Re-evaluating the impact of alternative RNA splicing on proteomic diversity

Jeru Manoj Manuel[1,2], Noé Guilloy[2], Inès Khatir[1,2], Xavier Roucou[2,3,4] and Benoit Laurent[1,2]*

[1]Research Center on Aging, Centre Intégré Universitaire de Santé et Services Sociaux de l'Estrie-Centre Hospitalier Universitaire de Sherbrooke, Sherbrooke, QC, Canada, [2]Department of Biochemistry and Functional Genomics, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada, [3]Centre de Recherche du Centre Hospitalier Universitaire de Sherbrooke (CRCHUS), Sherbrooke, QC, Canada, [4]Quebec Network for Research on Protein Function Structure and Engineering, PROTEO, Québec, QC, Canada

Alternative splicing (AS) constitutes a mechanism by which protein-coding genes and long non-coding RNA (lncRNA) genes produce more than a single mature transcript. From plants to humans, AS is a powerful process that increases transcriptome complexity. Importantly, splice variants produced from AS can potentially encode for distinct protein isoforms which can lose or gain specific domains and, hence, differ in their functional properties. Advances in proteomics have shown that the proteome is indeed diverse due to the presence of numerous protein isoforms. For the past decades, with the help of advanced high-throughput technologies, numerous alternatively spliced transcripts have been identified. However, the low detection rate of protein isoforms in proteomic studies raised debatable questions on whether AS contributes to proteomic diversity and on how many AS events are really functional. We propose here to assess and discuss the impact of AS on proteomic complexity in the light of the technological progress, updated genome annotation, and current scientific knowledge.

KEYWORDS

alternative splicing, RNA, isoform proteins, alternative proteins, ghost proteome

Introduction

Alternative splicing (AS) is a key process by which genes produce more than a single mRNA, hence contributing to the transcriptome complexity. In this process, specific exons of a gene can be included or excluded in the final RNA. Protein-coding genes and lncRNA genes can generate multiple splice variants from one gene through AS (Mercer et al., 2011; Khan, Wellinger, and Laurent 2021). From plants to humans, AS is a powerful mechanism that increases transcriptome plasticity and can control the expression level of certain genes (Castle et al., 2008; Gueroussov et al., 2015; Muhammad et al., 2022). Indeed, RNA splice variants arising from AS can exhibit different mRNA stabilities and structures. In humans, it is estimated that 95% of genes undergo AS, which underscores its importance (Castle et al., 2008; Pan et al., 2008; Nilsen and Graveley 2010). Three transcripts are produced in average from each protein-coding gene (Khan, Wellinger, and Laurent 2021). Importantly, splice variants produced from protein-coding genes can potentially encode for distinct protein isoforms. For a given gene, the most expressed transcript is usually defined as coding for the canonical protein. This canonical status is determined based on the transcript expression across different tissues of an organism, the conservation of its exon combination with other species, and/or the existence of a functional role for the protein (Osmanli et al., 2022). Compared to their canonical proteins, isoform

proteins can lose or gain certain domains and, therefore, can differ in their functional properties by the alteration of localization signals, sequences for post-translational modifications, or interaction with other proteins (Kriventseva et al., 2003; Stamm et al., 2005; Leoni et al., 2011; Light and Elofsson 2013). Advances in proteomics have shown that the proteome is indeed diverse due to the presence of numerous protein isoforms. For the past decades, with the help of advanced high-throughput technologies such as RNA sequencing (RNA-seq), a full catalog of alternatively spliced transcripts has been established, but the functional significance of most AS events remains still largely unknown. Hence, the identification of numerous alternatively spliced transcripts raises important and debatable questions: how many AS events are real and not mere artefacts of splicing machinery? How many AS products are functional? Does AS really expand proteomic diversity? We propose here to re-evaluate and discuss the impact of AS on proteomic diversity in the light of the technological progress, updated genome annotation, and current scientific knowledge.

Alternative splicing and proteomic diversity: Two different visions

Whether AS is a major source of proteome complexity has always been a contentious issue in the field. For example, on this debatable question, Benjamin J. Blencowe and Michael L. Tress et al. have mutually expressed their contrasting opinions few years ago (Tress et al., 2017a; Blencowe 2017; Tress et al., 2017b).

Michael L. Tress and colleagues claimed that AS might not be the key to proteome complexity. They argued that most genes only expressed one main transcript across multiple cell lines (Gonzalez-Porta et al., 2013), and hence, one single main protein isoform can be detected by high-resolution mass spectroscopy (Abascal et al., 2015; Ezkurdia et al., 2015). The abundance of alternatively spliced variants identified from more than 100 different tissues at various developmental stages was, therefore, in contrast with the low number of multiple protein isoforms per gene. They found that only 2% of genes had multiple isoform proteins (246 genes with splice event-specific peptide evidence over 12,716 human genes for which at least two peptides have been detected) (Abascal et al., 2015). As few genes provided reliable evidence for more than one isoform, the authors stated that alternative variants were not abundant at the protein level (Tress, Abascal, and Valencia 2017a). One possible reason could be the misidentification of a good peptide spectrum with multiple assigned peaks. However, the discrepancies between transcriptomics and proteomics experiments are difficult to explain solely on a technical issue. They described that alternatively spliced exons were not under selective pressure and are evolving neutrally (Tress, Abascal, and Valencia 2017a). This observation suggested in their opinion that AS events were not evolutionary innovations and that most alternatively spliced variants were not functionally important if translated.

In response to Tress et al. (2017a) and Tress et al. (2017b), Benjamin J Blencowe agreed that AS events were mostly specific to species and, hence, are under relaxed selection pressure (Blencowe 2017). However, he pointed out that even though alternatively spliced transcripts were expressed at lower levels than their corresponding main protein isoforms, it did not mean that these splice variants were not translated or did not have a relevant function in a given cell or

tissue type. Blencowe argued that protein abundance was predominantly related to transcript abundance (Liu, Beyer, and Aebersold 2016) and that many splice variants identified by transcriptomics have been detected in polysome fractions and were likely translated (Weatheritt, Sterne-Weiler, and Blencowe 2016). Finally, Blencowe stated that the low detection rate of protein isoforms by LC-MS/MS cannot be interpreted since their identification is limited by the coverage and sensitivity of the technology. Indeed, the peptide number largely exceeds the number of sequencing cycles provided using a mass spectrometer, thereby limiting the detection of splice variants compared to a constitutively expressed sequence (Blencowe 2017).

Different perspectives: Right and wrong at the same time?

These two visions highlight the AS potential role in proteomic diversity on two different ends of the spectrum. The limitations to the available technology and the scientific knowledge at the time the studies were conducted have potentially skewed the interpretation to opposite ends. In this section, we discuss critical points that should be considered to assess the impact of AS on proteomic diversity.

Alternative splicing: Real or artefact of splicing machinery?

The widespread presence of alternatively spliced transcripts has raised the question of whether they are artefacts of splicing machinery or have a biological purpose (Graveley 2001). With the high complexity of eukaryotic genes and the level of splice-site conservation, numerous AS events are expected to happen along the processing of pre-mRNAs, regardless of their functional relevance (Modrek and Lee 2002). However, having a reduced fidelity of the spliceosome to promote proteome diversification could be problematic for a cell since basic molecular mechanisms cannot afford to jeopardize levels of essential proteins (Hsu and Hertel 2009). Consequently, high degrees of specificity and fidelity are required for pre-mRNA splicing to ensure the correct expression of critical functional mRNAs. Indeed, even though the frequency of aberrant spliced transcripts varies widely among loci, tissues, and species, the minimum splicing error rate in vertebrates is around 0.1% aberrant transcripts per intron (Skandalis 2016). The spliceosome is extremely accurate in selecting splice junctions with error frequencies as low as one per 10^5 splicing events (Fox-Walsh and Hertel 2009). This estimation was only performed on specific transcripts (i.e., UBA52, RPL23, HPRT, POLB, and TRPV1), so the extent to which the spliceosome is error-prone remains to be globally assessed. Although the spliceosome is prone to errors, mis-spliced mRNAs can be degraded from cells through nonsense-mediated RNA decay (NMD) or other RNA quality control steps (Saudemont et al., 2017; Garcia-Moreno and Romao 2020). Therefore, the spliceosome is unlikely responsible for generating artefactual splice variants.

An evolutionary perspective of alternative splicing

The importance and functionality of AS events are often associated on whether these events are conserved during evolution. Generally,

95% of human multiexon genes undergo AS (Pan et al., 2008), but this ratio is 60.7% in the fruit fly (*Drosophila melanogaster*) (Graveley et al., 2011), 25% in the nematode (*Caenorhabditis elegans*) (Ramani et al., 2011), and only 2.9% in the green alga (*Volvox carteri*) (Kianianmomeni et al., 2014). Organisms with more complexity tend to have a higher ratio of AS events. There is a strong positive correlation between the number of unique cell types—referred as organism complexity—and the number of AS events (Chen et al., 2014). The study of the evolutionary landscape of AS over ~350 million years of evolution in vertebrates showed significant differences in AS complexity among vertebrate species, with primates harboring the highest complexity (Barbosa-Morais et al., 2012; Merkin et al., 2012). These studies demonstrated that the variation in gene expression was conserved at the tissue-specific level, while AS was conserved at the species-specific level, suggesting that AS diverged faster than gene expression. Moreover, AS event types varied in their frequency among different organisms. In animals, exon skipping is the most common AS event, which represents around 50% of all AS events (Pan et al., 2008), while in plants, intron retention is the most abundant AS event type (Reddy et al., 2013). Most AS events have variable tissue specificities and appear to be evolving neutrally (Wang et al., 2008). However, a subset of AS events is conserved between species and displays tissue specificity. For example, around 20% of alternative exons are conserved between humans and mice (Modrek and Lee 2002; Abascal et al., 2015). These conserved events are significantly enriched in genes that function in common biological processes and pathways. Alternative exons in these splicing "networks" allow the tissue-specific rewiring of protein–protein interaction networks (Buljan et al., 2012; Ellis et al., 2012; Irimia et al., 2014; Tapial et al., 2017). Investigating these networks in different tissues and organs has revealed that these conserved isoforms play a prominent role in the regulation of neuronal development (Boutz et al., 2007; Jiao et al., 2008; Laurent et al., 2015; Fiszbein et al., 2016), immunity (Zikherman and Weiss 2008), and muscle differentiation (Nakka et al., 2018). However, this evolutionary conservation does not mean that alternative exons, which are not evolutionarily conserved, are not significant and do not participate in proteomic diversity. These isoforms could be expressed in a lineage-specific manner, or they might have just recently evolved. For instance, the exonization of intronic sequences such as repetitive elements is now widely documented in many genomes. In primate and human genomes, Alu elements are the most abundant transposable elements that can generate new exons (i.e., Alu exons) and lead to novel spliced transcripts (Krull, Brosius, and Schmitz 2005). Ribosome profiling and proteomics data from human tissues and cell lines showed that some Alu-derived exons can be translated and present in human proteins (Lin et al., 2016), suggesting that some Alu exons can contribute to proteomic diversity. However, in primates and humans, the high number and complexity of AS events might not reflect the functional expansion of the transcriptome but could be explained by the nearly neutral theory (Ohta 1992). Weak selection results in an excess of neutral or slightly deleterious mutations, including those affecting AS regulation. A reduction of intron splicing accuracy, mutations introducing cryptic splicing signals, and transposable element insertion events can generate novel AS events that produce non-functional spliced transcripts (Pickrell et al., 2010). Since these mutations are not removed by purging selection, they can persist and some of them can selectively give

novel functional entities, for example, AS events that become functional.

Correlation between transcription and translation

One common argument supporting AS contribution to proteomic complexity is that protein abundance is predominantly related to transcript abundance (Liu, Beyer, and Aebersold 2016). Therefore, even low levels of alternatively spliced transcripts have a chance to be translated into functional proteins. However, there are many regulatory mechanisms that can balance the level of protein expression: the translation rate, the degradation rate, the protein synthesis rate, and transport (Vogel and Marcotte 2012; McManus, Cheng, and Vogel 2015). Different subsets of genes exhibit different types of regulation. At a steady state, mRNA levels correlate with protein levels even during dynamic processes such as proliferation or differentiation (Hsieh et al., 2012; Vogel and Marcotte 2012; Kristensen, Gsponer, and Foster 2013; Li et al., 2014). However, the mRNA levels of some genes are proxies for the corresponding protein levels because of post-transcriptional and translational mechanisms (Liu and Aebersold 2016; Liu, Beyer, and Aebersold 2016). For short-term adaptation such as stress response, the regulation of the transcript level of specific genes is unadapted to the cellular response and post-transcriptional mechanisms (e.g., increase of translation or increase of protein degradation) are thereby more efficient. For instance, changes in the translation rate could positively or negatively affect the mRNA–protein ratio (Lackner et al., 2012; Cheng et al., 2016) and, hence, foster a significant contribution of alternatively spliced transcripts to proteomic diversity.

Another argument supporting AS contribution to proteomic complexity is that many splice variants identified by transcriptomics have been detected in polysome fractions and, hence, are likely to be translated (Weatheritt, Sterne-Weiler, and Blencowe 2016). However, there may be significant levels of alternatively spliced transcripts that do not pass co-translational quality control mechanisms and are degraded. Aberrant polypeptides and mRNAs can be detected and eliminated by mRNA quality control systems while engaging the ribosome (Inada 2017). Because the ribosome has a central role in quality control processes, alternatively spliced transcripts associated with the ribosome are not necessarily translated into proteins.

What is new on proteomic diversity?

Re-evaluating the impact of AS on proteomic diversity necessitates examining the newest developments in this field of investigation, more specifically the technological progress, the update of genome annotation, and the latest advances in scientific knowledge.

Technological and technical advances

As highlighted by Blencowe, LC-MS/MS has some limitations in identifying all potential protein isoforms in a complex sample. The number of peptides exceeds the number of sequencing cycles provided using a mass spectrometer, and hence, the detection of alternative

splice isoforms present in low quantities is limited and could potentially explain why so few alternative isoforms can be detected in proteomics experiments (Blencowe 2017). To address this issue, the integration of RNA-seq with a data-independent acquisition method acquiring all theoretical spectra has been implemented to reduce peptide mapping uncertainty, improve quantitative accuracy, and detect novel peptides (Liu et al., 2017; Jeong, Kim, and Paik 2018; Agosto et al., 2019). This proteogenomic approach yielded high reproducibility between technical and biological replicates and enabled the quantification of a large fraction of the proteome with quantitative accuracy (Poulos et al., 2020). Another limitation to the detection of alternative splice isoforms is also attributed to enzymes used to digest protein samples. The standard protease used in shotgun proteomics is the trypsin that digests at K or R residues, hereby producing short peptides (around six amino acids) and limiting the proteome coverage and detection of isoform proteins (Wang et al., 2018). Other proteases (e.g., chymotrypsin, LysC, LysN, AspN, GluC, and ArgC) have been used to cover complementary fractions of the proteome and improve the detection of specific peptides (Giansanti et al., 2016). A combination of several enzymes could be the best approach to reach comprehensive peptide identification.

Another challenge is to improve the identification of potentially functional transcripts. The development of long-read sequencing technologies has transformed the field since we can now obtain the entire RNA sequence in a single read (Marx 2023). The full-length transcript recovery and quantification helped advance transcript-level analyses of AS processes, distinguish novel isoform changes, and improve the ability to identify functional isoforms (Uapinyoying et al., 2020; De Paoli-Iseppi, Gleeson, and Clark 2021; Hu et al., 2021; Troskie et al., 2021; Wright et al., 2022). For instance, alternative isoforms and tumor-specific isoforms arising from aberrant splicing during liver tumorigenesis were recently identified by single-molecule real-time long-read RNA sequencing (Chen et al., 2019). Another study combined long-read sequencing with polysome profiling and ribosome foot printing data to predict isoform-specific translational status in the rat hippocampus (Wang X et al., 2019). Indeed, single-molecule sequencing also provides the opportunity to improve ribosome profiling quantification by adapting existing methods for translation studies. For example, quantification of the translation of individual transcript isoforms using ribosome-protected mRNA fragments revealed evolutionary conserved impacts of differential splicing on the proteome (Reixachs-Sole et al., 2020). Finally, the single-cell revolution could also help address more accurately the impact of AS on proteomic diversity. Single-cell differential splicing analyses revealed novel differentially expressed splicing junctions (Liu et al., 2021). Single-cell proteomics is now taking the center stage. Novel quantitative single-cell proteomics approaches are capable of consistently quantifying thousands of proteins per cell across thousands of individual cells using limited instrument time and display ultra-high sensitivity to detect changes in a single-cell proteome (Schoof et al., 2021; Brunner et al., 2022). The technology could be applied for detecting specific protein isoforms in a particular cell type and, hence, could give unprecedented insights into the isoform proteome in health and disease. Interestingly, there are now integrated strategies that can profile single-cell proteome and transcriptome in a single reaction, highlighting the promising potential of highly multiplexed single-cell analyses (Genshaft et al., 2016; Specht et al., 2021).

Finally, an additional challenge is that most proteomic data were focused on the identification of proteins derived from alternatively spliced transcripts in steady-state conditions (Blakeley et al., 2010; Ezkurdia et al., 2012; Alfaro et al., 2017). However, most RNA splicing changes have been associated with changes in physiological conditions (e.g., stress response and hypoxia) or between normal and disease states (Ly et al., 2014). Some studies have also addressed the issue of whether targeted perturbations in RNA splicing patterns manifest as changes in the proteomic composition. For example, by depleting a spliceosome component (i.e., PRPF8) and using quantitative proteomics, it was established that significant changes in RNA relative abundance showed consistent changes in protein production (Liu et al., 2017). Using a similar approach, it would be interesting to determine more broadly how changes in AS for a subset of transcripts reflect in differential protein expression and assess the contribution of AS to proteomic complexity.

Genome annotation

Historically, mRNAs were defined as monocistronic and expected to encode a single protein. In addition, open reading frames (ORFs) shorter than 100 codons were automatically discarded from genome annotations as proteins of this length were deemed too short to be functional (Cheng et al., 2011). However, the annotation rules have considerably limited the exploration of the proteome. Based on the potential polycistronic nature of genes, a deeper ORF annotation from an exhaustive transcriptome has predicted all possible alternative ORFs (altORFs), which are defined as potential protein-coding ORFs located either in UTRs of transcripts, in alternative reading frames within the coding sequence of mRNAs, or in non-coding RNAs (Samandi et al., 2017; Brunet et al., 2018; Brunet et al., 2019). Numerous altORFs were identified to be both in-frame and out-of-frame of annotated ORFs. Many annotated altORFs are conserved in eukaryotes, suggesting that alternative proteins encoded from these alternative start codons might have a function across species. The community used ribosome profiling to capture all translation events across the genome and confirmed the translation of many altORFs (Bazzini et al., 2014; Ji et al., 2015; Samandi et al., 2017; Weaver et al., 2019). Combined with large-scale proteomics, these studies have led to the identification and functional relevance of alternative proteins translated from many altORFs located within mature transcripts (Saghatelian and Couso 2015; Na et al., 2018; Rothnagel and Menschaert 2018; Orr et al., 2020). Many functional studies showed that alternative proteins play central functions in the maintenance of cellular homeostasis (Delcourt et al., 2018; Cardon et al., 2020; Vergara et al., 2020; Brunet et al., 2021a; Cao et al., 2021; Ichihara, Nakayama, and Matsumoto 2022). In humans, mutations creating or deleting altORFs have been associated with physiopathological conditions such as amyotrophic lateral sclerosis (ALS) (Brunet et al., 2021b), craniofrontonasal syndrome (Tavares et al., 2019), and thrombocythemia (Wiestner et al., 1998). Interestingly, mutations found in cancers that are silent for reference proteins can impact the expression of alternative proteins resulting from the mutated mRNA, suggesting that alternative proteins could be new biomarkers of pathologies (Child, Miller, and Geballe 1999; Liu et al., 1999; Barbosa, Peixeiro, and Romao 2013; Sendoel et al., 2017; Schulz et al., 2018).
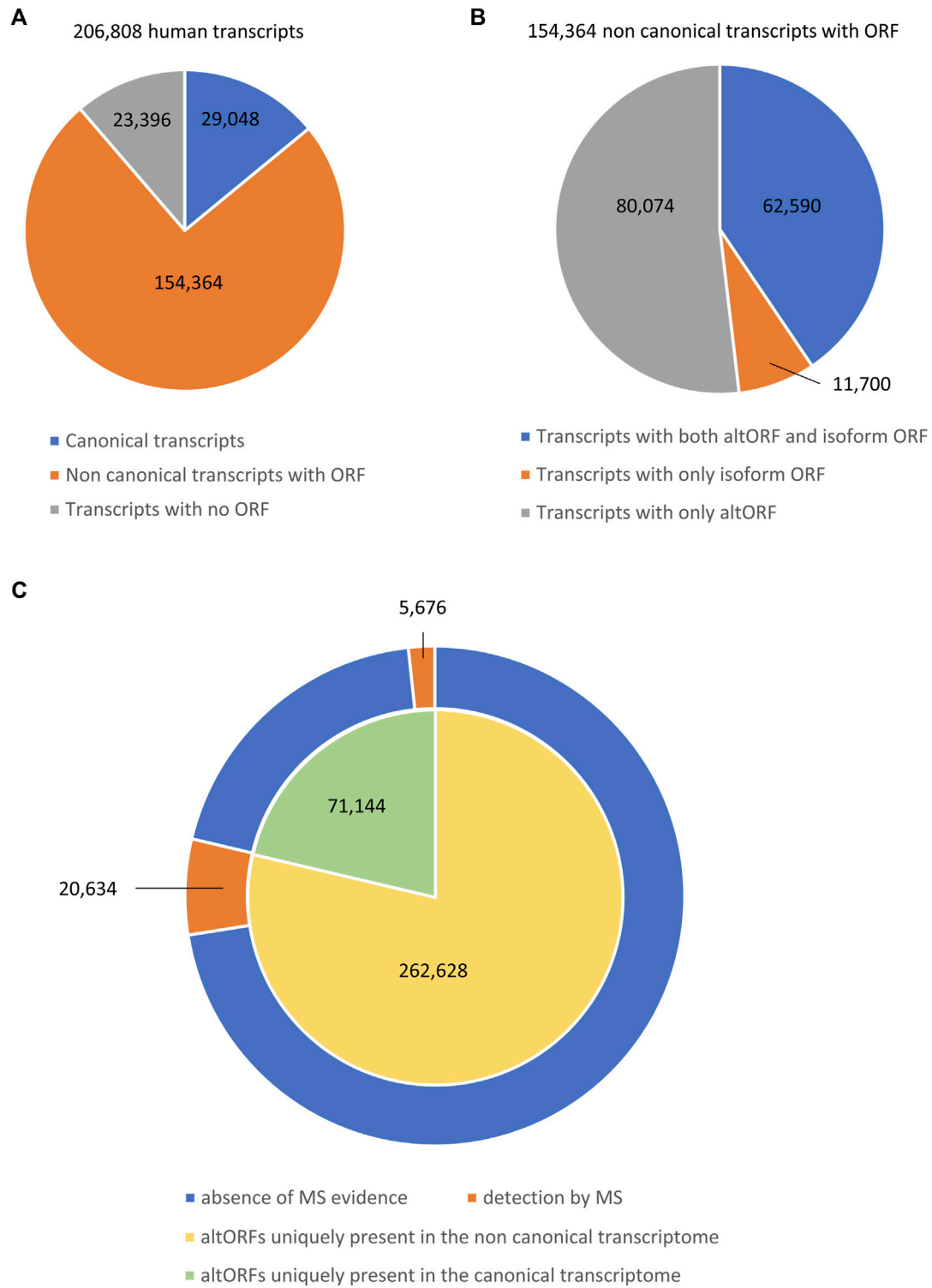
FIGURE 1
Composition of the human transcriptome. **(A)** Pie chart showing the number of different transcripts from the human reference genome (GRCh38 v95). Three types of transcripts are represented: canonical transcripts encoding a reference protein (blue), non-canonical transcripts generated through alternative splicing that contain an ORF (orange), and transcripts that do not have an annotated ORF (gray). **(B)** Pie chart showing the proportion of different sub-types of non-canonical transcripts containing an ORF. Three sub-types of transcripts are represented: non-canonical transcripts with both an alternative ORF (altORF) and an isoform ORF (blue), non-canonical transcripts with only an isoform ORF (orange), and non-canonical transcripts with only an altORF (gray). **(C)** Double pie chart representing the distribution of altORFs uniquely present in the canonical transcriptome (green) or the non-canonical transcriptome (yellow). Using the OpenProt database (Brunet et al., 2019), the evidence obtained by mass spectrometry (MS) of altORF-related proteins is represented in orange in the ring, while the absence of evidence is represented in blue.

A major problem is that alternative proteins expressed from these altORFs are usually not represented in the conventional protein databases (Brunet, Leblanc, and Roucou 2020; Cardon, Fournier, and Salzet 2021). Therefore, these alternative proteins represent a "ghost proteome" that was not considered until recently. Data-driven tools such as the sORF repository (Olexiouk, Van Criekinge, and Menschaert 2018) or the OpenProt database (Brunet et al., 2021a) have now been developed to offer a broader view of proteomes. The existence of thousands of altORFs hidden within known coding sequence of mRNAs raises the question of whether AS could also contribute to proteomic diversity through these small alternative proteins. To address this question, we performed a computational analysis using Ensembl human genome annotation (GRCh38 v95) and the OpenProt database (version 1.6) to determine the impact of AS on this hidden proteome. We identified a total of 206,808 transcripts including 29,048 transcripts defined as canonical as they encode reference proteins (Figure 1A). These transcripts might contain altORFs coding for alternative proteins. We also identified 154,364 transcripts (74.6%) that we categorized as non-canonical since they derive from AS but are not referenced to encode for reference proteins (Figure 1A). However, these transcripts may encode isoforms of reference proteins and/or contain an altORF. Finally, we identified 23,396 transcripts (11.3%) with no ORF according to the OpenProt database (Figure 1A). We next analyzed the non-canonical coding transcriptome landscape. Among these 154,364 transcripts, we identified 62,590 transcripts (40.5%) that contain both an ORF coding for an isoform of a reference protein and an altORF (Figure 1B). We found 80,074 transcripts (51.9%) only containing altORFs and 11,700 transcripts (7.6%) only containing an ORF coding for an isoform of a reference protein (Figure 1B). Our analysis highlights that AS generates numerous transcripts that do not encode for an isoform of a reference protein, supporting the claim by Tress and colleagues that AS might not be the key to proteomic complexity (Tress, Abascal, and Valencia 2017a). However, these transcripts contain altORFs that can potentially code for alternative proteins. These altORFs might also be commonly present in the related canonical transcripts as they could be located in the exons that are not directly affected by AS. We analyzed the distribution of these altORFs and identified 71,144 altORFs that were uniquely present in the canonical transcriptome (29,048 transcripts), while 262,628 altORFs were uniquely present in the non-canonical transcriptome (154, 364 transcripts) (Figure 1C). It represents an average of 2.4 unique altORFs per canonical transcript and 1.7 unique altORFs per non-canonical coding transcript. Using the OpenProt database that encompasses 87 ribosome profiling and 114 mass spectrometry studies from several species, tissues, and cell lines (Brunet et al., 2019), we looked for mass spectrometry evidence for all these altORFs. We found that 5,676 unique altORFs (7.98%) in canonical transcripts had evidence in mass spectrometry, while 20,634 unique altORFs (7.85%) in non-canonical transcripts produced alternative proteins detected by mass spectrometry (Figure 1C). This result clearly indicates that AS can indeed contribute to the human proteomic diversity through the translation of altORFs within mature RNAs.

Contribution of long non-coding RNAs and circular RNAs

Long non-coding RNAs (lncRNAs) represent an important part of the transcriptome (Liu et al., 2005; Derrien et al., 2012). LncRNAs are transcripts of 200 nucleotides or more that should not harbor protein-encoding ORFs (Dinger et al., 2008; Khalil et al., 2009; Derrien et al., 2012). Genome-wide translation profiling has recently revealed that small ORFs identified in lncRNA genes can code for micropeptides, polypeptides with a length of less than 100 amino acids essential for cellular growth (Chen et al., 2020). Other small peptides produced from lncRNAs have also been reported in functional studies (Odermatt et al., 1997; MacLennan and Kranias 2003; Slavoff et al., 2013; Ruiz-Orera et al., 2014; Pang, Mao, and Liu 2018; Wang J et al., 2019; Hartford Corrine and Lal, 2020; Nita et al., 2021; Mise et al., 2022). Eukaryotic lncRNA genes are usually composed of multiple exons with an average of 2.49 exons per human lncRNA gene (Khan, Wellinger, and Laurent 2021). LncRNA transcripts are efficiently spliced with a very similar distribution of AS-type events to that of protein-coding transcripts (Khan, Wellinger, and Laurent 2021). Hence, lncRNAs also generate multiple splice variants whose functional relevance can be associated with RNA-based differential functions (Khan, Wellinger, and Laurent 2021). Although the majority of alternatively spliced lncRNAs are likely non-functional, some of them can produce micropeptides. Indeed, specific splice variants of lncRNAs have the unique capability to produce functional micropeptides that are not encoded by the lncRNA of reference, that is, HOXB-AS3 lncRNA (Huang et al., 2017), LINC00948 lncRNA (Anderson et al., 2015), and LINC00665 lncRNA (Guo et al., 2020). Therefore, the proteomic diversity also depends on AS of lncRNAs. With a total of 354,855 lncRNA genes identified in 17 different species, the exact contribution of lncRNA splice variants to the proteomic complexity remains to be precisely determined and will be a major challenge in the field.

Circular RNAs (circRNAs) are produced from the back-splicing of linear RNAs where upstream splice-acceptor sites are covalently linked to downstream splice-donor sites to form an RNA loop structure (Kristensen et al., 2019). CircRNAs can be conserved during evolution and exhibit a tissue- or cell-specific expression (Kristensen et al., 2019; Santer, Bär, and Thum 2019). CircRNAs are functionally important as they act as microRNA decoys or scaffolds that sequester specific proteins (Chen et al., 2020). Due to their circular shape, circRNAs were not predicted to be translated, but there is growing evidence that circRNAs containing small ORFs can produce micropeptides that have a functional relevance (Legnini et al., 2017; Pamudurti et al., 2017; Liang et al., 2019; Lei et al., 2020; Sinha et al., 2022). It has been hypothesized that AS, particularly exon skipping, drives the formation of circRNAs. However, *in silico* analyses of AS and circRNA production in the human heart revealed that only 10% of circRNAs are produced from alternatively spliced exons, while 90% of circRNAs come from constitutive exons (Aufiero et al., 2018). Therefore, it is possible that AS can also impact the proteomic composition *via* circRNAs containing small ORFs, even though this contribution probably remains limited since circRNAs are described to largely be non-functional products of splicing errors (Xu and Zhang 2021). Future studies on circRNA translation will help uncover the circRNA-driven hidden proteome and enlighten on the functional importance of these novel proteins.

Perspectives

Although MS combined with long-read sequencing and ribosome profiling data has significantly improved the identification of new isoform proteins, many MS fragment spectra still remain unidentified and could potentially result from alternative proteins, micropeptides translated from lncRNAs, circRNAs, or other RNAs (Makarewich and

Olson 2017). Moreover, identifying isoform proteins or small proteins using "bottom–up" MS is challenging. An alternative form of a protein must have a tryptic peptide with more than eight amino acids in the region that differs from the canonical protein to be identified correctly. In addition, this peptide must be suitable for ionization and fragmentation. For small proteins with less than 100 amino acids, the chance to have unique detected peptides is strongly reduced compared to large proteins. Size selection, enrichment of small-size proteins, and careful selection of proteases may improve detection of low abundant proteins and micropeptides. Furthermore, matching MS spectra with custom databases will also help successfully identify novel isoform proteins or small-size micropeptides. "Top–down" proteomics, which characterizes intact proteins in complex mixtures without prior digestion, could be a good alternative approach. However, this method requires long ion accumulation, activation, and detection times and has not been achieved on a large scale due to lack of methods integrated with tandem MS. Despite significant advances, identifying new isoform proteins in the proteome complexity remains a challenge, and further improvements (e.g., methodology, filtering criteria, and database) will be required to substantially improve this situation in the future.

Determining which alternatively spliced transcripts produce proteins with important biological functions (i.e., isoform proteins, alternative proteins, and micropeptides) is the key to confirm the real impact of AS on proteomic complexity. To date, relatively few isoform and alternative proteins have been studied at the functional level, and the biological significance of AS-derived proteome remains obscure. For some AS events, functional consequences can be easily inferred based on changes in the protein sequence. Some alternatively spliced transcripts can encode protein isoforms, which lose or gain specific domains. Interestingly, 50% of AS events in the human transcriptome preserve the ORF and 65% of these frame-preserving splice variants are detected in polysome fractions and, hence, are likely translated (Weatheritt, Sterne-Weiler, and Blencowe 2016). This observation indicates that alternatively spliced transcripts with no frame preservation are potentially eliminated by quality control processes such as NMD. Indeed, some AS events can lead to the inclusion of highly conserved "poison" exons, which contain a premature truncation codon (Leclair et al., 2020). Although these exons do not contribute to the protein-coding capacity, their AS coupled to NMD plays an autoregulatory role in gene expression and protein abundance. Hence, the functional consequences of AS are not always obvious, and many studies failed to detect any differences in the activity of isoform proteins. However, the absence of functional relevance does not mean that there are no functional differences. Therefore, determining the biological function of a single AS event or an AS-derived product will be a major challenge of the proteomic era in the upcoming years.

AS also has a strong clinical relevance since dysregulations of AS have been associated with many chronic diseases including cancer (Ouyang et al., 2021; Zhang et al., 2021). It is, therefore, critical to advance the functional characterization of the AS-derived proteome, but the identification of AS events without regard to their contribution to proteomic diversity is also essential. Indeed, it is key to further study any potential AS alterations in diseases or pathological conditions as they could be valuable prognostic and diagnostic biomarkers. Such investigations could also provide tools for the development of therapeutics. Two splicing-based therapeutic agents are currently tested in clinical trials: small-molecule splicing modulators and antisense oligonucleotides (ASOs). Small-molecule drugs modulate the splicing activity by directly targeting the spliceosome and splicing factors. Surprisingly, these compounds do not induce global splicing inhibition but rather selective changes in AS for genes related to cell proliferation and apoptosis (Folco, Coil, and Reed 2011; Vigevani et al., 2017). However, potential problems of off-target effects require that AS mechanisms are fully understood before further clinical use. In contrast, ASOs are emerging as more secure therapeutic agents to modulate splicing. ASOs can specifically neutralize splice sites, inhibit the recruitment of specific RNA-binding proteins or inhibit the expression of specific splice variants (Rinaldi and Wood 2018). For instance, clinical applicability of ASO-based strategies has been successful in the treatment of patients with spinal muscular atrophy (Hua et al., 2008). ASOs could be used to specifically target specific disease-related splice variants, but advancing knowledge on the functional roles of isoform proteins is, hence, critical for efficient clinical interventions. Regardless of its contribution to proteomic diversity, targeting AS is now recognized an important area for clinical intervention.

Conclusion

On the contentious question "Does alternative splicing really expand proteomic diversity?," we can hereby affirm that AS indeed participates to proteomic complexity in many ways, that is, isoform proteins, alternative proteins, and micropeptides. In the light of this re-evaluation, the AS-related ghost proteome fills a gap and enlarges our vision of the current proteome. Importantly, the remaining limitations on the original question should be taken in consideration in future research endeavors. To continue assessing AS contribution to proteomic complexity, deeper ORF annotation and improvement of technologies and methodologies will be key to functional proteomic discoveries. With a repertoire of alternatively spliced transcripts now significantly expanded, more extensive functional studies on AS and its related proteome are necessary to unravel their unexpected implications in a variety of biological processes.

Author contributions

JMM and IK wrote the manuscript, and NG performed the bioinformatics analysis. XR supervised the bioinformatics analysis and revised the manuscript. BL designed, supervised the experiments, wrote, and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abascal, F., Ezkurdia, I., Rodriguez-Rivas, J., Rodriguez, J. M., del Pozo, A., Vazquez, J., et al. (2015). 'Alternatively spliced homologous exons have ancient origins and are highly expressed at the protein level. *PLoS Comput. Biol.* 11, e1004325. doi:10.1371/journal.pcbi.1004325

Agosto, L. M., Gazzara, M. R., Radens, C. M., Sidoli, S., Baeza, J., Garcia, B. A., et al. (2019). Deep profiling and custom databases improve detection of proteoforms generated by alternative splicing. *Genome Res.* 29, 2046–2055. doi:10.1101/gr.248435.119

Alfaro, J. A., Ignatchenko, A., Ignatchenko, V., Sinha, A., Boutros, P. C., and Kislinger, T. (2017). Detecting protein variants by mass spectrometry: A comprehensive study in cancer cell-lines. *Genome Med.* 9, 62. doi:10.1186/s13073-017-0454-9

Anderson, D. M., Anderson, K. M., Chang, C. L., Makarewich, C. A., Nelson, B. R., McAnally, J. R., et al. (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* 160, 595–606. doi:10.1016/j.cell.2015.01.009

Aufiero, S., van den Hoogenhof, M. M. G., Reckman, Y. J., Beqqali, A., van der Made, I., Kluin, J., et al. (2018). Cardiac circRNAs arise mainly from constitutive exons rather than alternatively spliced exons. *RNA* 24, 815–827. doi:10.1261/rna.064394.117

Barbosa, C., Peixeiro, I., and Romao, L. (2013). 'Gene expression regulation by upstream open reading frames and human disease. *PLoS Genet.* 9, e1003529. doi:10.1371/journal.pgen.1003529

Barbosa-Morais, N. L., Irimia, M., Pan, Q., Xiong, H. Y., Gueroussov, S., Lee, L. J., et al. (2012). The evolutionary landscape of alternative splicing in vertebrate species. *Science* 338, 1587–1593. doi:10.1126/science.1230612

Bazzini, A. A., Johnstone, T. G., Christiano, R., Mackowiak, S. D., Obermayer, B., Fleming, E. S., et al. (2014). 'Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J.* 33, 981–993. doi:10.1002/embj.201488411

Blakeley, P., Siepen, J. A., Lawless, C., and Hubbard, S. J. (2010). Investigating protein isoforms via proteomics: A feasibility study. *Proteomics* 10, 1127–1140. doi:10.1002/pmic.200900445

Blencowe, B. J. (2017). 'The relationship between alternative splicing and proteomic complexity. *Trends Biochem. Sci.* 42, 407–408. doi:10.1016/j.tibs.2017.04.001

Boutz, P. L., Stoilov, P., Li, Q., Lin, C. H., Chawla, G., Ostrow, K., et al. (2007). A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev.* 21, 1636–1652. doi:10.1101/gad.1558107

Brunet, M. A., Brunelle, M., Lucier, J. F., Delcourt, V., Levesque, M., Grenier, F., et al. (2019). 'OpenProt: A more comprehensive guide to explore eukaryotic coding potential and proteomes. *Nucleic Acids Res.* 47, D403–D10. doi:10.1093/nar/gky936

Brunet, M. A., Jacques, J. F., Nassari, S., Tyzack, G. E., McGoldrick, P., Zinman, L., et al. (2021a). 'The FUS gene is dual-coding with both proteins contributing to FUS-mediated toxicity. *EMBO Rep.* 22, e50640. doi:10.15252/embr.202050640

Brunet, M. A., Leblanc, S., and Roucou, X. (2020). 'Reconsidering proteomic diversity with functional investigation of small ORFs and alternative ORFs. *Exp. Cell Res.* 393, 112057. doi:10.1016/j.yexcr.2020.112057

Brunet, M. A., Levesque, S. A., Hunting, D. J., Cohen, A. A., and Roucou, X. (2018). 'Recognition of the polycistronic nature of human genes is critical to understanding the genotype-phenotype relationship. *Genome Res.* 28, 609–624. doi:10.1101/gr.230938.117

Brunet, M. A., Lucier, J. F., Levesque, M., Leblanc, S., Jacques, J. F., Al-Saedi, H. R. H., et al. (2021b). 'OpenProt 2021: Deeper functional annotation of the coding potential of eukaryotic genomes. *Nucleic Acids Res.* 49, D380–D388. doi:10.1093/nar/gkaa1036

Brunner, A. D., Thielert, M., Vasilopoulou, C., Ammar, C., Coscia, F., Mund, A., et al. (2022). 'Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol. Syst. Biol.* 18, e10798. doi:10.15252/msb.202110798

Buljan, M., Chalancon, G., Eustermann, S., Wagner, G. P., Fuxreiter, M., Bateman, A., et al. (2012). 'Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol. Cell* 46, 871–883. doi:10.1016/j.molcel.2012.05.039

Cao, X., Khitun, A., Luo, Y., Na, Z., Phoodokmai, T., Sappakhaw, K., et al. (2021). Alt-RPL36 downregulates the PI3K-AKT-mTOR signaling pathway by interacting with TMEM24. *Nat. Commun.* 12, 508. doi:10.1038/s41467-020-20841-6

Cardon, T., Fournier, I., and Salzet, M. (2021). 'Shedding light on the ghost proteome. *Trends Biochem. Sci.* 46, 239–250. doi:10.1016/j.tibs.2020.10.003

Cardon, T., Franck, J., Coyaud, E., Laurent, E. M. N., Damato, M., Maffia, M., et al. (2020). 'Alternative proteins are functional regulators in cell reprogramming by PKA activation. *Nucleic Acids Res.* 48, 7864–7882. doi:10.1093/nar/gkaa277

Castle, J. C., Zhang, C., Shah, J. K., Kulkarni, A. V., Kalsotra, A., Cooper, T. A., et al. (2008). 'Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat. Genet.* 40, 1416–1425. doi:10.1038/ng.264

Chen, H., Gao, F., He, M., Ding, X. F., Wong, A. M., Sze, S. C., et al. (2019). 'Long-Read RNA sequencing identifies alternative splice variants in hepatocellular carcinoma and tumor-specific isoforms. *Hepatology* 70, 1011–1025. doi:10.1002/hep.30500

Chen, J., Zachery Cogan, J., Nuñez, J. K., Fields, A. P., Britt Adamson, D. N., Matthias Mann, L., et al. (2020). 'Pervasive functional translation of noncanonical human open reading frames. *Science* 367, 1140–1146.

Chen, L., Bush, S. J., Tovar-Corona, J. M., Castillo-Morales, A., and Urrutia, A. O. (2014). 'Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. *Mol. Biol. Evol.* 31, 1402–1413. doi:10.1093/molbev/msu083

Cheng, H., Chan, W. S., Li, Z., Wang, D., Liu, S., and Zhou, Y. (2011). 'Small open reading frames: Current prediction techniques and future prospect. *Curr. Protein Pept. Sci.* 12, 503–507. doi:10.2174/138920311796957667

Cheng, Z., Teo, G., Krueger, S., Rock, T. M., Koh, H. W., Choi, H., et al. (2016). Differential dynamics of the mammalian mRNA and protein expression response to misfolding stress. *Mol. Syst. Biol.* 12, 855. doi:10.15252/msb.20156423

Child, S. J., Miller, M. K., and Geballe, A. P. (1999). Translational control by an upstream open reading frame in the HER-2/neu transcript. *J. Biol. Chem.* 274, 24335–24341. doi:10.1074/jbc.274.34.24335

De Paoli-Iseppi, R., Gleeson, J., and Clark, M. B. (2021). 'Isoform age - splice isoform profiling using long-read technologies. *Front. Mol. Biosci.* 8, 711733. doi:10.3389/fmolb.2021.711733

Delcourt, V., Brunelle, M., Roy, A. V., Jacques, J. F., Salzet, M., Fournier, I., et al. (2018). The protein coded by a short open reading frame, not by the annotated coding sequence, is the main gene product of the dual-coding gene MIEF1. *Mol. Cell Proteomics* 17, 2402–2411. doi:10.1074/mcp.RA118.000593

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* 22, 1775–1789. doi:10.1101/gr.132159.111

Dinger, M. E., Paulo AmaralTim, R. M., Marjan, E., and Askarian-Amiri, P. (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation'Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* 18, 1433–1445. doi:10.1101/gr.078378.108

Ellis, J. D., Barrios-Rodiles, M., Colak, R., Irimia, M., Kim, T., Calarco, J. A., et al. (2012). 'Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol. Cell* 46, 884–892. doi:10.1016/j.molcel.2012.05.037

Ezkurdia, I., del Pozo, A., Frankish, A., Rodriguez, J. M., Harrow, J., Ashman, K., et al. (2012). 'Comparative proteomics reveals a significant bias toward alternative protein isoforms with conserved structure and function. *Mol. Biol. Evol.* 29, 2265–2283. doi:10.1093/molbev/mss100

Ezkurdia, I., Rodriguez, J. M., Carrillo-de Santa Pau, E., Vazquez, J., Valencia, A., and Tress, M. L. (2015). 'Most highly expressed protein-coding genes have a single dominant isoform. *J. Proteome Res.* 14, 1880–1887. doi:10.1021/pr501286b

Fiszbein, A., Giono, L. E., Quaglino, A., Berardino, B. G., Sigaut, L., von Bilderling, C., et al. (2016). Alternative splicing of G9a regulates neuronal differentiation. *Cell Rep.* 14, 2797–2808. doi:10.1016/j.celrep.2016.02.063

Folco, E. G., Coil, K. E., and Reed, R. (2011). 'The anti-tumor drug E7107 reveals an essential role for SF3b in remodeling U2 snRNP to expose the branch point-binding region. *Genes & Dev.* 25, 440–444. doi:10.1101/gad.2009411

Fox-Walsh, K. L., and Hertel, K. J. (2009). 'Splice-site pairing is an intrinsically high fidelity process. *Proc. Natl. Acad. Sci. U. S. A.* 106, 1766–1771. doi:10.1073/pnas.0813128106

Garcia-Moreno, J. F., and Romao, L. (2020). 'Perspective in alternative splicing coupled to nonsense-mediated mRNA decay. *Int. J. Mol. Sci.* 21, 9424. doi:10.3390/ijms21249424

Genshaft, A. S., Li, S., Gallant, C. J., Darmanis, S., Prakadan, S. M., Ziegler, C. G., et al. (2016). 'Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. *Genome Biol.* 17, 188. doi:10.1186/s13059-016-1045-6

Giansanti, P., Tsiatsiani, L., Low, T. Y., and Heck, A. J. (2016). 'Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat. Protoc.* 11, 993–1006. doi:10.1038/nprot.2016.057

Gonzalez-Porta, M., Frankish, A., Rung, J., Harrow, J., and Brazma, A. (2013). 'Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. *Genome Biol.* 14, R70. doi:10.1186/gb-2013-14-7-r70

Graveley, B. R. (2001). Alternative splicing: Increasing diversity in the proteomic world. *Trends Genet.* 17, 100–107. doi:10.1016/s0168-9525(00)02176-4

Graveley, B. R., Brooks, A. N., Carlson, J. W., Duff, M. O., Landolin, J. M., Yang, L., et al. (2011). The developmental transcriptome of *Drosophila melanogaster. Nature* 471, 473–479. doi:10.1038/nature09715

Gueroussov, S., Gonatopoulos-Pournatzis, T., Irimia, M., Raj, B., Lin, Z. Y., Gingras, A. C., et al. (2015). An alternative splicing event amplifies evolutionary differences between vertebrates. *Science* 349, 868–873. doi:10.1126/science.aaa8381

Guo, B., Wu, S., Zhu, X., Zhang, L., Deng, J., Li, F., et al. (2020). 'Micropeptide CIP2A-BP encoded by LINC00665 inhibits triple-negative breast cancer progression. *EMBO J.* 39, e102190. doi:10.15252/embj.2019102190

Hartford Corrina, R., and Lal., A. (2020). 'When long noncoding becomes protein coding. *Mol. Cell. Biol.* 40, e00528-19. doi:10.1128/MCB.00528-19

Hsieh, A. C., Liu, Y., Edlind, M. P., Ingolia, N. T., Janes, M. R., Sher, A., et al. (2012). The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* 485, 55–61. doi:10.1038/nature10912

Hsu, S. N., and Hertel, K. J. (2009). 'Spliceosomes walk the line: Splicing errors and their impact on cellular function. *RNA Biol.* 6, 526–530. doi:10.4161/rna.6.5.9860

Hu, Y., Fang, L., Chen, X., Zhong, J. F., Li, M., and Wang, K. (2021). 'LIQA: Long-read isoform quantification and analysis. *Genome Biol.* 22, 182. doi:10.1186/s13059-021-02399-8

Hua, Y., Vickers, T. A., Okunola, H. L., Bennett, C. F., and Krainer, A. R. (2008). 'Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice. *Am. J. Hum. Genet.* 82, 834–848. doi:10.1016/j.ajhg.2008.01.014

Huang, J-Z., Chen, M., Chen, D., Gao, X- C., Zhu, S., Huang, H., et al. (2017). A peptide encoded by a putative lncRNA HOXB-AS3 suppresses colon cancer growth. *Mol. Cell* 68, 171–184. doi:10.1016/j.molcel.2017.09.015

Ichihara, K., Nakayama, K. I., and Matsumoto, A. (2022). 'Identification of unannotated coding sequences and their physiological functions. *J. Biochem.*, mvac064. doi:10.1093/jb/mvac064

Inada, T. (2017). 'The ribosome as a platform for mRNA and nascent polypeptide quality control. *Trends Biochem. Sci.* 42, 5–15. doi:10.1016/j.tibs.2016.09.005

Irimia, M., Weatheritt, R. J., Ellis, J. D., Parikshak, N. N., Gonatopoulos-Pournatzis, T., Babor, M., et al. (2014). A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* 159, 1511–1523. doi:10.1016/j.cell.2014.11.035

Jeong, S. K., Kim, C. Y., and Paik, Y. K. (2018). 'ASV-ID, a proteogenomic workflow to predict candidate protein isoforms on the basis of transcript evidence. *J. Proteome Res.* 17, 4235–4242. doi:10.1021/acs.jproteome.8b00548

Ji, Z., Song, R., Regev, A., and Struhl, K. (2015). 'Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *Elife* 4, e08890. doi:10.7554/eLife.08890

Jiao, Y., Robison, A. J., Bass, M. A., and Colbran, R. J. (2008). 'Developmentally regulated alternative splicing of densin modulates protein-protein interaction and subcellular localization. *J. Neurochem.* 105, 1746–1760. doi:10.1111/j.1471-4159.2008.05280.x

Khalil, Ahmad M., Guttman, Mitchell, Huarte, Maite, Garber, Manuel, Raj, Arjun, Rivea Morales, Dianali, et al. (2009). 'Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 106, 11667–11672. doi:10.1073/pnas.0904715106

Khan, M. R., Wellinger, R. J., and Laurent, B. (2021). 'Exploring the alternative splicing of long noncoding RNAs. *Trends Genet.* 37, 695–698. doi:10.1016/j.tig.2021.03.010

Kianianmomeni, A., Ong, C. S., Ratsch, G., and Hallmann, A. (2014). 'Genome-wide analysis of alternative splicing in Volvox carteri. *BMC Genomics* 15, 1117. doi:10.1186/1471-2164-15-1117

Kristensen, A. R., Gsponer, J., and Foster, L. J. (2013). 'Protein synthesis rate is the predominant regulator of protein expression during differentiation. *Mol. Syst. Biol.* 9, 689. doi:10.1038/msb.2013.47

Kristensen, Lasse S., MariaAndersenStagstedHansen, S. Lotte V. W. Karoline K. Ebbesen, Thomas B., Kjems, Jørgen, Hansen, T. B., and Kjems, J. (2019). 'The biogenesis, biology and characterization of circular RNAs. *Nat. Rev. Genet.* 20, 675–691. doi:10.1038/s41576-019-0158-7

Kriventseva, E. V., Koch, I., Apweiler, R., Vingron, M., Bork, P., Gelfand, M. S., et al. (2003). 'Increase of functional diversity by alternative splicing. *Trends Genet.* 19, 124–128. doi:10.1016/S0168-9525(03)00023-4

Krull, M., Brosius, J., and Schmitz, J. (2005). 'Alu-SINE exonization: En route to protein-coding function. *Mol. Biol. Evol.* 22, 1702–1711. doi:10.1093/molbev/msi164

Lackner, D. H., Schmidt, M. W., Wu, S., Wolf, D. A., and Bahler, J. (2012). 'Regulation of transcriptome, translation, and proteome in response to environmental stress in fission yeast. *Genome Biol.* 13, R25. doi:10.1186/gb-2012-13-4-r25

Laurent, B., Ruitu, L., Murn, J., Hempel, K., Ferrao, R., Xiang, Y., et al. (2015). A specific LSD1/KDM1A isoform regulates neuronal differentiation through H3K9 demethylation. *Mol. Cell* 57, 957–970. doi:10.1016/j.molcel.2015.01.010

Leclair, N. K., Brugiolo, M., Urbanski, L., Lawson, S. C., Thakar, K., Yurieva, M., et al. (2020). 'Poison exon splicing regulates a coordinated network of SR protein expression during differentiation and tumorigenesis. *Mol. Cell* 80, 648–665. doi:10.1016/j.molcel.2020.10.019

Legnini, I., Di Timoteo, G., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., et al. (2017). 'Circ-ZNF609 is a circular RNA that can Be translated and functions in myogenesis. *Mol. Cell* 66, 22–37. doi:10.1016/j.molcel.2017.02.017

Lei, M., Zheng, G., Ning, Q., Zheng, J., and Dong, D. (2020). 'Translation and functional roles of circular RNAs in human cancer. *Mol. Cancer* 19, 30. doi:10.1186/s12943-020-1135-7

Leoni, G., Le Pera, L., Ferre, F., Raimondo, D., and Tramontano, A. (2011). 'Coding potential of the products of alternative splicing in human. *Genome Biol.* 12, R9. doi:10.1186/gb-2011-12-1-r9

Li, G. W., Burkhardt, D., Gross, C., and Weissman, J. S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* 157, 624–635. doi:10.1016/j.cell.2014.02.033

Liang, W. C., Wong, C. W., Liang, P. P., Shi, M., Cao, Y., Rao, S. T., et al. (2019). Translation of the circular RNA circβ-catenin promotes liver cancer cell growth through activation of the Wnt pathway. *Genome Biol.* 20, 84. doi:10.1186/s13059-019-1685-4

Light, S., and Elofsson, A. (2013). 'The impact of splicing on protein domain architecture. *Curr. Opin. Struct. Biol.* 23, 451–458. doi:10.1016/j.sbi.2013.02.013

Lin, L., Jiang, P., Park, J. W., Wang, J., Lu, Z. X., Lam, M. P., et al. (2016). 'The contribution of Alu exons to the human proteome. *Genome Biol.* 17, 15. doi:10.1186/s13059-016-0876-5

Liu, C., Bai, B., Skogerbø, G., Cai, L., Deng, W., Zhang, Y., et al. (2005). 'NONCODE: An integrated knowledge database of non-coding RNAs. *Nucleic Acids Res.* 33, D112–D115. doi:10.1093/nar/gki041

Liu, L., Dilworth, D., Gao, L., Monzon, J., Summers, A., Lassam, N., et al. (1999). 'Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and predisposes to melanoma. *Nat. Genet.* 21, 128–132. doi:10.1038/5082

Liu, S., Zhou, B., Wu, L., Sun, Y., Chen, J., and Liu, S. (2021). Single-cell differential splicing analysis reveals high heterogeneity of liver tumor-infiltrating T cells. *Sci. Rep.* 11, 5325. doi:10.1038/s41598-021-84693-w

Liu, Y., and Aebersold, R. (2016). 'The interdependence of transcript and protein abundance: New data--new complexities. *Mol. Syst. Biol.* 12, 856. doi:10.15252/msb.20156720

Liu, Y., Beyer, A., and Aebersold, R. (2016). On the dependency of cellular protein levels on mRNA abundance. *Cell* 165, 535–550. doi:10.1016/j.cell.2016.03.014

Liu, Y., Gonzalez-Porta, M., Santos, S., Brazma, A., Marioni, J. C., Aebersold, R., et al. (2017). Impact of alternative splicing on the human proteome. *Cell Rep.* 20, 1229–1241. doi:10.1016/j.celrep.2017.07.025

Ly, T., Ahmad, Y., Shlien, A., Soroka, D., Mills, A., Emanuele, M. J., et al. (2014). 'A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. *Elife* 3, e01630. doi:10.7554/eLife.01630

MacLennan, D. H., and Kranias, E. G. (2003). Phospholamban: A crucial regulator of cardiac contractility. *Nat. Rev. Mol. Cell Biol.* 4, 566–577. doi:10.1038/nrm1151

Makarewich, C. A., and Olson, E. N. (2017). 'Mining for micropeptides. *Trends Cell Biol.* 27, 685–696. doi:10.1016/j.tcb.2017.04.006

Marx, V. (2023). 'Method of the year: Long-read sequencing. *Nat. Methods* 20, 6–11. doi:10.1038/s41592-022-01730-w

McManus, J., Cheng, Z., and Vogel, C. (2015). 'Next-generation analysis of gene expression regulation--comparing the roles of synthesis and degradation. *Mol. Biosyst.* 11, 2680–2689. doi:10.1039/c5mb00310e

Mercer, T. R., Gerhardt, D. J., Dinger, M. E., Crawford, J., Trapnell, C., Jeddeloh, J. A., et al. (2011). 'Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.* 30, 99–104. doi:10.1038/nbt.2024

Merkin, J., Russell, C., Chen, P., and Burge, C. B. (2012). 'Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* 338, 1593–1599. doi:10.1126/science.1228186

Mise, S., Matsumoto, A., Shimada, K., Hosaka, T., Takahashi, M., Ichihara, K., et al. (2022). 'Kastor and Polluks polypeptides encoded by a single gene locus cooperatively regulate VDAC and spermatogenesis. *Nat. Commun.* 13, 1071. doi:10.1038/s41467-022-28677-y

Modrek, B., and Lee, C. (2002). 'A genomic view of alternative splicing. *Nat. Genet.* 30, 13–19. doi:10.1038/ng0102-13

Muhammad, S., Xu, X., Zhou, W., and Wu, L. (2022e1758). *Alternative splicing: An efficient regulatory approach towards plant developmental plasticity.* Wiley Interdiscip Rev RNA.

Na, C. H., Barbhuiya, M. A., Kim, M. S., Verbruggen, S., Eacker, S. M., Pletnikova, O., et al. (2018). 'Discovery of noncanonical translation initiation sites through mass spectrometric analysis of protein N termini. *Genome Res.* 28, 25–36. doi:10.1101/gr.226050.117

Nakka, K., Ghigna, C., Gabellini, D., and Dilworth, F. J. (2018). 'Diversification of the muscle proteome through alternative splicing. *Skelet. Muscle* 8, 8. doi:10.1186/s13395-018-0152-3

Nilsen, T. W., and Graveley, B. R. (2010). 'Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463, 457–463. doi:10.1038/nature08909

Nita, A., Matsumoto, A., Tang, R., Shiraishi, C., Ichihara, K., Saito, D., et al. (2021). 'A ubiquitin-like protein encoded by the "noncoding" RNA TINCR promotes keratinocyte proliferation and wound healing. *PLoS Genet.* 17, e1009686. doi:10.1371/journal.pgen.1009686

Odermatt, A., Taschner, P. E., Scherer, S. W., Beatty, B., Khanna, V. K., Cornblath, D. R., et al. (1997). 'Characterization of the gene encoding human sarcolipin (SLN), a proteolipid associated with SERCA1: Absence of structural mutations in five patients with brody disease. *Genomics* 45, 541–553. doi:10.1006/geno.1997.4967

Ohta, T. (1992). 'Theoretical study of near neutrality. II. Effect of subdivided population structure with local extinction and recolonization. *Genetics* 130, 917–923. doi:10.1093/genetics/130.4.917

Olexiouk, V., Van Criekinge, W., and Menschaert, G. (2018). 'An update on sORFs.org: A repository of small ORFs identified by ribosome profiling. *Nucleic Acids Res.* 46, D497–D502. doi:10.1093/nar/gkx1130

Orr, M. W., Mao, Y., Storz, G., and Qian, S. B. (2020). 'Alternative ORFs and small ORFs: Shedding light on the dark proteome. *Nucleic Acids Res.* 48, 1029–1042. doi:10.1093/nar/gkz734

Osmanli, Z., Falgarone, T., Samadova, T., Aldrian, G., Leclercq, J., Shahmuradov, I., et al. (2022). *The Difference in Structural States between Canonical Proteins and Their Isoforms Established by Proteome-Wide Bioinformatics Analysis*, 12.Biomolecules

Ouyang, J., Zhang, Y., Xiong, F., Zhang, S., Gong, Z., Yan, Q., et al. (2021). 'The role of alternative splicing in human cancer progression. *Am. J. Cancer Res.* 11, 4642–4667.

Pamudurti, N. R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., et al. (2017). 'Translation of CircRNAs. *Mol. Cell* 66, 9–21. doi:10.1016/j.molcel.2017.02.021

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., and Blencowe, B. J. (2008). 'Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–1415. doi:10.1038/ng.259

Pang, Y., Mao, C., and Liu, S. (2018). 'Encoding activities of non-coding RNAs. *Theranostics* 8, 2496–2507. doi:10.7150/thno.24677

Pickrell, J. K., Pai, A. A., Gilad, Y., and Pritchard, J. K. (2010). Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet.* 6, e1001236. doi:10.1371/journal.pgen.1001236

Poulos, R. C., Hains, P. G., Shah, R., Lucas, N., Xavier, D., Manda, S. S., et al. (2020). 'Strategies to enable large-scale proteomics for reproducible research. *Nat. Commun.* 11, 3793. doi:10.1038/s41467-020-17641-3

Ramani, A. K., Calarco, J. A., Pan, Q., Mavandadi, S., Wang, Y., Nelson, A. C., et al. (2011). Genome-wide analysis of alternative splicing in *Caenorhabditis elegans*. *Genome Res.* 21, 342–348. doi:10.1101/gr.114645.110

Reddy, A. S., Marquez, Y., Kalyna, M., and Barta, A. (2013). 'Complexity of the alternative splicing landscape in plants. *Plant Cell* 25, 3657–3683. doi:10.1105/tpc.113.117523

Reixachs-Sole, M., Ruiz-Orera, J., Alba, M. M., and Eyras, E. (2020). 'Ribosome profiling at isoform level reveals evolutionary conserved impacts of differential splicing on the proteome. *Nat. Commun.* 11, 1768. doi:10.1038/s41467-020-15634-w

Rinaldi, C., and Wood, M. J. A. (2018). 'Antisense oligonucleotides: The next frontier for treatment of neurological disorders. *Nat. Rev. Neurol.* 14, 9–21. doi:10.1038/nrneurol.2017.148

Rothnagel, J., and Menschaert, G. (2018). Short open reading frames and their encoded peptides. *Proteomics* 18, e1700035.

Ruiz-Orera, J., Messeguer, X., Subirana, J. A., and Alba, M. M. (2014). 'Long non-coding RNAs as a source of new peptides. *Elife* 3, e03523. doi:10.7554/eLife.03523

Saghatelian, A., and Couso, J. P. (2015). 'Discovery and characterization of smORF-encoded bioactive polypeptides. *Nat. Chem. Biol.* 11, 909–916. doi:10.1038/nchembio.1964

Samandi, S., Roy, A. V., Delcourt, V., Lucier, J. F., Gagnon, J., Beaudoin, M. C., et al. (2017). 'Deep transcriptome annotation enables the discovery and functional characterization of cryptic small proteins. *Elife* 6.

Santer, L., Bär, C., and Thomas, T. (2019). 'Circular RNAs: A novel class of functional RNA molecules with a therapeutic perspective. *Mol. Ther.* 27, 1350–1363. doi:10.1016/j.ymthe.2019.07.001

Saudemont, B., Popa, A., Parmley, J. L., Rocher, V., Blugeon, C., Necsulea, A., et al. (2017). 'The fitness cost of mis-splicing is the main determinant of alternative splicing patterns. *Genome Biol.* 18, 208. doi:10.1186/s13059-017-1344-6

Schoof, E. M., Furtwangler, B., Uresin, N., Rapin, N., Savickas, S., Gentil, C., et al. (2021). 'Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. *Nat. Commun.* 12, 3341. doi:10.1038/s41467-021-23667-y

Schulz, J., Mah, N., Neuenschwander, M., Kischka, T., Ratei, R., Schlag, P. M., et al. (2018). 'Loss-of-function uORF mutations in human malignancies. *Sci. Rep.* 8, 2395. doi:10.1038/s41598-018-19201-8

Sendoel, A., Dunn, J. G., Rodriguez, E. H., Naik, S., Gomez, N. C., Hurwitz, B., et al. (2017). 'Translation from unconventional 5' start sites drives tumour initiation. *Nature* 541, 494–499. doi:10.1038/nature21036

Sinha, T., Panigrahi, C., Das, D., and Chandra Panda, A. (2022). Circular RNA translation, a path to hidden proteome. *Wiley Interdiscip. Rev. RNA* 13, e1685.

Skandalis, A. (2016). 'Estimation of the minimum mRNA splicing error rate in vertebrates. *Mutat. Res.* 784-785, 34–38. doi:10.1016/j.mrfmmm.2016.01.002

Slavoff, S. A., Mitchell, A. J., Schwaid, A. G., Cabili, M. N., Ma, J., Levin, J. Z., et al. (2013). Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat. Chem. Biol.* 9, 59–64. doi:10.1038/nchembio.1120

Specht, H., Emmott, E., Petelski, A. A., Huffman, R. G., Perlman, D. H., Serra, M., et al. (2021). 'Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biol.* 22, 50. doi:10.1186/s13059-021-02267-5

Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., et al. (2005). Function of alternative splicing. *Gene* 344, 1–20. doi:10.1016/j.gene.2004.10.022

Tapial, J., Ha, K. C. H., Sterne-Weiler, T., Gohr, A., Braunschweig, U., Hermoso-Pulido, A., et al. (2017). 'An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Res.* 27, 1759–1768. doi:10.1101/gr.220962.117

Tavares, R., Kague, E., Musso, C. M., Alegria, T. G. P., Freitas, R. S., Bertola, D. R., et al. (2019). 'Craniofrontonasal syndrome caused by introduction of a novel uATG in the 5'UTR of EFNB1. *Mol. Syndromol.* 10, 40–47. doi:10.1159/000490635

Tress, M. L., Abascal, F., and Valencia, A. (2017a). 'Alternative splicing may not Be the key to proteome complexity. *Trends Biochem. Sci.* 42, 98–110. doi:10.1016/j.tibs.2016.08.008

Tress, M. L., Abascal, F., and Valencia, A. (2017b). 'Most alternative isoforms are not functionally important. *Trends Biochem. Sci.* 42, 408–410. doi:10.1016/j.tibs.2017.04.002

Troskie, R. L., Jafrani, Y., Mercer, T. R., Ewing, A. D., Faulkner, G. J., and Cheetham, S. W. (2021). 'Long-read cDNA sequencing identifies functional pseudogenes in the human transcriptome. *Genome Biol.* 22, 146. doi:10.1186/s13059-021-02369-0

Uapinyoying, P., Goecks, J., Knoblach, S. M., Panchapakesan, K., Bonnemann, C. G., Partridge, T. A., et al. (2020). 'A long-read RNA-seq approach to identify novel transcripts of very large genes. *Genome Res.* 30, 885–897. doi:10.1101/gr.259903.119

Vergara, D., Verri, T., Damato, M., Trerotola, M., Simeone, P., Franck, J., et al. (2020). 'A hidden human proteome signature characterizes the epithelial mesenchymal transition program. *Curr. Pharm. Des.* 26, 372–375. doi:10.2174/1381612826666200129091610

Vigevani, L., Gohr, A., Webb, T., Irimia, M., and Valcarcel, J. (2017). 'Molecular basis of differential 3' splice site sensitivity to anti-tumor drugs targeting U2 snRNP. *Nat. Commun.* 8, 2100. doi:10.1038/s41467-017-02007-z

Vogel, C., and Marcotte, E. M. (2012). 'Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* 13, 227–232. doi:10.1038/nrg3185

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476. doi:10.1038/nature07509

Wang, J., Zhu, S., Meng, N., He, Y., Lu, R., and Yan, G. R. (2019). 'ncRNA-Encoded peptides or proteins and cancer. *Mol. Ther.* 27, 1718–1725. doi:10.1016/j.ymthe.2019.09.001

Wang, X., Codreanu, S. G., Wen, B., Li, K., Chambers, M. C., Liebler, D. C., et al. (2018). Detection of proteome diversity resulted from alternative splicing is limited by trypsin cleavage specificity. *Mol. Cell Proteomics* 17, 422–430. doi:10.1074/mcp.RA117.000155

Wang, X., You, X., Langer, J. D., Hou, J., Rupprecht, F., Vlatkovic, I., et al. (2019). 'Full-length transcriptome reconstruction reveals a large diversity of RNA and protein isoforms in rat hippocampus. *Nat. Commun.* 10, 5009. doi:10.1038/s41467-019-13037-0

Weatheritt, R. J., Sterne-Weiler, T., and Blencowe, B. J. (2016). 'The ribosome-engaged landscape of alternative splicing. *Nat. Struct. Mol. Biol.* 23, 1117–1123. doi:10.1038/nsmb.3317

Weaver, J., Mohammad, F., Buskirk, A. R., and Storz, G. (2019). *Identifying small proteins by ribosome profiling with stalled initiation complexes*, 10. mBio.

Wiestner, A., Schlemper, R. J., van der Maas, A. P., and Skoda, R. C. (1998). 'An activating splice donor mutation in the thrombopoietin gene causes hereditary thrombocythaemia. *Nat. Genet.* 18, 49–52. doi:10.1038/ng0198-49

Wright, D. J., Hall, N. A. L., Irish, N., Man, A. L., Glynn, W., Mould, A., et al. (2022). 'Long read sequencing reveals novel isoforms and insights into splicing regulation during cell state changes. *BMC Genomics* 23, 42. doi:10.1186/s12864-021-08261-2

Xu, C., and Zhang, J. (2021). 'Mammalian circular RNAs result largely from splicing errors. *Cell Rep.* 36, 109439. doi:10.1016/j.celrep.2021.109439

Zhang, Y., Qian, J., Gu, C., and Yang, Y. (2021). 'Alternative splicing and cancer: A systematic review. *Signal Transduct. Target Ther.* 6, 78. doi:10.1038/s41392-021-00486-7

Zikherman, J., and Weiss, A. (2008). Alternative splicing of CD45: The tip of the iceberg. *Immunity* 29, 839–841. doi:10.1016/j.immuni.2008.12.005