Letter

The Relationship between Alternative Splicing and Proteomic Complexity

Benjamin J. Blencowe[1,*]

In their Opinion article published recently in *Trends in Biochemical Sciences* [1], Tress and colleagues proposed that only a minor fraction of splice variants detected in transcriptome profiling data are translated and, therefore, that most splice variants have little or no function. Their proposal stems primarily from an attempt to detect splice variants at the protein level using over 100 published liquid chromatography-coupled to tandem mass-spectrometry (LC-MS/MS) data sets from diverse cell and tissue types, as well as inferences drawn from evolutionary comparisons of alternative exons [2]. In this Letter, I discuss issues with the authors' analyses and interpretations. I also summarize recent results supporting the contrary conclusion, namely that a major fraction of alternative splicing (AS) events detected in transcriptome profiling data is translated.

It is estimated that approximately 95% of multiexonic human genes give rise to transcripts containing more than 100 000 distinct AS events [3,4]. The majority of these AS events display tissue-dependent variation and 10–30% are subject to pronounced cell, tissue, or condition-specific regulation [4]. The latter events are significantly enriched for evolutionary conservation and frame-preserving potential. Moreover, dozens of independent studies have shown that subsets of differentially regulated AS events are significantly enriched in genes that function in common biological processes and pathways, and that, where characterized, exons in these splicing 'networks' have important functions [5,6]. For example, AS networks function extensively in the remodeling of cytoskeletal interactions, signaling cascades, and gene regulatory pathways, and the literature contains hundreds of examples in which translated splice variants contribute important roles in development, cell and tissue homeostasis, animal behavior, diseases, as well as other processes. Nevertheless, given the intrinsic challenges of assessing the functionality of protein isoforms, determining the overall fraction of biologically relevant AS remains an important and timely question.

In their analyses, Tress and colleagues applied stringent mapping and filtering criteria requiring detection of one or more peptides that distinguished splice isoforms from a gene, and where the peptides also had to be detected in more than one independent data set by at least two different calling methods [2]. The authors found peptide evidence for only 246 alternative isoforms (or a total of 282 AS events). They then simulated the number of splice variants they expected to detect in the LC-MS/MS data based on assumptions about relative isoform expression (Box 4 in their article and as reported in [2]). One model assumed that all transcripts are expressed equally, while the other allowed for 50-fold greater abundance of one isoform relative to other isoform(s). Based on these models, the observed numbers of genes expressing multiple isoforms fell short of the expected numbers by 15.5 and 5.7-fold, respectively.

Important questions relating to the authors' findings are: (i) are their observations surprising given inherent technical limitations associated with the detection of splice variants using LC-MS/MS data, the modeling approaches they used, and in the context of previously published knowledge of splice variant abundance?; (ii) are their conclusions justified based on their findings?; and (iii) what is the actual fraction of splice variants that is translated?

A major challenge associated with using LC-MS/MS data for the detection of splice variants is their limited coverage and sensitivity. For complex samples, such as those analyzed by the authors, the number of peptides far exceeds the number of sequencing cycles provided by a mass spectrometer, resulting in undersampling and a consequent lack of detection reproducibility between samples [7]. This limitation is especially relevant to the detection of splice variants since the probability of detecting a peptide that maps specifically to a variably expressed, alternatively spliced sequence (i.e., an exon–exon junction, exon, or intron) is significantly less than the probability of detecting a peptide that maps to a constitutively expressed sequence. Further compounding negative detection rates is that subsets of AS events most likely to be translated, in particular those that are cell, tissue, or condition-differentially regulated (see above), will be represented by peptides that are even more sparsely distributed across the LC-MS/MS data analyzed by the authors, and, therefore, even less likely to be detected by their methods. For example, what is not mentioned in their Opinion article, but is mentioned in their previously published work [2], is that their methods detect cell and/or tissue-specific splicing events in a meager 14 genes. Unfortunately, the authors' models for estimating expected numbers of detected splice variants fail to take these and possible additional sources of false negative detection into account (see also below). Consequently, the authors' conclusion that most genes express a single predominant isoform is compromised by the inherent limitations of the data and analysis methods they have used.

It is well established that most mammalian AS events are species specific and under relaxed selection pressure, and that the corresponding alternatively spliced

transcripts are typically expressed at lower levels in a given cell or tissue type (but not necessarily in every cell or tissue type, for the reasons summarized above) compared with main or 'principal' isoforms [8]. However, this does not mean that such events are not translated or lack function. Given that absolute quantification strategies in mammalian cells have indicated that most protein variance can be explained by mRNA variance [9], it is important to recognize that, aside from the sources of false negative detection including but not limited to those summarized above, the lack of detection in LC-MS/MS data of a large fraction of splice variants with potential to code for protein may be because they fall below the abundance range modeled by the authors. For example, in marked contrast to the findings and assertions of the authors, tens of thousands of splice variant transcripts have been detected in polysome fractions [10,11]. Moreover, by using ribosome profiling data from different cell types, it was shown recently that splice junction sequences representing approximately 75% of total exon skipping events detected in medium to high abundance transcripts by RNA-Seq were engaged by ribosomes [12]. The majority of these ribosome-engaged exon-skipping events have frame-preservation potential, and all of them correspond to AS events detected previously in transcript sequence data from diverse cell and tissue types [13].

In summary, when collectively considering multiple sources of false negative detection rates for splice variants in LC-MS/MS data, previous results demonstrating that protein abundance is predominately related to transcript abundance, and recent results from detecting splice variant sequences associated with ribosomes, it is apparent that most splice variants detected in transcriptome profiling data are likely translated. Therefore, it is possible that most splice variants contribute to cellular function. Unfortunately, the authors of [1] have unnecessarily dismissed as a 'theory' a large body of experimental data from numerous laboratories demonstrating extensive roles for AS in the remodeling of cellular networks that have diverse roles in critical biological processes. Nevertheless, the authors of [1] have drawn attention to the importance of obtaining reliable, quantitative measurements of translated splice variants to assess possible biological significance. In this regard, data-independent acquisition mass-spectrometry strategies that are less prone to the stochasticity of the shotgun proteomic data sets analyzed by Tress and colleagues have already shown promise for the more comprehensive detection of translated splice variants and will undoubtedly prove to be valuable in future studies [7,14]. Moreover, it is important to appreciate that it can take a single research group years of effort to determine the biological function of a single AS event. As such, an important goal for future studies will be to further develop high-throughput methods for interrogating the functions of splice variants. In the meantime, one should be mindful of the old aphorism, 'absence of evidence is not evidence of absence'.

Acknowledgments

I thank Anne-Claude Gingras, Serge Gueroussov, Manuel Irimia, Robert Weatheritt, and Juan Valcarcel for helpful discussions and comments on this Letter.

[1]Donnelly Centre and Department of Molecular Genetics, University of Toronto, 160 College Street, Toronto, ONT, M5S 3E1, Canada

*Correspondence:
b.blencowe@utoronto.ca (B.J. Blencowe).

http://dx.doi.org/10.1016/j.tibs.2017.04.001

References

1. Tress, M.L. *et al.* (2017) Alternative splicing may not be the key to proteome complexity. *Trends Biochem.* 42, 98–110
2. Abascal, F. *et al.* (2015) Alternatively spliced homologous exons have ancient origins and are highly expressed at the protein level. *PLoS Comput. Biol.* 11, e1004325
3. Pan, Q. *et al.* (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–1415
4. Wang, E.T. *et al.* (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476
5. Ellis, J.D. *et al.* (2012) Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol. Cell* 46, 884–892
6. Kalsotra, A. and Cooper, T.A. (2011) Functional consequences of developmentally regulated alternative splicing. *Nat. Rev. Genet.* 12, 715–729
7. Rost, H.L. *et al.* (2015) Reproducible quantitative proteotype data matrices for systems biology. *Mol. Biol. Cell* 26, 3926–3931
8. Xing, Y. and Lee, C. (2006) Alternative splicing and RNA selection pressure – evolutionary consequences for eukaryotic genomes. *Nat. Rev. Genet.* 7, 499–509
9. Liu, Y. *et al.* (2016) On the dependency of cellular protein levels on mRNA abundance. *Cell* 165, 535–550
10. Floor, S.N. and Doudna, J.A. (2016) Tunable protein synthesis by transcript isoforms in human cells. *Elife* 5, e10921
11. Sterne-Weiler, T. (2013) Frac-seq reveals isoform-specific recruitment to polyribosomes. *Genome Res.* 23, 1615–1623
12. Weatheritt, R.J. *et al.* (2016) The ribosome-engaged landscape of alternative splicing. *Nat. Struct. Mol. Biol.* 23, 1117–1123
13. Irimia, M. *et al.* (2014) A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* 159, 1511–1523
14. Schreiner, D. *et al.* (2015) Quantitative isoform-profiling of highly diversified recognition molecules. *Elife* 4, e07794

Letter

Most Alternative Isoforms Are Not Functionally Important

Michael L. Tress,[1,*] Federico Abascal,[2] and Alfonso Valencia[3]

There are two assertions that might be considered polemical in our Opinion article [1]: the first is that most protein-coding genes have a single main isoform, and the second is that most annotated alternative transcripts do not generate functionally important gene products. While alternative isoforms can have important cellular roles, we believe that alternative splicing is not the key to proteome complexity.

In his letter [2] Professor Blencowe asks three rhetorical questions. The first is whether our results are surprising given the limitations of shotgun mass spectrometry (MS). Here Professor Blencowe