 harbor evolutionarily conserved events and 78 do not. Therefore, for frame-preserving events, there is a significant association between MS detection and evolutionary conservation ($P = 0.01$, Fisher's exact test).

The detection frequency of random junctions was 4.0 percent (730/18320) in humans (Fig. 2a, dashed line) and 3.5 percent (599/16889) in mouse (Fig. 2b, dashed line). These frequencies are not significantly different from the detection frequencies of the exon-skipping events involving frame-preserving or frame-disrupting exons, for both species ($P > 0.5$ for all comparisons; χ^2-test). Therefore, we conclude that the estimation of the detection frequencies were consistent among larger and smaller exon-exon junction databases, and were similar regardless of the splicing type. Importantly, these conclusions hold even when only PSMs with $\mathrm{FDR} \leq 1\%$ are analyzed.

Supplementary Table S2, available online, summarizes features of the exon-skipping events we detected at the protein level; further information about each gene is also available in Supplementary Text 1, available online.

3.4 Low Frequency of Detected Exon-Exclusion Is Not Due to Low Gene Expression

The existence of a positive correlation between the gene expression and the level of protein detection on MS data have been described elsewhere [19], [28], [29]. By using Pearson's correlation we observe positive and significant correlations between gene expression (measured by RPKM) and protein detection (measured by the number of PSMs) ($P < 1.0 \times 10^{-7}$) for 6 tissue samples (Brain, Colon, Heart, Kidney, Liver, Lung) (Supplementary Dataset 4, available online). The best correlation was observed for the Liver sample (Fig. S3, available in the online supplemental material) where $R^2 = 0.19$, Pearson's correlation $= 0.44$, $P < 2.2 \times 10^{-16}$. Moreover, for these six tissues, the human genes containing the exon-skipping events (309 genes) showed a significantly higher level of expression than 100 sets of 309 genes randomly chosen from 20,048 human genes (Wilcox test, $P < 1.0 \times 10^{-24}$ for all comparisons). Additionally, the genes with conserved exon-skipping events are more highly expressed than the non-conserved ones. These results suggests that low gene expression cannot explain the low frequency of detected exon-skipping events in MS data.

4 Discussion

Since its discovery 40 years ago, alternative splicing has been seen as a biological mechanism for generating protein variability. Through RNA-Seq experiments [15], [17] hundreds of exon-skipping events shared among mammalian species have been discovered, with approximately 17 percent of them causing frame-disruption in both human and mouse genes. These events cause severe shortening of reference genes, producing highly truncated proteins.

Here we searched mass spectrometry data and found evidence for the translation of several of these out-of-frame isoforms (Table S2, available in the online supplemental material). Many of the exon exclusion events detected here at the protein level were previously detected at the transcript level by distinct methods, not only by RNA-seq (Supplementary text 2, available online). There is therefore good evidence that many of the events we have detected are real and biologically replicable.

The low detection frequency of truncated proteins found in mass spectrometry data must be considered against the low detection frequency of all splicing junctions, which was below 5 percent for both alternative and constitutive exons. The reasons for this low detection rate may be biological or methodological. Among the biological reasons, low protein abundance could account, at least in part, for this result. However, we found that most alternatively spliced genes analyzed here are highly expressed in several tissues, which likely indicates high protein abundance [29], [30] (Supplementary Dataset 4, Fig. S3, available online). Methodological limitations–such as the short length of the amino acid sequence required to detect an alternative exon-exon junction or low coverage of the proteome—could also explain the low detection frequency. Consistent with low protein coverage in bottom-up proteomics [31], our search of millions of spectra only resulted in the detection of about 30 percent of the human and mouse reference proteins when considering PSMs that mapped to any part of the protein sequence (Tables S3a and S3b, available in the online supplemental material). It should therefore not be surprising that when restricting the target matches to exon-exon junctions only, the detection frequency decreases to 4 percent.

Interestingly, this frequency is similar for alternative and constitutive exon-exon junctions. This surprising result suggests that translation of alternative transcripts may be more common than previously thought, at least in human and mouse. Importantly, the results presented here do not mean that frame-disrupting events are comparable in frequency to frame-preserving events, but rather that the rate at which junctions are detected in MS data seems to be proportional to their frequencies. This observation holds regardless of whether AS events cause frame disruption or whether the junction is constitutive or alternative.

Many authors have proposed that alternative splicing may be a mechanism for regulating transcript level via NMD by introducing PTCs into mature mRNAs [4], [6]. While there is good evidence supporting this mechanism [8], [32], our data suggest that some of these transcripts are translated (see also [33]). Regarding unproductive translation, we believe that the discovery of evolutionarily conserved truncated proteins suggests functional roles for these short proteins. Among the possible biological mechanisms regulated by truncated proteins are changes in cellular localization, for instance by the removal of nuclear signal peptides (e.g., [9]); or the creation of dominant negatives, for instance by proteins that can form heterodimers but are not enzymatically active (e.g., [34]). Any such proteins that can "quench" signals by receiving but not transmitting information are sure to be a useful tool in the regulation of cellular function.

Acknowledgments

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) Postdoc fellowship 275912/8 to Rodrigo F. Ramalho. National Institutes of Health grant R01 GM103725-04 to Predrag Radivojac.

References

[1] E. Kim, A. Goren, and G. Ast, "Alternative splicing: Current perspectives," *Bioessays*, vol. 30, no. 1, pp. 38–47, 2008.

[2] K. Thorsen, K. D. Sorensen, A. S. Brems-Eskildsen, C. Modin, M. Gaustadnes, A. M. Hein, M. Kruhoffer, S. Laurberg, M. Borre, K. Wang, S. Brunak, A. R. Krainer, N. Torring, L. Dyrskjot, C. L. Andersen, and T. F. Orntoft, "Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis," *Mol. Cell Proteomics*, vol. 7, no. 7, pp. 1214–24, 2008.

[3] A. Resch, Y. Xing, A. Alekseyenko, B. Modrek, and C. Lee, "Evidence for a subpopulation of conserved alternative splicing events under selection pressure for protein reading frame preservation," *Nucleic Acids Res.*, vol. 32, no. 4, pp. 1261–9, 2004.

[4] B. P. Lewis, R. E. Green, and S. E. Brenner, "Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans," in *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 1, pp. 189–92, 2003.

[5] J. T. Mendell, N. A. Sharifi, J. L. Meyers, F. Martinez-Murillo, and H. C. Dietz, "Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise," *Nat. Genet.*, vol. 36, no. 10, pp. 1073–8, 2004.

[6] D. Baek and P. Green, "Sequence conservation, relative isoform frequencies, and nonsense-mediated decay in evolutionarily conserved alternative splicing," in *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 36, pp. 12 813–8, 2005.

[7] L. F. Lareau, R. E. Green, R. S. Bhatnagar, and S. E. Brenner, "The evolving roles of alternative splicing," *Curr. Opin. Struct. Biol.*, vol. 14, no. 3, pp. 273–82, 2004.

[8] L. F. Lareau, M. Inada, R. E. Green, J. C. Wengrod, and S. E. Brenner, "Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements," *Nature*, vol. 446, no. 7138, pp. 926–9, 2007.

[9] Y. Zhou, S. Liu, G. Liu, A. Ozturk, and G. G. Hicks, "ALS-associated FUS mutations result in compromised FUS alternative splicing and autoregulation," *PLoS Genet.*, vol. 9, no. 10, p. e1003895, 2013.

[10] H. Liang and L. F. Landweber, "A genome-wide study of dual coding regions in human alternatively spliced genes," *Genome. Res.*, vol. 16, no. 2, pp. 190–6, 2006.

[11] R. Szklarczyk, J. Heringa, S. K. Pond, and A. Nekrutenko, "Rapid asymmetric evolution of a dual-coding tumor suppressor INK4a/ARF locus contradicts its function," *Proc. Nat. Acad. Sci. USA*, vol. 104, no. 31, pp. 12 807–12, 2007.

[12] Y. Xing and C. Lee, "Alternative splicing and RNA selection pressure–evolutionary consequences for eukaryotic genomes," *Nat. Rev. Genet.*, vol. 7, no. 7, pp. 499–509, 2006.

[13] J. K. Pickrell, A. A. Pai, Y. Gilad, and J. K. Pritchard, "Noisy splicing drives mRNA isoform diversity in human cells," *PLoS Genet.*, vol. 6, no. 12, p. e1001236, 2010.

[14] A. Magen and G. Ast, "The importance of being divisible by three in alternative splicing," *Nucleic Acids Res.*, vol. 33, no. 17, pp. 5574–82, 2005.

[15] J. Merkin, C. Russell, P. Chen, and C. B. Burge, "Evolutionary dynamics of gene and isoform regulation in mammalian tissues," *Science*, vol. 338, no. 6114, pp. 1593–9, 2012.

[16] R. Sorek, R. Shamir, and G. Ast, "How prevalent is functional alternative splicing in the human genome?" *Trends Genet.*, vol. 20, no. 2, pp. 68–71, 2004.

[17] N. L. Barbosa-Morais, M. Irimia, Q. Pan, H. Y. Xiong, S. Gueroussov, L. J. Lee, V. Slobodeniuc, C. Kutter, S. Watt, R. Colak, T. Kim, C. M. Misquitta-Ali, M. D. Wilson, P. M. Kim, D. T. Odom, B. J. Frey, and B. J. Blencowe, "The evolutionary landscape of alternative splicing in vertebrate species," *Science*, vol. 338, no. 6114, pp. 1587–93, 2012.

[18] G. M. Sheynkman, M. R. Shortreed, B. L. Frey, and L. M. Smith, "Discovery and mass spectrometric analysis of novel splice-junction peptides using RNA-Seq," *Mol. Cell Proteomics*, vol. 12, no. 8, pp. 2341–53, 2013.

[19] I. Ezkurdia, A. del Pozo, A. Frankish, J. M. Rodriguez, J. Harrow, K. Ashman, A. Valencia, and M. L. Tress, "Comparative proteomics reveals a significant bias toward alternative protein isoforms with conserved structure and function," *Mol. Biol. Evol.*, vol. 29, no. 9, pp. 2265–83, 2012.

[20] E. T. Wang, R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S. F. Kingsmore, G. P. Schroth, and C. B. Burge, "Alternative isoform regulation in human tissue transcriptomes," *Nature*, vol. 456, no. 7221, pp. 470–6, 2008.

[21] P. Jones, R. Cote, L. Martens, A. Quinn, C. Taylor, W. Derache, H. Hermjakob, and R. Apweiler, "PRIDE: A public repository of protein and peptide identifications for the proteomics community," *Nucleic Acids Res.*, vol. 34, no. Database issue, pp. D659–D663, 2006.

[22] F. Desiere, E. W. Deutsch, N. L. King, A. I. Nesvizhskii, P. Mallick, J. Eng, S. Chen, J. Eddes, S. N. Loevenich, and R. Aebersold, "The PeptideAtlas project," *Nucleic Acids Res.*, vol. 34, no. Database issue, pp. D655–8, 2006.

[23] J. E. Elias and S. P. Gygi, "Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry," *Nat Methods*, vol. 4, no. 3, pp. 207–214, 2007.

[24] M. Krupp, J. U. Marquardt, U. Sahin, P. R. Galle, J. Castle, and A. Teufel, "RNA-Seq Atlas–a reference database for gene expression profiling in normal tissue by next-generation sequencing," *Bioinformatics*, vol. 28, no. 8, pp. 1184–5, 2012.

[25] Q. Pan, M. A. Bakowski, Q. Morris, W. Zhang, B. J. Frey, T. R. Hughes, and B. J. Blencowe, "Alternative splicing of conserved exons is frequently species-specific in human and mouse," *Trends Genet.*, vol. 21, no. 2, pp. 73–7, 2005.

[26] K. Jeong, S. Kim, and N. Bandeira, "False discovery rates in spectral identification," *BMC Bioinf.*, vol. 13, no. Suppl 16, p. S2, 2012.

[27] J. E. Elias and S. P. Gygi, "Target-decoy search strategy for mass spectrometry-based proteomics," *Methods Molecular Biol.*, vol. 604, pp. 55–71, 2010.

[28] K. Ning and A. I. Nesvizhskii, "The utility of mass spectrometry-based proteomic data for validation of novel alternative splice forms reconstructed from RNA-Seq data: A preliminary assessment," *BMC Bioinf.*, vol. 11, no. Suppl. 11, p. S14, 2010.

[29] T. Wang, Y. Cui, J. Jin, J. Guo, G. Wang, X. Yin, Q. Y. He, and G. Zhang, "Translating mRNAs strongly correlate to proteins in a multivariate manner and their translation ratios are phenotype specific," *Nucleic Acids Res.*, vol. 41, no. 9, pp. 4743–54, 2013.

[30] Y. Gunawardana and M. Niranjan, "Bridging the gap between transcriptome and proteome measurements identifies post-translationally regulated genes," *Bioinformatics*, vol. 29, no. 23, pp. 3060–3066, 2013.

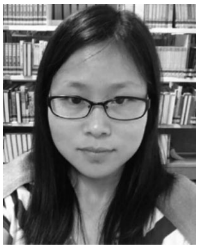
[31] Y. F. Li and P. Radivojac, "Computational approaches to protein inference in shotgun proteomics," *BMC Bioinf.*, vol. 13, no. Suppl 16, p. S4, 2012.

[32] A. Sureau, R. Gattoni, Y. Dooghe, J. Stevenin, and J. Soret, "SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs," *EMBO J.*, vol. 20, no. 7, pp. 1785–1796, 2001.

[33] T. Sterne-Weiler, R. T. Martinez-Nunez, J. M. Howard, I. Cvitovik, S. Katzman, M. A. Tariq, N. Pourmand, and J. R. Sanford, "Frac-seq reveals isoform-specific recruitment to polyribosomes," *Genome. Res.*, vol. 23, no. 10, pp. 1615–23, 2013.

[34] Y. Stasiv, M. Regulski, B. Kuzin, T. Tully, and G. Enikolopov, "The Drosophila nitric-oxide synthase gene (dNOS) encodes a family of proteins that can modulate NOS activity by acting as dominant negative regulators," *J. Biol. Chem.*, vol. 276, no. 45, pp. 42241–42251, 2001.

Rodrigo F. Ramalho received the BS degree from the University of Sao Paulo in 2003. He received the MS degree in 2008 and the PhD degree in biological sciences in 2012 under the mentorship of Professor Diogo Meyer. From 2013 to 2014, he held a US National Science Foundation (CAPES) postdoctoral fellowship to work at the Indiana University, with Professors Matthew Hahn and Predrag Radivojac. He is a scientific researcher at A.C. Camargo Cancer Center in Sao Paulo, Brazil. His interests include bioinformatics, human genomics, and population genetics.



Sujun Li received the BS degree in medicine from the Tongji Medical College at Huazhong University of Science and Technology in 2002. In 2008, she received the PhD degree in bioinformatics from Shanghai Institute for Biological Sciences, Chinese Academy of Sciences under the direction of Prof. Yixue Li. In 2009, she held a postdoctoral position in Prof. Predrag Radivojac's Lab at Indiana University, Bloomington. In 2011, she took a position of scientist at Biogen Idec, after which she returned back to Indiana University, Bloomington, as an assistant scientist. Her research interests focus on computational proteomics and disease-related proteogenomics studies.

Predrag Radivojac received the BS degree from the University of Novi Sad in 1994 and the MS degree in 1997 from the University of Belgrade, both in electrical engineering. In 2003, he received the PhD degree in computer and information sciences from Temple University under the direction of Prof. Zoran Obradovic. In 2004, he held a postdoctoral position in Prof. Keith Dunker's lab at Indiana University's School of Medicine, after which he joined Indiana University, Bloomington, as a faculty member. His research interests span the areas of computational biology, biomedical informatics, and machine learning. He received the US National Science Foundation CAREER Award in 2007 dedicated to understanding protein post-translational modifications. Currently, he is an Editorial Board member for the journal *Bioinformatics*, an associate editor for *PLoS Computational Biology*, and serves on the Board of Directors of the International Society for Computational Biology (ISCB).

Matthew W. Hahn received the BS degree from Cornell University in 1998. In 2003, he received the PhD degree in biology from Duke University under the mentorship of Professor Mark Rausher. From 2003 to 2005, he held a US National Science Foundation postdoctoral fellowship to work at the University of California, Davis, with Professors Charles Langley and John Gillespie. He is a professor of biology and informatics at Indiana University, where he has held a faculty position since 2005. His research interests include bioinformatics, genomics, population genetics, and phylogenetics. He is an author or coauthor on 97 publications in these areas, as well as a coauthor of one book, *Introduction to Computational Genomics* (Cambridge University Press, 2007). He has received the US National Science Foundation CAREER award and a fellowship from the Alfred P. Sloan Foundation.