Alternative Transcripts Diversify Genome Function for Phenome Relevance to Health and Diseases

Shane A. Carrion, Jennifer J. Michal and Zhihua Jiang *

Department of Animal Sciences and Center for Reproductive Biology, Washington State University, Pullman, WA 99164-7620, USA; shane.carrion@wsu.edu (S.A.C.); jennifer_michal@wsu.edu (J.J.M.)
* Correspondence: jiangz@wsu.edu; Tel.: +1-509-335-8761

Abstract: Manipulation using alternative exon splicing (AES), alternative transcription start (ATS), and alternative polyadenylation (APA) sites are key to transcript diversity underlying health and disease. All three are pervasive in organisms, present in at least 50% of human protein-coding genes. In fact, ATS and APA site use has the highest impact on protein identity, with their ability to alter which first and last exons are utilized as well as impacting stability and translation efficiency. These RNA variants have been shown to be highly specific, both in tissue type and stage, with demonstrated importance to cell proliferation, differentiation and the transition from fetal to adult cells. While alternative exon splicing has a limited effect on protein identity, its ubiquity highlights the importance of these minor alterations, which can alter other features such as localization. The three processes are also highly interwoven, with overlapping, complementary, and competing factors, RNA polymerase II and its CTD (C-terminal domain) chief among them. Their role in development means dysregulation leads to a wide variety of disorders and cancers, with some forms of disease disproportionately affected by specific mechanisms (AES, ATS, or APA). Challenges associated with the genome-wide profiling of RNA variants and their potential solutions are also discussed in this review.

Keywords: RNA variants; genome–phenome bridges; health and disease relevance; challenges and solutions

1. Introduction

The number of genes in the human genome has been an open question in biology for decades. Historically, there were several significant waves of interest from the scientific communities, speculating answers to this question. The earliest recorded attempt can be attributed to James Spuhler, who in 1948 published an article titled "On the Number of Genes in Man", where he proposed two estimates, 42 k genes, extrapolating from the chromosomal length of fruit fly genes, and 20–30 k, based on loci count derived from X-linked lethal mutations [1]. Vogel produced the next estimate in 1964 in "A Preliminary Estimate of the Number of Human Genes". By assuming that the entire genome was protein-coding and genes were of comparable length (reasonable assumptions at the time), he used the molecular weight of hemoglobin to calculate the DNA weight of haploid chromosomes and divided it by his "standard" gene size, predicting an enormous 6.7 million genes, which he acknowledged seemed "disturbingly high". He then posited that instead using the gene length of Dipteran giant chromosomes (~50 k nucleotides) would place the number of genes instead at 67 k, a number he was much more comfortable with [2].

The announcement of the Human Genome Project (HGP) appeared to re-galvanize interest in the subject, with a slate of papers on the subject being published in the 1990's. The HGP was launched in 1990, with the goal of constructing the first full sequence of the human genome and identifying all protein-coding genes. The first prediction released by the Human Genome Project was 100,000 genes, driven by an assumption that that standard gene size was 30 kb, and was used as a baseline for many years afterwards [3–5].

The estimates that came after covered the gamut, from lows of 14,000 all the way up to 312,000, utilizing a variety of methods such as estimating from ESTs, chromosomes, and genome homologies [6–9]. Even as late as 2000, after the first rough draft of the genome assembly had been released, estimates varied significantly, from 26 k to 120 k, highlighting the difficulty of identifying protein-coding genes [10–14]. Though we now know the true number is likely close to 20 k, dependent on the stringency of filtering, it still fluctuates as our understanding evolves and more experimental evidence is found [15,16].

With that, the question became how our systems were able to develop and display such complexity with a protein-coding gene count (~20 k genes in 3000 Mb) a scant few hundred more genes than the simple nematode *C. elegans* (~19 k genes in 97 Mb) [7,17]. Human's complex system requires a diversity of gene products to build and maintain the ~30 trillion cells through a rapid progression of tissue- and stage-specific proliferation, differentiation, and development [18]. Decades prior, the discovery of alternative splicing (AS), alternative transcription start (ATS) and alternative polyadenylation (APA) sites hinted at potential explanations. At the time of their discovery, however, their significance was not well understood. Most protein-coding genes possess a dominant isoform, the product most prevalent in cells and tissues across time points and development stages, but these mechanisms allowed the creation of alternative transcripts from these same genes [19].

For many years, AS was considered the dominant mechanism, contributing to transcriptome diversity and publications focused on the subject climbing 425% over 10 years, from 243 a year in 1990 when the HGP was announced to 1276 by the year 2000 with the completion of the first draft of the human genome assembly. Alternative splicing refers to a process where exons from the same gene are combined into different mRNA transcripts, allowing for multiple but related proteins with distinct structures and functions. Here we will refer to AS as alternative exon splicing (AES) to distinguish it from ATS and APA, which are also forms of alternative splicing. However, researchers have mounting evidence that AES is not the primary driver of transcriptome and proteome diversity, but our understanding of these processes (especially ATS and APA which have received relatively minimal attention comparatively) is still shallow [20,21].

Estimates for transcript variants have also ranged significantly over the years, from as low as ~46 k to over 300,000. Advancements in transcripts have led to more accurate gene databases (such as CHESS, Ensembl, Gencode, and RefSeq). Although the differences between them are shrinking, many discrepancies still exist [22–26]. The question we face now is how many transcripts have biological functions and how many produce a protein product? Here we review and summarize key findings about AES, ATS and APA site usages, exploring their prevalence, tissue use, motifs, and disease trends. We also discuss the challenges associated with RNA variant profiling, propose some solutions, and catalogue a significant number of experimentally verified isoforms and isoform functions from the available literature, comparing them to currently understood motifs.

2. Alternative Exon Splicing (AES) Events: Features and Functions

2.1. AES—Increasingly Prevalent in Complex Organisms

Splicing (at the pre-mRNA stage, co- or post-transcriptionally) was discovered in an adenovirus experiment in 1977, which also discussed possible regulation of alternative splicing [27]. Alternative splicing was formally proposed as a theory by Walter Gilbert in 1978 [28]. The application of AES is influenced by the strength of the splicing signal, intronic/exonic enhancers or silencers, RNA binding proteins (RBPs), epigenetic modifications, genetic mutations and more. These factors can be fine-tuned by the organism to accommodate the development stage, differentiation, tissue/cell type, and other environmental elements [29,30]. AES sites are present in 95% of genes, with the average human gene producing three or more alternative transcripts, but is considered the least impactful in its contribution to RNA and protein diversity [20,21,30]. Despite this, its prevalence in higher eukaryotes has increased substantially from primitive eukaryotic organisms such

as *C. elegans*, with an average of 6.4 exons per transcript compared to ~11 in humans. *C. elegans* also undergoes alterative exon splicing in only 25% of their protein-coding genes, indicating an evolutionary advantage to incorporating additional splicing elements [17].

There are currently five recognized forms of alternative splicing: exon skipping (aka cassette alternative), alternative 5′ and 3′ splice site within exons (where one side has a constitutive splice site and the other has 2+ alternative splice sites—meaning there are alternate regions that can be included or excluded), intron retention, and mutually exclusive alternative exons (two exons where one or the other, but not both, can be included) [31–33]. Exon skipping (~30%) and alternative 5′ or 3′ splice site within exons (~25%) account for most AES events in eukaryotes. Alterations to the mRNA may introduce premature stop codons (PTCs) resulting in truncated proteins, which frequently ends in decay of the RNA products through nonsense-mediated RNA decay (NMD) pathways. Multiple studies have shown the majority of alternatively spliced transcripts are either not expected to or do not produce a protein product or express it at such a minor level as to be undetectable using mass spectrometry [20,34]. These transcripts may be producing micropeptides (some new research suggests this may be the case), peptides transcribed from a short open reading frame (sORF) with a length of 100 or less amino acids (AA), or have regulatory functions as RNA [20,35,36]. Shorter isoforms are most often missing one or more exons (whole or partial), leading to potential domain loss, such as localization signals, regulatory domains, and binding sites [32].

2.2. AES—Dominantly Located in the Cytoplasm

Changes that lead to altered localization signals can affect the ability of an RNA or protein to be properly positioned, either by causing the transporter protein to be unable to dock or removing the localization signal entirely, as is the case in the isoform c-FLIP$_S$. The *CFLAR* gene isoforms (called c-FLIP) are Death Effector domain containing proteins that are recruited to the DISC complex and regulate caspase-8 and 10 as well as DR5, playing a role in FAS-mediated apoptosis and necroptosis as well as T-cell proliferation. While the long form contains a catalytically inactive caspase-like domain that contains a nuclear localization signal, resulting in a large proportion of that isoform in the nucleus, the short form includes exon 7, which contains a stop codon. This truncated protein is missing the domain containing the localization signal and is restricted to the cytoplasm where it acts in an anti-apoptotic manner, as opposed to the long form that can be either pro- or anti-apoptotic in function [37,38].

NUMA1 is another example of short isoforms localizing in the cytosol due to alternative exon splicing. The full (long) form of *NUMA1* is a large protein (~238 kDa) consisting of N and C terminal globular domains, with a long central coiled-coil domain, and acts as a structural hub in the nuclear matrix, interacting with microtubules and involved in the formation and positioning of mitotic spindles. The nuclear localization signal in its C-terminal region allows this isoform to perform its function. The short isoform, *NUMA1-s*, consists of only the N-terminal globular region of the long isoform, and though its function has been only marginally explored compared to the long form, it appears to have strong tumor-suppressing effects, inhibiting the proliferation of HeLa, heavily impeding the formation of cell colonies and suppressing the expression of *MYBL2*, a gene known for being overexpressed in the development of multiple cancer types [39,40].

Short isoforms also commonly have either an antagonist effect to the long isoform, as seen in prolactin receptors (*PRLRs*), or a complementary effect as displayed by the short form of *OPA1* [41]. *PRLRs* have short and long form isoforms, which can act as dominant negatives towards the other. This prevents excessive signaling of one form, with the short form operating different signaling pathways than the long form [42,43]. In the case of *OPA1*, which regulates mitochondrial stability and energetics, the long and short forms work together to balance function. The long forms are fusion-competent but poor at energetics whereas the short forms are competent at energetics and poor at fusion. The ratio of isoforms allows the fine-tuning of mitochondrial performance [44].

This is reinforced by the experimentally verified group of alternatively spliced isoforms we collected from the available literature. Out of the genes for which we cataloged splicing isoforms for, nearly all had verified localization of those isoforms, and the majority of those had verified "short" forms with shorter lengths than the canonical isoform. Of the ~75% genes whose isoforms had both verified localization and short forms, almost half of those short forms localized to the cytoplasm (*NUMA1-s*, *IGF-1ea*, *c-FLIPs/r*, and *CD33-s*). These isoforms showed high degrees of tissue specificity, concentrated primarily in brain and muscle tissues [Table S1].

2.3. AES—Commonly Expressed in Tissue-Specific Manner

AES is commonly tissue- and development-stage-specific, allowing myriad cell types to efficiently use their resources by fine-tuning expression. Tissue-specific AES events can make up as many as 65% of total splicing events, with the major transcript expressed varying in up to 60% of coding genes [45,46]. The prevalence of these events also differs by tissue, with splicing in nervous, muscle (particularly cardiac), testis and blood tissues comprising the majority, and these events often extending to the protein level [45,47–49]. Recent studies have shown the presence of microexons, exons comprising 1–9 AA, produced by splicing events in neuronal tissues involved in cell differentiation, synaptic function, and axon guidance. Found on surface-accessible domains, especially in charged regions, these microexons are located in close proximity to or overlapping protein domains, providing an additional level of regulation [50].

Besides being tissue-specific, AES events are often developmentally regulated. This makes them a key factor in highly region-specific cell differentiation and morphogenesis in multiple tissues such as embryonic neurons, spermatozoa, skeletal muscle myoblasts and stem/progenitor cells, among others. The precisely timed swapping of splicing regulators, and the dominantly expressed isoform, is integral for the transition from fetal tissue to adult tissue [51]. An example of this is the transition of dominant *PTBP1* expression to *PTBP2* expression during differentiation of progenitor cells into postmitotic neurons [47,48,52]. These changes can also differ between regions of the same tissues, as occurs with the gene *LIMK2*, a member of the LIM kinase (*LIMK*) family that regulates actin dynamics through cofilin phosphoregulation. *LIMK2* encodes two isoforms: *LIMK2a*, the primary isoform, which is expressed evenly through the brain, and *LIMK2b*, which is highly expressed in the thalamus and cerebellum [45,53,54].

2.4. AES—Highly Dysregulated in Neurological Diseases

While these mechanisms allow for great diversity in the transcriptome and proteome, their dysregulation can have serious health consequences. Mutations in ~50% of the known RNA-modifying enzymes have been linked to human disease [55]. Splicing defects can arise from mutations to splicing elements, which are present throughout the genes in large numbers and can result in the deletion or creation of splicing elements, or to the splicing machinery itself. They are highly associated with nearly every aspect of cancer development, developmental syndromes like Prader Willi, and degenerative diseases such as retinitis pigmentosa [56,57]. Alternative exon splicing-related diseases fall into two broad categories: mutations within the transcript itself and mutations within the splicing machinery or regulatory elements. Mutations anywhere within the ORF may lead to frameshifts that result in transcripts often consigned to NMD, while changes in the CDS (SNPs/INDELS) can also lead to changes in amino acid identity. Mutations in introns, the UTR (particularly those regions closest to the coding sequence) and exon/intron borders can alter splicing elements, potentially leading to deleterious transcripts/proteins [31,52,56,58]. Splicing errors in the 3′ UTR can affect the stability and translation efficiency of transcripts, creating imbalances that lead to disease [46,49,52,59].

As only ~10% of a gene is comprised of exon coding sequences and changes to coding sequences most often have inconsequential effects, it should come as no surprise that ~85–90% of disease-causing splicing errors occur outside of exon regions [20,60]. Errors

in brain tissue-specific networks are responsible for several known neurological disorders such as autism spectrum disorder (ASD) [48,56]. Mutations in RBFOX proteins, which are local regulatory factors, subsequently cause the alternative splicing of *SHANK3*, *CACNA1C*, and *TSC2*, all of which are involved in ASD. In addition, the mis-splicing of microexons by Ser/Arg RBPs is known to be involved in ASDs. Splicing errors are also highly associated with various forms of cancer [57,60].

3. Alternative Transcription Start (ATS) Events: Features and Functions

3.1. ATS—Genomically Aligned by Sequence Structures, Clustering Patterns, and Promoter Motifs

As the name suggests, alternative transcription start events, first noted by Zitomer et al. in 1984, employ ATS sites and promoters in order to create transcripts with alternate first exons (AFE) or alter the length of the 5′ end [61,62]. As with AES, many factors contribute to transcription start site selection, including the presence or absence of motifs like TATA boxes, sequence structures, and internal ribosome entry sites (IRES), and a wide variety of transcription factors, including tissue and promoter exclusive factors [61–63]. The use of alternative transcription start sites can produce isoforms with different amino acid compositions, potentially altering function, or change the available regulatory regions of the 5′ UTR by adjusting its length. Genome-wide analyses have shown that 50% (though likely more, due to limitations of technology at the time of the studies) of human genes have at least two transcription start sites, with nearly five promoter peaks per gene on average [61,63].

Like AES, ATS events can be divided into several categories, based on the proximity of ATS sites, the levels of expression within a cluster of ATS sites, the location of the ATS site (3 or 5′) compared to the primary transcription start (PTS) site, and the proximity of the ATS sites to the PTS site. When compared by proximity, most ATS sites can be found in clusters within the core promoter region, with single or distal ATS sites the minority. These clusters are categorized by the level of transcription initiation relative to each ATS site. Clusters with more ubiquitous expression profiles across their start sites are broad, while clusters that show dominant expression from one start site are categorized as sharp or peaked. Peaked transcription start site (TSS) clusters show strong correlation with tissue-specific expression, while broad profiles show correlation with ubiquitous expression [63]. Initiations from ATS sites distant from the primary site tend to produce transcripts with AFE, while those in close proximity tend to produce transcripts with altered 5′ UTR lengths [62].

ATS sites and promoters can also affect the translation of these transcripts, through strong secondary structures, IRES or the inclusion of additional upstream open reading frames (uORFs) through the extension of 5′ UTRs [64]. Translation can be initiated upstream or downstream of the primary ORF, which affects their translation level, Kozak sequence, start codon motifs and more. The Kozak sequence is highly important in start site selection, particularly the nucleotides at the −3 and +4 positions. While a start site with a strong Kozak sequence will primarily produce the dominant isoform, one with a weaker Kozak sequence will progressively utilize alternative start sites (also known as leaky scanning) [65–68]. Start sites that are upstream of the primary ORF and distal (not overlapping the primary ORF) are strongly correlated with short ORFs that can potentially encode small peptide products like micropeptides. This setup can also lead to the reinitiation of transcription further downstream, with the small ribosome 40 s subunit remaining associated with the mRNA after termination at the stop codon [65,68]. Both distal uORFs as well as proximal uORFs (which overlap with the primary ORF) instead often act as regulators of the primary translation initiation site (TIS), suppressing its activation, either partially or completely (Figure 1) [65–67].
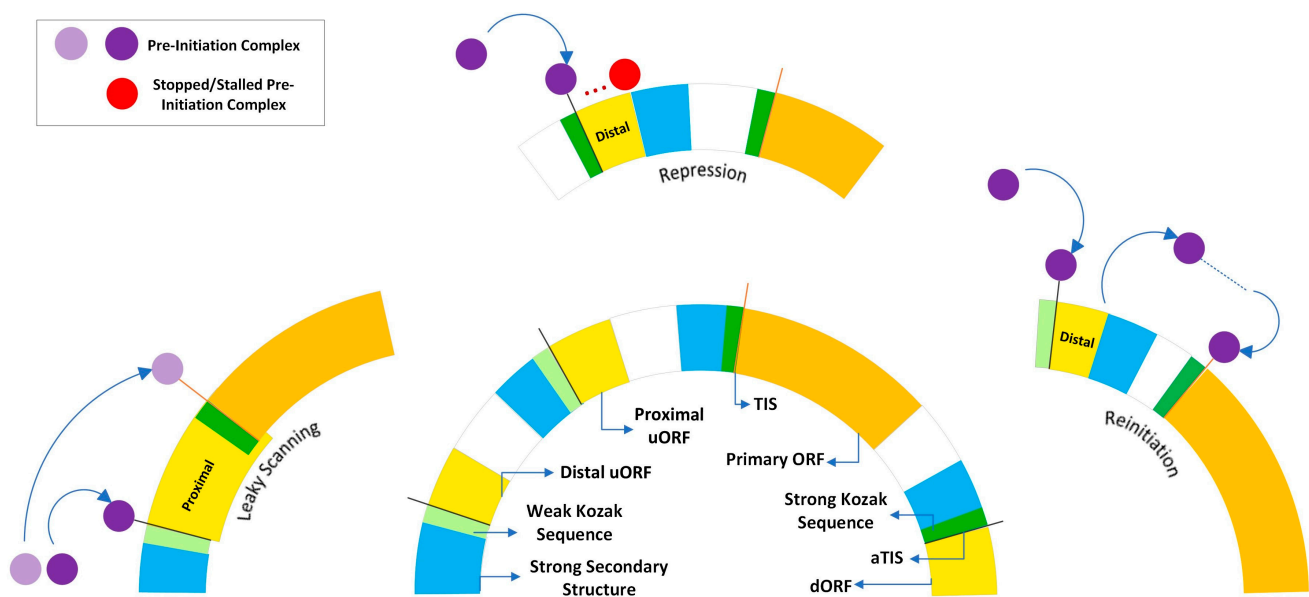
Figure 1. Three mechanisms impacted by ATS sites and ORFs. Leaky scanning can occur when translation begins at a start codon with a weak Kozak sequence, allowing a portion of the ribosomal subunits to skip over (leak) and initiate translation at an alternative start codon. Upstream ORFs and strong secondary structures from ATS sites/promoters can stop or stall the PIC complex, causing it to fall off and repressing gene expression downstream of it. Lastly, short uORFs separated from the main ORF can cause the ribosomal subunit to remain associated after termination and resume scanning, which is termed reinitiation. Abbreviations: ORF (open reading frame), dORF/uORF (downstream/upstream ORF), TIS (transcription initiation site), aTIS (alternative TIS).

Proximal ORFs often utilize non-canonical start codons, particularly CUG, with correspondingly weaker Kozak sequences, allowing for leaky scanning where some proportion of the ribosomal subunits fail to initialize at the start codon and continue scanning, allowing for multiple transcripts to be translated. Distal ORFs instead utilize more standard AUG codons, with strong Kozak sequences and secondary structures, in order to block or stall the transcription machinery before it can reach the primary TIS site [64–66]. Downstream ORFs also utilize stronger AUG codons and Kozak sequences in comparison to the primary TIS and are responsible for the N-terminal truncated proteins attributed to alternative start sites [66]. The most common ORFs are proximal to the primary TIS, acting in a repressive regulatory capacity [65].

3.2. ATS—Highly Involved in Altered N-Terminal Proteins, Localization, Stability, and Complementary Functions

ATS influences which uORFs are available in a transcript and as uORFs reside upstream of the canonical start site, most products will take the form of N-terminal extended transcripts or N-terminal truncated proteins, particularly for repressive uORF clusters. Extended N-terminal domains contribute to stability and translation efficiency without affecting the protein sequence [68]. This can be seen in the long isoform of *MAPKAPK2* (a gene whose primary isoform regulates the biosynthesis of pro-inflammatory cytokines), which uses a CUG start site in the 5′ UTR, displays markedly improved stability and is constitutively expressed [69].

ATS sites are also the primary means of changing isoform localization due to their manipulation of the N-terminal, the predominant location of localization signals. For example, in the gene *PTEN*, a well-known tumor suppressor with nuclear localization where *PTENα* and *PTENβ*, with extended N-terminal domains, localize in the mitochondria and the nucleolus, respectively. N-terminal truncated transcripts on the other hand can vary significantly, depending on the distance from the ATS sites to the canonical start

site [70]. Close downstream start sites may yield products identical in function to the primary protein, while distant start sites produce proteins with different functions.

An example of isoforms with divergent functions can be found in the isoforms of *ADK*, which acts as a sensor and regulator of the energy equilibrium in cell. The long form of *ADK* is prominent during early brain development, is nucleus localized, and is associated with the increased methylation of DNA and histones. The short form is prominent in the adult phase of brain development, particularly glial cells, and controls adenosine receptor activation [71].

Our collection of ATS site isoforms yielded a table of predominantly truncated isoforms. The localization of short and long isoforms differed in nearly every case (the few exceptions predominantly where the localization signal was not located on the N-terminal side, such as in *FRQ* in *Lachnellula* whose signal is in the C-terminal), though the location of the short isoforms varied. The majority of isoforms where specific tissues were noted were located predominantly in the brain, muscle, and heart tissues, with representation in liver, spleen, and pancreas.

While no previous trend has been noted, in our dataset these isoforms demonstrated complementary instead of antagonistic functions, as in the genes *UL138*, *FRQ,* and *NR3C1*. For instance, both isoforms of viral UL*138* suppress immediate early (IE) gene transcription and generation of infectious CMV virions during latency. The long form is more effective at suppressing virion production during early stages, while the short form is more effective in later stages [72] [Table S2].

3.3. ATS—Frequently Tissue-Specific, Heavy Use of Intronic Enhancers

Like AES, ATS sites show significant usage of tissue-dependent isoforms [20,63,64]. Transcription start sites show a degree of tissue preference in up to 80% of genes surveyed. Among protein-coding genes, 23% have two or more active promoters that contribute more than 10% of the gene's expression. While most alterative promoters produce limited transcripts compared to the constitutively expressed primary isoform, there are a small percentage (~15%) where switches result in the alternative start site producing the dominant transcript [73]. The presence of CpG islands and absence of TATA boxes near and in promoter sequences are most associated with ubiquitous tissue expression and a tendency towards nuclear or mitochondrial proteins, whereas the opposite is found in tissue specific genes, with a tendency towards extracellular proteins [74].

Though our catalog of alternative start site isoforms shows a preference for the same tissues as alternative splicing, most of the literature discussing tissue specificity in ATS sites does not discuss specific tissue preferences, with the few exceptions pointing to higher cerebellar, muscle, heart, liver and testicular tissue use [20,75]. Intronic enhancers are common features in tissue-specific genes, and ~70% of enhancers in cardiac/muscle tissues mapped to the first intron [76]. Tissue-specific ATS sites are highly enriched in regulatory pathways of transcription and development, particularly along cell lines rather than cell types [64,77]. Distal upstream uORFs, upstream uORFs with AUG codons and optimal sequences, and secondary structures upstream of the canonical TIS all typically act as translational repressors, reducing the level of protein production.

PTPRJ, which encodes a tumor-suppressing protein, utilizes alternate promoters with difficult-to-translate sequences, attenuating production [78]. This sort of regulation can change with external signaling, such as in ATF4, where the uORFs act in a repressive manner under normal conditions but become more permissive under stress conditions, allowing increased expression by the primary start site [35]. Several studies have noted increased activity of uORFs (particularly in response to eIF2α phosphorylation) during conditions of high stress that are highly conserved, indicating this is an evolutionary adaptation [68]. Alternative start sites also demonstrate temporal regulation, demonstrated by genes like *TEX101*, which produces a germ cell-specific protein involved in gonadal cells and is strongly involved in male fertility. While the first transcript is constitutively expressed in the gonads of both sexes, the second and third transcripts (which possess

distinct 5′ terminal sequences) are expressed in males only after spermatogenesis [79]. More so than any of the other forms of regulation, uORFs have the capacity for the generation of regulatory micropeptides. This is demonstrated in *scl* in *Drosophila*, which encodes two micropeptides, each less than 30 AA, that regulate calcium transport impacting heart contraction [36].

3.4. ATS—Commonly Linked with Tumor-Specific Oncogenesis, Invasion, and Metastasis

Errors in ATS sites and subsequent translation can lead to frameshifts or highly irregular amino acid conformations due to mutations. The silencing of ATS sites is also responsible for the expression of several diseases. While transcripts derived from ATS have had comparatively little research applied to their specific impact on the topic of disease association, studies show involvement in cancer phenotypes such as *CDC6* in breast cancer and *NRXN1* in neurological conditions [80,81]. Multiple types of associated diseases have been discovered in recent years that could be categorized as mutations within alternate ORFs, mutations that create new uORFs, aberrant promoter use, and change in imprinting status [67]. Oncogenesis and cancer progression are highly correlated with altered promoter use, shifting the transcript ratio to facilitate invasion, motility, metastasis, and more [82,83]. This deregulation of promoters is not only tissue-specific but tumor-specific, with different kidney tumors demonstrating different alternative promoter use [73]. The use of certain alternate promoters in *LEF1*, *TP73*, *NAT1* and other genes generate oncogenic transcripts.

High levels of β-catenin/TCF complexes in colon cancer cells are capable of activating the promoter for the full-length transcript, setting up a positive feedback loop for WNT signaling, which is a hallmark of many colon cancers [83]. Mutations in the 5′ UTR can also result in translational errors. An example is the creation of a uORF in the human clotting factor 12 gene FXII, a coagulation protein. A single C to T SNP in the 5′ leader sequence results in the creation of a two-codon uORF, which also alters the strength of the Kozak sequence. While this change does not result in a change in mRNA levels it results in a marked decrease in protein expression, predisposing that individual to thrombosis [68]. Another example is a point mutation of G to T just upstream of the canonical start site in the gene *CDKN2A* (a strong tumor suppressor) creating a new AUG codon with a similar Kozak sequence, resulting in the primarily truncated gene product by effectively blocking translation of the canon AUG. The loss of this transcript results in increased motility, invasion, and metastasis in melanoma cells [61,82].

Finally, certain cancer lines can alter the imprinting status of genes, in particular genes with tumor-suppressor properties like *IGF2* and *PEG3*. In the case of *IGF2*, loss of imprinting allows the transcription of the normally silenced maternal allele, leading to the overexpression of IGF2 in some cancers such as bladder cancer [84]. Alternately, *PEG3* undergoes epigenetic silencing via hypermethylation of its promoters in many cervical and ovarian cancer lines, preventing transcription [85].

4. Alternative Polyadenylation (APA) Events: Features and Functions

4.1. APA—High Contribution to Transcript Diversity

Alternative polyadenylation, discovered by multiple independent labs in 1980, has proven to be a major contributor to transcript diversity [86,87]. Polyadenylation site selection is a dynamic process, determined by predominantly cis elements, such as genetic motifs like the AAUAAA hexamer, the upstream UGUA, or downstream U/GU elements and their respective subunits APA site usage can be proximal or distal, with the former potentially changing the composition of the respective protein and distal usage conferring varying lengths of 3′ UTR [88,89]. APA usage is widespread, present in approximately 70% of 3′ UTRs in human genes, with ~50% of genes containing three or more polyadenylation sites [90,91]. This mechanism also appears to be highly conserved among eukaryotes, appearing in mammals, plants, and surprisingly even ~70% of yeast genes, which undergo nearly no alternative splicing, evidence of its evolutionary significance [92]. APA use is also

prevalent in ncRNAs, with one genome-wide mouse study finding at least one significant APA isoform in ~79% of mRNA genes and 66% of lncRNA genes [93].

The existing literature has discussed different types of polyadenylation sites (PAS) utilizing a variety of nomenclature and schema, but here we categorize them broadly as tandem APA and upstream region (UR) APA. Tandem 3′ UTR APA occurs when both the proximal and distal APA sites reside within the 3′ UTR, changing the length of the 3′ UTR but leaving the gene product identical. UR-APA truncates the protein product to varying degrees and can be further classified as alternative last exon (ALE), intronic, or internal exons. ALE is the result of upstream splicing, resulting in a new terminal exon and PAS selection. Intronic APAs are utilized by bypassing or blocking of the 5′ splice site, causing an internal exon to extend into its adjacent intron. Finally, internal exon APAs are rare PASs that occur inside of an internal exon, producing a transcript with no stop codon and no 3′ UTRs [88,92]. The majority of APA PAS sites in multi-exon genes are tandem 3′ UTR, comprising approximately 67% of all PAS sites. Multi-UTR genes have markedly longer 3′ UTRs (nearly 4 × longer) than genes with single UTRs, with ubiquitously expressed genes exhibiting longer 3′ UTRs than tissue-specific genes, even longer than those expressed by neural tissues [94].

A few recent papers have suggested that the impact of APA on translation and stability are not as significant as previously thought; for example, finding APA site choice only influences ~10% of miRNA targeting [91,95,96]. Though these findings contrast with the previous literature, it should be noted that much of the regulation that occurs within 3′ UTRs is defined by regions (most notably AU-rich regions) that can be concentrated into adjacent motifs or spread throughout the entirety of the UTR, making identification difficult. On top of this, RBP binding sites can typically bind multiple different RBPs, denying easy regulatory identity, and even single RBPs have been shown to recruit different *trans*-factors (with a wide range of effects) when exposed to different stimuli [93,97]. Given confounding factors, including primary cell vs. cell line experimental setups, context-dependent results should be expected.

Collectively, UR-APA is significantly rarer than tandem, at up to 33% of total PASs, with ALE as the most frequent event, a coupling of alternative splicing and APA to produce an isoform with a different last exon and the use of an internal PAS [93]. As a result, ALE isoforms have different terminal coding sequences and 3′ UTRs. Intronic and internal exon APAs are the least common types of UR-APA, and have the highest probability of producing truncated proteins, with internal exon isoforms typically rapidly degraded by either the no-stop decay or nonsense-mediated decay pathways due to missing stop codons and/or UTR regions [88,91]. Evolutionarily, distal PASs are the most highly conserved, with strong consensus sequences and features while proximal PASs are poorly conserved between species, with weaker features [89,92]. PolyA signals have proven more similar in the same tissue across different species than for different tissues within that species [90]. Generally, longer isoforms tend to localize more to the nuclear fraction than the cytoplasmic one, with ~10% of all detected isoforms showing significant differences in nuclear/cytosolic abundance according to a recent study [89,91]. Overall, short isoforms are weakly correlated with higher protein production (without impacting mRNA expression), potentially due to their ability to more efficiently form polysomes. As in many cases with APA, however, it is highly context-dependent, with a small subset of short isoforms displaying marked increases in protein production of 40–100× (Figure 2) [98–100].
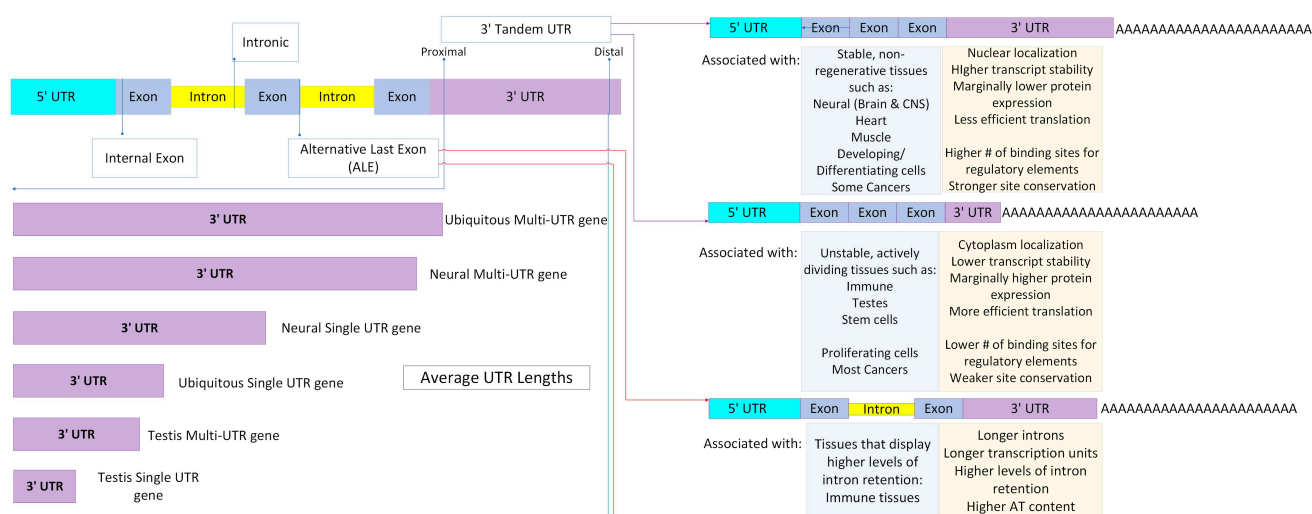
Figure 2. Different forms of polyadenylation, locations of typical polyadenylation signals (PASs) for each, and average lengths of untranslated regions (UTRs) based on tissue type and number of UTRs. The two most common forms, 3′ tandem UTR and alternative last exon (ALE), display common tissue associations and traits. The upper left shows the typical intragenic location of each form of alternative polyadenylation (APA) site. The lower left displays the average UTR length for single- and multi-UTR genes as well as ubiquitous versus tissue-specific genes. The right shows common associations for 3′ tandem UTR and ALE APA isoforms. Some ALE isoforms are the result of intronic APA site selection, as detailed on the right.

4.2. APA—Involved in Transcript Stability and Translation Efficiency

Tandem 3′ UTR modifications do not change the protein-coding sequence but still affect translation efficiency and localization through regulatory elements in their UTRs, including RBP binding sites, miRNA binding sites and scaffolding for RNA or protein transport [88,94,101]. As an example, *AAMDC-W*, an APA-derived isoform of AAMDC, is expressed at lower levels than both the L and S isoforms due to miRNA regulation [102]. Tandem 3′ UTR APA has also demonstrated the ability to regulate stability with 3′ UTR elements that contain destabilizing elements, such as miRNAs and RBPs that recruit decapping or deadenylating factors, and secondary structures that influence stability [88,94]. Isoforms of the gene *CALM1*, a calcium sensor and regulator, are an example of APA affecting stability. It expresses a short and long isoform, where the long form exhibits lower stability, with an expected half-life of ~50% of the short form [103].

Isoforms *RUNX1-A* and *CDC42(E6)* and *(E7)* are examples of ALE, having spliced alternate terminal exons and proximal PAS sites. *RUNX1A* is functionally antagonistic to its alternatively spliced isoforms, *RUNX1B/C*, balancing between differentiation and self-renewal in hemopoietic stem cells [104]. *CDC42*, a GTPase that regulates cell morphology and regulates multiple functions in the brain, has two isoforms called E6 and E7 [105]. The E6 isoform is both prenylated and palmitoylated—an indicator of strong membrane affinity. It is also brain tissue-specific and has mRNA localized to the soma, while the protein localizes to dendritic spines and plays a role in their formation. E7 is prenylated, giving it a hydrophobic c terminus, has mRNA localized to neurites and is expressed ubiquitously, while having a role in axonogenesis in neural tissues [76,77]. Several studies have suggested that non-coding transcripts generated by APA can act as scaffolding for the transport, production and regulation of other APA isoforms, particularly in the nuclear matrix [89,94]. The transmembrane *CD47*, which is associated with immune response, has a long and a short isoform. The long isoform contains a binding site for a complex of HUR-SET, which relocates it to the plasma membrane while the short isoform lacks these sequences and is localized to the ER [94].

Our collection of experimentally verified APA isoforms (where isoform tissue is noted) are all tissue-specific variants, with all but one expressed in brain tissues. The

sole exception is *RUNX1A*, which is expressed in immature hematopoietic and progenitor cells. We found that in all but one case, short and long isoforms demonstrated different localization patterns, with long isoforms preferentially localizing to nuclei and to distal sites in neural tissues, whereas short isoforms preferentially localized to cytoplasm/cytoplasmic organelles and proximal sites in neural tissues [Table S3]. As with ATS sites, isoform function was overwhelmingly either complementary (as in *MCL1*) or, largely unique to APA, nearly identical but localized to a different tissue (as in *IMPA1*). The two isoforms of *MCL1*, *MCL1pa1* and *pa2*, show similar localization in the mitochondria, nucleus and cytoplasm, where they regulate apoptosis (anti- and pro-, respectively), mitochondria morphology, and cell proliferation but different translation efficiencies between the two isoforms keep the basal level of *MCL1* stable while allowing for quick adjustments based on cell requirements [106]. *IMPA1* on the other hand produces three isoforms, L/S/C, which localize to axons, but the L isoform enriches in distal axons where the S form enriches in proximal axons. In distal axons, a portion of the L isoform undergoes cleavage by an AGO2 complex to form the C isoform. All participate in the regulation of NGF-dependent pathways and are involved in the survival of neuron axons [107].

4.3. APA—Tissue-Specific Processes in Response to Proliferation and Differentiation

Unsurprisingly, alternative polyadenylation has demonstrated significant tissue and temporal specificity in eukaryotes, with several clear motifs. There is clear evidence of tandem 3′ UTR and ALE regulation in neural, male and female reproductive, blood, muscle, stem cells and cancer tissues, despite the notable paucity of tissue-specific regulatory factors [91,92,108,109]. Tissue-specific regulation appears to rely on the prevalence/composition of core polyadenylation factors and competition with splicing factors (in the case of ALEs) in context-dependent fashion [91,93,109]. Globally, the enhanced use of proximal PAS is associated with proliferating cells where the use of distal PASs is linked to developing or differentiating cells [88]. This association extends to the respective tissues, with highly proliferative tissues, such as blood, showing overall preference for short isoforms, while more stable, non-regenerative tissues, such as heart tissue, show a preference for long isoforms [101].

The shorter, proximal PAS using form is also widely associated with cancers, though this has proven to be more a general association, with a selection of tissues and cancers (such as certain breast and thyroid) preferentially producing longer distal PAS-associated isoforms [88]. Thus, it may be more accurately stated that cancer cells display broad changes in PAS use, that can be proximal or distal dependent on tissue and cancer specificity. The best-known example of global APA regulation comes in neural tissues, where multiple studies have shown enriched translation of long 3′ isoforms, utilizing distal APA sites. In fact, neural tissue-specific genes demonstrate the longest 3′ UTRs out of all tissue-specific genes by a significant margin [94]. These isoforms show a preferred localization to dendrite and axon regions and, in several cases, demonstrate enzymatic cleavage of these long forms into shorter isoforms upon arrival. Shorter isoforms instead often localize to the soma but can be found elsewhere in the neuron, such as the axon, in response to stressors such as depolarization [107]. By contrast, hematopoietic cells (which have a high and constant rate of turnover) are known for their preference for shorter isoforms. Some of the oldest cells known to undergo APA are B cells, which produce an even ratio of the long and short heavy-chain isoforms in mature B cells, but shift progressively over to the short, secreted isoform in plasma cells as immunoglobulin secretion increases [88,100,110,111].

Initially, nearly all studies of polyadenylation focused on tandem 3′ UTR due to difficulties in identifying and isolating internal PAS usage, but in recent years several studies of ALE sites have emerged. Universally, these studies found ALE isoforms are regulated in a similar manner to tandem 3′ UTR isoforms in regard to PAS usage and the 3′ length in the tissues examined, with significant impacts on localization, and, in neural tissues, distal ALE isoforms preferentially residing in neurites [108,109,112]. They display longer introns, longer transcription units, and a higher AT content than other

forms of APA, and tissues with higher ALE usage also demonstrate higher levels of intron retention, which suggests these factors play a part in their regulation. ALE isoforms are especially prominent in immune cells, and one study suggests that the prevalence of intron retention in blood cells provides the transcription complex the time to select ALE PAS sites, as happens in tandem 3′ UTR APA, and affects the selection of transcription start sites. An unexpected finding was that immune cells were enriched for intronic PAS in the 5′ end, producing transcripts that were less than 100 AA, indicating either non-coding or micropeptide transcripts [112]. Though the function of these transcripts is currently unknown, previous research suggests they may act as regulators themselves or act as scaffolding for the further regulation and production of APA as discussed prior [91,92]. This suggests similar findings might be made in other highly proliferative tissue types such as the gonads, particularly male gametes.

4.4. APA—Significantly Associated with Diseases in High Differentiation or Proliferation Profiles

APA sites can affect amino acid composition by truncating mRNAs at various points, affect their localization, and even affect their stability and regulation. Mutations in APA sites are commonly linked to hematological diseases, certain cancers such as breast, brain, and colorectal cancers, immune syndromes like immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX), and neurological diseases [20,113,114]. Collectively, shifts in isoform production that result in the overabundance or minimization of an isoform can also result in disease states. We can categorize current disease causes as the loss or gain of individual APA sites and remodeling of APA site usage. Mutations as small as SNPs can introduce gain or loss of function in PAS, often the canonical PAS in the case of disease-causing mutations. Gain-of-function mutations, as seen in cases of systematic lupus erythematous (SLE), are the best known, with an A to G mutation in an alternate APA site leading to increased expression of a more stable form, resulting in increased levels in the cell. Individuals with this mutation appear to be susceptible to SLE [113,114].

Loss-of-function mutations are also prevalent, such as those that can occur in type 2 diabetes. A change in the C allele to a T allele in *TCF7L2* leads to the increased use of an intronic PAS that produces a truncated protein. This protein appears to repress TCF/LEF-dependent genes and is associated with increased risk for a particular form of diabetes [115]. The remodeling of APA site usage is commonly associated with cancer but is also the cause of several neurological and autoimmune disorders. Manipulation of APA regulators such as CSTF, CFI/CFII, and CPSF, pathways such as mTOR/Rho, and other factors such as PAP (poly(A) polymerase) are vital to the complex control of PAS site usage during normal development, and can contribute to the initiation or continuation of disease states when perturbed [89,91,92]. It has been proposed that short isoforms are favored in cancer due to the relative lack of RBP and miRNA binding sites, allowing them to escape regulation [91,92]. It is important to note, however, that different cancer lines can be highly variable, with different dependencies, sensitivities, and enriched pathways [88]. Thus, the effects of cancer on cell regulation have proven to be largely context-dependent, with several cancer lines preferentially expressing longer transcripts [91].

Studies of gastric cancer using the MKN28 cell line (regulated by the Rho GTPase pathway) have shown a switch in usage from the distal to the proximal PAS site, demonstrating strong transcriptional activity in the reporter gene, proving involvement in metastasis [88]. Conversely, studies in breast cancer using the MB231 cell line have shown many genes display preferential expression from distal APA sites, producing long isoforms [114]. Many of these are from genes associated with apoptosis and programmed cell death, indicating the use of the longer isoform may allow them to escape apoptosis of cancerous cells. Oculopharyngeal muscular dystrophy (OPMD) has been linked to a GCG expansion in the n-terminus of the *PABPN1* gene [88]. This mutated protein has been shown to sequester normal proteins in nuclear inclusions, instead of dispersing throughout the nucleoplasm like normal. This sequestering is common in neurological diseases, and this build-up of PABPN1 in nuclear inclusions has toxic effects [114].

5. AES, ATS and APA Events: Evidence of Cooperation and Antagonism

Though we have, to this point, discussed these three forms of modification as though they were largely independent, there is a significant amount of coordination, competition, and overlap between them. More than 80% of murine multi-transcript genes display interdependence between alternative splicing and the choice of transcription start site, with 37% of genes showing links between all types of features (TSS/Alternative exon, Alternative exon/APA, and TSS/APA) [116]. All three share regulatory factors, most significantly RNA polymerase II, particularly the C-terminal domain (CTD) of the largest subunit [32,117]. This CTD undergoes a number of modifications such as phosphorylation or methylation and can be further altered by cis/trans prolyl isomerases, which can alter its size and structure significantly and change which binding sites are presented and factor binding affinity [117]. As part of the complex, it recruits factors that help determine transcription start sites, splice sites, and polyadenylation sites used.

U1 snRNP, one of the factors recruited by the RNA pol II CTD, impacts both splicing and polyadenylation. This protein suppresses premature 3' end cleavage and polyadenylation, particularly in intronic cryptic PAS. Pushing the majority of PAS site usage to distal sites enriches primarily full-length transcripts in a process termed telescripting and lowers the proportion of transcripts being degraded. U1 snRNP also recognizes 5' splice sites and base-pairing with the pre-mRNA, and is crucial for intron removal during splicing and alternative splicing [113]. Elongation or stalling of the pol II machinery has been associated with the selection of rarer AES sites (especially intronic), and the selection of proximal APA sites over stronger distal ones, while ATS-influenced translation reinitiation and leaky scanning rely on manipulating the PIC [59,65,115]. This can be caused by factors such as the secondary structure, chromatin state, histone modifications and a high AT or AU content, all of which can themselves interact with ATS, AES, and APA regulators [59,117]. Studies have shown that pausing of the PIC is more likely to occur at the uORFs of genes whose primary isoform is being repressed and near the proximal APA sites of highly expressed genes [115].

The extension of isoforms towards the 5' or 3' ends by ATS or APA provides more extensive regions for AES, particularly in the case of ATS. The selection of start sites impacts 5' splicing patterns while promoter/enhancer activity can affect exon choice, including inducing exon skipping or the use of mutually exclusive exons. Coordination has been demonstrated between AES and APA in the cases of both tandem 3' UTR and especially ALE, with the 3' splice site and PAS communicating early in the transcription process, through PAS cleavage, the addition of the polyA tail, and concluding with splicing of the terminal exon [116,118].

An example of the coupled regulation of splicing and APA is represented in the gene *ACHE*. This gene, which terminates synaptic transmission, encodes multiple isoforms and is regulated by a combination of hnRNP H (another ribonucleoprotein) and CstF64. Here, hnRNP H causes distal PAS site use, inhibiting the selection of any proximal APA sites. Conversely, the inhibition of hnRNP H allows its antagonist CstF64 to bind intronic PAS sites, creating a truncated protein [110]. This short transcript can instead use the proximal 3' APA site or retain intron 4. More generally, CstF64 acts in the same manner across a host of genes by binding intronic APA sites to activate their expression. CstF64 is, broadly, suppressed by U1 snRNPs, which, as previously discussed, are integral parts of the splicing machinery, recognizing 5' splice sites [59].

The MBNL (Musclebind-like) family of proteins, regulators of RNA metabolism, is a known component of both the alternative splicing and alternative polyadenylation systems, particularly in muscle and neural tissues where it has been shown to bind splicing elements in nascent transcripts and 3' UTR binding sites. Disruption of MBNL is highly associated with the altered localization and stability of transcripts and is a primary cause of the neuromuscular disease myotonic dystrophy (DM) [48,51,107,118]. The presence of co-transcriptional splicing has been proven experimentally, with spliced mRNA, spliceosome components, and splicing factors in the chromatin fractions (fragments created during the

process of chromatin immunoprecipitation, more commonly known as ChIP) of actively transcribed genes before they are released into the nucleoplasm [119,120]. This was further confirmed in mammals, revealing that Ser/ARG-rich proteins (that bind the spliceosome), which would prevent hnRNPs (which transport pre-mRNA to degradation complexes) binding, were only effective when added before transcription. It was also shown that weakening of 3′ splice sites and inhibitory factors that bind intronic 3′ sites cause alternative splicing to occur post-transcriptionally instead of co-transcriptionally [119].

The composition of these elements within the genes is an integral component of the complex interactions that allow for such a diverse transcriptome and proteome, with these elements falling into the 5′ UTR, coding sequence (CDS), and 3′ UTR areas. Of note is the region of the 5′ UTR between the first and last start codon, and the area between the first and last stop codon in the 3′ UTR, which for ease of use we will refer to as the start rich regions (STRR) and stop rich regions (SPRR). The elements of AS are most common in the CDS, STRR and 5′ UTR of genes, from highest to lowest. Their presence in the 3′ UTR and SPRR is low as these regions have minimal intron content. ATS elements are most common in the 5′ UTR and STRRs but still prevalent in CDS, while APA elements favor the 3′ UTR and SPRRs but are also prominent in the CDS (Figure 3) [21,58].
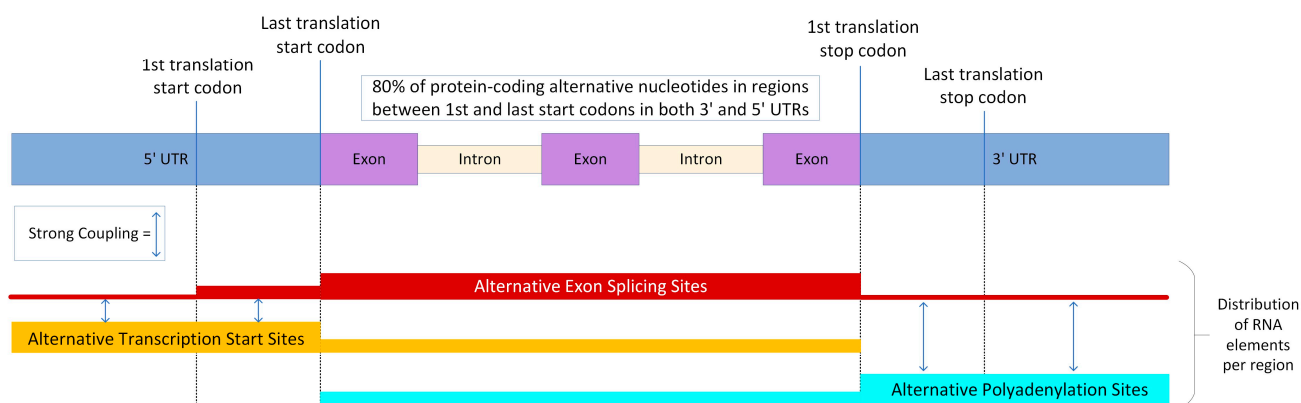

Figure 3. Distribution of post-transcriptional modification RNA elements. The layout of gene regions and distribution of the RNA elements within those regions for the three types of modification discussed here, alternative exon splicing, alternative transcription start sites, and alternative polyadenylation. It provides a visual representation of start rich and stop rich regions as the regions between the 1st and last start codon, and the 1st and last stop codon, respectively. It also denotes areas of coupling between these elements. Abbreviations: UTR (untranslated region).

The intron content has proven to be a good indicator of polymorphic genes, which are more enriched for them than monomorphic genes. While they are especially prevalent in CDS, the STRR and SPRR also contain multiple introns, with the average being ~4 introns each. Correlations have been made between the number of introns and alternative nucleotide content, with more than 80% of protein-coding alternative nucleotides located in the STRR and SPRR in humans, and a similar 76% in the murine genome [21]. This placement shows the importance of these terminal extensions for transcript variance. Interestingly, tissue-dependent splicing is enriched amongst non-coding transcripts, and in non-coding exons generally. Combined with the weak expression of exons that display tissue-dependent splicing, only ~15% of this form of splicing is expected to involve primary transcripts, though this 15% could have significant impacts as they may proportionally produce altered proteins [20]. These data outline why AES is considered to have the least impact on transcriptome and proteome diversity, compared with ATS and APA.

Broadly, highly expressed genes are intron-dense, and their distributions differ reliably, dependent on the functional area being examined (5′ UTR, CDS, etc.) [20,21,58]. N-terminal regions of proteins are enriched for intrinsically disordered regions (IDRs), which are segments with a higher proportion of charged amino acids and so lack a single unique

3D structure. This flexibility allows them to change conformation from an extended coil all the way to a collapsed globule (and any form in between) based on environmental contexts, allowing them to fulfill several purposes, like exposing and hiding motifs that mediate interactions with other proteins. Their adaptability makes them ideal regions for post-translational modification, facilitating substrate engagement and degradation, and for regulating binding partners by the adoption of different conformations. These IDRs, which are heavily involved in protein–protein interactions, are also enriched in alternative protein ends, suggesting more regulatory roles for proteins derived from alternative transcripts [121]. ATS and AES have shown strong coupling in 5′ UTR and STRRs, where AES occurs at a relatively high frequency but rarely without the co-occurrence of ATI. AES and APA also demonstrate strong coupling in the 3′ UTR but here they are not strongly correlated in SPRRs. Interestingly, the occurrence of AS in the coding sequence has shown a strong inverse relationship to the occurrence of AS or ATS in the 5′ UTR region [20]. On the other hand, AS in the coding sequence shows a positive relationship with APA events in the 3′ UTR and SPRRs. In fact, APA events occur in the 3′ UTR almost exclusively in the presence of AS in the CDS, further implicating the strong relationship between AS and APA [24,57,58,80].

6. Genome-Wide Profiling of RNA Variants: Challenges and Solutions

One challenge that is sometimes overlooked when considering the study of these mechanisms is the choice of appropriate tools. The massive influx of available RNA-seq data that accompanied next-generation sequencing methods and the increasing availability of long-read data from platforms like ONT and Pac-Bio have been accompanied by a proportionate increase in the release of detection and analysis tools designed to handle this data, many of which are tailored towards specific tasks. Hundreds of algorithms/potential pipelines exist for RNA-seq analysis, where the variability and quality of the data can range significantly. Each step of the process (trimming, alignment, counting, normalization, pseudoalignment (an alternative that combines alignment, counting and normalization into one step), and differential expression (DE)) has multiple tools available. Several benchmark studies have been conducted over the years, attempting to narrow down which tools produce the best combination of accuracy and precision. Though these studies could not find consensus, the general trend suggests that the selection of trimming and alignment algorithms were the least impactful, with counting and normalization being critical steps [122–124]. Despite this, the effects of normalization on the effectiveness of DE tools have been contested [125,126]. DE tools have shown high similarity in performance, with almost universally improved performance in precision, recall, and FDR as sample numbers increase [123,125,127]. One benchmark study found that exon-based methods (DEXSeq, edgeR, limma etc.) demonstrated higher precision and that exon- and event-based methods were generally low-FDR, high-precision and had moderate recall [128–130]. It was also found that the highest overlap of detected DE genes was among exon-based methods [131]. Several studies have suggested that DESeq2, limma and edgeR are overall reliable, non-biased, computationally light tools for differential expression [123,125,127,132].

As alternative splicing has long been considered integral to gene variation, it also possesses the most tools specific to the three mechanisms we have talked about in this paper. Unfortunately, benchmarking studies performed on these tools have suggested that the performance of individual tools is relatively unreliable, with high rates of false positives and a very low overlap of detected AES events between algorithms. Suggested means to combat this have included the use of more than one tool to increase the validity of the analysis or to utilize specific tools for the detection of particular types of events [133–135]. One study advised a combination of rMATS and Whippet, and provided a pipeline for users [135–137]. Recommendations for event-based tool selection varied from known annotated events (ASpli, Whippet, SGSeq) to intron retention (IRFinder) to de novo events (combination of tools like splAdder and Whippet/MAJIQ) [133,134,138–142]. Overall Whippet and rMATs are the tools most recommended for use in AS event detection, in combination with more

specialized tools [133–135]. Tools specifically meant for AS detection/analysis utilizing scRNA-seq, like BRIE2, have not yet been benchmarked [143].

Tools for the detection and analysis of alternative transcription start sites and alternative polyadenylation sites are sparser than AES tools, though APA tools have been released with increasing frequency in recent years. For transcription start site detection and analysis CAGEfightR and TSRexploreR are known quantities, with SEASTAR, mountainClimber (RNA-seq data) and CamoTSS (scRNA-seq data) tailored towards alternative start site detection and usage rates [62,144–147]. We were unable to find any benchmarking studies for tools geared towards alternative transcription start sites. The development of APA detection and analysis tools has been prolific in the last decade, with algorithms covering the gamut from bulk RNA-seq to scRNA-seq, including machine learning and deep learning models. Many of these tools were developed for bulk RNA-seq, with the intention of leveraging the enormous stores of data in this format, but these tools consistently perform inferior to tools designed for 3′ seq or long-read (Pac Bio or ONT) data [148–151]. This is consistent with the difficulty and computational complexity of deriving APA sites and usage from data biased against 3′ ends. Benchmarking studies have found, like AES tools, that each algorithm returns highly individual APA results. These results have minimal overlap with each other, even amongst tools that utilize the same general method of detection (such as changes in read density as in TAPAS or APAtrap) and false positives are a common issue [148,149,151–153]. One study found that all RNA-seq input tools produced comparable numbers of APA sites, where sites found demonstrated the characteristics expected of polyA sites, while acknowledging the potentially high number of false positives [148]. A separate study promoted the use of ML and DL models for prediction from DNA, such as DeepPASTA, PASNet, or PASS [149,154–156]. The availability of annotations can play a significant role in the accuracy and precision of many of these tools, which often rely on databases for comparisons. All studies agreed that running any single tool was unlikely to perform for every task, with some suggesting a combination of tools, such as QAPA with a small number of Iso-Seq or 3′ Seq annotations to bridge the gap [148,149,151,157]. Other suggestions were picking the right tool for the job, such as TAPAS or DaPars2 for the de novo detection from RNA-seq data or APAtrap for plant data [149,151,158]. TAPAS, DaPars2, APAtrap, QAPA for bulk data and scAPAtrap for single-cell data were the most suggested tools in these studies [159]. An overall point of emphasis is that the use of any of these tools (AS or APA) requires independent validations of a random sampling of your findings [148–151]. Recommended tools can be found in Table 1.

Table 1. Recommended method by mechanism. A list of tools recommended by benchmarked studies by purpose, name, programming language and article source [129,130,132,136,137,152,153,157–164].

Method Type	Method Name	Program Environment	Introductory Article
Recommended DE Tools	limma	Bioconductor R package	https://doi.org/10.1093/nar/gkv007, accessed on 1 November 2023
	edgeR	Bioconductor R package	https://doi.org/10.1093/bioinformatics/btp616, accessed on 1 November 2023
	DESeq2	Bioconductor R package	https://doi.org/10.1186/s13059-014-0550-8, accessed on 1 November 2023
Recommended AES Tools	rMATs	R package	https://rnaseq-mats.sourceforge.io/, accessed on 1 November 2023
	Whippet	Julia	https://github.com/timbitz/Whippet.jl, accessed on 1 November 2023
Recommended APA Tools RNA-seq	TAPAS	R package	https://doi.org/10.1093/bioinformatics/bty110, accessed on 1 November 2023
	DaPars2	Python	https://doi.org/10.1038/ncomms6274, accessed on 1 November 2023

Table 1. *Cont.*

Method Type	Method Name	Program Environment	Introductory Article
	APAtrap	R package/PERL	https://doi.org/10.1093/bioinformatics/bty029, accessed on 1 November 2023
	QAPA	R package/Python	https://doi.org/10.1186/s13059-018-1414-4, accessed on 1 November 2023
Recommended APA Tools scRNA-seq	scAPA	R package	https://doi.org/10.1093/nar/gkz781, accessed on 1 November 2023
	scAPAtrap	R package	https://doi.org/10.1093/bib/bbaa273, accessed on 1 November 2023
Web Tools	eVITTA		https://doi.org/10.1093/nar/gkab366, accessed on 1 November 2023
	SpliceTools	PERL for download	https://doi.org/10.1093/nar/gkad111, accessed on 1 November 2023
	APAview	Jinja/Python	https://doi.org/10.3389/fgene.2022.928862, accessed on 1 November 2023
	Cas-Viewer		https://doi.org/10.1186/s12920-018-0348-8, accessed on 1 November 2023

Many resources are integral to effective research, but few are as important as databases, particularly in motif or characteristic driven detection. As with tools, there are hundreds of databases, though many of them are no longer maintained or have been folded into more established databases. Several major databases contain variant transcript annotations, including Ensembl (ensembl.org, accessed on 1 November 2023), Gencode (gencodegenes. org, accessed on 1 November 2023), Genotype Tissue Expression portal (gtexportal.org, accessed on 1 November 2023) and NCBIs (https://www.ncbi.nlm.nih.gov/, accessed on 1 November 2023) Consensus Coding Sequence and RefSeq. Analysis of these database sequences forms the basis for many specialized databases, such as APAatlas, which was developed for the study of APA events in human tissues [165]. AES databases can be very specialized, from cataloging the effects of mutation on splice sites to clinical phenotypes to exon skipping [166–168]. ATS and APA databases are more generalized, documenting TSS and polyA sites, respectively, with some APA databases adding events and conservation, but the biggest variance is the species covered. A collection of some of these specialized databases can be found in Table 2. Many of the tools described above reference database annotations to locate events and sites, and accurate sequencing underpins all the discoveries in this paper, highlighting the importance of both improving current data and adding new sequenced genome data, samples, and tissues to expand our understanding and efficiency.

A relatively recent category of tool–webtools sees sporadic entries. These are often aimed at enabling researchers with less bioinformatics or programming experience to take advantage of these advances. Though they lack the full flexibility of the modules they utilize, they offer ease of use with some flexible parameters, normally in a pipeline, leveraging multiple tools while mitigating the memory/cpu drain of running these same tools manually. These tools may also provide examples or previews of pipeline structures for a multitude of purposes. While webtools that are solely web-based only have limited functions, they provide a valuable resource suited for exploratory purposes and those with less programming experience.

Table 2. Databases by mechanism. A list of databases by primary mechanism, with database name, brief description, and url [166–175].

Database Type	Database Name	Description	Website
AES	MutSpliceDB	Effects of mutation on splicing	https://brb.nci.nih.gov/splicing/, accessed on 1 November 2023
	VastDB	Splicing in multiple species	https://vastdb.crg.eu, accessed on 1 November 2023
	HEXEvent	Human exon splicing	https://hexevent.mmg.uci.edu, accessed on 1 November 2023
	ExonSkipDB	Exon-skipping events	https://ccsm.uth.edu/ExonSkipDB/, accessed on 1 November 2023
	ClinVar	Variants with clinical phenotypes	https://www.ncbi.nlm.nih.gov/clinvar/, accessed on 1 November 2023
ATS	DBTSS	Human adult and embryonic tissues	https://dbtss.hgc.jp/, accessed on 1 November 2023
	refTSS	Human and mouse	https://reftss.riken.jp/, accessed on 1 November 2023
APA	PolyASite 2.0	Sites and usage in human, mouse and worm	https://www.polyasite.unibas.ch/, accessed on 1 November 2023
	PolyA DB3	Sites, cleavage, and conservation	https://exon.apps.wistar.org/PolyA_DB/, accessed on 1 November 2023
	scAPAdb	Sites and usage in multiple species, single-cell data	http://www.bmibig.cn/scAPAdb/, accessed on 1 November 2023

Finally, many methods have been developed to directly profile genome-wide expressed RNA variants, such as whole-transcriptome start and termini site-sequencing (WTSS-seq and WTTS-seq) methods [176,177]. These assays have advantages over the conventional RNA-seq. For example, WTSS-seq and WTTS-seq methods do not synthesize full-length cDNA, so there is no bias against long transcripts. The "5′ adapter—transcript target—3′ adapter" constructs are synthesized, avoiding the low-efficiency issues associated with ligation. There are no primer/template switches and/or changes in the protocol, which significantly minimizes amplification biases or detours. These assays involve just one run of PCR, thus minimizing the over-amplification of abundantly expressed transcripts and maximizing the transcriptome coverage. However, both WTSS-seq and WTTS-seq methods are tag-based approaches, so they cannot produce full-length transcripts. This should be easily overcome by using a long-read sequencing approach by linking the short tags to the full-length transcripts for functional characterization.

7. Conclusions

Pre- and post- transcriptional modifications are extremely common and critical for transcriptome and proteome diversity. APA and ATS are responsible for most of this diversity, with AES's contribution being widespread but minor. These phenomena utilize complex mechanisms and a variety of factors to alter the stability, location, and efficiency in a tissue- and stage-dependent manner. Some of these mechanisms and factors are unique to individual modifications but there is also overlap between them, notably the RNA polymerase II complex, where outcomes like proximal APA site selection benefit from pausing of the complex. These features can be antagonistic or cooperative, particularly between ATS and AES or APA and AES, influencing transcription start site and PAS selection. Collectively, dysregulation of these phenomena contributes to a large proportion of human disease, especially in neurological, blood, immune and muscle diseases, as well as various cancers. Given the integral role of these modifications in all levels of cell function, and their role in disease when dysregulated, it is important that we continue to study these phenomena. The study of these mechanisms requires careful consideration, the strengths of the tools selected need to match the aims of the research and often the use of

more than one algorithm for detection and analysis is recommended. We have still only identified very limited numbers of their factors (ex: RBPs), their environmental context, or their networks (protein or otherwise). With improvements in sequencing techniques, identification methods, tissue profiling, and advancements in machine learning, we will be able to link these profiles to modifications, phenotypes, and clinical applications.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/genes14112051/s1. Table S1: Alternate Exons Splicing Isoforms, Table S2: Alternative Start Site Isoforms, Table S3: Alternative Polyadenylation Isoforms. Refs [178–238] are cited in the Supplementary File.

Author Contributions: S.A.C. and Z.J. contributed to the study conception and design. Data collection and analysis were performed by S.A.C. The first draft of the manuscript was written by S.A.C. and all authors commented on previous versions of the manuscript. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: All data generated or analyzed during this study are included in this published article and its Supplementary Materials (Tables S1–S3).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Spuhler, J.N. On the Number of Genes in Man. *Science* **1948**, *108*, 279–280. [CrossRef] [PubMed]
2. Vogel, F. A Preliminary Estimate of the Number of Human Genes. *Nature* **1964**, *201*, 847. [CrossRef] [PubMed]
3. Clancy, S. RNA Splicing: Introns, Exons and Spliceosome. Available online: http://www.nature.com/scitable/topicpage/rna-splicing-introns-exons-and-spliceosome-12375 (accessed on 2 January 2023).
4. The Human Genome Project. Available online: https://www.genome.gov/human-genome-project (accessed on 2 January 2023).
5. *Understanding Our Genetic Inheritance: The US Human Genome Project, The First Five Years FY 1991–1995*; National Center for Human Genome Research: Bethesda, MD, USA; USDOE Office of Energy Research: Washington, DC, USA; Office of Health and Environmental Research: Washington, DC, USA, 1990.
6. Dunham, I.; Hunt, A.R.; Collins, J.E.; Bruskiewich, R.; Beare, D.M.; Clamp, M.; Smink, L.J.; Ainscough, R.; Almeida, J.P.; Babbage, A.; et al. The DNA Sequence of Human Chromosome 22. *Nature* **1999**, *402*, 489–495. [CrossRef] [PubMed]
7. Fields, C.; Adams, M.D.; White, O.; Venter, J.C. How Many Genes in the Human Genome? *Nat. Genet.* **1994**, *7*, 345–346. [CrossRef]
8. Adams, M.D.; Kerlavage, A.R.; Fleischmann, R.D.; Fuldner, R.A.; Bult, C.J.; Lee, N.H.; Kirkness, E.F.; Weinstock, K.G.; Gocayne, J.D.; White, O. Initial Assessment of Human Gene Diversity and Expression Patterns Based upon 83 Million Nucleotides of cDNA Sequence. *Nature* **1995**, *377*, 3–174.
9. Antequera, F.; Bird, A. Number of CpG Islands and Genes in Human and Mouse. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11995–11999. [CrossRef]
10. Liang, F.; Holt, I.; Pertea, G.; Karamycheva, S.; Salzberg, S.L.; Quackenbush, J. Gene Index Analysis of the Human Genome Estimates Approximately 120,000 Genes. *Nat. Genet.* **2000**, *25*, 239–240. [CrossRef]
11. Wright, F.A.; Lemon, W.J.; Zhao, W.D.; Sears, R.; Zhuo, D.; Wang, J.-P.; Yang, H.-Y.; Baer, T.; Stredney, D.; Spitzner, J.; et al. A Draft Annotation and Overview of the Human Genome. *Genome Biol.* **2001**, *2*, 0025.1–0025.18. [CrossRef]
12. Daly, M.J. Estimating the Human Gene Count. *Cell* **2002**, *109*, 283–284. [CrossRef]
13. International Human Genome Sequencing Consortium. Finishing the Euchromatic Sequence of the Human Genome. *Nature* **2004**, *431*, 931–945. [CrossRef]
14. Venter, J.C.; Adams, M.D.; Myers, E.W.; Li, P.W.; Mural, R.J.; Sutton, G.G.; Smith, H.O.; Yandell, M.; Evans, C.A.; Holt, R.A.; et al. The Sequence of the Human Genome. *Science* **2001**, *291*, 1304–1351. [CrossRef] [PubMed]
15. Abascal, F.; Juan, D.; Jungreis, I.; Martinez, L.; Rigau, M.; Rodriguez, J.M.; Vazquez, J.; Tress, M.L. Loose Ends: Almost One in Five Human Genes Still Have Unresolved Coding Status. *Nucleic Acids Res.* **2018**, *46*, 7070–7084. [CrossRef] [PubMed]
16. Nurk, S.; Koren, S.; Rhie, A.; Rautiainen, M.; Bzikadze, A.V.; Mikheenko, A.; Vollger, M.R.; Altemose, N.; Uralsky, L.; Gershman, A.; et al. The Complete Sequence of a Human Genome. *Science* **2022**, *376*, 44–53. [CrossRef] [PubMed]
17. Zahler, A.M. *Pre-mRNA Splicing and Its Regulation in Caenorhabditis Elegans*; WormBook: Pasadena, CA, USA, 2018.

18. Sender, R.; Fuchs, S.; Milo, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol.* **2016**, *14*, e1002533. [CrossRef] [PubMed]

19. Rodriguez, J.M.; Pozo, F.; Cerdán-Vélez, D.; Di Domenico, T.; Vázquez, J.; Tress, M.L. APPRIS: Selecting Functionally Important Isoforms. *Nucleic Acids Res.* **2022**, *50*, D54–D59. [CrossRef] [PubMed]

20. Reyes, A.; Huber, W. Alternative Start and Termination Sites of Transcription Drive Most Transcript Isoform Differences across Human Tissues. *Nucleic Acids Res.* **2018**, *46*, 582–592. [CrossRef]

21. Shabalina, S.A.; Ogurtsov, A.Y.; Spiridonov, N.A.; Koonin, E.V. Evolution at Protein Ends: Major Contribution of Alternative Transcription Initiation and Termination to the Transcriptome and Proteome Diversity in Mammals. *Nucleic Acids Res.* **2014**, *42*, 7132–7144. [CrossRef]

22. Cunningham, F.; Allen, J.E.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.R.; Armean, I.M.; Austine-Orimoloye, O.; Azov, A.G.; Barnes, I.; Bennett, R.; et al. Ensembl 2022. *Nucleic Acids Res.* **2022**, *50*, D988–D995. [CrossRef]

23. Frankish, A.; Diekhans, M.; Jungreis, I.; Lagarde, J.; Loveland, J.E.; Mudge, J.M.; Sisu, C.; Wright, J.C.; Armstrong, J.; Barnes, I.; et al. GENCODE 2021. *Nucleic Acids Res.* **2021**, *49*, D916–D923. [CrossRef]

24. Pertea, M.; Shumate, A.; Pertea, G.; Varabyou, A.; Breitwieser, F.P.; Chang, Y.-C.; Madugundu, A.K.; Pandey, A.; Salzberg, S.L. CHESS: A New Human Gene Catalog Curated from Thousands of Large-Scale RNA Sequencing Experiments Reveals Extensive Transcriptional Noise. *Genome Biol.* **2018**, *19*, 208. [CrossRef]

25. Piovesan, A.; Antonaros, F.; Vitale, L.; Strippoli, P.; Pelleri, M.C.; Caracausi, M. Human Protein-Coding Genes and Gene Feature Statistics in 2019. *BMC Res. Notes* **2019**, *12*, 315. [CrossRef] [PubMed]

26. Tress, M.L.; Abascal, F.; Valencia, A. Alternative Splicing May Not Be the Key to Proteome Complexity. *Trends Biochem. Sci.* **2017**, *42*, 98–110. [CrossRef] [PubMed]

27. Berget, S.M.; Moore, C.; Sharp, P.A. Spliced Segments at the 5′ Terminus of Adenovirus 2 Late mRNA. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 3171–3175. [CrossRef] [PubMed]

28. Gilbert, W. Why Genes in Pieces? *Nature* **1978**, *271*, 501. [CrossRef] [PubMed]

29. Chao, Y.; Jiang, Y.; Zhong, M.; Wei, K.; Hu, C.; Qin, Y.; Zuo, Y.; Yang, L.; Shen, Z.; Zou, C. Regulatory Roles and Mechanisms of Alternative RNA Splicing in Adipogenesis and Human Metabolic Health. *Cell Biosci.* **2021**, *11*, 66. [CrossRef]

30. Lee, Y.; Rio, D.C. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu. Rev. Biochem.* **2015**, *84*, 291–323. [CrossRef] [PubMed]

31. Blencowe, B.J. Alternative Splicing: New Insights from Global Analyses. *Cell* **2006**, *126*, 37–47. [CrossRef]

32. Wang, Y.; Liu, J.; Huang, B.; Xu, Y.-M.; Li, J.; Huang, L.-F.; Lin, J.; Zhang, J.; Min, Q.-H.; Yang, W.-M.; et al. Mechanism of Alternative Splicing and Its Regulation (Review). *Biomed. Rep.* **2015**, *3*, 152–158. [CrossRef]

33. Koren, E.; Lev-Maor, G.; Ast, G. The Emergence of Alternative 3′ and 5′ Splice Site Exons from Constitutive Exons. *PLoS Comput. Biol.* **2007**, *3*, e95. [CrossRef]

34. Pozo, F.; Martinez-Gomez, L.; Walsh, T.A.; Rodriguez, J.M.; Di Domenico, T.; Abascal, F.; Vazquez, J.; Tress, M.L. Assessing the Functional Relevance of Splice Isoforms. *NAR Genom. Bioinform.* **2021**, *3*, lqab044. [CrossRef]

35. Orr, M.W.; Mao, Y.; Storz, G.; Qian, S.-B. Alternative ORFs and Small ORFs: Shedding Light on the Dark Proteome. *Nucleic Acids Res.* **2020**, *48*, 1029–1042. [CrossRef] [PubMed]

36. Magny, E.G.; Pueyo, J.I.; Pearl, F.M.G.; Cespedes, M.A.; Niven, J.E.; Bishop, S.A.; Couso, J.P. Conserved Regulation of Cardiac Calcium Uptake by Peptides Encoded in Small Open Reading Frames. *Science* **2013**, *341*, 1116–1120. [CrossRef] [PubMed]

37. Zhang, J.; Chen, Y.; Huang, Q.; Cheng, W.; Kang, Y.; Shu, L.; Yin, W.; Hua, Z.-C. Nuclear Localization of C-FLIP-L and Its Regulation of AP-1 Activity. *Int. J. Biochem. Cell Biol.* **2009**, *41*, 1678–1684. [CrossRef] [PubMed]

38. Hillert, L.K.; Ivanisenko, N.V.; Espe, J.; König, C.; Ivanisenko, V.A.; Kähne, T.; Lavrik, I.N. Long and Short Isoforms of C-FLIP Act as Control Checkpoints of DED Filament Assembly. *Oncogene* **2020**, *39*, 1756–1772. [CrossRef]

39. Qin, W.-S.; Wu, J.; Chen, Y.; Cui, F.-C.; Zhang, F.-M.; Lyu, G.-T.; Zhang, H.-M. The Short Isoform of Nuclear Mitotic Apparatus Protein 1 Functions as a Putative Tumor Suppressor. *Chin. Med. J. (Engl.)* **2017**, *130*, 1824–1830. [CrossRef]

40. Identification and Characterization of Novel NuMA Isoforms. *Biochem. Biophys. Res. Commun.* **2014**, *454*, 387–392. [CrossRef]

41. Bhuiyan, S.A.; Ly, S.; Phan, M.; Huntington, B.; Hogan, E.; Liu, C.C.; Liu, J.; Pavlidis, P. Systematic Evaluation of Isoform Function in Literature Reports of Alternative Splicing. *BMC Genom.* **2018**, *19*, 637. [CrossRef]

42. Sangeeta Devi, Y.; Halperin, J. Reproductive Actions of Prolactin Mediated through Short and Long Receptor Isoforms. *Mol. Cell. Endocrinol.* **2014**, *382*, 400–410. [CrossRef]

43. Bouilly, J.; Sonigo, C.; Auffret, J.; Gibori, G.; Binart, N. Prolactin Signaling Mechanisms in Ovary. *Mol. Cell. Endocrinol.* **2012**, *356*, 80–87. [CrossRef]

44. Del Dotto, V.; Fogazza, M.; Carelli, V.; Rugolo, M.; Zanna, C. Eight Human OPA1 Isoforms, Long and Short: What Are They For? *Biochim. Biophys. Acta BBA Bioenerg.* **2018**, *1859*, 263–269. [CrossRef]

45. Rodriguez, J.M.; Pozo, F.; di Domenico, T.; Vazquez, J.; Tress, M.L. An Analysis of Tissue-Specific Alternative Splicing at the Protein Level. *PLoS Comput. Biol.* **2020**, *16*, e1008287. [CrossRef] [PubMed]

46. Baralle, F.E.; Giudice, J. Alternative Splicing as a Regulator of Development and Tissue Identity. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 437–451. [CrossRef] [PubMed]

47. Su, C.-H.; D, D.; Tarn, W.-Y. Alternative Splicing in Neurogenesis and Brain Development. *Front. Mol. Biosci.* **2018**, *5*, 12. [CrossRef] [PubMed]

48. Vuong, C.K.; Black, D.L.; Zheng, S. The Neurogenetics of Alternative Splicing. *Nat. Rev. Neurosci.* **2016**, *17*, 265–281. [CrossRef] [PubMed]

49. Nikonova, E.; Kao, S.-Y.; Spletter, M.L. Contributions of Alternative Splicing to Muscle Type Development and Function. *Semin. Cell Dev. Biol.* **2020**, *104*, 65–80. [CrossRef] [PubMed]

50. Gehring, N.H.; Roignant, J.-Y. Anything but Ordinary—Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. *Trends Genet.* **2021**, *37*, 355–372. [CrossRef] [PubMed]

51. Kalsotra, A.; Xiao, X.; Ward, A.J.; Castle, J.C.; Johnson, J.M.; Burge, C.B.; Cooper, T.A. A Postnatal Switch of CELF and MBNL Proteins Reprograms Alternative Splicing in the Developing Heart. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 20333–20338. [CrossRef] [PubMed]

52. Naro, C.; Cesari, E.; Sette, C. Splicing Regulation in Brain and Testis: Common Themes for Highly Specialized Organs. *Cell Cycle* **2021**, *20*, 480–489. [CrossRef]

53. Johnson, M.B.; Kawasawa, Y.I.; Mason, C.E.; Krsnik, Ž.; Coppola, G.; Bogdanović, D.; Geschwind, D.H.; Mane, S.M.; State, M.W.; Šestan, N. Functional and Evolutionary Insights into Human Brain Development through Global Transcriptome Analysis. *Neuron* **2009**, *62*, 494–509. [CrossRef]

54. Ribba, A.-S.; Fraboulet, S.; Sadoul, K.; Lafanechère, L. The Role of LIM Kinases during Development: A Lens to Get a Glimpse of Their Implication in Pathologies. *Cells* **2022**, *11*, 403. [CrossRef]

55. Jonkhout, N.; Tran, J.; Smith, M.A.; Schonrock, N.; Mattick, J.S.; Novoa, E.M. The RNA Modification Landscape in Human Disease. *RNA* **2017**, *23*, 1754–1769. [CrossRef] [PubMed]

56. Cooper, T.A.; Wan, L.; Dreyfuss, G. RNA and Disease. *Cell* **2009**, *136*, 777–793. [CrossRef] [PubMed]

57. Tazi, J.; Bakkour, N.; Stamm, S. Alternative Splicing and Disease. *Biochim. Biophys. Acta BBA Mol. Basis Dis.* **2009**, *1792*, 14–26. [CrossRef]

58. Shabalina, S.A.; Spiridonov, A.N.; Spiridonov, N.A.; Koonin, E.V. Connections between Alternative Transcription and Alternative Splicing in Mammals. *Genome Biol. Evol.* **2010**, *2*, 791–799. [CrossRef] [PubMed]

59. Nazim, M.; Masuda, A.; Rahman, M.A.; Nasrin, F.; Takeda, J.; Ohe, K.; Ohkawara, B.; Ito, M.; Ohno, K. Competitive Regulation of Alternative Splicing and Alternative Polyadenylation by hnRNP H and CstF64 Determines Acetylcholinesterase Isoforms. *Nucleic Acids Res.* **2017**, *45*, 1455–1468. [CrossRef] [PubMed]

60. Scotti, M.M.; Swanson, M.S. RNA Mis-Splicing in Disease. *Nat. Rev. Genet.* **2016**, *17*, 19–32. [CrossRef]

61. Davuluri, R.V.; Suzuki, Y.; Sugano, S.; Plass, C.; Huang, T.H.-M. The Functional Consequences of Alternative Promoter Use in Mammalian Genomes. *Trends Genet.* **2008**, *24*, 167–177. [CrossRef]

62. Qin, Z.; Stoilov, P.; Zhang, X.; Xing, Y. SEASTAR: Systematic Evaluation of Alternative Transcription Start Sites in RNA. *Nucleic Acids Res.* **2018**, *46*, e45. [CrossRef]

63. A Promoter-Level Mammalian Expression Atlas. *Nature* **2014**, *507*, 462–470. [CrossRef]

64. Wang, X.; Hou, J.; Quedenau, C.; Chen, W. Pervasive Isoform-specific Translational Regulation via Alternative Transcription Start Sites in Mammals. *Mol. Syst. Biol.* **2016**, *12*, 875. [CrossRef]

65. Lee, S.; Liu, B.; Lee, S.; Huang, S.-X.; Shen, B.; Qian, S.-B. Global Mapping of Translation Initiation Sites in Mammalian Cells at Single-Nucleotide Resolution. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E2424–E2432. [CrossRef] [PubMed]

66. Benitez-Cantos, M.S.; Yordanova, M.M.; O'Connor, P.B.F.; Zhdanov, A.V.; Kovalchuk, S.I.; Papkovsky, D.B.; Andreev, D.E.; Baranov, P.V. Translation Initiation Downstream from Annotated Start Codons in Human mRNAs Coevolves with the Kozak Context. *Genome Res.* **2020**, *30*, 974–984. [CrossRef] [PubMed]

67. Calvo, S.E.; Pagliarini, D.J.; Mootha, V.K. Upstream Open Reading Frames Cause Widespread Reduction of Protein Expression and Are Polymorphic among Humans. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 7507–7512. [CrossRef] [PubMed]

68. Barbosa, C.; Peixeiro, I.; Romão, L. Gene Expression Regulation by Upstream Open Reading Frames and Human Disease. *PLoS Genet.* **2013**, *9*, e1003529. [CrossRef] [PubMed]

69. Trulley, P.; Snieckute, G.; Bekker-Jensen, D.; Menon, M.B.; Freund, R.; Kotlyarov, A.; Olsen, J.V.; Diaz-Muñoz, M.D.; Turner, M.; Bekker-Jensen, S.; et al. Alternative Translation Initiation Generates a Functionally Distinct Isoform of the Stress-Activated Protein Kinase MK2. *Cell Rep.* **2019**, *27*, 2859–2870.e6. [CrossRef]

70. Liang, H.; Chen, X.; Yin, Q.; Ruan, D.; Zhao, X.; Zhang, C.; McNutt, M.A.; Yin, Y. PTENβ Is an Alternatively Translated Isoform of PTEN That Regulates rDNA Transcription. *Nat. Commun.* **2017**, *8*, 14771. [CrossRef]

71. Murugan, M.; Fedele, D.; Millner, D.; Alharfoush, E.; Vegunta, G.; Boison, D. Adenosine Kinase: An Epigenetic Modulator in Development and Disease. *Neurochem. Int.* **2021**, *147*, 105054. [CrossRef]

72. Lee, S.H.; Caviness, K.; Albright, E.R.; Lee, J.-H.; Gelbmann, C.B.; Rak, M.; Goodrum, F.; Kalejta, R.F. Long and Short Isoforms of the Human Cytomegalovirus UL138 Protein Silence IE Transcription and Promote Latency. *J. Virol.* **2016**, *90*, 9483–9494. [CrossRef]

73. Demircioğlu, D.; Cukuroglu, E.; Kindermans, M.; Nandi, T.; Calabrese, C.; Fonseca, N.A.; Kahles, A.; Lehmann, K.-V.; Stegle, O.; Brazma, A.; et al. A Pan-Cancer Transcriptome Analysis Reveals Pervasive Regulation through Alternative Promoters. *Cell* **2019**, *178*, 1465–1477.e17. [CrossRef]

74. Schug, J.; Schuller, W.-P.; Kappen, C.; Salbaum, J.M.; Bucan, M.; Stoeckert, C.J. Promoter Features Related to Tissue Specificity as Measured by Shannon Entropy. *Genome Biol.* **2005**, *6*, R33. [CrossRef]

75. Jacox, E.; Gotea, V.; Ovcharenko, I.; Elnitski, L. Tissue-Specific and Ubiquitous Expression Patterns from Alternative Promoters of Human Genes. *PLoS ONE* **2010**, *5*, e12274. [CrossRef] [PubMed]

76. Gacita, A.M.; Dellefave-Castillo, L.; Page, P.G.T.; Barefield, D.Y.; Wasserstrom, J.A.; Puckelwartz, M.J.; Nobrega, M.A.; McNally, E.M. Altered Enhancer and Promoter Usage Leads to Differential Gene Expression in the Normal and Failed Human Heart. *Circ. Heart Fail.* **2020**, *13*, e006926. [CrossRef] [PubMed]

77. Floor, S.N.; Doudna, J.A. Tunable Protein Synthesis by Transcript Isoforms in Human Cells. *eLife* **2016**, *5*, e10921. [CrossRef] [PubMed]

78. Karagyozov, L.; Godfrey, R.; Böhmer, S.-A.; Petermann, A.; Hölters, S.; Östman, A.; Böhmer, F.-D. The Structure of the 5′-End of the Protein-Tyrosine Phosphatase PTPRJ mRNA Reveals a Novel Mechanism for Translation Attenuation. *Nucleic Acids Res.* **2008**, *36*, 4443–4453. [CrossRef]

79. Tsukamoto, H.; Takizawa, T.; Takamori, K.; Ogawa, H.; Araki, Y. Genomic Organization and Structure of the 5′-Flanking Region of the TEX101 Gene: Alternative Promoter Usage and Splicing Generate Transcript Variants with Distinct 5′-Untranslated Region. *Mol. Reprod. Dev.* **2007**, *74*, 154–162. [CrossRef]

80. Akman, B.H.; Can, T.; Erson-Bensan, A.E. Estrogen-Induced Upregulation and 3′-UTR Shortening of CDC6. *Nucleic Acids Res.* **2012**, *40*, 10679–10688. [CrossRef]

81. Kiese, K.; Jablonski, J.; Boison, D.; Kobow, K. Dynamic Regulation of the Adenosine Kinase Gene during Early Postnatal Brain Development and Maturation. *Front. Mol. Neurosci.* **2016**, *9*, 99. [CrossRef]

82. Liu, L.; Dilworth, D.; Gao, L.; Monzon, J.; Summers, A.; Lassam, N.; Hogg, D. Mutation of the CDKN2A 5′ UTR Creates an Aberrant Initiation Codon and Predisposes to Melanoma. *Nat. Genet.* **1999**, *21*, 128–132. [CrossRef]

83. Sendoel, A.; Dunn, J.G.; Rodriguez, E.H.; Naik, S.; Gomez, N.C.; Hurwitz, B.; Levorse, J.; Dill, B.D.; Schramek, D.; Molina, H.; et al. Translation from Unconventional 5′ Start Sites Drives Tumour Initiation. *Nature* **2017**, *541*, 494–499. [CrossRef]

84. Honda, S.; Arai, Y.; Haruta, M.; Sasaki, F.; Ohira, M.; Yamaoka, H.; Horie, H.; Nakagawara, A.; Hiyama, E.; Todo, S.; et al. Loss of Imprinting of IGF2 Correlates with Hypermethylation of the H19 Differentially Methylated Region in Hepatoblastoma. *Br. J. Cancer* **2008**, *99*, 1891–1899. [CrossRef]

85. Dowdy, S.C.; Gostout, B.S.; Shridhar, V.; Wu, X.; Smith, D.I.; Podratz, K.C.; Jiang, S.-W. Biallelic Methylation and Silencing of Paternally Expressed Gene 3 (PEG3) in Gynecologic Cancer Cell Lines. *Gynecol. Oncol.* **2005**, *99*, 126–134. [CrossRef] [PubMed]

86. Rogers, J.; Early, P.; Carter, C.; Calame, K.; Bond, M.; Hood, L.; Wall, R. Two mRNAs with Different 3′ Ends Encode Membrane-Bound and Secreted Forms of Immunoglobulin μ Chain. *Cell* **1980**, *20*, 303–312. [CrossRef] [PubMed]

87. Setzer, D.R.; McGrogan, M.; Nunberg, J.H.; Schimke, R.T. Size Heterogeneity in the 3′ End of Dihydrofolate Reductase Messenger RNAs in Mouse Cells. *Cell* **1980**, *22*, 361–370. [CrossRef] [PubMed]

88. Zhang, Y.; Liu, L.; Qiu, Q.; Zhou, Q.; Ding, J.; Lu, Y.; Liu, P. Alternative Polyadenylation: Methods, Mechanism, Function, and Role in Cancer. *J. Exp. Clin. Cancer Res.* **2021**, *40*, 51. [CrossRef]

89. Tang, P.; Yang, Y.; Li, G.; Huang, L.; Wen, M.; Ruan, W.; Guo, X.; Zhang, C.; Zuo, X.; Luo, D.; et al. Alternative Polyadenylation by Sequential Activation of Distal and Proximal PolyA Sites. *Nat. Struct. Mol. Biol.* **2022**, *29*, 21–31. [CrossRef]

90. Derti, A.; Garrett-Engele, P.; MacIsaac, K.D.; Stevens, R.C.; Sriram, S.; Chen, R.; Rohl, C.A.; Johnson, J.M.; Babak, T. A Quantitative Atlas of Polyadenylation in Five Mammals. *Genome Res.* **2012**, *22*, 1173–1183. [CrossRef]

91. Tian, B.; Manley, J.L. Alternative Polyadenylation of mRNA Precursors. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 18–30. [CrossRef]

92. Shi, Y. Alternative Polyadenylation: New Insights from Global Analyses. *RNA* **2012**, *18*, 2105–2117. [CrossRef]

93. Hoque, M.; Ji, Z.; Zheng, D.; Luo, W.; Li, W.; You, B.; Park, J.Y.; Yehia, G.; Tian, B. Analysis of Alternative Cleavage and Polyadenylation by 3′ Region Extraction and Deep Sequencing. *Nat. Methods* **2013**, *10*, 133–139. [CrossRef]

94. Evolution and Biological Roles of Alternative 3′UTRs. *Trends Cell Biol.* **2016**, *26*, 227–237. [CrossRef]

95. Gruber, A.R.; Martin, G.; Müller, P.; Schmidt, A.; Gruber, A.J.; Gumienny, R.; Mittal, N.; Jayachandran, R.; Pieters, J.; Keller, W.; et al. Global 3′ UTR Shortening Has a Limited Effect on Protein Abundance in Proliferating T Cells. *Nat. Commun.* **2014**, *5*, 5465. [CrossRef]

96. Spies, N.; Burge, C.B.; Bartel, D.P. 3′ UTR-Isoform Choice Has Limited Influence on the Stability and Translational Efficiency of Most mRNAs in Mouse Fibroblasts. *Genome Res.* **2013**, *23*, 2078–2090. [CrossRef] [PubMed]

97. Geisberg, J.V.; Moqtaderi, Z.; Fan, X.; Ozsolak, F.; Struhl, K. Global Analysis of mRNA Isoform Half-Lives Reveals Stabilizing and Destabilizing Elements in Yeast. *Cell* **2014**, *156*, 812–824. [CrossRef] [PubMed]

98. Mittleman, B.E.; Pott, S.; Warland, S.; Zeng, T.; Mu, Z.; Kaur, M.; Gilad, Y.; Li, Y. Alternative Polyadenylation Mediates Genetic Regulation of Gene Expression. *eLife* **2020**, *9*, e57492. [CrossRef] [PubMed]

99. Kakoki, M.; Pochynyuk, O.M.; Hathaway, C.M.; Tomita, H.; Hagaman, J.R.; Kim, H.-S.; Zaika, O.L.; Mamenko, M.; Kayashima, Y.; Matsuki, K.; et al. Primary Aldosteronism and Impaired Natriuresis in Mice Underexpressing TGFβ1. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 5600–5605. [CrossRef]

100. Yeh, H.-S.; Yong, J. Alternative Polyadenylation of mRNAs: 3′-Untranslated Region Matters in Gene Expression. *Mol. Cells* **2016**, *39*, 281–285. [CrossRef]

101. Agarwal, V.; Lopez-Darwin, S.; Kelley, D.R.; Shendure, J. The Landscape of Alternative Polyadenylation in Single Cells of the Developing Mouse Embryo. *Nat. Commun.* **2021**, *12*, 5101. [CrossRef]

102. Xiao, R.; Li, C.; Wang, C.; Cao, Y.; Zhang, L.; Guo, Y.; Xin, Y.; Zhang, H.; Zhou, G. Adipogenesis Associated Mth938 Domain Containing (AAMDC) Protein Expression Is Regulated by Alternative Polyadenylation and microRNAs. *FEBS Lett.* **2019**, *593*, 1724–1734. [CrossRef]

103. Bae, B.; Gruner, H.N.; Lynch, M.; Feng, T.; So, K.; Oliver, D.; Mastick, G.S.; Yan, W.; Pieraut, S.; Miura, P. Elimination of Calm1 Long 3′-UTR mRNA Isoform by CRISPR–Cas9 Gene Editing Impairs Dorsal Root Ganglion Development and Hippocampal Neuron Activation in Mice. *RNA* **2020**, *26*, 1414–1430. [CrossRef]

104. Davis, A.G.; Einstein, J.M.; Zheng, D.; Jayne, N.D.; Fu, X.-D.; Tian, B.; Yeo, G.W.; Zhang, D.-E. A CRISPR RNA-Binding Protein Screen Reveals Regulators of RUNX1 Isoform Generation. *Blood Adv.* **2021**, *5*, 1310–1323. [CrossRef]

105. GUAN, X.; FIERKE, C.A. Understanding Protein Palmitoylation: Biological Significance and Enzymology. *Sci. China Chem.* **2011**, *54*, 1888–1897. [CrossRef] [PubMed]

106. Pereira-Castro, I.; Garcia, B.C.; Curinha, A.; Neves-Costa, A.; Conde-Sousa, E.; Moita, L.F.; Moreira, A. MCL1 Alternative Polyadenylation Is Essential for Cell Survival and Mitochondria Morphology. *Cell. Mol. Life Sci.* **2022**, *79*, 164. [CrossRef] [PubMed]

107. Andreassi, C.; Luisier, R.; Crerar, H.; Darsinou, M.; Blokzijl-Franke, S.; Lenn, T.; Luscombe, N.M.; Cuda, G.; Gaspari, M.; Saiardi, A.; et al. Cytoplasmic Cleavage of IMPA1 3′ UTR Is Necessary for Maintaining Axon Integrity. *Cell Rep.* **2021**, *34*, 108778. [CrossRef] [PubMed]

108. Taliaferro, J.M.; Vidaki, M.; Oliveira, R.; Olson, S.; Zhan, L.; Saxena, T.; Wang, E.T.; Graveley, B.R.; Gertler, F.B.; Swanson, M.S.; et al. Distal Alternative Last Exons Localize mRNAs to Neural Projections. *Mol. Cell* **2016**, *61*, 821–833. [CrossRef]

109. Lee, S.; Chen, Y.-C.; Gillen, A.E.; Taliaferro, J.M.; Deplancke, B.; Li, H.; Lai, E.C. Diverse Cell-Specific Patterns of Alternative Polyadenylation in Drosophila. *Nat. Commun.* **2022**, *13*, 5372. [CrossRef]

110. Calame, K.L.; Lin, K.-I.; Tunyaplin, C. Regulatory Mechanisms That Determine the Development and Function of Plasma Cells. *Annu. Rev. Immunol.* **2003**, *21*, 205–230. [CrossRef]

111. Galli, G.; Guise, J.W.; McDevitt, M.A.; Tucker, P.W.; Nevins, J.R. Relative Position and Strengths of Poly(A) Sites as Well as Transcription Termination Are Critical to Membrane versus Secreted Mu-Chain Expression during B-Cell Development. *Genes Dev.* **1987**, *1*, 471–481. [CrossRef]

112. Singh, I.; Lee, S.-H.; Sperling, A.S.; Samur, M.K.; Tai, Y.-T.; Fulciniti, M.; Munshi, N.C.; Mayr, C.; Leslie, C.S. Widespread Intronic Polyadenylation Diversifies Immune Cell Transcriptomes. *Nat. Commun.* **2018**, *9*, 1716. [CrossRef]

113. Gruber, A.J.; Zavolan, M. Alternative Cleavage and Polyadenylation in Health and Disease. *Nat. Rev. Genet.* **2019**, *20*, 599–614. [CrossRef]

114. Curinha, A.; Oliveira Braz, S.; Pereira-Castro, I.; Cruz, A.; Moreira, A. Implications of Polyadenylation in Health and Disease. *Nucleus* **2014**, *5*, 508–519. [CrossRef]

115. Rehfeld, A.; Plass, M.; Krogh, A.; Friis-Hansen, L. Alterations in Polyadenylation and Its Implications for Endocrine Disease. *Front. Endocrinol.* **2013**, *4*, 53. [CrossRef] [PubMed]

116. Anvar, S.Y.; Allard, G.; Tseng, E.; Sheynkman, G.M.; de Klerk, E.; Vermaat, M.; Yin, R.H.; Johansson, H.E.; Ariyurek, Y.; den Dunnen, J.T.; et al. Full-Length mRNA Sequencing Uncovers a Widespread Coupling between Transcription Initiation and mRNA Processing. *Genome Biol.* **2018**, *19*, 46. [CrossRef] [PubMed]

117. Hsin, J.-P.; Manley, J.L. The RNA Polymerase II CTD Coordinates Transcription and RNA Processing. *Genes Dev.* **2012**, *26*, 2119–2137. [CrossRef] [PubMed]

118. Batra, R.; Charizanis, K.; Manchanda, M.; Mohan, A.; Li, M.; Finn, D.J.; Goodwin, M.; Zhang, C.; Sobczak, K.; Thornton, C.A.; et al. Loss of MBNL Leads to Disruption of Developmentally Regulated Alternative Polyadenylation in RNA-Mediated Disease. *Mol. Cell* **2014**, *56*, 311–322. [CrossRef] [PubMed]

119. Kornblihtt, A.R.; Schor, I.E.; Alló, M.; Dujardin, G.; Petrillo, E.; Muñoz, M.J. Alternative Splicing: A Pivotal Step between Eukaryotic Transcription and Translation. *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 153–165. [CrossRef]

120. Listerman, I.; Sapra, A.K.; Neugebauer, K.M. Cotranscriptional Coupling of Splicing Factor Recruitment and Precursor Messenger RNA Splicing in Mammalian Cells. *Nat. Struct. Mol. Biol.* **2006**, *13*, 815–822. [CrossRef]

121. Babu, M.M. The Contribution of Intrinsically Disordered Regions to Protein Function, Cellular Complexity, and Human Disease. *Biochem. Soc. Trans.* **2016**, *44*, 1185–1200. [CrossRef]

122. Schaarschmidt, S.; Fischer, A.; Zuther, E.; Hincha, D.K. Evaluation of Seven Different RNA-Seq Alignment Tools Based on Experimental Data from the Model Plant Arabidopsis Thaliana. *Int. J. Mol. Sci.* **2020**, *21*, 1720. [CrossRef]

123. Corchete, L.A.; Rojas, E.A.; Alonso-López, D.; De Las Rivas, J.; Gutiérrez, N.C.; Burguillo, F.J. Systematic Comparison and Assessment of RNA-Seq Procedures for Gene Expression Quantitative Analysis. *Sci. Rep.* **2020**, *10*, 19737. [CrossRef]

124. Robert, C.; Watson, M. Errors in RNA-Seq Quantification Affect Genes of Relevance to Human Disease. *Genome Biol.* **2015**, *16*, 177. [CrossRef]

125. Assefa, A.T.; De Paepe, K.; Everaert, C.; Mestdagh, P.; Thas, O.; Vandesompele, J. Differential Gene Expression Analysis Tools Exhibit Substandard Performance for Long Non-Coding RNA-Sequencing Data. *Genome Biol.* **2018**, *19*, 96. [CrossRef] [PubMed]

126. Zyprych-Walczak, J.; Szabelska, A.; Handschuh, L.; Górczak, K.; Klamecka, K.; Figlerowicz, M.; Siatkowski, I. The Impact of Normalization Methods on RNA-Seq Data Analysis. *BioMed Res. Int.* **2015**, *2015*, e621690. [CrossRef] [PubMed]

127. Wang, T.; Li, B.; Nelson, C.E.; Nabavi, S. Comparative Analysis of Differential Gene Expression Analysis Tools for Single-Cell RNA Sequencing Data. *BMC Bioinform.* **2019**, *20*, 40. [CrossRef] [PubMed]

128. Anders, S.; Reyes, A.; Huber, W. Detecting Differential Usage of Exons from RNA-Seq Data. *Genome Res.* **2012**, *22*, 2008–2017. [CrossRef] [PubMed]

129. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* **2010**, *26*, 139–140. [CrossRef]

130. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* **2015**, *43*, e47. [CrossRef]

131. Mehmood, A.; Laiho, A.; Venäläinen, M.S.; McGlinchey, A.J.; Wang, N.; Elo, L.L. Systematic Evaluation of Differential Splicing Tools for RNA-Seq Studies. *Brief. Bioinform.* **2020**, *21*, 2052–2065. [CrossRef]

132. Love, M.I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [CrossRef]

133. Fenn, A.; Tsoy, O.; Faro, T.; Rössler, F.; Dietrich, A.; Kersting, J.; Louadi, Z.; Lio, C.T.; Völker, U.; Baumbach, J.; et al. Alternative Splicing Analysis Benchmark with DICAST. *NAR Genom. Bioinform.* **2022**, *5*, lqad044. [CrossRef]

134. Jiang, M.; Zhang, S.; Yin, H.; Zhuo, Z.; Meng, G. A Comprehensive Benchmarking of Differential Splicing Tools for RNA-Seq Analysis at the Event Level. *Brief. Bioinform.* **2023**, *24*, bbad121. [CrossRef]

135. Olofsson, D.; Preußner, M.; Kowar, A.; Heyd, F.; Neumann, A. One Pipeline to Predict Them All? On the Prediction of Alternative Splicing from RNA-Seq Data. *Biochem. Biophys. Res. Commun.* **2023**, *653*, 31–37. [CrossRef] [PubMed]

136. Shen, S.; Park, J.W.; Lu, Z.; Lin, L.; Henry, M.D.; Wu, Y.N.; Zhou, Q.; Xing, Y. rMATS: Robust and Flexible Detection of Differential Alternative Splicing from Replicate RNA-Seq Data. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E5593–E5601. [CrossRef] [PubMed]

137. Sterne-Weiler, T.; Weatheritt, R.J.; Best, A.; Ha, K.C.H.; Blencowe, B.J. Whippet: An Efficient Method for the Detection and Quantification of Alternative Splicing Reveals Extensive Transcriptomic Complexity. *bioRxiv* **2017**. [CrossRef]

138. Mancini, E.; Rabinovich, A.; Iserte, J.; Yanovsky, M.; Chernomoretz, A. ASpli: Integrative Analysis of Splicing Landscapes through RNA-Seq Assays. *Bioinformatics* **2021**, *37*, 2609–2616. [CrossRef]

139. Goldstein, L.D.; Cao, Y.; Pau, G.; Lawrence, M.; Wu, T.D.; Seshagiri, S.; Gentleman, R. Prediction and Quantification of Splice Events from RNA-Seq Data. *PLoS ONE* **2016**, *11*, e0156132. [CrossRef]

140. Middleton, R.; Gao, D.; Thomas, A.; Singh, B.; Au, A.; Wong, J.J.-L.; Bomane, A.; Cosson, B.; Eyras, E.; Rasko, J.E.J.; et al. IRFinder: Assessing the Impact of Intron Retention on Mammalian Gene Expression. *Genome Biol.* **2017**, *18*, 51. [CrossRef]

141. Kahles, A.; Ong, C.S.; Zhong, Y.; Rätsch, G. SplAdder: Identification, Quantification and Testing of Alternative Splicing Events from RNA-Seq Data. *Bioinformatics* **2016**, *32*, 1840–1847. [CrossRef]

142. Vaquero-Garcia, J.; Barrera, A.; Gazzara, M.R.; González-Vallinas, J.; Lahens, N.F.; Hogenesch, J.B.; Lynch, K.W.; Barash, Y. A New View of Transcriptome Complexity and Regulation through the Lens of Local Splicing Variations. *eLife* **2016**, *5*, e11752. [CrossRef]

143. Huang, Y.; Sanguinetti, G. BRIE2: Computational Identification of Splicing Phenotypes from Single-Cell Transcriptomic Experiments. *Genome Biol.* **2021**, *22*, 251. [CrossRef]

144. Thodberg, M.; Thieffry, A.; Vitting-Seerup, K.; Andersson, R.; Sandelin, A. CAGEfightR: Analysis of 5′-End Data Using R/Bioconductor. *BMC Bioinform.* **2019**, *20*, 487. [CrossRef]

145. Policastro, R.A.; McDonald, D.J.; Brendel, V.P.; Zentner, G.E. Flexible Analysis of TSS Mapping Data and Detection of TSS Shifts with TSRexploreR. *NAR Genom. Bioinform.* **2021**, *3*, lqab051. [CrossRef] [PubMed]

146. Cass, A.A.; Xiao, X. mountainClimber Identifies Alternative Transcription Start and Polyadenylation Sites in RNA-Seq. *Cell Syst.* **2019**, *9*, 393–400.e6. [CrossRef] [PubMed]

147. Hou, R.; Hon, C.-C.; Huang, Y. CamoTSS: Analysis of Alternative Transcription Start Sites for Cellular Phenotypes and Regulatory Patterns from 5′ scRNA-Seq Data. *bioRxiv* **2023**. [CrossRef]

148. Shah, A.; Mittleman, B.E.; Gilad, Y.; Li, Y.I. Benchmarking Sequencing Methods and Tools That Facilitate the Study of Alternative Polyadenylation. *Genome Biol.* **2021**, *22*, 291. [CrossRef] [PubMed]

149. Ye, W.; Lian, Q.; Ye, C.; Wu, X. A Survey on Methods for Predicting Polyadenylation Sites from DNA Sequences, Bulk RNA-Seq, and Single-Cell RNA-Seq. *Genom. Proteom. Bioinform.* **2023**, *21*, 67–83. [CrossRef]

150. Bryce-Smith, S.; Burri, D.; Gazzara, M.R.; Herrmann, C.J.; Danecka, W.; Fitzsimmons, C.M.; Wan, Y.K.; Zhuang, F.; Fansler, M.M.; Fernández, J.M.; et al. Extensible Benchmarking of Methods That Identify and Quantify Polyadenylation Sites from RNA-Seq Data. *RNA* **2023**, *Online ahead of print*. [CrossRef]

151. Chen, M.; Ji, G.; Fu, H.; Lin, Q.; Ye, C.; Ye, W.; Su, Y.; Wu, X. A Survey on Identification and Quantification of Alternative Polyadenylation Sites from RNA-Seq Data. *Brief. Bioinform.* **2020**, *21*, 1261–1276. [CrossRef]

152. Ye, C.; Long, Y.; Ji, G.; Li, Q.Q.; Wu, X. APAtrap: Identification and Quantification of Alternative Polyadenylation Sites from RNA-Seq Data. *Bioinformatics* **2018**, *34*, 1841–1849. [CrossRef]

153. Arefeen, A.; Liu, J.; Xiao, X.; Jiang, T. TAPAS: Tool for Alternative Polyadenylation Site Analysis. *Bioinformatics* **2018**, *34*, 2521–2529. [CrossRef]

154. Arefeen, A.; Xiao, X.; Jiang, T. DeepPASTA: Deep Neural Network Based Polyadenylation Site Analysis. *Bioinformatics* **2019**, *35*, 4577–4585. [CrossRef]

155. Hao, J.; Kim, Y.; Kim, T.-K.; Kang, M. PASNet: Pathway-Associated Sparse Deep Neural Network for Prognosis Prediction from High-Throughput Data. *BMC Bioinform.* **2018**, *19*, 510. [CrossRef] [PubMed]

156. Ji, G.; Zheng, J.; Shen, Y.; Wu, X.; Jiang, R.; Lin, Y.; Loke, J.C.; Davis, K.M.; Reese, G.J.; Li, Q.Q. Predictive Modeling of Plant Messenger RNA Polyadenylation Sites. *BMC Bioinform.* **2007**, *8*, 43. [CrossRef] [PubMed]

157. Ha, K.C.H.; Blencowe, B.J.; Morris, Q. QAPA: A New Method for the Systematic Analysis of Alternative Polyadenylation from RNA-Seq Data. *Genome Biol.* **2018**, *19*, 45. [CrossRef] [PubMed]

158. Xia, Z.; Donehower, L.A.; Cooper, T.A.; Neilson, J.R.; Wheeler, D.A.; Wagner, E.J.; Li, W. Dynamic Analyses of Alternative Polyadenylation from RNA-Seq Reveal a 3′-UTR Landscape across Seven Tumour Types. *Nat. Commun.* **2014**, *5*, 5274. [CrossRef]

159. Wu, X.; Liu, T.; Ye, C.; Ye, W.; Ji, G. scAPAtrap: Identification and Quantification of Alternative Polyadenylation Sites from Single-Cell RNA-Seq Data. *Brief. Bioinform.* **2021**, *22*, bbaa273. [CrossRef]

160. Shulman, E.D.; Elkon, R. Cell-Type-Specific Analysis of Alternative Polyadenylation Using Single-Cell Transcriptomics Data. *Nucleic Acids Res.* **2019**, *47*, 10027–10039. [CrossRef]

161. Cheng, X.; Yan, J.; Liu, Y.; Wang, J.; Taubert, S. eVITTA: A Web-Based Visualization and Inference Toolbox for Transcriptome Analysis. *Nucleic Acids Res.* **2021**, *49*, W207–W215. [CrossRef]

162. Flemington, E.K.; Flemington, S.A.; O'Grady, T.M.; Baddoo, M.; Nguyen, T.; Dong, Y.; Ungerleider, N.A. SpliceTools, a Suite of Downstream RNA Splicing Analysis Tools to Investigate Mechanisms and Impact of Alternative Splicing. *Nucleic Acids Res.* **2023**, *51*, e42. [CrossRef]

163. Hu, X.; Song, J.; Chyr, J.; Wan, J.; Wang, X.; Du, J.; Duan, J.; Zhang, H.; Zhou, X.; Wu, X. APAview: A Web-Based Platform for Alternative Polyadenylation Analyses in Hematological Cancers. *Front. Genet.* **2022**, *13*, 928862. [CrossRef]

164. Han, S.; Kim, D.; Kim, Y.; Choi, K.; Miller, J.E.; Kim, D.; Lee, Y. CAS-Viewer: Web-Based Tool for Splicing-Guided Integrative Analysis of Multi-Omics Cancer Data. *BMC Med. Genom.* **2018**, *11*, 25. [CrossRef]

165. Hong, W.; Ruan, H.; Zhang, Z.; Ye, Y.; Liu, Y.; Li, S.; Jing, Y.; Zhang, H.; Diao, L.; Liang, H.; et al. APAatlas: Decoding Alternative Polyadenylation across Human Tissues. *Nucleic Acids Res.* **2020**, *48*, D34–D39. [CrossRef] [PubMed]

166. Palmisano, A.; Vural, S.; Zhao, Y.; Sonkin, D. MutSpliceDB: A Database of Splice Sites Variants with RNA-Seq Based Evidence on Effects on Splicing. *Hum. Mutat.* **2021**, *42*, 342–345. [CrossRef] [PubMed]

167. Landrum, M.J.; Lee, J.M.; Benson, M.; Brown, G.R.; Chao, C.; Chitipiralla, S.; Gu, B.; Hart, J.; Hoffman, D.; Jang, W.; et al. ClinVar: Improving Access to Variant Interpretations and Supporting Evidence. *Nucleic Acids Res.* **2018**, *46*, D1062–D1067. [CrossRef] [PubMed]

168. Kim, P.; Yang, M.; Yiya, K.; Zhao, W.; Zhou, X. ExonSkipDB: Functional Annotation of Exon Skipping Event in Human. *Nucleic Acids Res.* **2020**, *48*, D896–D907. [CrossRef]

169. Tapial, J.; Ha, K.C.H.; Sterne-Weiler, T.; Gohr, A.; Braunschweig, U.; Hermoso-Pulido, A.; Quesnel-Vallières, M.; Permanyer, J.; Sodaei, R.; Marquez, Y.; et al. An Atlas of Alternative Splicing Profiles and Functional Associations Reveals New Regulatory Programs and Genes That Simultaneously Express Multiple Major Isoforms. *Genome Res.* **2017**, *27*, 1759–1768. [CrossRef]

170. Busch, A.; Hertel, K.J. HEXEvent: A Database of Human EXon Splicing Events. *Nucleic Acids Res.* **2013**, *41*, D118–D124. [CrossRef]

171. Yamashita, R.; Sugano, S.; Suzuki, Y.; Nakai, K. DBTSS: DataBase of Transcriptional Start Sites Progress Report in 2012. *Nucleic Acids Res.* **2012**, *40*, D150–D154. [CrossRef]

172. Abugessaisa, I.; Noguchi, S.; Hasegawa, A.; Kondo, A.; Kawaji, H.; Carninci, P.; Kasukawa, T. refTSS: A Reference Data Set for Human and Mouse Transcription Start Sites. *J. Mol. Biol.* **2019**, *431*, 2407–2422. [CrossRef]

173. Herrmann, C.J.; Schmidt, R.; Kanitz, A.; Artimo, P.; Gruber, A.J.; Zavolan, M. PolyASite 2.0: A Consolidated Atlas of Polyadenylation Sites from 3′ End Sequencing. *Nucleic Acids Res.* **2020**, *48*, D174–D179. [CrossRef]

174. Wang, R.; Nambiar, R.; Zheng, D.; Tian, B. PolyA_DB 3 Catalogs Cleavage and Polyadenylation Sites Identified by Deep Sequencing in Multiple Genomes. *Nucleic Acids Res.* **2018**, *46*, D315–D319. [CrossRef]

175. Zhu, S.; Lian, Q.; Ye, W.; Qin, W.; Wu, Z.; Ji, G.; Wu, X. scAPAdb: A Comprehensive Database of Alternative Polyadenylation at Single-Cell Resolution. *Nucleic Acids Res.* **2022**, *50*, D365–D370. [CrossRef] [PubMed]

176. Zhou, X.; Li, R.; Michal, J.J.; Wu, X.-L.; Liu, Z.; Zhao, H.; Xia, Y.; Du, W.; Wildung, M.R.; Pouchnik, D.J.; et al. Accurate Profiling of Gene Expression and Alternative Polyadenylation with Whole Transcriptome Termini Site Sequencing (WTTS-Seq). *Genetics* **2016**, *203*, 683–697. [CrossRef] [PubMed]

177. Zhou, X.; Zhang, Y.; Michal, J.J.; Qu, L.; Zhang, S.; Wildung, M.R.; Du, W.; Pouchnik, D.J.; Zhao, H.; Xia, Y.; et al. Alternative Polyadenylation Coordinates Embryonic Development, Sexual Dimorphism and Longitudinal Growth in Xenopus Tropicalis. *Cell. Mol. Life Sci. CMLS* **2019**, *76*, 2185–2198. [CrossRef]

178. Agrotis, A.; Pengo, N.; Burden, J.J.; Ketteler, R. Redundancy of Human ATG4 Protease Isoforms in Autophagy and LC3/GABARAP Processing Revealed in Cells. *Autophagy* **2019**, *15*, 976–997. [CrossRef]

179. Foltran, R.B.; Diaz, S.L. BDNF Isoforms: A Round Trip Ticket between Neurogenesis and Serotonin? *J. Neurochem.* **2016**, *138*, 204–221. [CrossRef] [PubMed]

180. Cheng, A.; Coksaygan, T.; Tang, H.; Khatri, R.; Balice-Gordon, R.J.; Rao, M.S.; Mattson, M.P. Truncated Tyrosine Kinase B Brain-Derived Neurotrophic Factor Receptor Directs Cortical Neural Stem Cells to a Glial Cell Fate by a Novel Signaling Mechanism. *J. Neurochem.* **2007**, *100*, 1515–1530. [CrossRef]

181. Alsarraj, J.; Faraji, F.; Geiger, T.R.; Mattaini, K.R.; Williams, M.; Wu, J.; Ha, N.-H.; Merlino, T.; Walker, R.C.; Bosley, A.D.; et al. BRD4 Short Isoform Interacts with RRP1B, SIPA1 and Components of the LINC Complex at the Inner Face of the Nuclear Membrane. *PLoS ONE* **2013**, *8*, e80746. [CrossRef]

182. Wu, S.-Y.; Lee, C.-F.; Lai, H.-T.; Yu, C.-T.; Lee, J.-E.; Zuo, H.; Tsai, S.Y.; Tsai, M.-J.; Ge, K.; Wan, Y.; et al. Opposing Functions of BRD4 Isoforms in Breast Cancer. *Mol. Cell* **2020**, *78*, 1114–1132.e10. [CrossRef]

183. Drumond-Bock, A.L.; Bieniasz, M. The Role of Distinct BRD4 Isoforms and Their Contribution to High-Grade Serous Ovarian Carcinoma Pathogenesis. *Mol. Cancer* **2021**, *20*, 145. [CrossRef]

184. Bhattacherjee, A.; Jung, J.; Zia, S.; Ho, M.; Eskandari-Sedighi, G.; St. Laurent, C.D.; McCord, K.A.; Bains, A.; Sidhu, G.; Sarkar, S.; et al. The CD33 Short Isoform Is a Gain-of-Function Variant That Enhances Aβ1–42 Phagocytosis in Microglia. *Mol. Neurodegener.* **2021**, *16*, 19. [CrossRef]

185. Ciolli Mattioli, C.; Rom, A.; Franke, V.; Imami, K.; Arrey, G.; Terne, M.; Woehler, A.; Akalin, A.; Ulitsky, I.; Chekulaeva, M. Alternative 3′ UTRs Direct Localization of Functionally Diverse Protein Isoforms in Neuronal Compartments. *Nucleic Acids Res.* **2019**, *47*, 2560–2573. [CrossRef] [PubMed]

186. Phizicky, D.V.; Bell, S.P. Transcriptional Repression of CDC6 and SLD2 during Meiosis Is Associated with Production of Short Heterogeneous RNA Isoforms. *Chromosoma* **2018**, *127*, 515–527. [CrossRef] [PubMed]

187. Kiriyama, S.; Yokoyama, S.; Ueno, M.; Hayami, S.; Ieda, J.; Yamamoto, N.; Yamaguchi, S.; Mitani, Y.; Nakamura, Y.; Tani, M.; et al. CEACAM1 Long Cytoplasmic Domain Isoform Is Associated with Invasion and Recurrence of Hepatocellular Carcinoma. *Ann. Surg. Oncol.* **2014**, *21*, 505–514. [CrossRef] [PubMed]

188. Sadekova, S.; Lamarche-Vane, N.; Li, X.; Beauchemin, N. The CEACAM1-L Glycoprotein Associates with the Actin Cytoskeleton and Localizes to Cell–Cell Contact through Activation of Rho-like GTPases. *Mol. Biol. Cell* **2000**, *11*, 65–77. [CrossRef] [PubMed]

189. Ascenzi, F.; Barberi, L.; Dobrowolny, G.; Villa Nova Bacurau, A.; Nicoletti, C.; Rizzuto, E.; Rosenthal, N.; Scicchitano, B.M.; Musarò, A. Effects of IGF-1 Isoforms on Muscle Growth and Sarcopenia. *Aging Cell* **2019**, *18*, e12954. [CrossRef]

190. Annibalini, G.; Contarelli, S.; De Santi, M.; Saltarelli, R.; Di Patria, L.; Guescini, M.; Villarini, A.; Brandi, G.; Stocchi, V.; Barbieri, E. The Intrinsically Disordered E-Domains Regulate the IGF-1 Prohormones Stability, Subcellular Localisation and Secretion. *Sci. Rep.* **2018**, *8*, 8919. [CrossRef]

191. Philippou, A.; Maridaki, M.; Pneumaticos, S.; Koutsilieris, M. The Complexity of the IGF1 Gene Splicing, Posttranslational Modification and Bioactivity. *Mol. Med.* **2014**, *20*, 202–214. [CrossRef]

192. Lewis-Tuffin, L.J.; Jewell, C.M.; Bienstock, R.J.; Collins, J.B.; Cidlowski, J.A. Human Glucocorticoid Receptor β Binds RU-486 and Is Transcriptionally Active. *Mol. Cell. Biol.* **2007**, *27*, 2266–2282. [CrossRef]

193. Dierolf, J.G.; Watson, A.J.; Betts, D.H. Differential Localization Patterns of Pyruvate Kinase Isoforms in Murine Naïve, Formative, and Primed Pluripotent States. *Exp. Cell Res.* **2021**, *405*, 112714. [CrossRef]

194. Taniguchi, K.; Ito, Y.; Sugito, N.; Kumazaki, M.; Shinohara, H.; Yamada, N.; Nakagawa, Y.; Sugiyama, T.; Futamura, M.; Otsuki, Y.; et al. Organ-Specific PTB1-Associated microRNAs Determine Expression of Pyruvate Kinase Isoforms. *Sci. Rep.* **2015**, *5*, 8647. [CrossRef]

195. Kiviluoto, S.; Decuypere, J.-P.; De Smedt, H.; Missiaen, L.; Parys, J.B.; Bultynck, G. STIM1 as a Key Regulator for Ca²⁺ Homeostasis in Skeletal-Muscle Development and Function. *Skelet. Muscle* **2011**, *1*, 16. [CrossRef] [PubMed]

196. Knapp, M.L.; Alansary, D.; Poth, V.; Förderer, K.; Sommer, F.; Zimmer, D.; Schwarz, Y.; Künzel, N.; Kless, A.; Machaca, K.; et al. A Longer Isoform of Stim1 Is a Negative SOCE Regulator but Increases cAMP-Modulated NFAT Signaling. *EMBO Rep.* **2022**, *23*, e53135. [CrossRef] [PubMed]

197. Ramesh, G.; Jarzembowski, L.; Schwarz, Y.; Poth, V.; Konrad, M.; Knapp, M.L.; Schwär, G.; Lauer, A.A.; Grimm, M.O.W.; Alansary, D.; et al. A Short Isoform of STIM1 Confers Frequency-Dependent Synaptic Enhancement. *Cell Rep.* **2021**, *34*, 108844. [CrossRef]

198. Goldsmith, J.F.; Hall, C.G.; Atkinson, T.P. Identification of an Alternatively Spliced Isoform of the Fyn Tyrosine Kinase. *Biochem. Biophys. Res. Commun.* **2002**, *298*, 501–504. [CrossRef] [PubMed]

199. Brignatz, C.; Paronetto, M.P.; Opi, S.; Cappellari, M.; Audebert, S.; Feuillet, V.; Bismuth, G.; Roche, S.; Arold, S.T.; Sette, C.; et al. Alternative Splicing Modulates Autoinhibition and SH3 Accessibility in the Src Kinase Fyn. *Mol. Cell. Biol.* **2009**, *29*, 6438–6448. [CrossRef] [PubMed]

200. Picard, C.; Gabert, J.; Olive, D.; Collette, Y. Altered Splicing in Hematological Malignancies Reveals a Tissue-Specific Translational Block of the Src-Family Tyrosine Kinase Fyn Brain Isoform Expression. *Leukemia* **2004**, *18*, 1737–1739. [CrossRef]

201. Toutant, M.; Costa, A.; Studler, J.-M.; Kadaré, G.; Carnaud, M.; Girault, J.-A. Alternative Splicing Controls the Mechanisms of FAK Autophosphorylation. *Mol. Cell. Biol.* **2002**, *22*, 7731–7743. [CrossRef]

202. Jereb, S.; Hwang, H.-W.; Van Otterloo, E.; Govek, E.-E.; Fak, J.J.; Yuan, Y.; Hatten, M.E.; Darnell, R.B. Differential 3′ Processing of Specific Transcripts Expands Regulatory and Protein Diversity across Neuronal Cell Types. *eLife* **2018**, *7*, e34042. [CrossRef]

203. Wang, J.-Z.; Fu, X.; Fang, Z.; Liu, H.; Zong, F.-Y.; Zhu, H.; Yu, Y.-F.; Zhang, X.-Y.; Wang, S.-F.; Huang, Y.; et al. QKI-5 Regulates the Alternative Splicing of Cytoskeletal Gene ADD3 in Lung Cancer. *J. Mol. Cell Biol.* **2020**, *13*, 347–360. [CrossRef]

204. Wang, Y.; Vogel, G.; Yu, Z.; Richard, S. The QKI-5 and QKI-6 RNA Binding Proteins Regulate the Expression of MicroRNA 7 in Glial Cells. *Mol. Cell. Biol.* **2013**, *33*, 1233–1243. [CrossRef]

205. Larocque, D.; Fragoso, G.; Huang, J.; Mushynski, W.E.; Loignon, M.; Richard, S.; Almazan, G. The QKI-6 and QKI-7 RNA Binding Proteins Block Proliferation and Promote Schwann Cell Myelination. *PLoS ONE* **2009**, *4*, e5867. [CrossRef] [PubMed]

206. Pattwell, S.S.; Arora, S.; Cimino, P.J.; Ozawa, T.; Szulzewsky, F.; Hoellerbauer, P.; Bonifert, T.; Hoffstrom, B.G.; Boiani, N.E.; Bolouri, H.; et al. A Kinase-Deficient NTRK2 Splice Variant Predominates in Glioma and Amplifies Several Oncogenic Signaling Pathways. *Nat. Commun.* **2020**, *11*, 2977. [CrossRef] [PubMed]

207. Clark, K.; Hammond, E.; Rabbitts, P. Temporal and Spatial Expression of Two Isoforms of the Dutt1/Robo1 Gene in Mouse Development. *FEBS Lett.* **2002**, *523*, 12–16. [CrossRef]

208. Camurri, L.; Mambetisaeva, E.; Davies, D.; Parnavelas, J.; Sundaresan, V.; Andrews, W. Evidence for the Existence of Two Robo3 Isoforms with Divergent Biochemical Properties. *Mol. Cell. Neurosci.* **2005**, *30*, 485–493. [CrossRef]

209. Ruedel, A.; Schott, M.; Schubert, T.; Bosserhoff, A.K. Robo3A and Robo3B Expression Is Regulated via Alternative Promoters and mRNA Stability. *Cancer Cell Int.* **2016**, *16*, 71. [CrossRef] [PubMed]

210. Ikebe, C.; Ohashi, K.; Fujimori, T.; Bernard, O.; Noda, T.; Robertson, E.J.; Mizuno, K. Mouse LIM-Kinase 2 Gene: cDNA Cloning, Genomic Organization, and Tissue-Specific Expression of Two Alternatively Initiated Transcripts. *Genomics* **1997**, *46*, 504–508. [CrossRef] [PubMed]

211. Subcellular Localization and Protein Interaction of the Human LIMK2 Gene Expressing Alternative Transcripts with Tissue-Specific Regulation. *Biochem. Biophys. Res. Commun.* **1996**, *229*, 582–589. [CrossRef]

212. Ahn, J.-Y.; Rong, R.; Kroll, T.G.; Meir, E.G.V.; Snyder, S.H.; Ye, K. PIKE (Phosphatidylinositol 3-Kinase Enhancer)-A GTPase Stimulates Akt Activity and Mediates Cellular Invasion*. *J. Biol. Chem.* **2004**, *279*, 16441–16451. [CrossRef]

213. Ahn, J.-Y.; Ye, K. PIKE GTPase Signaling and Function. *Int. J. Biol. Sci.* **2005**, *1*, 44–50. [CrossRef]

214. Han, C.; Zhao, R.; Kroger, J.; Qu, M.; Wani, A.A.; Wang, Q.-E. Caspase-2 Short Isoform Interacts with Membrane-Associated Cytoskeleton Proteins to Inhibit Apoptosis. *PLoS ONE* **2013**, *8*, e67033. [CrossRef]

215. Logette, E.; Wotawa, A.; Solier, S.; Desoche, L.; Solary, E.; Corcos, L. The Human Caspase-2 Gene: Alternative Promoters, Pre-mRNA Splicing and AUG Usage Direct Isoform-Specific Expression. *Oncogene* **2003**, *22*, 935–946. [CrossRef] [PubMed]

216. Benassayag, C.; Montero, L.; Colombié, N.; Gallant, P.; Cribbs, D.; Morello, D. Human C-Myc Isoforms Differentially Regulate Cell Growth and Apoptosis in Drosophila Melanogaster. *Mol. Cell. Biol.* **2005**, *25*, 9897–9909. [CrossRef] [PubMed]

217. Diernfellner, A.; Colot, H.V.; Dintsis, O.; Loros, J.J.; Dunlap, J.C.; Brunner, M. Long and Short Isoforms of Neurospora Clock Protein FRQ Support Temperature Compensated Circadian Rhythms. *FEBS Lett.* **2007**, *581*, 5759–5764. [CrossRef]

218. Cha, J.; Yuan, H.; Liu, Y. Regulation of the Activity and Cellular Localization of the Circadian Clock Protein FRQ. *J. Biol. Chem.* **2011**, *286*, 11469–11478. [CrossRef] [PubMed]

219. Brubaker, S.W.; Gauthier, A.E.; Mills, E.W.; Ingolia, N.T.; Kagan, J.C. A Bicistronic MAVS Transcript Highlights a Class of Truncated Variants in Antiviral Immunity. *Cell* **2014**, *156*, 800–811. [CrossRef]

220. Vazquez, C.; Beachboard, D.C.; Horner, S.M. Methods to Visualize MAVS Subcellular Localization. *Methods Mol. Biol.* **2017**, *1656*, 131–142. [CrossRef]

221. Qi, N.; Shi, Y.; Zhang, R.; Zhu, W.; Yuan, B.; Li, X.; Wang, C.; Zhang, X.; Hou, F. Multiple Truncated Isoforms of MAVS Prevent Its Spontaneous Aggregation in Antiviral Innate Immune Signalling. *Nat. Commun.* **2017**, *8*, 15676. [CrossRef] [PubMed]

222. Kadmiel, M.; Cidlowski, J.A. Glucocorticoid Receptor Signaling in Health and Disease. *Trends Pharmacol. Sci.* **2013**, *34*, 518–530. [CrossRef]

223. Oakley, R.H.; Cidlowski, J.A. Cellular Processing of the Glucocorticoid Receptor Gene and Protein: New Mechanisms for Generating Tissue-Specific Actions of Glucocorticoids. *J. Biol. Chem.* **2011**, *286*, 3177–3184. [CrossRef]

224. Jenkins, A.K.; Paterson, C.; Wang, Y.; Hyde, T.M.; Kleinman, J.E.; Law, A.J. Neurexin 1 (NRXN1) Splice Isoform Expression During Human Neocortical Development and Aging. *Mol. Psychiatry* **2016**, *21*, 701–706. [CrossRef]

225. Vieler, M.; Sanyal, S. P53 Isoforms and Their Implications in Cancer. *Cancers* **2018**, *10*, 288. [CrossRef] [PubMed]

226. Bourdon, J.-C.; Fernandes, K.; Murray-Zmijewski, F.; Liu, G.; Diot, A.; Xirodimas, D.P.; Saville, M.K.; Lane, D.P. P53 Isoforms Can Regulate P53 Transcriptional Activity. *Genes Dev.* **2005**, *19*, 2122–2137. [CrossRef] [PubMed]

227. Khoury, M.P.; Bourdon, J.-C. The Isoforms of the P53 Protein. *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a000927. [CrossRef] [PubMed]

228. Tan, L.-Y.; Whitfield, P.; Llorian, M.; Monzon-Casanova, E.; Diaz-Munoz, M.D.; Turner, M.; Smith, C.W.J. Generation of Functionally Distinct Isoforms of PTBP3 by Alternative Splicing and Translation Initiation. *Nucleic Acids Res.* **2015**, *43*, 5586–5600. [CrossRef]

229. Huang, S.N.; Dalla Rosa, I.; Michaels, S.A.; Tulumello, D.V.; Agama, K.; Khiati, S.; Jean, S.R.; Baechler, S.A.; Factor, V.M.; Varma, S.; et al. Mitochondrial Tyrosyl-DNA Phosphodiesterase 2 and Its TDP2S Short Isoform. *EMBO Rep.* **2018**, *19*, e42139. [CrossRef]

230. Graham, R.R.; Kyogoku, C.; Sigurdsson, S.; Vlasova, I.A.; Davies, L.R.L.; Baechler, E.C.; Plenge, R.M.; Koeuth, T.; Ortmann, W.A.; Hom, G.; et al. Three Functional Variants of IFN Regulatory Factor 5 (IRF5) Define Risk and Protective Haplotypes for Human Lupus. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 6758–6763. [CrossRef]

231. Abbas, W.; Kumar, A.; Herbein, G. The eEF1A Proteins: At the Crossroads of Oncogenesis, Apoptosis, and Viral Infections. *Front. Oncol.* **2015**, *5*, 75. [CrossRef]

232. Manzo, M.; Wirz, J.; Ambrosi, C.; Villaseñor, R.; Roschitzki, B.; Baubec, T. Isoform-Specific Localization of DNMT3A Regulates DNA Methylation Fidelity at Bivalent CpG Islands. *EMBO J.* **2017**, *36*, 3421–3434. [CrossRef]

233. Lax, E.; Sapozhnikov, D.M. Dnmt3a2 in the Nucleus Accumbens Shell Mediates Cue-Induced Cocaine-Seeking Behavior. *J. Neurosci.* **2019**, *39*, 2574–2576. [CrossRef]

234. Zhang, Z.; So, K.; Peterson, R.; Bauer, M.; Ng, H.; Zhang, Y.; Kim, J.H.; Kidd, T.; Miura, P. Elav-Mediated Exon Skipping and Alternative Polyadenylation of the Dscam1 Gene Are Required for Axon Outgrowth. *Cell Rep.* **2019**, *27*, 3808–3817.e7. [CrossRef]

235. Jeon, S.; Kim, Y.; Jeong, Y.M.; Bae, J.S.; Jung, C.K. CCND1 Splice Variant as A Novel Diagnostic and Predictive Biomarker for Thyroid Cancer. *Cancers* **2018**, *10*, 437. [CrossRef] [PubMed]

236. Wang, Q.; He, G.; Hou, M.; Chen, L.; Chen, S.; Xu, A.; Fu, Y. Cell Cycle Regulation by Alternative Polyadenylation of CCND1. *Sci. Rep.* **2018**, *8*, 6824. [CrossRef] [PubMed]

237. Carr, H.S.; Morris, C.A.; Menon, S.; Song, E.H.; Frost, J.A. Rac1 Controls the Subcellular Localization of the Rho Guanine Nucleotide Exchange Factor Net1A To Regulate Focal Adhesion Formation and Cell Spreading. *Mol. Cell. Biol.* **2013**, *33*, 622–634. [CrossRef] [PubMed]

238. Winter, C.; Pawel, B.; Seiser, E.; Zhao, H.; Raabe, E.; Wang, Q.; Judkins, A.R.; Attiyeh, E.; Maris, J.M. Neural cell adhesion molecule (NCAM) isoform expression is associated with neuroblastoma differentiation status. *Pediatr. Blood Cancer* **2008**, *51*, 10–16. [CrossRef]