

Review

Illuminating the Transcriptome through the Genome

David J. Elliott

Institute of Genetic Medicine, Newcastle University, Newcastle, NE1 3BZ, UK;

E-Mail: David.Elliott@ncl.ac.uk; Tel.: +44-0191-241-8694; Fax: +44-0191-241-8666

Received: 17 January 2014; in revised form: 3 March 2014 / Accepted: 5 March 2014 /

Published: 14 March 2014

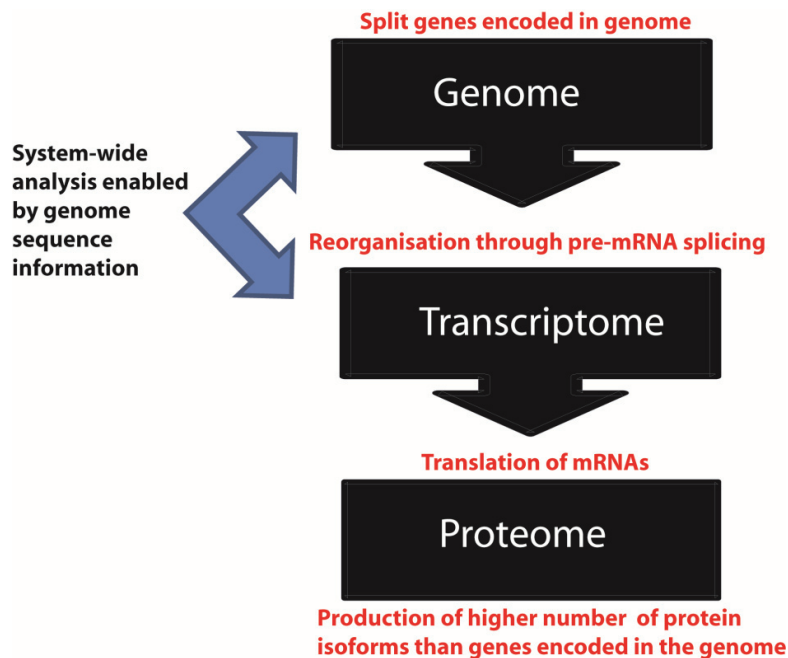
Abstract: Sequencing the human genome was a huge milestone in genetic research that revealed almost the total DNA sequence required to create a human being. However, in order to function, the DNA genome needs to be expressed as an RNA transcriptome. This article reviews how knowledge of genome sequence information has led to fundamental discoveries in how the transcriptome is processed, with a focus on new system-wide insights into how pre-mRNAs that are encoded by split genes in the genome are rearranged by splicing into functional mRNAs. These advances have been made possible by the development of new post-genome technologies to probe splicing patterns. Transcriptome-wide approaches have characterised a “splicing code” that is embedded within and has a significant role in deciphering the genome, and is deciphered by RNA binding proteins. These analyses have also found that most human genes encode multiple mRNA isoforms, and in some cases proteins, leading in turn to a re-assessment of what exactly a gene is. Analysis of the transcriptome has given insights into how the genome is packaged and transcribed, and is helping to explain important aspects of genome evolution.

Keywords: transcriptome; RNA splicing; post-genome

1. Introduction

The completion of the human genome sequence [1,2] brought together key scientific and philosophical questions, including exactly what we are as a species and individuals. However, in order to function, the genome has to be expressed. The primary expression product of the genome is RNA, and the complete set of all RNA molecules made through copying the genome into RNA is called the transcriptome (Figure 1). After transcription in the nucleus, mRNAs are translated into protein in the cytoplasm to yield the proteome while other RNAs have noncoding functions [3].

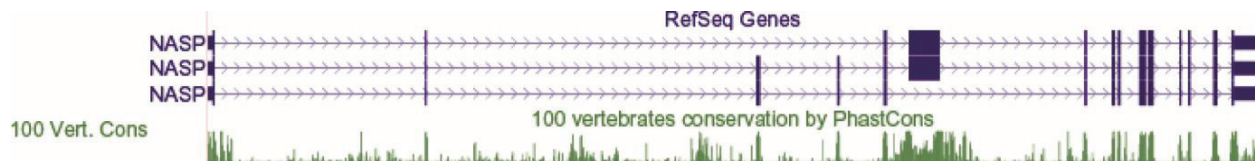
Figure 1. Information flow from the genome to the proteome. The genome represents an archive of information embedded in DNA. This information is transcribed as RNA to give the transcriptome, and then translated into protein to give the proteome.



A key feature of human (and other eukaryotic) genes is their split exon-intron structure [4,5]. Figure 2 shows the exon-intron structure of a typical human gene displayed on a genome browser [6]. The exons include the protein coding information of the gene while introns are the intervening sequences between them. The term exon refers to the fact that exon sequences are expressed in the mRNA made from the gene, as opposed to introns which are removed (intron refers to intragenic regions) [7]. The presence of introns within genes and the long intergenic sequences between genes mean that only a small fraction of the human genome is truly protein-encoding. To put some figures on this, human protein-encoding genes contribute ~33.5% of the human genome sequence [1,2] but exons alone comprise 2.94% of the genome [8]. Protein coding exons make up a smaller proportion still (1.2%) of the genome. This is because there are at least partially untranslated exons in every mRNA (some of which can have important regulatory roles), and some exons remain entirely untranslated (see below).

Genes need to be transcribed over their full length (including both introns and exons) to generate precursor mRNAs (pre-mRNAs). Transcription of long genes represents a considerable energy investment by the cell. One of the longest genes in the human genome, *DYSTROPHIN* takes in the order of 16 hours to transcribe, yet produces a final mRNA of just ~14 kb that would have just taken ~7 minutes to transcribe by itself, assuming an elongation rate of 2 kb/minute [9]. It is calculated that ~95% of RNA does not leave the nucleus [10]. Nuclear-retained RNA includes intron sequences and some long ncRNAs that are also spliced but retained in the nucleus.

Figure 2. The intron-exon structure of a typical human gene displayed on the UCSC genome browser [6]. Introns are shown as lines (the “arrowheads” in the lines indicate the direction of transcription). Exons are shown as vertical bars. Coding exons are shown as thicker vertical bars than non-coding exon sequence. This example shows the *NASP* gene. The gene structures shown are “Refseq genes” that represent known human protein-coding and non-protein-coding genes taken from the NCBI RNA reference sequences collection. Notice that this single gene locus contains three distinct Refseq annotations containing different exon structures. Conserved sequences detected by comparative genomic information from 100 vertebrate genome sequences are shown at the bottom as a Phastcons plot—the higher values are most conserved, and often correspond to exons.



The splicing reaction is catalysed by the spliceosome. While introns are generally discarded after splicing, some introns can yield functional RNAs after splicing (e.g., miRNAs) [11,12]. Spliceosomes themselves are multi-component machines containing five snRNAs and at least 200 proteins, making them one of the most complicated assemblies in the cell [13,14]. Spliceosomes assemble *de novo* around each intron to be removed. A typical gene containing eight exons would require the assembly of eight spliceosomes to create a functional mRNA. Input of energy is required for spliceosomes to properly assemble through multiple ATP-dependent RNA helicases and other energy consuming proteins (including GTPases) [15].

Prior to completion of the human genome sequence, research into splicing typically looked at single genes and exons as models. However, while these detailed studies continue to be very important, the advances in genomics catalysed by genome sequencing projects have spawned parallel advances in transcriptomics, enabling a much broader system-wide dissection of RNA processing pathways. Here I review some of these new insights. While the focus here is on pre-mRNA splicing, transcriptome-wide analyses have also been directed at other aspects of genome expression, including RNA editing, RNA stability, expression of ncRNAs, polyadenylation and translation.

2. The Human Genome Sequence Has Led to New Global Insights into the Control of Splicing

In the 1980s examination of a limited number of genes led to the identification of short conserved sequences called 5' and 3' splice sites at exon-intron junctions [16]. The availability of the human and other genome sequences have enabled these studies to be extended genome-wide [17]. Most human exons are spliced together by a single kind of major spliceosome that recognises most 5' and 3' splice sites. In addition a second minor spliceosome exists in parallel that decodes a smaller subset of intron–exon junctions [18]. This minor spliceosome has a different but overlapping complement of snRNAs to the major spliceosome. Recent transcriptome-wide data show a key snRNA component in the minor spliceosome (called U6_{ATAC}) is an important gene expression switch controlling patterns of splicing [19]. In the rest of this review the activities of the major and minor spliceosomes are not separately distinguished.

The conserved 5' and 3' splice site sequences encoded in the genome at exon-intron junctions are quite short. Furthermore, scattered within introns are short sequence elements called pseudoexons. Pseudoexons “look like” exons in that they have 5' and 3' splice sites, but are not selected as exons by the spliceosome. Estimates from model human genes suggest pre-mRNA splicing is remarkably accurate [20]. However, transcriptome-wide analyses of splicing patterns using RNAseq do detect some errors in splicing (at a rate of around 0.7% errors/intron)—these errors have been termed “noisy splicing”, and might contribute to gene and protein evolution by enabling new mRNA isoforms to be made even at low frequencies [21].

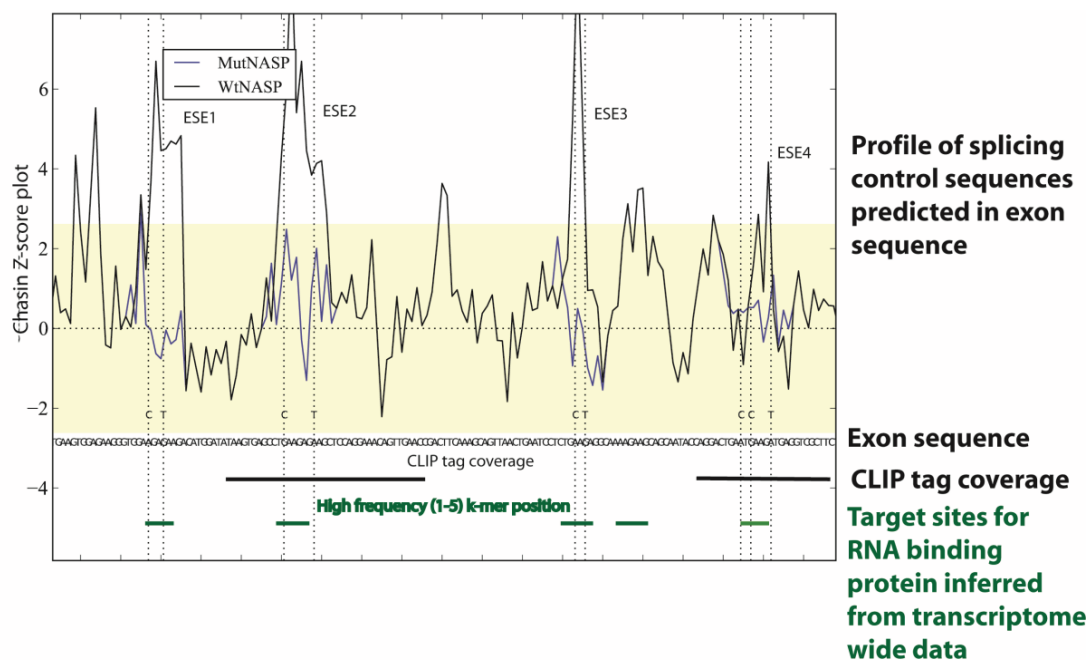
The spliceosome uses several mechanisms to accurately decode exon/intron structure using information embedded in the transcriptome. Firstly, in humans and most vertebrates, spliceosomes recognise exons from introns through a process called exon definition [22–24]. The advantage of exon definition is that since exons are quite small they should be easier to identify as discrete units compared most (considerably longer) vertebrate introns. In exon definition, early spliceosome components bind to the pre-mRNA and “flag” exons to be spliced together. Following exon definition, pairs of exons are then joined together by splicing which removes the intervening intron sequences. Amongst the early binding components of the spliceosome involved in exon definition are the U1 snRNP RNA-protein complex that recognises the 5' splice site, and a protein dimer which recognises 3' splice sites called U2AF (U2AF65 and U2AF35 are the two proteins in the dimer).

A second mechanism that facilitates accurate decoding of the genome is the presence of a splicing code that helps to differentiate between exons and introns in pre-mRNA. Before the sequencing of the human and other eukaryotic genomes, the important sequences controlling splicing of an exon were usually worked out on a gene by gene basis, using a finite number of model exons. The importance of exon sequences outside the splice junctions for splicing were first identified in pioneering experiments using model exons in the *FIBRONECTIN* and β globin genes [25,26]. It is now known that pre-mRNAs each contain multiple short nucleotide sequences that can enhance or silence splicing of their associated exons [27–29]. Exon sequences can function as exonic splicing enhancers (abbreviated ESEs) that help the spliceosome to recognise exons, or exonic splicing silencers (abbreviated ESSs) that inhibit spliceosome recognition by the spliceosome. Similarly, intron associated sequences can function as intronic splicing enhancers (abbreviated ISE) or Intronic Splicing Silencers (abbreviated ISSs). Splicing enhancers are bound by proteins or complexes of proteins, including the SR proteins that contain domains enriched in serine and arginine residues, and splicing silencers are frequently bound by heterogeneous ribonucleoproteins (abbreviated hnRNPs).

The availability of genome sequences have allowed system-wide approaches to identify splicing enhancers and silencers that control splicing and led to significant insights into the “splicing code” [29]. These approaches have included machine learning approaches to utilize hundreds of features in pre-mRNAs including motifs bound by RNA binding proteins and RNA secondary structure predictions to predict *in vivo* splicing decisions [30]. Computer programmes have also been devised that can computationally predict positions of predicted splicing enhancers and silencers and the target sequences of RNA binding proteins in an input genomic sequence (e.g., Figure 3) [30–33]. The combination of these system-wide experimental and bioinformatic analyses show the splicing code is maintained as nucleotide information in the genome. The splicing code has similar importance to the genetic code that is deciphered to read amino acid sequences from mRNAs. However, the splicing

code is much more complex than the genetic code. While the genetic code uses triplet codons to specify amino acids, in the splicing code multiple sequence elements act in combination to decipher the exon/intron structure of pre-mRNAs [30].

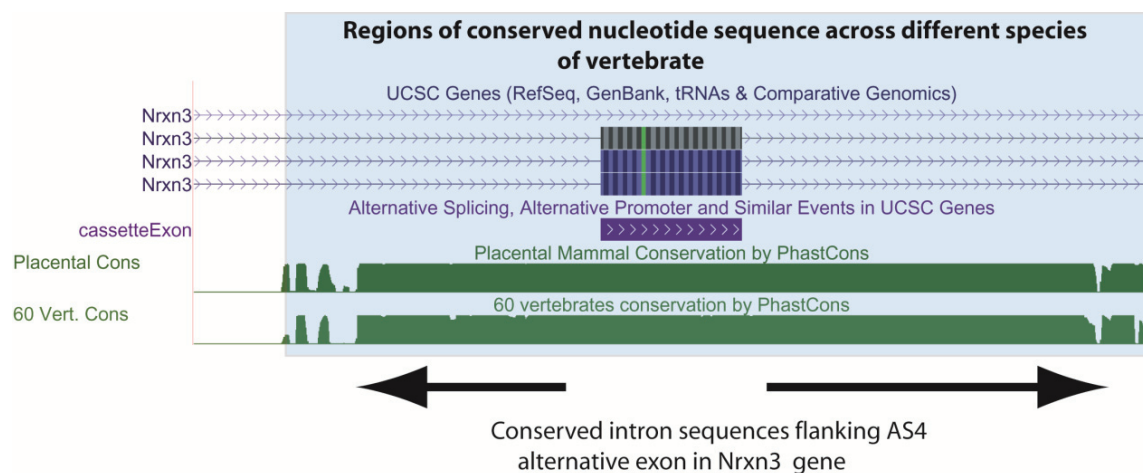
Figure 3. Transcriptome-wide data can be used to predict the splicing code in specific genes. In this example sequences within a cassette exon in the mouse *NASP* gene have been analysed using genome and transcriptome wide datasets to pinpoint splicing control sequences. Firstly a Chasin Z-score plot was used that can identify sequences predictive of exonic splicing enhancers and silencers [32,34]: the four exonic splicing enhancer (ESE) sequences identified are shown as peaks in the plot above the sequence. In this example, these ESE sequences were individually mutated to test function in minigenes (the Chasin profiles of the mutants M1-M4 are shown compared to the wild type sequence: notice the change in the Z-score plot removes predicted ESE activity for each mutant). The positions of these ESEs mapped to binding sites for the splicing factor Tra2 β , both individually identified using cross linking immunoprecipitation (CLIP) in the mouse testis and predicted from the *in vivo* binding site generated from transcriptome-wide Tra2 β binding data from the mouse testis. This figure is adapted from [32].



Because they are needed for exons to be spliced into mRNAs, ESE sequences have been maintained in exons as well as the codons that specify amino acid sequences [35]. The intronic sequences that flank exons are often also strongly conserved between species (Figure 4 shows as an example conserved nucleotide sequences flanking an exon in the mouse *Neurexin3* gene downloaded from the UCSC genome browser). Comparison of the human and mouse genomes [1,36] which diverged 75 million years ago show that alternative exons are usually flanked by much longer stretches of conserved intron sequences than constitutive exons, consistent with more elaborate control mechanisms [37,38]. Conservation in these exon flanking intron sequences in some cases is much

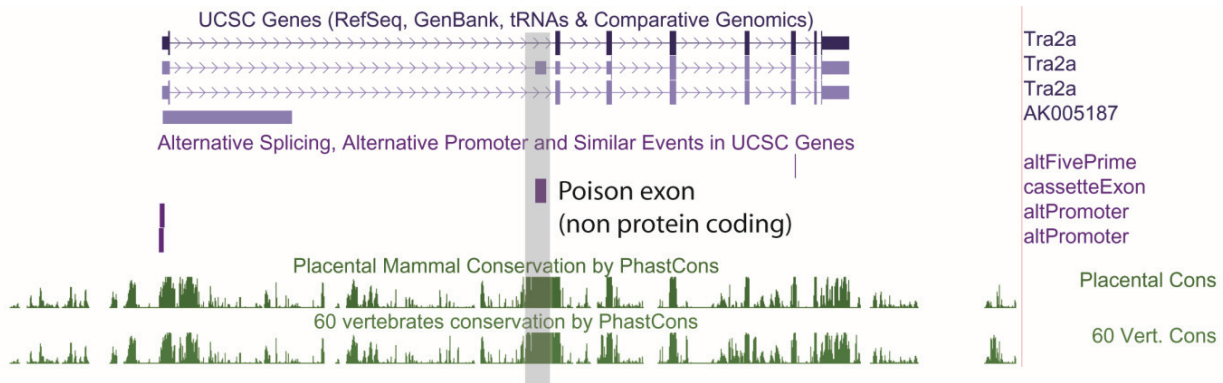
higher than in promoters, suggesting that one of the main functions of conserved noncoding sequences between mouse and human is the regulation of alternative splicing [37]. Even non-protein coding exons can be highly conserved in the genome. Noncoding exons include highly conserved “poison exons” (for example see Figure 5), that when included insert premature translational termination codons and lead to mRNA decay [32,39]. “Poison exons” are very important for auto-regulation of RNA binding proteins that control splicing [39,40]. Together these studies show the maintenance of splicing control sequences has had a significant impact in constraining genome evolution.

Figure 4. A functional requirement to maintain splicing control sequences constrains evolution of the genome. This screenshot is downloaded from the UCSC genome browser [6] and shows conserved intron sequences flanking the alternatively spliced AS4 exon from the mouse *Nrxn3* gene. The conserved sequences are shaded. At the top, the UCSC gene annotations show that this cassette exon is included in $\frac{3}{4}$ mRNA isoforms made from this gene. At the bottom the Phastcons plot shows that the flanking intron sequences are also highly conserved as well as the exon sequence (exons might be conserved because of their protein-coding content). Conservation of these intron sequences are likely important to control tissue-specific splicing of this exon by the spliceosome. Known alternative events are annotated on the UCSC track “Alt events”, and can be shown alongside the gene structure (here the cassette exon is annotated in the alt events track, and is in purple).



Genome sequences have been used to help develop technologies aimed to globally dissect RNA processing pathways [41]. These technologies can identify the target sites of RNA binding proteins transcriptome-wide. In cross linking immunoprecipitation (abbreviated CLIP) experiments RNA binding proteins are cross-linked *in situ* to their target RNAs within intact cells using ultra violet irradiation, followed by immunoprecipitation of the RNA protein complexes and amplification by PCR [42–44]. Once unique target sites are identified by next generation sequencing (these are called CLIP tags), these can be mapped onto genome sequences to reveal the initial binding sites in the transcriptome. Transcriptome-wide CLIP analyses have enabled maps to be drawn of the target sites of RNA binding proteins relative to regulated exons, and these maps can be used to predict mechanisms of splicing control [45]. For example, CLIP tags of the RNA binding protein Tra2 β that is needed for splicing inclusion for a regulated cassette exon in the NASP gene are shown in Figure 3 [32].

Figure 5. Non protein coding exons are conserved in the genome. Some non-coding exons are highly conserved in the human genome and play important roles in controlling the expression levels of splicing regulator proteins. The *TRA2A* gene encodes the splicing regulator protein Tra2 α and contains a poison exon that contains multiple stop codons and is only inserted into some mRNAs. Despite not containing coding information, the *TRA2A* poison exon is highly conserved across species (indicated by the Phastcons score). Notice that the *TRA2A* gene encodes also alternative 5' splice sites and uses alternative promoters. This screenshot was downloaded from the UCSC genome browser [6].



The completion of the human genome sequence enabled the development of comprehensive microarrays to interrogate gene expression. These global techniques include the development of splice-sensitive microarrays. These microarrays either detect specific exons in the transcriptome or the use of specific splice junctions and report splicing patterns in mRNA [46,47]. Transcriptome-wide patterns of alternative splicing can also be detected by RNAseq [48,49]. These technologies have been used to search for exons mis-spliced after depletion of particular RNA binding proteins from cells. By analysing thousands of exons in parallel the splicing events that are regulated by specific RNA binding proteins can be comprehensively identified. These experiments have shown some individual RNA binding proteins bind to and regulate similar mRNAs. For example, the NOVA protein regulates splicing of a functionally coherent set of genes involved in synapse function [50]. Other splicing regulators might similarly have coherent RNA functional targets, including T-STAR and SAM68 which regulate regional alternative splicing of the synaptic neurexin genes in the brain [51,52].

Knowing which splicing events are regulated by what proteins at a global level has been used to derive general rules. For example, binding of the NOVA proteins upstream of exons tends to block splicing, while binding of the same proteins downstream of exons enhance splicing [44,45]. These rules governing the RNA motifs bound by NOVA and their role in activating or repressing associated exon inclusion are conserved between flies and mammals, although the actual target mRNAs are different [53]. Similar rules have also been uncovered for PTB and some other splicing regulator proteins [45,54,55]. Recent developments to understand the splicing code have compared binding maps for different RNA binding proteins, and show that some functionally collaborate with each other to generate tissue specific splicing patterns [56].

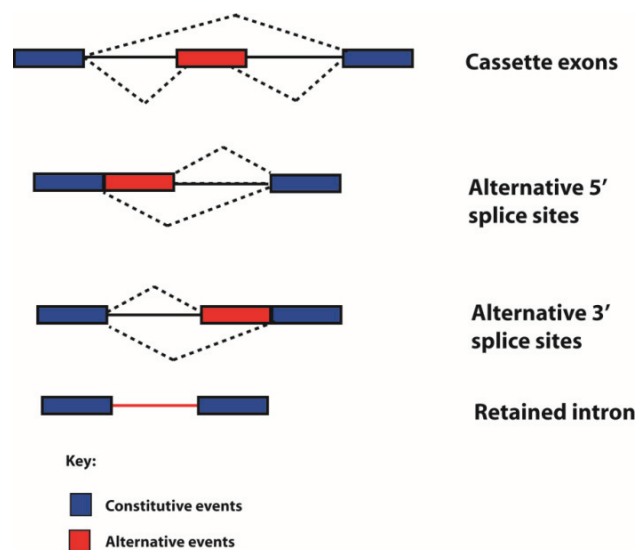
Transcriptome-wide insights have also revealed the involvement of RNA binding proteins in other aspects of genome biology. Alu sequences are retro-transposable elements that frequently insert

into introns, and have sequence similarities to exons [57]. There are over 15,000 Alu sequences in the human genome, many of which are inserted into introns. The RNA binding protein hnRNP C has an important role for in protecting the transcriptome from potential mis-splicing caused by the insertion of Alu retrotransposable elements into genomic introns [58]. Depletion of hnRNPC leads to aberrant inclusion of around 1000 Alu-derived exons into the transcriptome. Alu sequences contain polypyrimidine tracts that potentially bind U2AF65, leading to them being aberrantly included into mRNAs, but this splicing is blocked by hnRNPC.

3. Analysis of the Human Transcriptome Led to the Realisation That Most Human Genes Encode Alternatively Spliced mRNAs

Historically genes have been defined in different ways by different people at different times. Following the human genome sequencing project, the fundamental definition of what a gene actually is has been enriched, and to some extent clouded, by comparison of genome and transcriptome sequences [59,60]. Alternative mRNA isoforms can originate from the same genetic loci through use of alternative promoters and polyadenylation sites, and by the process of alternative splicing through which exons can be spliced into different combinations to give multiple mRNAs. Hence a “single gene” can encode multiple products. Alternative splicing fits into four different categories, depending on how variable splice junctions are utilised (Figure 6). In the simplest form of alternative splicing, called exon skipping, whole exons are either spliced into the mRNA or skipped (ignored by the spliceosome). Exon sequences can also be spliced into mRNA using different splice sites (alternative 5' splice site or 3' splice sites can be selected). Whole introns can also be left in the mRNA (this is called intron retention). Transcriptome-wide analyses show that in humans exon skipping is the most frequent form of alternative splicing, and intron retention the least frequent [61].

Figure 6. Types of alternative splicing events detected in the human transcriptome. Exons are shown as boxes, introns as lines, and splicing patterns as broken lines. mRNAs can be made from individual genes can using multiple alternative events, including different types of splicing, to build up complex patterns of alternative spliced mRNAs.



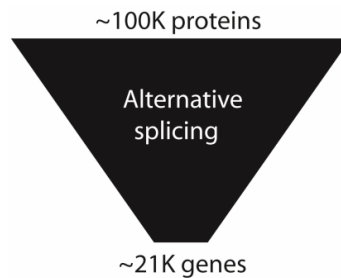
Evidence provided by comparative genome and transcriptome sequences has shown alternative splicing to be extremely frequent. Before the human genome was published, random sequencing of human mRNAs from their 3' ends using oligo dT priming suggested ~40% of human genes encoded alternatively spliced mRNAs [62,63]. During the human genome sequencing project, reconstruction of mRNAs from the gene rich chromosomes 19 and 22 upped this estimate to 59% of genes encode alternative mRNAs, with 2–3 mRNA isoforms made/gene [1]. The use of microarrays to detect global patterns of alternative splicing indicated 73%–74% of human genes express alternatively spliced mRNA isoforms [46,64]. The most recently reported RNAseq analyses of the human transcriptome is consistent with ~95% of multi-exonic genes expressing variant mRNA isoforms, with a plateau of 10–11 isoforms/gene/cell line [65]. Usually one or two major mRNA isoforms predominate in a given cell line so many cell types will just express one major mRNA isoform [60,65,66]. 86% of genes have a minor isoform frequency of 15% or more, and more than 50% of alternative exons are tissue specific in expression [48,67]. Alternative events are now annotated on genome browsers like the UCSC genome browser (e.g., Figure 4) [6].

4. To What Extent Can Human Complexity Be Ascribed to Alternative Splicing?

Proteins make up large components of human bodies. Prior to the genome sequence estimates of human gene numbers went as high as 100,000. An initial surprise from the human genome sequence was a much lower protein coding gene number, initially counted at around 23,000 [1,2]. The most recent gene counts from the ENCODE consortia suggest only 20,687 human protein coding genes [8]. The number of proteins expressed in a human cell is in contrast estimated to be ~100,000 [68,69]. This protein number represents an amplification factor of 5-fold compared to the counted number of genes. This total gene number in humans does not seem to be exceptionally higher than seemingly less sophisticated organisms. The genome of *Haemophilus influenza* contains 1743 predicted genes, the yeast *Saccharomyces cerevisiae* ~6000 genes, *Drosophila melanogaster* ~13,600 genes, and nematode worms 18,425 genes [70–72].

This apparent failure of gene numbers to correlate with complexity has been called the gene number paradox, and counted as one of the major surprises from the human genome sequence. To what extent might alternative splicing and the resulting expansion in protein coding information help explain developmental and physiological complexity in humans (Figure 7)? Until recently this question has been difficult to address, since the higher number of mRNA and EST sequences available from humans made comparisons of alternative splicing frequency with other species biased. However, a recent modencode project based on RNAseq analysis to look at alternative splicing in *C. elegans* found ~25% alternative splicing in 5000 genes, with around 30% of these being alternatively spliced between different developmental stages [73]. In depth experiments using RNAseq and tiling arrays do show a lower frequency (60.7%) of genes in the fruitfly are alternatively spliced than in humans, often in a developmental or sex-specific fashion [74].

Figure 7. Alternative splicing amplifies genome information. Alternative splicing amplifies information in the human transcriptome relative to the genome.



Hence amongst multicellular animals investigated in detail, humans do seem to exhibit higher levels of alternative splicing, which might lend credence to the idea that alternative splicing may be a factor contributing to human sophistication. If phylogenetic differences in alternative splicing frequency correlate with complexity, one might expect a decreased level in single celled organisms compared with multicellular organisms. On the one level less introns are found in the single celled baker's yeast *Saccharomyces cerevisiae*: only 5% of genes contain introns in this yeast (290/6000 genes). However, these lower intron numbers are a bit misleading. The reason for a low overall intron number in this yeast is that many introns have been lost because of reverse transcriptase activity converting mRNAs into cDNAs, which then re-integrate into the genome through high levels of homologous recombination replacing originally intron containing genes [75]. While intron-containing genes generally are rare in yeast, alternative splicing of some of these introns are specifically utilised to control developmental timing of during meiosis that takes place under conditions of limiting nutrients [76,77] so alternative splicing is used to control a complex stage in the lifecycle of this single celled organism. Taken as a whole, it is difficult to draw general correlations between overall frequencies of alternative splicing and organism sophistication.

Alternative splicing patterns can also evolve rapidly and sometimes differ between closely related species. For example, despite almost 99% identity in protein coding information, comparative transcriptome analyses have shown that 6%–8% of alternative spliced exons have different inclusion patterns between humans and chimps. These observations are consistent with alternative splicing contributing to species-species differences, and transcriptomes being more distinct between species than protein coding information [78]. Although the major conclusion from evolutionary comparisons is that much alternative splicing is not conserved between species, comparative genomics show some alternative exons have been highly conserved during evolution [79–81]. These include the highly conserved “AS4” alternative exon in the *Neurexin3* gene (abbreviated *Nrxn3*, Figure 4) that is conserved across the vertebrate lineage [51]. The *Neurexin* genes have been linked with autism and schizophrenia, and mice genetically engineered to be unable to regulate this AS4 alternative exon in the *Nrxn3* gene have different synaptic activity in the brain [82].

Post-genome analyses have also addressed the question what alternative splicing does. Protein sequences encoded by alternatively spliced exons are frequently involved in protein-protein interaction networks and contain signalling domains [83,84]. Some alternatively spliced mRNAs encode proteins with clearly different functional activity. For example, different mRNA splice isoforms encoding the FOXP1 transcription factor are made between neural stem cells and differentiated cells, and these

encode proteins that activate different promoters [85]. Alternative splicing regulators have been implicated in human cognitive diseases like autism [86]. Different groups of genes are regulated by alternative splicing from those regulated by transcription [87].

Alternative splicing pathways have been shown to have important roles in controlling development, and some human diseases are caused by defects in alternative splicing including the multi-system disorder myotonic dystrophy [87]. However, individual RNA binding proteins likely regulate coordinated splicing programmes of many target exons, and each of these individual exons might only have somewhat subtle biological contributions. Furthermore, post-genome technologies are also starting to introduce a note of caution in the interpretation of high levels of alternative splicing. Some lower abundance splice variants might be non-functional isoforms that occur as a result of low frequency events mistakes in the splicing process if they are neither evolutionarily conserved nor protein-coding [21,88]. Hence the frequency of functional alternative splicing is likely to be lower than the total frequency of all alternative splicing events.

5. Human Genome Packaging into Chromatin Correlates with Its Intron/Exon Structure

Post genome analyses have shown that the genome and transcriptome are intimately linked. In particular the exon/intron structure of genes correlates with how the genome is packaged. Within the nucleus the genome is wrapped around protein complexes called nucleosomes to form chromatin. Each nucleosome is itself made up of eight positively charged histone molecules that can be modified by the addition of small chemical groups (typically methyl and acetyl groups) [89]. Packaging of the genome within chromatin is important to enable storage of chromosomes within the comparatively small space afforded by the nuclear volume.

After experimental treatment of chromatin with the enzyme DNase I, the DNA sequences wrapped around nucleosomes remain protected, while the DNA linkages joining nucleosomes together become degraded. Genome-wide analysis of sequences protected from DNase I digestion in humans and other species indicate a 1.5 fold enrichment of exon sequences over nucleosomes compared to intron sequences [90,91]. This means that in chromatin exons are preferentially (but not exclusively) associated with nucleosomes. This is likely to have a biochemical explanation: exon sequences are GC-enriched, while their flanking intron sequences are AT-rich. Nucleosomes interact more strongly with GC-rich sequences, which likely help anchor exons to nucleosomes. The association of exons with nucleosomes has in turn had important implications for genome evolution. A length of ~150 nucleotides of DNA is needed to wrap around a nucleosome, and this is also the average size of an exon. Hence nucleosome wrapping has placed an evolutionary constraint on exon size, while in contrast introns have been able to expand in size to thousands of nucleotides.

6. Most Splicing Occurs Co-Transcriptionally

Another important connection between the genome and the transcriptome is their physical proximity during important RNA processing steps. In several species including humans much pre-mRNA splicing has been shown to take place co-transcriptionally [92]. This means that “full length” pre-mRNA copies of genes are not made. Instead processing takes place as pre-mRNAs are produced on nascent pre-mRNAs still attached to RNA polymerases engaged in transcription. Deep sequencing

of fractionated nascent RNA in human cell lines and total RNA in the brain show that exons located more upstream in genes are most likely to be spliced on nascent RNA [65,93,94]. Transcription of the genome also functionally depends on components more “traditionally” thought to be involved in splicing. Transcriptome-wide analysis of gene expression following depletion of U1 snRNP (a component of the spliceosome that recognises 5' splice sites) has shown that high nuclear concentrations of U1 snRNP are needed to prevent premature intragenic polyadenylation at sites upstream of the proper 3' boundary of genes [95].

The fact that splicing takes place on chromatin during ongoing transcription has important implications for alternative splicing patterns. Single molecule experiments have shown that nucleosomes slow down the progress of transcription [96,97]. This means that the time taken to traverse nucleosomes can provide pauses in RNA polymerase II elongation, allowing the spliceosome a window to assemble on nascent pre-mRNA. Nucleosomes have thus been described as “speed bumps” [98–101]. Interestingly only true exons, and not pseudoexons, are associated with nucleosomes [90,91]. Exon association with nucleosomes is likely to help in their recognition by the spliceosome. Pauses in transcriptional elongation on bona fide exons might facilitate spliceosome assembly before potentially competing splice sites in downstream pseudoexons can be transcribed. Changes in chromatin structure can locally speed up or slow down transcription within genes and be important for controlling alternative splicing [99]. Local transcriptional pauses would provide spliceosomes a longer “window” of time in which to assemble and sometimes even carry out splicing of an exon before a competing downstream splice site appears [102,103]. Faster elongating RNA polymerase II enzymes would give spliceosomes less time to assemble on nascent pre-mRNA before competing exons were transcribed, and so would favour exon skipping. In some cases changes in histone modification can recruit or stabilise RNA binding proteins which regulate splicing of the pre-mRNAs made from the gene [104]. The interactions between the different processes in gene expression have been recently reviewed in [105].

7. Conclusions

The human genome sequence has provided a catalyst for understanding the transcriptome. System-wide approaches of the transcriptome have led to a fuller appreciation of how genomes work including how the human genome operates with a finite gene number and providing a system wide view of RNA processing.

Acknowledgments

I thank Julian Venables for comments on the manuscript. This work was supported by the Wellcome Trust (grant numbers WT080368MA and WT089225/Z/09/Z to DJE) and the BBSRC (grant numbers BB/D013917/1 and BB/I006923/1 to DJE).

Author Contributions

DJE wrote the manuscript.

Conflicts of Interest

The author declares no conflict of interest.

References

1. Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; *et al.* Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860–921.
2. 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.
3. Fatica, A.; Bozzoni, I. Long non-coding RNAs: New players in cell differentiation and development. *Nat. Rev. Genet.* **2014**, *15*, 7–21.
4. 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.
5. Chow, L.T.; Gelinis, R.E.; Broker, T.R.; Roberts, R.J. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **1977**, *12*, 1–8.
6. Karolchik, D.; Barber, G.P.; Casper, J.; Clawson, H.; Cline, M.S.; Diekhans, M.; Dreszer, T.R.; Fujita, P.A.; Guruvadoo, L.; Haeussler, M.; *et al.* The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res.* **2014**, *42*, D764–D770.
7. Gilbert, W. Why genes in pieces? *Nature* **1978**, *271*, 501.
8. Dunham, I.; Kundaje, A.; Aldred, S.F.; Collins, P.J.; Davis, C.; Doyle, F.; Epstein, C.B.; Frietze, S.; Harrow, J.; Kaul, R.; *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74.
9. Tennyson, C.N.; Klamut, H.J.; Worton, R.G. The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nat. Genet.* **1995**, *9*, 184–190.
10. Jackson, D.A.; Pombo, A.; Iborra, F. The balance sheet for transcription: An analysis of nuclear RNA metabolism in mammalian cells. *FASEB J.* **2000**, *14*, 242–254.
11. Okamura, K.; Hagen, J.W.; Duan, H.; Tyler, D.M.; Lai, E.C. The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* **2007**, *130*, 89–100.
12. Ruby, J.G.; Jan, C.H.; Bartel, D.P. Intronic microRNA precursors that bypass Drosha processing. *Nature* **2007**, *448*, 83–86.
13. Jurica, M.S.; Moore, M.J. Pre-mRNA splicing: Awash in a sea of proteins. *Mol. Cell* **2003**, *12*, 5–14.
14. Wahl, M.C.; Will, C.L.; Luhrmann, R. The spliceosome: Design principles of a dynamic RNP machine. *Cell* **2009**, *136*, 701–718.
15. Brow, D.A. Allosteric cascade of spliceosome activation. *Annu. Rev. Genet.* **2002**, *36*, 333–360.
16. Mount, S.M. A catalogue of splice junction sequences. *Nucleic Acids Res.* **1982**, *10*, 459–472.
17. Lim, L.P.; Burge, C.B. A computational analysis of sequence features involved in recognition of short introns. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 11193–11198.

18. Patel, A.A.; Steitz, J.A. Splicing double: Insights from the second spliceosome. *Nat. Rev. Mol. Cell. Biol.* **2003**, *4*, 960–970.
19. Younis, I.; Dittmar, K.; Wang, W.; Foley, S.W.; Berg, M.G.; Hu, K.Y.; Wei, Z.; Wan, L.; Dreyfuss, G. Minor introns are embedded molecular switches regulated by highly unstable U6atac snRNA. *Elife* **2013**, *2*, e00780.
20. Fox-Walsh, K.L.; Hertel, K.J. Splice-site pairing is an intrinsically high fidelity process. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 1766–1771.
21. Pickrell, J.K.; Pai, A.A.; Gilad, Y.; Pritchard, J.K. Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet.* **2010**, *6*, e1001236.
22. Berget, S.M. Exon recognition in vertebrate splicing. *J. Biol. Chem.* **1995**, *270*, 2411–2414.
23. Black, D.L. Finding splice sites within a wilderness of RNA. *RNA* **1995**, *1*, 763–771.
24. Robberson, B.L.; Cote, G.J.; Berget, S.M. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* **1990**, *10*, 84–94.
25. Mardon, H.J.; Sebastio, G.; Baralle, F.E. A role for exon sequences in alternative splicing of the human fibronectin gene. *Nucleic Acids Res.* **1987**, *15*, 7725–7733.
26. Reed, R.; Maniatis, T. A role for exon sequences and splice-site proximity in splice-site selection. *Cell* **1986**, *46*, 681–690.
27. Black, D.L. Protein diversity from alternative splicing: A challenge for bioinformatics and post-genome biology. *Cell* **2000**, *103*, 367–370.
28. Maniatis, T.; Tasic, B. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **2002**, *418*, 236–243.
29. Wang, Z.; Burge, C.B. Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA* **2008**, *14*, 802–813.
30. Barash, Y.; Calarco, J.A.; Gao, W.; Pan, Q.; Wang, X.; Shai, O.; Blencowe, B.J.; Frey, B.J. Deciphering the splicing code. *Nature* **2010**, *465*, 53–59.
31. Schwartz, S.; Hall, E.; Ast, G. SROOGLE: Webserver for integrative, user-friendly visualization of splicing signals. *Nucleic Acids Res.* **2009**, *37*, W189–W192.
32. Grellscheid, S.; Dalglish, C.; Storbeck, M.; Best, A.; Liu, Y.; Jakubik, M.; Mende, Y.; Ehrmann, I.; Curk, T.; Rossbach, K.; *et al.* Identification of evolutionarily conserved exons as regulated targets for the splicing activator tra2beta in development. *PLoS Genet.* **2011**, *7*, e1002390.
33. Paz, I.; Akerman, M.; Dror, I.; Kosti, I.; Mandel-Gutfreund, Y. SFmap: A web server for motif analysis and prediction of splicing factor binding sites. *Nucleic Acids Res.* **2010**, *38*, W281–W285.
34. Zhang, X.H.; Chasin, L.A. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev.* **2004**, *18*, 1241–1250.
35. Caceres, E.F.; Hurst, L.D. The evolution, impact and properties of exonic splice enhancers. *Genome Biol.* **2013**, *14*, R143.
36. Waterston, R.H.; Lindblad-Toh, K.; Birney, E.; Rogers, J.; Abril, J.F.; Agarwal, P.; Agarwala, R.; Ainscough, R.; Alexandersson, M.; An, P.; *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **2002**, *420*, 520–562.

37. Sorek, R.; Ast, G. Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Res.* **2003**, *13*, 1631–1637.
38. Sugnet, C.W.; Srinivasan, K.; Clark, T.A.; O'Brien, G.; Cline, M.S.; Wang, H.; Williams, A.; Kulp, D.; Blume, J.E.; Haussler, D.; *et al.* Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS Comput. Biol.* **2006**, *2*, e4.
39. Lareau, L.F.; Inada, M.; Green, R.E.; Wengrod, J.C.; Brenner, S.E. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **2007**, *446*, 926–929.
40. Ni, J.Z.; Grate, L.; Donohue, J.P.; Preston, C.; Nobida, N.; O'Brien, G.; Shiue, L.; Clark, T.A.; Blume, J.E.; Ares, M., Jr. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* **2007**, *21*, 708–718.
41. Buratti, E.; Baralle, M.; Baralle, F.E. From single splicing events to thousands: The ambiguous step forward in splicing research. *Brief. Funct. Genomics* **2013**, *12*, 3–12.
42. Licatalosi, D.D.; Mele, A.; Fak, J.J.; Ule, J.; Kayikci, M.; Chi, S.W.; Clark, T.A.; Schweitzer, A.C.; Blume, J.E.; Wang, X.; Darnell, J.C.; Darnell, R.B. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **2008**, *456*, 464–469.
43. Konig, J.; Zarnack, K.; Rot, G.; Curk, T.; Kayikci, M.; Zupan, B.; Turner, D.J.; Luscombe, N.M.; Ule, J. iCLIP—Transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *J. Vis. Exp.* **2011**, *50*, 2638.
44. Anko, M.L.; Neugebauer, K.M. RNA-protein interactions *in vivo*: Global gets specific. *Trends Biochem. Sci.* **2012**, *37*, 255–262.
45. Witten, J.T.; Ule, J. Understanding splicing regulation through RNA splicing maps. *Trends Genet.* **2011**, *27*, 89–97.
46. Johnson, J.M.; Castle, J.; Garrett-Engele, P.; Kan, Z.; Loerch, P.M.; Armour, C.D.; Santos, R.; Schadt, E.E.; Stoughton, R.; Shoemaker, D.D. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **2003**, *302*, 2141–2144.
47. Pan, Q.; Shai, O.; Misquitta, C.; Zhang, W.; Saltzman, A.L.; Mohammad, N.; Babak, T.; Siu, H.; Hughes, T.R.; Morris, Q.D.; *et al.* Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. *Mol. Cell* **2004**, *16*, 929–941.
48. Pan, Q.; Shai, O.; Lee, L.J.; Frey, B.J.; Blencowe, B.J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **2008**, *40*, 1413–1415.
49. Sultan, M.; Schulz, M.H.; Richard, H.; Magen, A.; Klingenhoff, A.; Scherf, M.; Seifert, M.; Borodina, T.; Soldatov, A.; Parkhomchuk, D.; *et al.* A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **2008**, *321*, 956–960.
50. Ule, J.; Ule, A.; Spencer, J.; Williams, A.; Hu, J.S.; Cline, M.; Wang, H.; Clark, T.; Fraser, C.; Ruggiu, M.; *et al.* Nova regulates brain-specific splicing to shape the synapse. *Nat. Genet.* **2005**, *37*, 844–852.

51. Ehrmann, I.; Dalglish, C.; Liu, Y.; Danilenko, M.; Crosier, M.; Overman, L.; Arthur, H.M.; Lindsay, S.; Clowry, G.J.; Venables, J.P.; *et al.* The tissue-specific RNA binding protein T-STAR controls regional splicing patterns of neurexin pre-mRNAs in the brain. *PLoS Genet.* **2013**, *9*, e1003474.
52. Iijima, T.; Wu, K.; Witte, H.; Hanno-Iijima, Y.; Glatter, T.; Richard, S.; Scheiffele, P. SAM68 regulates neuronal activity-dependent alternative splicing of neurexin-1. *Cell* **2011**, *147*, 1601–1614.
53. Brooks, A.N.; Yang, L.; Duff, M.O.; Hansen, K.D.; Park, J.W.; Dudoit, S.; Brenner, S.E.; Graveley, B.R. Conservation of an RNA regulatory map between Drosophila and mammals. *Genome Res.* **2011**, *21*, 193–202.
54. Llorian, M.; Schwartz, S.; Clark, T.A.; Hollander, D.; Tan, L.Y.; Spellman, R.; Gordon, A.; Schweitzer, A.C.; de la Grange, P.; Ast, G.; *et al.* Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. *Nat. Struct. Mol. Biol.* **2010**, *17*, 1114–1123.
55. Xue, Y.; Zhou, Y.; Wu, T.; Zhu, T.; Ji, X.; Kwon, Y.S.; Zhang, C.; Yeo, G.; Black, D.L.; Sun, H.; *et al.* Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol. Cell* **2009**, *36*, 996–1006.
56. Zhang, C.; Frias, M.A.; Mele, A.; Ruggiu, M.; Eom, T.; Marney, C.B.; Wang, H.; Licatalosi, D.D.; Fak, J.J.; Darnell, R.B. Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. *Science* **2010**, *329*, 439–443.
57. Sorek, R.; Lev-Maor, G.; Reznik, M.; Dagan, T.; Belinky, F.; Graur, D.; Ast, G. Minimal conditions for exonization of intronic sequences: 5' splice site formation in Alu exons. *Mol. Cell* **2004**, *14*, 221–231.
58. Zarnack, K.; Konig, J.; Tajnik, M.; Martincorena, I.; Eustermann, S.; Stevant, I.; Reyes, A.; Anders, S.; Luscombe, N.M.; Ule, J. Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* **2013**, *152*, 453–466.
59. Gerstein, M.B.; Bruce, C.; Rozowsky, J.S.; Zheng, D.; Du, J.; Korbel, J.O.; Emanuelsson, O.; Zhang, Z.D.; Weissman, S.; Snyder, M. What is a gene, post-ENCODE? History and updated definition. *Genome Res.* **2007**, *17*, 669–681.
60. Mudge, J.M.; Frankish, A.; Harrow, J. Functional transcriptomics in the post-ENCODE era. *Genome Res.* **2013**, *12*, 1961–1973.
61. Ast, G. How did alternative splicing evolve? *Nat. Rev. Genet.* **2004**, *5*, 773–782.
62. Brett, D.; Hanke, J.; Lehmann, G.; Haase, S.; Delbruck, S.; Krueger, S.; Reich, J.; Bork, P. EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. *FEBS Lett.* **2000**, *474*, 83–86.
63. Mironov, A.A.; Fickett, J.W.; Gelfand, M.S. Frequent alternative splicing of human genes. *Genome Res.* **1999**, *9*, 1288–1293.
64. Clark, T.A.; Schweitzer, A.C.; Chen, T.X.; Staples, M.K.; Lu, G.; Wang, H.; Williams, A.; Blume, J.E. Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biol.* **2007**, *8*, R64.

65. Djebali, S.; Davis, C.A.; Merkel, A.; Dobin, A.; Lassmann, T.; Mortazavi, A.; Tanzer, A.; Lagarde, J.; Lin, W.; Schlesinger, F.; *et al.* Landscape of transcription in human cells. *Nature* **2012**, *489*, 101–108.
66. Gonzalez-Porta, M.; Frankish, A.; Rung, J.; Harrow, J.; Brazma, A. Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. *Genome Biol.* **2013**, *14*, R70.
67. Wang, E.T.; Sandberg, R.; Luo, S.; Khrebtkova, I.; Zhang, L.; Mayr, C.; Kingsmore, S.F.; Schroth, G.P.; Burge, C.B. Alternative isoform regulation in human tissue transcriptomes. *Nature* **2008**, *456*, 470–476.
68. Harrison, P.M.; Kumar, A.; Lang, N.; Snyder, M.; Gerstein, M. A question of size: The eukaryotic proteome and the problems in defining it. *Nucleic Acids Res.* **2002**, *30*, 1083–1090.
69. Stamm, S. Signals and their transduction pathways regulating alternative splicing: A new dimension of the human genome. *Hum. Mol. Genet.* **2002**, *11*, 2409–2416.
70. Adams, M.D.; Celniker, S.E.; Holt, R.A.; Evans, C.A.; Gocayne, J.D.; Amanatides, P.G.; Scherer, S.E.; Li, P.W.; Hoskins, R.A.; Galle, R.F.; *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **2000**, *287*, 2185–2195.
71. Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.; *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **1995**, *269*, 496–512.
72. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* **1998**, *282*, 2012–2018.
73. Ramani, A.K.; Calarco, J.A.; Pan, Q.; Mavandadi, S.; Wang, Y.; Nelson, A.C.; Lee, L.J.; Morris, Q.; Blencowe, B.J.; Zhen, M.; *et al.* Genome-wide analysis of alternative splicing in *Caenorhabditis elegans*. *Genome Res.* **2011**, *21*, 342–348.
74. Graveley, B.R.; Brooks, A.N.; Carlson, J.W.; Duff, M.O.; Landolin, J.M.; Yang, L.; Artieri, C.G.; van Baren, M.J.; Boley, N.; Booth, B.W.; *et al.* The developmental transcriptome of *Drosophila melanogaster*. *Nature* **2011**, *471*, 473–479.
75. Roy, S.W. Intron-rich ancestors. *Trends Genet.* **2006**, *22*, 468–471.
76. Munding, E.M.; Igel, A.H.; Shiue, L.; Dorigi, K.M.; Trevino, L.R.; Ares, M., Jr. Integration of a splicing regulatory network within the meiotic gene expression program of *Saccharomyces cerevisiae*. *Genes Dev.* **2010**, *24*, 2693–2704.
77. Pleiss, J.A.; Whitworth, G.B.; Bergkessel, M.; Guthrie, C. Rapid, transcript-specific changes in splicing in response to environmental stress. *Mol. Cell* **2007**, *27*, 928–937.
78. Calarco, J.A.; Xing, Y.; Caceres, M.; Calarco, J.P.; Xiao, X.; Pan, Q.; Lee, C.; Preuss, T.M.; Blencowe, B.J. Global analysis of alternative splicing differences between humans and chimpanzees. *Genes Dev.* **2007**, *21*, 2963–2975.
79. Barbosa-Morais, N.L.; Irimia, M.; Pan, Q.; Xiong, H.Y.; Gueroussov, S.; Lee, L.J.; Slobodeniuc, V.; Kutter, C.; Watt, S.; Colak, R.; *et al.* The evolutionary landscape of alternative splicing in vertebrate species. *Science* **2012**, *338*, 1587–1593.
80. Venables, J.P.; Vignal, E.; Baghdiguian, S.; Fort, P.; Tazi, J. Tissue-specific alternative splicing of *Tak1* is conserved in deuterostomes. *Mol. Biol. Evol.* **2012**, *29*, 261–269.
81. Merkin, J.; Russell, C.; Chen, P.; Burge, C.B. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* **2012**, *338*, 1593–1599.

82. Aoto, J.; Martinelli, D.C.; Malenka, R.C.; Tabuchi, K.; Sudhof, T.C. Presynaptic neurexin-3 alternative splicing trans-synaptically controls postsynaptic AMPA receptor trafficking. *Cell* **2013**, *154*, 75–88.
83. Colak, R.; Kim, T.; Michaut, M.; Sun, M.; Irimia, M.; Bellay, J.; Myers, C.L.; Blencowe, B.J.; Kim, P.M. Distinct types of disorder in the human proteome: Functional implications for alternative splicing. *PLoS Comput. Biol.* **2013**, *9*, e1003030.
84. Ellis, J.D.; Barrios-Rodiles, M.; Colak, R.; Irimia, M.; Kim, T.; Calarco, J.A.; Wang, X.; Pan, Q.; O'Hanlon, D.; Kim, P.M.; *et al.* Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol. Cell* **2012**, *46*, 884–892.
85. Gabut, M.; Samavarchi-Tehrani, P.; Wang, X.; Slobodeniuc, V.; O'Hanlon, D.; Sung, H.K.; Alvarez, M.; Talukder, S.; Pan, Q.; Mazzoni, E.O.; *et al.* An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell* **2011**, *147*, 132–146.
86. Voineagu, I.; Wang, X.; Johnston, P.; Lowe, J.K.; Tian, Y.; Horvath, S.; Mill, J.; Cantor, R.M.; Blencowe, B.J.; Geschwind, D.H. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **2011**, *474*, 380–384.
87. Kalsotra, A.; Cooper, T.A. Functional consequences of developmentally regulated alternative splicing. *Nat. Rev. Genet.* **2011**, *12*, 715–729.
88. Hsu, S.N.; Hertel, K.J. Spliceosomes walk the line: Splicing errors and their impact on cellular function. *RNA Biol.* **2009**, *6*, 526–530.
89. Luger, K.; Mader, A.W.; Richmond, R.K.; Sargent, D.F.; Richmond, T.J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **1997**, *389*, 251–260.
90. Schwartz, S.; Meshorer, E.; Ast, G. Chromatin organization marks exon-intron structure. *Nat. Struct. Mol. Biol.* **2009**, *16*, 990–995.
91. Tilgner, H.; Nikolaou, C.; Althammer, S.; Sammeth, M.; Beato, M.; Valcarcel, J.; Guigo, R. Nucleosome positioning as a determinant of exon recognition. *Nat. Struct. Mol. Biol.* **2009**, *16*, 996–1001.
92. Brugiolo, M.; Herzel, L.; Neugebauer, K.M. Counting on co-transcriptional splicing. *F1000Prime Rep.* **2013**, *5*, 9.
93. Tilgner, H.; Knowles, D.G.; Johnson, R.; Davis, C.A.; Chakraborty, S.; Djebali, S.; Curado, J.; Snyder, M.; Gingeras, T.R.; Guigo, R. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* **2012**, *22*, 1616–1625.
94. Ameer, A.; Zaghlool, A.; Halvardson, J.; Wetterbom, A.; Gyllensten, U.; Cavellier, L.; Feuk, L. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat. Struct. Mol. Biol.* **2011**, *18*, U1435–U1157.
95. Kaida, D.; Berg, M.G.; Younis, I.; Kasim, M.; Singh, L.N.; Wan, L.; Dreyfuss, G. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **2010**, *468*, 664–668.
96. Hodges, C.; Bintu, L.; Lubkowska, L.; Kashlev, M.; Bustamante, C. Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science* **2009**, *325*, 626–628.
97. Kornblihtt, A.R.; Schor, I.E.; Allo, M.; Blencowe, B.J. When chromatin meets splicing. *Nat. Struct. Mol. Biol.* **2009**, *16*, 902–903.

98. Schwartz, S.; Ast, G. Chromatin density and splicing destiny: On the cross-talk between chromatin structure and splicing. *EMBO J.* **2010**, *29*, 1629–1636.
99. Carrillo Oesterreich, F.; Bieberstein, N.; Neugebauer, K.M. Pause locally, splice globally. *Trends Cell. Biol.* **2011**, *21*, 328–335.
100. Schor, I.E.; Allo, M.; Kornblihtt, A.R. Intragenic chromatin modifications: A new layer in alternative splicing regulation. *Epigenetics* **2010**, *5*, 174–179.
101. Tilgner, H.; Guigo, R. From chromatin to splicing: RNA-processing as a total artwork. *Epigenetics* **2010**, *5*, 180–184.
102. De la Mata, M.; Alonso, C.R.; Kadener, S.; Fededa, J.P.; Blaustein, M.; Pelisch, F.; Cramer, P.; Bentley, D.; Kornblihtt, A.R. A slow RNA polymerase II affects alternative splicing *in vivo*. *Mol. Cell* **2003**, *12*, 525–532.
103. Schor, I.E.; Rascovan, N.; Pelisch, F.; Allo, M.; Kornblihtt, A.R. Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 4325–4330.
104. Luco, R.F.; Pan, Q.; Tominaga, K.; Blencowe, B.J.; Pereira-Smith, O.M.; Misteli, T. Regulation of alternative splicing by histone modifications. *Science* **2010**, *327*, 996–1000.
105. Braunschweig, U.; Gueroussov, S.; Plocik, A.M.; Graveley, B.R.; Blencowe, B.J. Dynamic integration of splicing within gene regulatory pathways. *Cell* **2013**, *152*, 1252–1269.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).