Review: Alternative Splicing (AS) of Genes As An Approach for Generating Protein Complexity

Bishakha Roy, Larisa M. Haupt and Lyn R. Griffiths*

Genomics Research Centre, Griffith Health Institute, Griffith University Gold Coast, Queensland 4222, Australia

Abstract: Prior to the completion of the human genome project, the human genome was thought to have a greater number of genes as it seemed structurally and functionally more complex than other simpler organisms. This along with the belief of "one gene, one protein", were demonstrated to be incorrect. The inequality in the ratio of gene to protein formation gave rise to the theory of alternative splicing (AS). AS is a mechanism by which one gene gives rise to multiple protein products. Numerous databases and online bioinformatic tools are available for the detection and analysis of AS. Bioinformatics provides an important approach to study mRNA and protein diversity by various tools such as expressed sequence tag (EST) sequences obtained from completely processed mRNA. Microarrays and deep sequencing approaches also aid in the detection of splicing events. Initially it was postulated that AS occurred only in about 5% of all genes but was later found to be more abundant. Using bioinformatic approaches, the level of AS in human genes was found to be fairly high with 35-59% of genes having at least one AS form. Our ability to determine and predict AS is important as disorders in splicing patterns may lead to abnormal splice variants resulting in genetic diseases. In addition, the diversity of proteins produced by AS poses a challenge for successful drug discovery and therefore a greater understanding of AS would be beneficial.

1. INTRODUCTION

With the completion of the human genome project, new insights into genomic data have been made possible. Humans produce around 90,000 different types of proteins, hence it was assumed that a corresponding number of genes would be present in the human genome [1]. It was also reasoned that since the human genome is structurally and functionally more complex than other simpler organisms, this would correspond to a proportionally higher number of genes. However it came as a surprise to discover that humans had fewer than 25,000 genes while a much simpler organism such as corn (*Zea mays)* contains approximately 40,000 genes [1]. This acknowledged mismatch in the ratio of gene to protein formation and the accompanying 'one gene' to 'one protein' paradigm resulted in the concept of alternative splicing (AS) of genes.

In order to explain the small number of human genes discovered, several hypotheses have been postulated. One suggestion was that since the human genome is more complex than other multicellular organisms, a greater level of regulation of genes and pathways would be involved [2, 3]. Another hypothesis suggests that post-translational modifications, which consist of more than 200 different types, could produce a number of different protein products. A third

hypothesis relates to alternative splicing. AS is another mechanism by which a single gene could give rise to multiple protein products adding to protein diversity [3, 4].

Although the concept of AS was first introduced in the 1970s, only a few hundred AS genes have been recognized to date. In 1977, introns and exons were discovered in the adenovirus *hexon* gene. It was predicted by Walter Gilbert that different arrangements of exons could be spliced together (AS) to give diverse mRNA isoforms [5]. It was assumed that around 5% of genes in higher eukaryotes undergo AS [6]. High-throughput sequencing of the human genome, especially via analysis of expressed sequence tag (EST) sequences has provided a different approach for the analysis of AS through bioinformatics. As ESTs are obtained from completely processed mRNA, they present extensive mRNA diversity. From the year 2000 onwards, bioinformatic studies have progressively identified a magnitude of AS in genes than were previously known and are beginning to provide a global view of AS in humans [5]. More than 90% of pre-mRNA is detached as introns and only about 10% of the average pre-mRNA is connected as exonic sequence by the splicing of pre-mRNA. Although AS was initially assumed to occur in about 5% of all genes [6], it was soon realized that this mechanism was vastly more abundant [7]. AS occurs in all tissues, but the most common occurrence of tissue-specific splicing is observed in brain cells [8, 9]. Via EST data comparisons, the process of AS can be seen in evolutionarily discrete species such as humans, mice, *Drosophila* and *C. elegans*, signifying a plausible role for AS in evolution [3, 7].

*Address correspondence to this author at the Genomics Research Centre, Griffith Health Institute, Griffith University, Gold Coast, Queensland, Australia, 4222; Tel: +61-7-55528664; Fax: +61-7-55948908; E-mail: l.griffiths@griffith.edu.au

Much research has focused on the genome-wide identification of AS events in various cell and tissue types, under different conditions to ascertain the degree of functionally significant AS events. Sequence-based and microarray-based analyses have shown that AS events are more commonly observed in transcripts present in those genes of functionally complex tissues having varied cell types like the brain and testis [10]. These AS events also commonly occur in genes in cell types that participate in the selection and supply of diverse functions, such as the immune system [11-13]. Several complex examples have been identified including the neurexin and CD44 loci, which have two or more AS sites, with the autonomy of AS cassettes thought to form several protein accumulations encoded by a single locus [14, 15]. Another example is the *Drosophila* gene Dscam, which encodes the axon guidance receptor and is able to produce approximately 38, 016 different protein isoforms [14, 16].

Importantly, when cells alter their AS pattern in response to a signal, protein variants with varied biological functions may be formed. When protein complexes are formed on processed pre-mRNA, it helps in the selection of AS sites. It has now been demonstrated that reversible phosphorylation shows a strong effect on pre-mRNA processing [17]. It is likely that AS variants that code for unstable protein domains are nonfunctional because the encoded protein is unlikely to fold to its expected unique structure. Reverse transcriptase-polymerase chain reaction (RT-PCR) experiments have demonstrated that splice variants coding for unstable proteins are less abundant than those that code for stable proteins [18]. From the abundance of AS seen in higher eukaryotes, it is probable that many features of the cell phenotype, including those that lead to tumour development, are monitored by the relative expression of the AS isoforms of several genes [19].

1.1. Mechanism of Alternative Splicing

The process of splicing occurs in the spliceosome, a complex of five small nuclear RNAs, associated core proteins and many hundreds of proteins that gather on nascent pre-mRNAs during transcription. The splicing mechanism is an organized series of congregation and conformational reorganization events, interrupted by the chemical transformations of the cleavage of phosphodiester bonds at exon/intron junctions and the formation of a phosphodiester bond when the exon is ligated [20]. This model, referred to as the combinatorial model, suggests that AS in mammalian cells is mainly controlled by the binding of general splicing factors to pre-mRNA molecules when the spliceosome is produced. The spliceosome has most of these factors, including a class of proteins called serine/arginine-rich (SR) proteins, which have one or two RNA-binding domains and a protein-protein interaction domain, rich in serine and arginine amino acids. This model also suggests that AS transcripts employ various combinations of pre-mRNA splicing factors *in vivo*. The selection of splicing factors employed for each spliceosome depends on both the concentration of each factor in individual cell types and the regulatory elements present in each pre-mRNA [21].

The detection of splice sites and pre-mRNA splicing are dynamic processes that require steady remodelling of pro-

teins and ribonuclear protein units on the pre-mRNA being processed. They are related to other processing links including transcription, 5'-end capping, 3'-end polyadenylation and nuclear export. When splicing regulatory proteins remain attached to their pre-mRNA, they frequently participate in several processing events [17, 22]. Splice sites bordering exons in higher eukaryotes are vastly degenerate and are inadequate for accurate exon recognition. Thus protein complexes are generated on the pre-mRNA and aid in the detection of splice sites with high precision. The majority of these splicing regulatory proteins are defined as two major classes namely heterogeneous nuclear ribonucleoproteins (hnRNPs) and SR proteins. These proteins include domains for RNA-binding and protein/protein interactions. They bind with low specificity to the available and generally single-stranded component of pre-mRNA [23]. For enhancing RNA binding specificity, splicing regulatory proteins utilize their protein interaction domains to bind to each other. Protein/protein interactions in these proteins also experience a low specificity. After the formation of these protein complexes around an exon, ribonuclear protein components of the core spliceosome generate RNA/RNA interactions at the 5'-splice site and at the branch point [17]. These mechanisms have aided our understanding of the factors affecting splicing regulation.

1.2. Types of AS

Currently, there are seven main types of AS described. Of these, the cassette-type alternative exon account for approximately one-third of all AS events. The alternative 5' or 3' splice sites commonly occur together, forming approximately 25% of all AS events and are proficient enough to bring about changes in the coding sequence in as little as one codon. All AS events summarised may occur in the translated as well as the untranslated regions of the transcripts [24] (Fig. **1**).

2. CLINICAL RELEVANCE

Abnormal splice variants can be formed due to disorders in the splicing patterns and this is observed in more than 50% of genetic diseases including cancer. Tumorous growth is favoured by distortion of the splice sites in a neoplastic cell. Due to mutations or polymorphisms in the gene, RNA splicing can be dysregulated and this becomes part of the acquired pathology. Around 15% of point mutations linked with genetic disorders are credited to aberrant splicing [25], with several diseases linked to mutations that contribute to abnormal regulation of AS resulting in the expression of irregular protein isoforms for the specific cell type [26, 27].

It has become evident in recent years that successful drug discovery needs an understanding of the complexities associated with the associated biological background. In terms of proteomics, AS generates an added layer of complexity. In recent years, there has been great progress in obtaining information about AS. Despite these advancements, core drug discovery processes involve techniques unable to differentiate between splice variants and are thus trapped in the 'one gene - one protein' theory. As such, when AS is not taken into consideration, drug discovery is limited to a small proportion of the targets discoverable via proteomics and thus
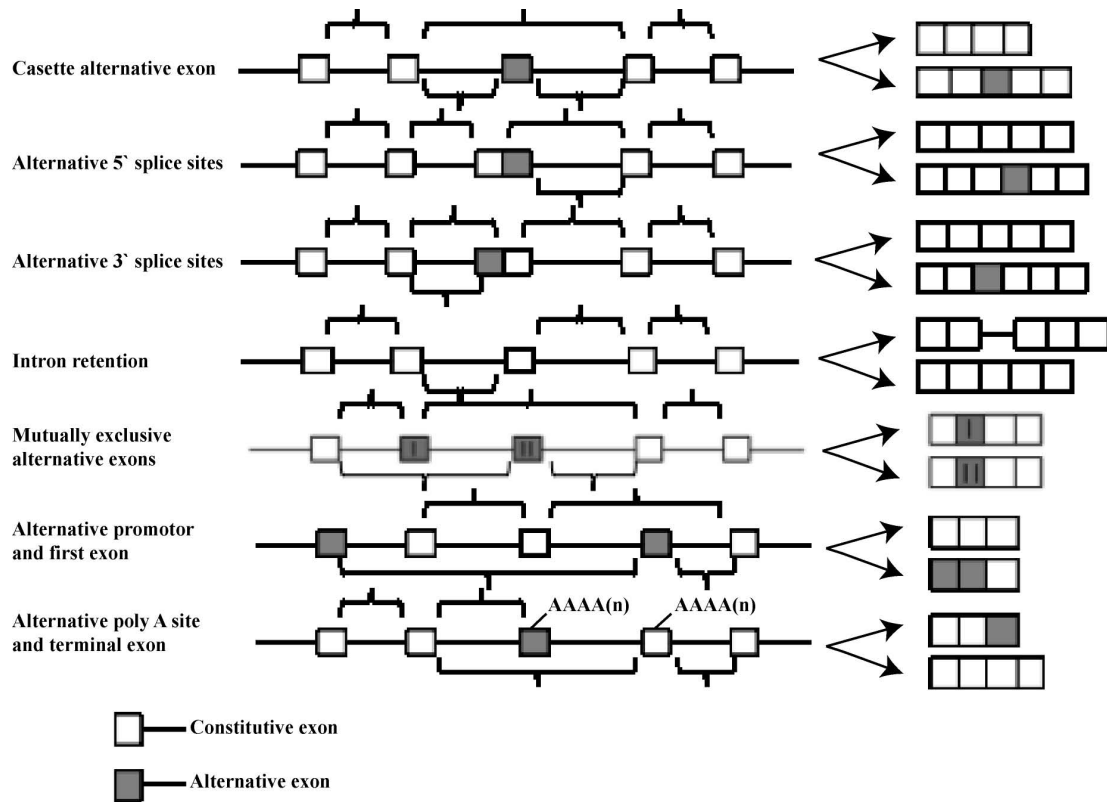
Fig. (1). Types of AS commonly observed adapted from Blencowe [10].

may miss many possible protein targets. Even when studies are carried out focusing on a specific gene, results will be prejudiced if AS is ignored [28]. Alteration in splicing account for many diseases including β-thalassemia (an autosomal recessive disorder) related to mutations in the second intron of the β-globin gene [29]. Mutations near splice sites may alter the transcript splicing pattern either by abnormal inclusion or exclusion of exon(s) and/or variations of 5' or 3' sites [30]. Variation in the 5' splice site activates a common 3' cryptic site upstream of the mutations. It also stimulates inclusion of a fragment of the intron-containing stop codon resulting in a decreased amount of functional β-globin protein [29].

Cancer is often a result of errors in pre-mRNA splicing leading to altered gene expression patterns. Affected proteins can include transcription factors, cell signal transducers and components of the extracellular matrix. For diagnostic purposes, changes in splicing patterns can be used as markers of the cell linked with disease. Some therapeutic tactics making use of AS include the use of protein over-expression to alter splicing of the affected exon [31], using antisense oligonucleotides [32], siRNA-based drugs for silencing gene expression [33], using compounds affecting phosphorylation of splicing factors [34], high-throughput screening to recognize compounds affecting splicing efficiencies of target pre-mRNAs [35] and restoring mutated exons with wild-type exons [36].

Another example of a disease caused by alterations in the splicing patterns is the microtubule-associated protein tau (encoded by the *MAPT* gene). This protein undergoes AS to produce six protein isoforms in the human brain [37, 38]. Mutations that cause abnormality in the splicing elements in the *MAPT* gene produce irregular ratios of *MAPT* transcripts observed in patients affected with an inherited form of dementia. In 1994, an inherited form of fronto-temporal dementia with Parkinsonism was related to chromosome 17 (FTDP-17) [39]. It was later found that mutations in *MAPT* produced abnormal tau protein causing neurodegeneration, demonstrating a role for *MAPT* in FTDP-17. There are two types of mutations in tau causing FTDP-17 of which the first is either a missense or deletion mutation in the tau protein and the second, affects AS of exon 10 in tau, causing an increase or decrease in exon 10 ultimately leading to FTDP-17 [40-42]. In addition, exons 2 and 3 along with exon 10 in the tau pre-mRNA, undergo AS to generate isoforms that differ in their N- and C- terminals. Exons 2 and 3 undergo AS in the N- terminal half, generating three separate mRNA molecules, which skip both exon 2 and 3, include only exon 2 or include both exons 2 and 3. In the C-terminal half, exons 9 to 12 encode four microtubule-binding domains which are repeats of 31 or 32 amino acids [43]. In the N-terminal half, exon 10 may be included or skipped independently of AS, generating three-microtubule repeats (3R, exon 10⁻) or the four-microtubule repeat (4R, exon 10⁺) isoforms of the protein. Under normal circumstances, the ratio of 4R to 3R is approximately equal to one and this balance is essential for proper neuronal function [41, 42, 44].

To further understand the significance of precise AS patterns required for the normal functioning of a cell, *MAPT* pre-mRNA serves as a good model as when alternatively

spliced, six isoforms of the tau protein are produced [45]. Specifically, AS in exon 10 of this gene is regulated through the complex interaction of exonic splicing enhancers (ESEs), exonic splicing silencers (ESSs), intronic splicing enhancers (ISEs) and intronic splicing silencers (ISS) [46, 47]. The mutation (N279K) in exon 10 causes transversion from T→G (U→G in the transcript). N279K increases exon 10 inclusion by fortifying a purine-rich ESE, which enhances the sequence from AA<u>U</u>AAGAAG to AA<u>G</u>AAGAAG [46, 48]. Another mutation, Del280K further assists the function of the enhancer. This mutation causes a deletion of an AAG repeat from the ESE, leading to decreased exon 10 inclusion and also a decreased 4R/3R tau ratio [46]. The ESE engages Tra2β (transformer 2 beta homolog - Drosophila) to bind the AAG repeat and augments the inclusion in exon 10 [48]. A silent mutation (L284L) occurs downstream of the ESE and causes a disorder of the ESS (U<u>U</u>AG to U<u>C</u>AG), enhancing exon 10 inclusion [46]. Intron 10 is also a host to many other mutations occurring immediately downstream of the 5' splice site. These mutations may cause destabilization of a stem-loop structure that forms between the end of exon 10 and the 5' splice site. This destabilization is thought to lead to enhanced recognition of exon 10 by the spliceosomal U1 snRNP (U1 small nuclear ribonucleoprotein), leading to increased exon 10 inclusion [40-42, 46]. The significance of altered regulation of exon 10 inclusion in FTDP-17 cases is emphasized by both the number and the range of potential mutations [45].

Another example of the significance of AS in accurate cellular functioning is the cellular processes associated with cell survival through the production of the promoter of apoptosis Bcl-xS instead of a suppressor, the splice variant Bcl-xL. In the daily activities of cells, AS is ubiquitous, influencing and affecting central processes including intracellular signaling and structural phenotype [1]. It is also important to note that there is a high level of conservation in intron sequences that border conserved AS events. This highlights the need for regular inspection of these regions in diseased genes for mutations that alter splicing patterns.

Insight into Cancer Therapeutics

The occurrence of splicing events varies dependent upon cell type, developmental stage and disease state, including cancer. Metastasis in cancer cells is an active process and may occur as a result of genome variability and flexible proteomes caused by AS affecting the cell growth and survival mechanisms. In addition, cancer cells may themselves produce AS events in order to sustain their population and/or metastasise. As such, it is conceivable that AS events may play an important role at all stages of tumourigenesis. Our ability to identify and differentiate between AS variants as biomarkers in normal and cancerous cells will aid in better understanding and treating all cancers [49].

In 2008, Thorsen *et al.* [50] performed whole genome exon expression of 102 normal and cancer tissue samples of different stages from colon, urinary bladder and prostate cancers using the GeneChip Human Exon 1.0 ST Array to examine tissue and tumour-specific AS. In the study, 2069 candidate AS events were detected between normal tissue samples from colon, bladder and prostate. From these, 115

splicing events were selected for validation using RT-PCR and sequencing, with 10 events successfully validated. They further examined several candidate tumour-specific variations from colon, bladder and prostate cancer for validation using RT-PCR in an independent set (n= 81) of normal and tumour tissue samples. Using this replication population of samples, they were able to detect seven genes with tumour-specific splice variants (ACTN1, CALD1, COL6A3, LRRFIP2, PIK4CB, TPM1 and VCL). The validated tumour-specific variations were found to be consistent to clearly distinguish between normal and cancer samples and in some cases different tumour stages. A subgroup of the tumour-specific splicing variations (ACTN1, CALD1, and VCL) was observed in all three organs and may indicate general cancer-related splicing events [50]. In other studies, antisense oligonucleotides have been applied to down-regulate mRNAs resulting in improved cancer cell survival and to modify splicing patterns in muscular dystrophy, revealing promising results in the clinical investigations [51].

Most chemotherapeutic interventions against cancer utilize the induction of an apoptotic response but to date, there is minimal information available about the effect of chemotherapeutic agents on the AS of the apoptotic regulatory genes. A study in 2008 by Shkreta *et al.* [52] tested 20 mainstream anticancer drugs for their ability to mediate the production of Bcl-x splice isoforms. Using 293 cells, several drugs shifted splicing toward the proapoptotic Bcl-x$_S$ splice variant. Several drugs also shifted Bcl-x in the cancer cell lines (MCF-7, HeLa, PC-3, PA-1, and SKOV-3) however the set of active drugs differed between cell lines. They also observed that almost every drug could modify a subset of AS events in each cell line and indicated cell line-specific differences may occur in the pathways that control AS [52]. A major problem in cancer treatment is resistance to chemotherapy. Tumours are frequently sensitive to chemotherapy upon early treatment but can develop resistance over repeated treatments, with some cells able to survive and promote cellular mechanisms boosting their resistance to subsequent chemotherapy. Acknowledging the role of AS in regulating the apoptotic response to chemotherapy, research is now examining novel ways of regulating cellular control of key AS genes. There have been some advances in the development of chemotherapeutics that directly target AS in specific genes to regulate the protein isoform that will be formed including splice switching oligonucleotides which are designed to attach to pre-mRNA and prevent splice site utilization at the binding site [53].

3. DETECTION OF AS

3.1. Using Expressed Sequence Tags (ESTs)

EST and cDNA sequences can be aligned to genomic sequences using programs that explore conserved splice-site consensus sequences flanking the gaps formed by intron sequences between the aligned exons [10]. Huge databases of AS events extracted this way have been created for several species including humans, mice and rats [5, 54, 55]. However, there is an essential limitation to AS analyses using transcript sequence data. EST coverage is classically biased towards the 3′ and 5′ ends of transcripts with a general inadequacy in the number of sequenced transcripts available to

determine the frequency of either the incorporation or exclusion of alternative exons in a given cell/tissue, or under specific experimental conditions [13]. The effect of AS-induced transcript variations may cause a complete loss of protein function along with other effects such as variation in transcript localization, stability and translation [56-58]. This suggests that AS not only diversifies protein function, but also gives a level of control during the development and function of healthy tissue [59].

3.2. Using Microarrays

Microarray data helps in predicting splicing events with these predictions able to be applied to direct RT-PCR at specific transcript locations. These are useful for the observation of tissue-specific gene expression at the genomic level along with the detection and quantification of isoforms [19]. It has also been demonstrated that microarrays can be used to examine pre m-RNA splicing [57, 58]. There are two main classes of microarrays that are used in probe design for detecting and quantifying diverse AS forms. The first includes 'annotate-to-design' methods [60, 61], the second the 'design-to-annotate' methods [62]. These can be further categorized into three subclasses of microarray platforms namely 'tiling', 'exon/junction' and 'focused', to detect and quantify mRNA splice variants (Fig. **2**).

Tiling Arrays

Tiling arrays offer a complete coverage of the genome and are thus useful in the discovery of new splice variants while the disadvantage of this platform is that it requires a large number of probes.

Exons and/or Junction

The exons and/or junction platform offers some flexibility in the position of probes allowing the probe characteristics to be more homogenous. The disadvantages of this system include that it is mainly restricted to type I deletions (cassette exons) and known information of exon boundaries is a prerequisite. Junction arrays are useful as they support tissue-specific AS in thousands of genes and allow the identification and verification of splice variants not currently represented by ESTs or mRNAs. However, there are some limitations to using junction arrays. Similar to ESTs, junction arrays cannot establish if two splicing events in one tissue are in the same or different transcripts, and novel isoform sequences are not described. Cross-hybridization may also generate false positives when genes with similar sequences have strong tissue-specific regulation [19].

Focussed Design

The focussed design platform utilises a more simple data analysis requiring a smaller number of probes and also allows the quantitative determination of known splice forms. Disadvantages of this platform include the limitation of the design to known splice forms and the inclusion of errors due to unknown splice forms. The common disadvantages of these three platforms are that the analysis is difficult, there may be variations in probe characteristics and the occurrence of cross hybridization [19].

3.3. Other Microarray Platforms

Splice-Variant Oligonucleotide Microarray

Splice-variant oligonucleotide microarray is a high-throughput platform used to monitor splice-variant expression for therapeutics and diagnostics. Isoform-specific differential expression data can be obtained through the observed difference in hybridization signal between two oligonucleotides within the same gene. Genomic tiling arrays and exon arrays are used to recognize co-regulated exons allowing the inference of variant mixtures [64]. Expression arrays with many probes are examined to recognize exons differentially added or excluded in a tissue-specific manner. The RNA-mediated ligation pooled with array technologies, gives a novel approach for identifying exon-exon junction information of known splice variants.

Spotted Oligonucleotide Microarrays

Spotted oligonucleotide microarrays using probes designed to identify unprocessed and processed RNA have been utilized to screen pre-mRNA splicing in yeast [65] and the processing of non-coding RNAs in yeast and mammalian cells [66].

3.4. Other Tools

Other methods including a fiber-optic based arrays [67], polymerase colony assay [68] and other microarray-based methods using spotted cDNA fragments or oligonucleotides
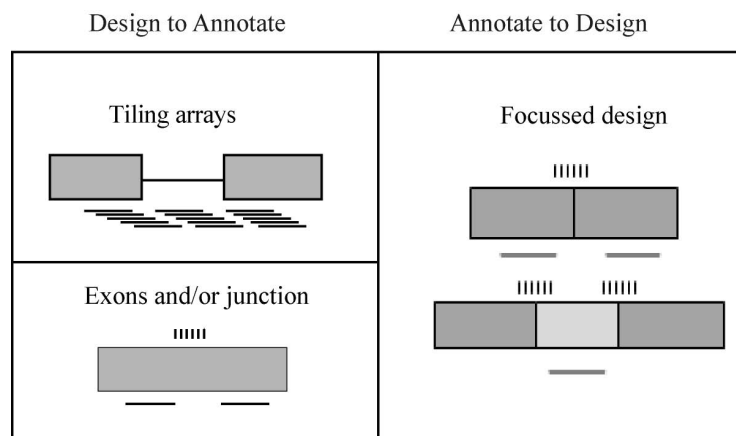

Fig. (2). Methods used to design AS microarray probes adapted from Cuperlovic-Culf *et al.* [63].

[40] have also proved useful for observing AS in mammalian cells. In order to detect AS in various tissues, it is imperative that experimental tools capable of identifying AS patterns directly and on a large scale are readily available [24]. A limitation of the microarray technique used for detecting AS is that long cDNAs cannot identify mRNA isoforms that differ by a micro exon or by the substitution of an exon of similar sequence [69].

RNA-Mediated Annealing, Selection and Ligation (RASL)

RNA-mediated annealing, selection and ligation (RASL) is an excellent tool to detect these minor sequence differences. RASL requires the annealing of total cellular RNA samples with a mixture of oligonucleotides complementary to sequences on both sides of the splice junction. This is followed by poly (A)$^+$ selection, ligation of flanking oligonucleotides and then amplification of the ligated oligonucleotides by PCR. PCR products are fluorescently labelled and examined on microarrays [24].

Surface Enhanced Raman Scattering (SERS)

Surface enhanced Raman scattering (SERS) based detection using nonfluorescent labels can be used to examine AS events. This method is a sensitive and selective tool for complete AS profiling of genes that are relevant to specific diseases [70].

Bead-Based Fiber-Optic Microarray System

In 2002, the novel technology of a bead-based fiber-optic microarray system able to analyze both gene expression and AS analogously was developed [67]. Due to its specificity, the assay is able to distinguish mRNA isoforms within a single gene. Limitations of this method include selection of the most suitable sequence for hybridization, difficulty in screening for novel transcript expression and difficulty in analyzing mRNA isoforms when many regions in a particular transcript undergo AS. This approach is most suitable for large-scale analysis of genomic surveys in large databases. The cDNA arrays also have their limitations as their long probes cannot identify the minor sequence differences that may occur due to insertion of nucleotides or the substitution of related exons [67]. Out of the techniques described for detecting AS, the ideal method is selected based on experimental design.

Deep Sequencing Technology

Unlike microarray methods, sequencing-based methods directly determine the cDNA sequence. Initially, sequencing reactions were performed using Sanger sequencing of cDNA or EST libraries [71]. However this method is comparatively low throughput, expensive and usually not quantitative. Tag-based approaches were then developed such as serial analysis of gene expression (SAGE) [72] cap analysis of gene expression (CAGE) [73] and massively parallel signature sequencing (MPSS) [74]. These methods are high-throughput and can offer precise, 'digital' gene expression levels. Nevertheless, most are still based on expensive Sanger sequencing technology and a large portion of the short tags cannot be uniquely mapped to the reference genome. In addition, only a portion of the transcript is analyzed and isoforms are commonly difficult to distinguish from each other. More recently, the development of novel high-throughput DNA sequencing approaches have offered new means for mapping and quantifying transcriptomes. This method, termed RNA-Seq (RNA sequencing) provides a more accurate measurement of transcript levels and their isoforms. In this technology, a library of cDNA fragments with adaptors attached to one or both ends are obtained from a population of RNA (total or fractioned, such as poly (A)+). High-throughput sequencing is carried out to obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing) from each molecule, with or without amplification [75]. Depending on the DNA sequencing technology used, including Illumina IG [76], Applied Biosystems SOLiD [77] and Roche 454 Life Science [78], the read lengths attained are generally in the 30-400bp range. After sequencing, the acquired reads are either aligned to a reference genome/reference transcripts or gathered de novo without the genomic sequence to generate a genome-scale transcription map that includes both the transcriptional structure and/or level of expression for each gene [75].

Advantages and Challenges of Deep Sequencing Technology

One of the main advantages of RNA-Seq over hybridisation-based methods is that it is not limited to identifying transcripts corresponding to existing genomic sequences. As an example, using the 454-based RNA sequencing, transcriptome sequencing of the Glanville fritillary butterfly has been attained [79]. RNA-Seq can also determine the exact location of transcription boundaries down to a single-base resolution. Using the 30-bp short reads obtained from RNA-Seq, information about the connectivity of two exons can be acquired and longer reads or pair-end short reads can reveal connectivity between multiple exons, making RNA-Seq a useful tool for analyzing complex transcriptomes. In addition, information about sequence variations (for example, SNPs) in the transcribed regions can also be obtained from RNA-Seq [80].

Another important advantage of RNA-Seq over DNA microarrays is that RNA-Seq has minimal background signal as DNA sequences can be clearly mapped to unique regions of the genome. This methodology has also been found to be accurate in quantifying expression levels by quantitative PCR (qPCR) [81], demonstrated by spike-in RNA controls [76]. RNA-Seq data also has a high level of reproducibility for technical as well as biological replicates [81]. In addition, with the absence of cloning steps as well as the absence of any amplification steps due to the Helicos technology, RNA-Seq also requires less starting RNA [75].

Although RNA-Seq involves only a few steps, there are several manipulation steps in the cDNA library preparation stage, which may complicate its use in profiling all types of transcripts. As a high-throughput sequencing method, RNA-seq faces the challenge to develop efficient retrieval and processing of large amounts of data while minimizing errors in image analysis, base-calling and low-quality reads. When processing large transcriptomes, a key problem is when a major portion of the sequence matches multiple locations within the genome. Short reads with high copy numbers (>100) and long portions of repetitive regions have proved to be challenging to this technology [75]. One solution to over-

come the multi-matching problem is to acquire longer reads or instead of a paired-end sequence approach where short sequences are obtained from both ends of a DNA fragment [82], extending the mapped fragment length to 200-500bp. Additionally, sequencing errors and polymorphisms can provide mapping challenges for all genomes, not only for repetitive DNA. Usually, single base changes are not problematic as most mapping algorithms accommodate one or two base differences. Nevertheless, larger differences may cause problems and will need better reference genome annotation for polymorphisms and deeper sequence coverage. Greater sequence coverage screening entails more sequence depth, while in order to determine a rare transcript or variant, extensive depth is essential. Usually, the larger the genome, the more complex the transcriptome and the more sequence depth is needed for adequate coverage [75].

4. BIOINFORMATIC APPROACHES

A significant focus of research in bioinformatics has centered on the development of useful AS databases for the analysis of AS patterns in genes. The most common bioinformatics method employed is the use of ESTs to determine differences such as large insertions or deletions. This enables further analysis by placing the ESTs within the gene sequence. In this way, the matched sequence will form the candidate exon while candidate splices will have large gaps in the EST-genomic alignment. Around 145 genes were described by 'The International Human Genome Sequencing Consortium' to be alternatively spliced, following widespread analysis on chromosome 22 by aligning ESTs with the genomic sequence [5, 83]. The level of AS in human genes was found to be between 35 to 59% of genes having at least one AS form [83] with estimates that 70–88% of AS alter the protein product [83] such as substitution of the amino or carboxyl terminus, in-frame addition and removal of a functional unit. As only a few ESTs have been sequenced for most genes, it is likely that many more AS forms may exist and are yet to be detected [5].

Bioinformatic Tools

Some of the existing databases and bioinformatics tools related to AS detection and analysis are described below. The online database ASAP (Alternative Splicing Annotation Project) is a valuable tool for biologists to obtain vast amounts of AS information from genomic and proteomic applications. ASAP commenced in 2002 and currently supplies AS data for human genes. A genome-wide analysis of several other organisms (Arabidopsis, mouse etc.) is currently being carried out. ASAP gives accurate gene exon-intron structure, AS analysis, tissue specificity of AS forms and the protein isoform sequences arising as a result of AS, providing protein isoform sequences for each splice form. ASAP employs EST alignment to the sequences for AS detection and exact sequence information for the resultant change in the sequence [55]. It also gives results from genome-wide analysis of tissue specific human AS forms [9].

The Alternative Splicing Gallery (ASG) represents the huge quantity of EST and cDNA data at the genomic level. It serves as an initiation point for the systematic examination of gene structure and the transcriptome. ASG has integrated

transcript data from RefSeq, Ensembl, UniGene, STACK and TIGR with Ensembl genes clearly displayed into splicing graphs for human genes. However, ASG does have several limitations. The database may over predict transcript number when there are dependences between AS events in a gene (e.g. events which always occur together or never occur together). Also, it does not show annotations of coding sequences, promoters, polyadenylation sites, strength of the splice sites and transcript truncations [84]. Another AS database called 'HOLLYWOOD' was constructed using the genomic annotation of splicing patterns of known genes obtained from the spliced alignment of cDNAs and ESTs. It shows many options including splice site sequences, exonic splicing enhancers and silencers, conserved and non-conserved patterns of splicing and complementary DNA (cDNA) library data for the queried alternative exons. It currently holds comprehensive information for human and mouse genes [85]. The alternative splicing database (ASD) consortium is methodically gathering and annotating information on AS and consists of three databases namely AltSplice, AltExtron and AEdb [86]. The largest part of ASD is AltSplice, a database of computationally defined AS events. This database has information regarding AS introns/exons, events, isoform splicing patterns and isoform peptide sequences. Data is gathered by investigating gene-transcript alignments. AEdb is the manually produced element of ASD. It contains sequence and properties of AS exons, a functional account of observed splicing events, characterization of observed splicing regulatory elements and a compilation of experimentally elucidated minigene constructs. ASD data has been included with the Ensembl genome annotation project as a 'Distributed Annotation System' (DAS) resource and can be accessed via the Ensembl genome browser [86].

The Manually Annotated Alternatively Spliced Events (MAASE) database system was formed to support splicing microarray functions. It consists of a manual/computational annotation tool for the proficient mining of crucial sequence and functional information of AS events and a user-friendly database of annotated events for the supply of data for microarray design and data analysis [54]. The AsMamDB database was generated to study AS genes in mammals. It provides information regarding AS patterns in genes, structure of genes, their location in the chromosome, and the gene products and tissues where expression takes place [87]. ASmodeler is a unique web-based tool for obtaining gene models including AS events from the genomic alignment of mRNA, EST and protein sequences [88]. The Alternative Splicing and Transcript Diversity database (ASTD) gives information about a huge collection of alternative transcripts combining transcription initiation, polyadenylation and splicing variant data and has now been intergrated into Ensembl genome browser [89].

ProSplicer is a database of acknowledged AS data obtained from the alignment of proteins, mRNA sequences and ESTs matched against human genomic sequences. Proteins, mRNA and ESTs provide important data that can divulge gene splice variants [90]. ASTRA is a database with a Java-based browser that can efficiently reorganize the order of displayed splicing patterns [91]. SpliceInfo is an information repository describing the incidence of the four main AS

modes in the human genome which include exon skipping, 5'-AS, 3'-AS and intron retention. Data is obtained by comparing nucleotide and protein sequences for a given gene for any evidence of AS. Added features including tissue specificity, protein domain contained by exons, GC-ratio of exons, repeats present within the exons and gene ontology are annotated computationally for each exonic region containing AS [92]. In 2001, three independent databases GeneNest, SpliceNest and SYSTERS which are available online, were linked to each other and to other major databases. This collaboration has helped researchers in exploring the entire sequences of many proteins, mRNA, ESTs and genomic DNA [93]. Putative alternate splicing database (PALS db) contains information from 19,936 human UniGene clusters and 16,615 mouse UniGene clusters. It predicts AS sites by using the longest mRNA sequence in each UniGene cluster as the reference sequence and then aligning it with a linked sequence in UniGene and dbEST [94]. Human-transcriptome DataBase for Alternative Splicing (H-DBAS) is a unique database of AS human transcripts [95].

ECgene provides functional annotation for alternatively spliced genes. It includes the genome-based transcript modeling for alternative splicing (AS), domain analysis with Gene Ontology (GO) annotation and expression analysis based on the EST and Serial analysis of gene expression (SAGE) data. ECgene has AS modeling and EST clustering includes nine organisms for which sufficient EST data is available in the GenBank. ECgene has also launched several new applications to analyze differential expression for the human genome [96]. H-InvDB was started in 2004 and in their latest version, H-InvDB 8.0, mapping of 244 709 human complementary DNA was performed onto the hg19 reference genome and 43 829 gene loci, including nonprotein-coding ones, were identified [97]. Alternative splicing transcriptional landscape visualization tool (ASTALAVISTA) utilizes an intuitive and complete notation system to clearly recognize AS events. It can characterize AS from data for entire transcriptomes from reference annotations (GENCODE, REFSEQ, ENSEMBL) and also for genes selected by the user based on common functional/structural attributes of interest [98]. Alternative splicing prediction database (ASPicDB) provides the means for splice-site detection and full-length transcript modelling by a genome-wide application of the ASPic algorithm that uses the multiple alignments of gene-related transcripts (typically a UniGene cluster) to the genomic sequence. It also provides information on tissue-specific splicing patterns of normal and cancer cells by looking at data on available EST sequences and their library source annotation [99]. DBASS3 and DBASS5 offer comprehensive data of new exon boundaries induced by pathogenic mutations in human disease genes. They currently contain approximately 900 records of cryptic and de novo 3'- and 5'-splice sites that were produced by over a thousand different mutations in approximately 360 genes [100].

Alternative splicing electronic RT-PCR (ASePCR) is a web-based tool that emulates RT–PCR in various tissues. NCBI provides a reverse e-PCR option which generates transcript models and ASePCR can then estimate the amplicon size for a given primer pair based on these transcript models. The tissue specificity of each PCR band is inferred from the

tissue data of EST sequences compatible with each transcript structure. The result generated in the output page is in the form of PCR bands in a gel electrophoresis in various tissues. Each band characterizes a putative isoform that may occur in a tissue-specific manner also providing the EST alignment and tissue information in the genome browser. Also, the AS patterns of orthologous genes in other species can be compared [101].

SNPSplicer helps in the quick analysis of experiments involving screening of SNPs located in sequences of splicing relevance. It proposes to use sets of matching DNA and complementary DNA (cDNA) as a screening tool to detect the potential splice effects of SNPs in RT-PCR experiments with tissue material from genotyped sources [102]. SplicingViewer is a very useful tool for the clear detection, annotation and visualization of splice junctions and alternative splicing events from RNA-Seq data [103]. SpliceGrapher is another tool that aids in discriminating between real and false splice sites and can improve the reliability of identification of AS for RNA-Seq data [104]. TopHat is another proficient algorithm programmed to align reads from an RNA-Seq experiment to a reference genome without depending on known splice sites [105]. Multivariate analysis of transcript splicing (MATS) is a Bayesian statistical framework used for RNA-Seq data analysis. It is used for flexible hypothesis testing of differential alternative splicing patterns. The MATS method is useful for almost any type of null hypotheses of interest, giving the flexibility to identify differential alternative splicing events that match a given user-defined pattern [106]. PASSion is a pattern growth algorithm-based pipeline useful in the detection of splice sites in paired-end RNA-Seq reads and can detect junctions that do not have known splicing motifs, which cannot be found by the other tools [107]. SpliceTrap is a method to quantify inclusion levels of exon utilizing paired-end RNA-seq data. Contrasting other tools that focus on full-length transcript isoforms, SpliceTrap looks at the expression-level estimation of each exon as an independent Bayesian inference problem. Also, it can identify key classes of alternative splicing events under a single cellular condition, without the need for a background set of reads to estimate relative splicing changes [108].

AltAnalyze and the Cytoscape plugin DomainGraph programs offer an intuitive and comprehensive solution for the analysis and visualization of AS data from Affymetrix exon and gene arrays at the level of proteins, domains, microRNA binding sites, molecular interactions and pathways [109]. Junction and Exon Toolkits for Transcriptome Analysis (JETTA) is an integrated software tool for calculating gene expression indices and also aids in the identification and visualization of AS events [110]. Cufflinks is another useful bioinformatic tool that does not use existing gene annotations during the assembly of transcripts but assembles a minimum set of transcripts that describe the reads in the dataset in the best way [111]. Alternative Splice Site Predictor (ASSP) is a web tool used for predicting alternative splice sites [112]. GeneSplicer is a flexible bioinformatic system employed in identifying splice sites in the genomic DNA of various eukaryotes [113]. MapSplice is a second generation splice detection algorithm targeting high sensitivity and specificity for detecting splices and also CPU and memory efficiency. It can be applied to both short (<75 bp)

and long reads (>75 bp) and does not depend on splice site features or intron length, thus it can identify novel canonical as well as non-canonical splices [114]. Mixture-of-isoforms (MISO) model is a statistical model used in the estimation of expression of alternatively spliced exons and isoforms and also evaluates confidence in these estimates. MISO is very valuable to infer isoform regulation from high-throughput sequencing of cDNA fragments (RNA-Seq) [115]. The PASA software [116] gathers and clusters spliced transcript alignments, offering transcript-based gene structures that are used to automatically improve existing gene annotations by adding untranslated regions (UTRs), adjusting intron and exon boundaries and adding new models representing alternative splicing from its many other functions [117]. DEXSeq is a statistical method used to analyze RNA-Seq data by testing for differential exon usage with high sensitivity [118]. HMMSplicer performs particularly well in compact genomes and on genes with low expression levels, alternative splice isoforms, or non-canonical splice junctions [119].

5. CONCLUSION

AS is a vital mechanism regulating gene expression in multicellular organisms. Transcriptional regulation of the promoter mainly results in variations in the amount of RNA generated, producing N-terminal protein variants by different transcriptional start sites. However, regulation through AS is much more flexible with alterations in the protein sequence potentially affecting almost all areas of protein function including binding properties, enzyme activity, intracellular localization, protein stability, phosphorylation and glycosylation patterns. AS-site regulation is obtained by combining many weak interactions between regulatory proteins and signals on the pre-mRNA [7].

Changes in the selection of splice sites may arise not only due to disease but also to the natural adaptation of a cell. Many stimuli including growth factors, cytokines, hormones and cell depolarization cause altered splice site selection [120]. Sequence analyses predict that only around 10% of AS events are likely to delete, insert or modify functional domains in proteins and these events give a fairly straightforward approach to design suitable experiments for addressing the functional consequences of AS [10, 121]. As for the remaining 90% of AS events found in coding regions, they do not overlap with boundaries of functional protein domains or regions which are vital for overall protein folding [10]. Most AS events influence coiled or "loop" regions of secondary protein structures and the greater part of these coding sequences are situated on protein surfaces [122]. These regions help perform protein-protein interactions or interactions with other types of ligands [10].

After sequencing of the genome by the human genome project, it is now possible to examine the evolutionary effects and limitations of AS. AS provides a vital strategy in understanding pathways that are applicable in the evolution of gene structures and in uncovering novel protein-coding sequences. Mammalian genomes give a broad indication of RNA selection pressure due to the restriction of AS regulation, thus highlighting the functional significance of AS patterns [123].

REFERENCES

[1] H.H. Valdivia, One gene, many proteins: alternative splicing of the ryanodine receptor gene adds novel functions to an already complex channel protein, Circ Res, 100 (**2007**) 761-763.

[2] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, J.D. Gocayne, P. Amanatides, R.M. Ballew, D.H. Huson, J.R. Wortman, Q. Zhang, C.D. Kodira, X.H. Zheng, L. Chen, M. Skupski, G. Subramanian, P.D. Thomas, J. Zhang, G.L. Gabor Miklos, C. Nelson, S. Broder, A.G. Clark, J. Nadeau, V.A. McKusick, N. Zinder, A.J. Levine, R.J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A.E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T.J. Heiman, M.E. Higgins, R.R. Ji, Z. Ke, K.A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G.V. Merkulov, N. Milshina, H.M. Moore, A.K. Naik, V.A. Narayan, B. Neelam, D. Nusskern, D.B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M.L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y.H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N.N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J.F. Abril, R. Guigo, M.J. Campbell, K.V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y.H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, X. Zhu, The sequence of the human genome, Science, 291 (**2001**) 1304-1351.

[3] D. Brett, H. Pospisil, J. Valcarcel, J. Reich, P. Bork, Alternative splicing and genome complexity, Nat Genet, 30 (**2002**) 29-30.

[4] B. Ewing, P. Green, Analysis of expressed sequence tags indicates 35,000 human genes, Nat Genet, 25 (**2000**) 232-234.

[5] B. Modrek, C. Lee, A genomic view of alternative splicing, Nat Genet, 30 (**2002**) 13-19.

[6] P.A. Sharp, Split genes and RNA splicing, Cell, 77 (**1994**) 805-815.

[7] S. Stamm, S. Ben-Ari, I. Rafalska, Y. Tang, Z. Zhang, D. Toiber, T.A. Thanaraj, H. Soreq, Function of alternative splicing, Gene, 344 (**2005**) 1-20.

[8]   S. Stamm, J. Zhu, K. Nakai, P. Stoilov, O. Stoss, M.Q. Zhang, An alternative-exon database and its statistical analysis, DNA Cell Biol, 19 (**2000**) 739-756.

[9]   Q. Xu, B. Modrek, C. Lee, Genome-wide detection of tissue-specific alternative splicing in the human transcriptome, Nucleic Acids Res, 30 (**2002**) 3754-3766.

[10]   B.J. Blencowe, Alternative splicing: new insights from global analyses, Cell, 126 (**2006**) 37-47.

[11]   B. Modrek, A. Resch, C. Grasso, C. Lee, Genome-wide detection of alternative splicing in expressed sequences of human genes, Nucleic Acids Res, 29 (**2001**) 2850-2859.

[12]   F.L. Watson, R. Puttmann-Holgado, F. Thomas, D.L. Lamar, M. Hughes, M. Kondo, V.I. Rebel, D. Schmucker, Extensive diversity of Ig-superfamily proteins in the immune system of insects, Science, 309 (**2005**) 1874-1878.

[13]   J.M. Johnson, J. Castle, P. Garrett-Engele, Z. Kan, P.M. Loerch, C.D. Armour, R. Santos, E.E. Schadt, R. Stoughton, D.D. Shoemaker, Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays, Science, 302 (**2003**) 2141-2144.

[14]   J. Zhu, J. Shendure, R.D. Mitra, G.M. Church, Single molecule profiling of alternative pre-mRNA splicing, Science, 301 (**2003**) 836-838.

[15]   M. Missler, T.C. Sudhof, Neurexins: three genes and 1001 products, Trends Genet, 14 (**1998**) 20-26.

[16]   D. Schmucker, J.C. Clemens, H. Shu, C.A. Worby, J. Xiao, M. Muda, J.E. Dixon, S.L. Zipursky, Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity, Cell, 101 (**2000**) 671-684.

[17]   S. Stamm, Regulation of alternative splicing by reversible protein phosphorylation, J Biol Chem, 283 (**2008**) 1223-1227.

[18]   K. Homma, R.F. Kikuno, T. Nagase, O. Ohara, K. Nishikawa, Alternative splice variants encoding unstable protein domains exist in the human brain, J Mol Biol, 343 (**2004**) 1207-1220.

[19]   A. Relogio, C. Ben-Dov, M. Baum, M. Ruggiu, C. Gemund, V. Benes, R.B. Darnell, J. Valcarcel, Alternative splicing microarrays reveal functional expression of neuron-specific regulators in Hodgkin lymphoma cells, J Biol Chem, 280 (**2005**) 4779-4784.

[20]   M.J. Fedor, Alternative splicing minireview series: combinatorial control facilitates splicing regulation of gene expression and enhances genome diversity, J Biol Chem, 283 (**2008**) 1209-1210.

[21]   S.A. Mabon, T. Misteli, Differential recruitment of pre-mRNA splicing factors to alternatively spliced transcripts in vivo, PLoS Biol, 3 (**2005**) e374.

[22]   M.J. Moore, From birth to death: the complex lives of eukaryotic mRNAs, Science, 309 (**2005**) 1514-1518.

[23]   M. Hiller, Z. Zhang, R. Backofen, S. Stamm, Pre-mRNA secondary structures influence exon recognition, PLoS Genet, 3 (**2007**) e204.

[24]   P. Grabowski, Alternative splicing in parallel, Nat Biotechnol, 20 (**2002**) 346-347.

[25]   M. Krawczak, J. Reiss, D.N. Cooper, The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences, Hum Genet, 90 (**1992**) 41-54.

[26]   W. Jin, I.E. McCutcheon, G.N. Fuller, E.S. Huang, G.J. Cote, Fibroblast growth factor receptor-1 alpha-exon exclusion and polypyrimidine tract-binding protein in glioblastoma multiforme tumors, Cancer Res, 60 (**2000**) 1221-1224.

[27]   L.H. Boise, M. Gonzalez-Garcia, C.E. Postema, L. Ding, T. Lindsten, L.A. Turka, X. Mao, G. Nunez, C.B. Thompson, bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death, Cell, 74 (**1993**) 597-608.

[28]   E.Y. Levanon, R. Sorek, The importance of alternative splicing in the drug discovery process, TARGETS, 2 (**2003**) 109-114.

[29]   M. Hagiwara, Alternative splicing: a new drug target of the post-genome era, Biochim Biophys Acta, 1754 (**2005**) 324-331.

[30]   N.A. Faustino, T.A. Cooper, Pre-mRNA splicing and human disease, Genes Dev, 17 (**2003**) 419-437.

[31]   J. Wang, E. Pegoraro, E. Menegazzo, M. Gennarelli, R.C. Hoop, C. Angelini, E.P. Hoffman, Myotonic dystrophy: evidence for a possible dominant-negative RNA mutation, Hum Mol Genet, 4 (**1995**) 599-606.

[32]   A. Herrera-Gayol, S. Jothy, Adhesion proteins in the biology of breast cancer: contribution of CD44, Exp Mol Pathol, 66 (**1999**) 149-156.

[33]   R.J. Sneath, D.C. Mangham, The normal structure and function of CD44 and its role in neoplasia, Mol Pathol, 51 (**1998**) 191-200.

[34]   B. Kwabi-Addo, F. Ropiquet, D. Giri, M. Ittmann, Alternative splicing of fibroblast growth factor receptors in human prostate cancer, Prostate, 46 (**2001**) 163-172.

[35]   S. Takaishi, M. Sawada, Y. Morita, H. Seno, H. Fukuzawa, T. Chiba, Identification of a novel alternative splicing of human FGF receptor 4: soluble-form splice variant expressed in human gastrointestinal epithelial cells, Biochem Biophys Res Commun, 267 (**2000**) 658-662.

[36]   I. Bieche, R. Lidereau, Increased level of exon 12 alternatively spliced BRCA2 transcripts in tumor breast tissue compared with normal tissue, Cancer Res, 59 (**1999**) 2546-2550.

[37]   M. Goedert, C.M. Wischik, R.A. Crowther, J.E. Walker, A. Klug, Cloning and sequencing of the cDNA encoding a core protein of the paired helical filament of Alzheimer disease: identification as the microtubule-associated protein tau, Proc Natl Acad Sci U S A, 85 (**1988**) 4051-4055.

[38]   M. Goedert, M.G. Spillantini, M.C. Potier, J. Ulrich, R.A. Crowther, Cloning and sequencing of the cDNA encoding an isoform of microtubule-associated protein tau containing four tandem repeats: differential expression of tau protein mRNAs in human brain, EMBO J, 8 (**1989**) 393-399.

[39]   T. Lynch, M. Sano, K.S. Marder, K.L. Bell, N.L. Foster, R.F. Defendini, A.A. Sima, C. Keohane, T.G. Nygaard, S. Fahn, et al., Clinical characteristics of a family with chromosome 17-linked disinhibition-dementia-parkinsonism-amyotrophy complex, Neurology, 44 (**1994**) 1878-1884.

[40]   M. Hutton, C.L. Lendon, P. Rizzu, M. Baker, S. Froelich, H. Houlden, S. Pickering-Brown, S. Chakraverty, A. Isaacs, A. Grover, J. Hackett, J. Adamson, S. Lincoln, D. Dickson, P. Davies, R.C. Petersen, M. Stevens, E. de Graaff, E. Wauters, J. van Baren, M. Hillebrand, M. Joosse, J.M. Kwon, P. Nowotny, L.K. Che, J. Norton, J.C. Morris, L.A. Reed, J. Trojanowski, H. Basun, L. Lannfelt, M. Neystat, S. Fahn, F. Dark, T. Tannenberg, P.R. Dodd, N. Hayward, J.B. Kwok, P.R. Schofield, A. Andreadis, J. Snowden, D. Craufurd, D. Neary, F. Owen, B.A. Oostra, J. Hardy, A. Goate, J. van Swieten, D. Mann, T. Lynch, P. Heutink, Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17, Nature, 393 (**1998**) 702-705.

[41]   M. Hong, V. Zhukareva, V. Vogelsberg-Ragaglia, Z. Wszolek, L. Reed, B.I. Miller, D.H. Geschwind, T.D. Bird, D. McKeel, A. Goate, J.C. Morris, K.C. Wilhelmsen, G.D. Schellenberg, J.Q. Trojanowski, V.M. Lee, Mutation-specific functional impairments in distinct tau isoforms of hereditary FTDP-17, Science, 282 (**1998**) 1914-1917.

[42]   M.G. Spillantini, J.R. Murrell, M. Goedert, M.R. Farlow, A. Klug, B. Ghetti, Mutation in the tau gene in familial multiple system tauopathy with presenile dementia, Proc Natl Acad Sci U S A, 95 (**1998**) 7737-7741.

[43]   G. Lee, R.L. Neve, K.S. Kosik, The microtubule binding domain of tau protein, Neuron, 2 (**1989**) 1615-1624.

[44]   M. Goedert, R. Jakes, Expression of separate isoforms of human tau protein: correlation with the tau pattern in brain and effects on tubulin polymerization, EMBO J, 9 (**1990**) 4225-4230.

[45]   M.A. Garcia-Blanco, A.P. Baraniak, E.L. Lasda, Alternative splicing in disease and therapy, Nat Biotechnol, 22 (**2004**) 535-546.

[46]   I. D'Souza, P. Poorkaj, M. Hong, D. Nochlin, V.M. Lee, T.D. Bird, G.D. Schellenberg, Missense and silent tau gene mutations cause frontotemporal dementia with parkinsonism-chromosome 17 type, by affecting multiple alternative RNA splicing regulatory elements, Proc Natl Acad Sci U S A, 96 (**1999**) 5598-5603.

[47]   Q.S. Gao, J. Memmott, R. Lafyatis, S. Stamm, G. Screaton, A. Andreadis, Complex regulation of tau exon 10, whose missplicing causes frontotemporal dementia, J Neurochem, 74 (**2000**) 490-500.

[48]   L.N. Clark, P. Poorkaj, Z. Wszolek, D.H. Geschwind, Z.S. Nasreddine, B. Miller, D. Li, H. Payami, F. Awert, K. Markopoulou, A. Andreadis, I. D'Souza, V.M. Lee, L. Reed, J.Q. Trojanowski, V. Zhukareva, T. Bird, G. Schellenberg, K.C. Wilhelmsen, Pathogenic implications of mutations in the tau gene in pallido-ponto-nigral degeneration and related neurodegenerative disorders linked to chromosome 17, Proc Natl Acad Sci U S A, 95 (**1998**) 13103-13107.

[49]   Y.J. Kim, H.S. Kim, Alternative splicing and its impact as a cancer diagnostic marker, Genomics Inform, 10 (**2012**) 74-80.

[50]   K. Thorsen, K.D. Sorensen, A.S. Brems-Eskildsen, C. Modin, M. Gaustadnes, A.M. Hein, M. Kruhoffer, S. Laurberg, M. Borre, K. Wang, S. Brunak, A.R. Krainer, N. Torring, L. Dyrskjot, C.L. An-

dersen, T.F. Orntoft, Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis, Mol Cell Proteomics, 7 (**2008**) 1214-1224.

[51]   C.A. Blair, X. Zi, Potential molecular targeting of splice variants for cancer treatment, Indian J Exp Biol, 49 (**2011**) 836-839.

[52]   L. Shkreta, U. Froehlich, E.R. Paquet, J. Toutant, S.A. Elela, B. Chabot, Anticancer drugs affect the alternative splicing of Bcl-x and other human apoptotic genes, Mol Cancer Ther, 7 (**2008**) 1398-1409.

[53]   S.T. Eblen, Regulation of chemoresistance via alternative messenger RNA splicing, Biochem Pharmacol, 83 (**2012**) 1063-1072.

[54]   C.L. Zheng, Y.S. Kwon, H.R. Li, K. Zhang, G. Coutinho-Mansfield, C. Yang, T.M. Nair, M. Gribskov, X.D. Fu, MAASE: an alternative splicing database designed for supporting splicing microarray applications, RNA, 11 (**2005**) 1767-1776.

[55]   C. Lee, L. Atanelov, B. Modrek, Y. Xing, ASAP: the Alternative Splicing Annotation Project, Nucleic Acids Res, 31 (**2003**) 101-105.

[56]   H. Wang, E. Hubbell, J.S. Hu, G. Mei, M. Cline, G. Lu, T. Clark, M.A. Siani-Rose, M. Ares, D.C. Kulp, D. Haussler, Gene structure-based splice variant deconvolution using a microarray platform, Bioinformatics, 19 Suppl 1 (**2003**) i315-322.

[57]   C. Lee, M. Roy, Analysis of alternative splicing with microarrays: successes and challenges, Genome Biol, 5 (**2004**) 231.

[58]   D.D. Shoemaker, E.E. Schadt, C.D. Armour, Y.D. He, P. Garrett-Engele, P.D. McDonagh, P.M. Loerch, A. Leonardson, P.Y. Lum, G. Cavet, L.F. Wu, S.J. Altschuler, S. Edwards, J. King, J.S. Tsang, G. Schimmack, J.M. Schelter, J. Koch, M. Ziman, M.J. Marton, B. Li, P. Cundiff, T. Ward, J. Castle, M. Krolewski, M.R. Meyer, M. Mao, J. Burchard, M.J. Kidd, H. Dai, J.W. Phillips, P.S. Linsley, R. Stoughton, S. Scherer, M.S. Boguski, Experimental annotation of the human genome using microarray technology, Nature, 409 (**2001**) 922-927.

[59]   A.A. Alizadeh, M.B. Eisen, R.E. Davis, C. Ma, I.S. Lossos, A. Rosenwald, J.C. Boldrick, H. Sabet, T. Tran, X. Yu, J.I. Powell, L. Yang, G.E. Marti, T. Moore, J. Hudson, Jr., L. Lu, D.B. Lewis, R. Tibshirani, G. Sherlock, W.C. Chan, T.C. Greiner, D.D. Weisenburger, J.O. Armitage, R. Warnke, R. Levy, W. Wilson, M.R. Grever, J.C. Byrd, D. Botstein, P.O. Brown, L.M. Staudt, Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling, Nature, 403 (**2000**) 503-511.

[60]   A. Hanamura, J.F. Caceres, A. Mayeda, B.R. Franza, Jr., A.R. Krainer, Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors, RNA, 4 (**1998**) 430-444.

[61]   E. Stickeler, F. Kittrell, D. Medina, S.M. Berget, Stage-specific changes in SR splicing factors and alternative splicing in mammary tumorigenesis, Oncogene, 18 (**1999**) 3574-3582.

[62]   C. Wang, K. Chua, W. Seghezzi, E. Lees, O. Gozani, R. Reed, Phosphorylation of spliceosomal protein SAP 155 coupled with splicing catalysis, Genes Dev, 12 (**1998**) 1409-1414.

[63]   M. Cuperlovic-Culf, N. Belacel, A.S. Culf, R.J. Ouellette, Data analysis of alternative splicing microarrays, Drug Discov Today, 11 (**2006**) 983-990.

[64]   C. Marwick, First "antisense" drug will treat CMV retinitis, JAMA, 280 (**1998**) 871.

[65]   S. Kawamoto, J. Yoshii, K. Mizuno, K. Ito, Y. Miyamoto, T. Ohnishi, R. Matoba, N. Hori, Y. Matsumoto, T. Okumura, Y. Nakao, H. Yoshii, J. Arimoto, H. Ohashi, H. Nakanishi, I. Ohno, J. Hashimoto, K. Shimizu, K. Maeda, H. Kuriyama, K. Nishida, A. Shimizu-Matsumoto, W. Adachi, R. Ito, S. Kawasaki, K.S. Chae, BodyMap: a collection of 3' ESTs for analysis of human gene expression information, Genome Res, 10 (**2000**) 1817-1827.

[66]   J.D. Cogan, M.A. Prince, S. Lekhakula, S. Bundey, A. Futrakul, E.M. McCarthy, J.A. Phillips, 3rd, A novel mechanism of aberrant pre-mRNA splicing in humans, Hum Mol Genet, 6 (**1997**) 909-912.

[67]   J.M. Yeakley, J.B. Fan, D. Doucet, L. Luo, E. Wickham, Z. Ye, M.S. Chee, X.D. Fu, Profiling alternative splicing on fiber-optic arrays, Nat Biotechnol, 20 (**2002**) 353-358.

[68]   T.I. Orban, E. Olah, Expression profiles of BRCA1 splice variants in asynchronous and in G1/S synchronized tumor cell lines, Biochem Biophys Res Commun, 280 (**2001**) 32-38.

[69]   G.K. Hu, S.J. Madore, B. Moldover, T. Jatkoe, D. Balaban, J. Thomas, Y. Wang, Predicting splice variant from DNA chip expression data, Genome Res, 11 (**2001**) 1237-1245.

[70]   L. Sun, C. Yu, J. Irudayaraj, Raman multiplexers for alternative gene splicing, Anal Chem, 80 (**2008**) 3342-3349.

[71]   D.S. Gerhard, L. Wagner, E.A. Feingold, C.M. Shenmen, L.H. Grouse, G. Schuler, S.L. Klein, S. Old, R. Rasooly, P. Good, M. Guyer, A.M. Peck, J.G. Derge, D. Lipman, F.S. Collins, W. Jang, S. Sherry, M. Feolo, L. Misquitta, E. Lee, K. Rotmistrovsky, S.F. Greenhut, C.F. Schaefer, K. Buetow, T.I. Bonner, D. Haussler, J. Kent, M. Kiekhaus, T. Furey, M. Brent, C. Prange, K. Schreiber, N. Shapiro, N.K. Bhat, R.F. Hopkins, F. Hsie, T. Driscoll, M.B. Soares, T.L. Casavant, T.E. Scheetz, M.J. Brown-stein, T.B. Usdin, S. Toshiyuki, P. Carninci, Y. Piao, D.B. Dudekula, M.S. Ko, K. Kawakami, Y. Suzuki, S. Sugano, C.E. Gruber, M.R. Smith, B. Simmons, T. Moore, R. Waterman, S.L. Johnson, Y. Ruan, C.L. Wei, S. Mathavan, P.H. Gunaratne, J. Wu, A.M. Garcia, S.W. Hulyk, E. Fuh, Y. Yuan, A. Sneed, C. Kowis, A. Hodgson, D.M. Muzny, J. McPherson, R.A. Gibbs, J. Fahey, E. Helton, M. Ketteman, A. Madan, S. Rodrigues, A. Sanchez, M. Whiting, A. Madari, A.C. Young, K.D. Wetherby, S.J. Granite, P.N. Kwong, C.P. Brinkley, R.L. Pearson, G.G. Bouffard, R.W. Blakesly, E.D. Green, M.C. Dickson, A.C. Rodriguez, J. Grimwood, J. Schmutz, R.M. Myers, Y.S. Butterfield, M. Griffith, O.L. Griffith, M.I. Krzywinski, N. Liao, R. Morin, D. Palmquist, A.S. Petrescu, U. Skalska, D.E. Smailus, J.M. Stott, A. Schnerch, J.E. Schein, S.J. Jones, R.A. Holt, A. Baross, M.A. Marra, S. Clifton, K.A. Makowski, S. Bosak, J. Malek, The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC), Genome Res, 14 (**2004**) 2121-2127.

[72]   M. Harbers, P. Carninci, Tag-based approaches for transcriptome research and genome annotation, Nat Methods, 2 (**2005**) 495-502.

[73]   R. Kodzius, M. Kojima, H. Nishiyori, M. Nakamura, S. Fukuda, M. Tagami, D. Sasaki, K. Imamura, C. Kai, M. Harbers, Y. Hayashizaki, P. Carninci, CAGE: cap analysis of gene expression, Nat Methods, 3 (**2006**) 211-222.

[74]   J. Reinartz, E. Bruyns, J.Z. Lin, T. Burcham, S. Brenner, B. Bowen, M. Kramer, R. Woychik, Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms, Brief Funct Genomic Proteomic, 1 (**2002**) 95-104.

[75]   Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, Nat Rev Genet, 10 (**2009**) 57-63.

[76]   A. Mortazavi, B.A. Williams, K. McCue, L. Schaeffer, B. Wold, Mapping and quantifying mammalian transcriptomes by RNA-Seq, Nat Methods, 5 (**2008**) 621-628.

[77]   N. Cloonan, A.R. Forrest, G. Kolle, B.B. Gardiner, G.J. Faulkner, M.K. Brown, D.F. Taylor, A.L. Steptoe, S. Wani, G. Bethel, A.J. Robertson, A.C. Perkins, S.J. Bruce, C.C. Lee, S.S. Ranade, H.E. Peckham, J.M. Manning, K.J. McKernan, S.M. Grimmond, Stem cell transcriptome profiling via massive-scale mRNA sequencing, Nat Methods, 5 (**2008**) 613-619.

[78]   W.B. Barbazuk, S.J. Emrich, H.D. Chen, L. Li, P.S. Schnable, SNP discovery via 454 transcriptome sequencing, Plant J, 51 (**2007**) 910-918.

[79]   J.C. Vera, C.W. Wheat, H.W. Fescemyer, M.J. Frilander, D.L. Crawford, I. Hanski, J.H. Marden, Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing, Mol Ecol, 17 (**2008**) 1636-1647.

[80]   R. Morin, M. Bainbridge, A. Fejes, M. Hirst, M. Krzywinski, T. Pugh, H. McDonald, R. Varhol, S. Jones, M. Marra, Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing, Biotechniques, 45 (**2008**) 81-94.

[81]   U. Nagalakshmi, Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein, M. Snyder, The transcriptional landscape of the yeast genome defined by RNA sequencing, Science, 320 (**2008**) 1344-1349.

[82]   L.W. Hillier, G.T. Marth, A.R. Quinlan, D. Dooling, G. Fewell, D. Barnett, P. Fox, J.I. Glasscock, M. Hickenbotham, W. Huang, V.J. Magrini, R.J. Richt, S.N. Sander, D.A. Stewart, M. Stromberg, E.F. Tsung, T. Wylie, T. Schedl, R.K. Wilson, E.R. Mardis, Whole-genome sequencing and variant discovery in C. elegans, Nat Methods, 5 (**2008**) 183-188.

[83]   E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J.P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, J. Clee, N. Carter, A. Coulson, R.

Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J.C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R.H. Waterston, R.K. Wilson, L.W. Hillier, J.D. McPherson, M.A. Marra, E.R. Mardis, L.A. Fulton, A.T. Chinwalla, K.H. Pepin, W.R. Gish, S.L. Chissoe, M.C. Wendl, K.D. Delehaunty, T.L. Miner, A. Delehaunty, J.B. Kramer, L.L. Cook, R.S. Fulton, D.L. Johnson, P.J. Minx, S.W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R.A. Gibbs, D.M. Muzny, S.E. Scherer, J.B. Bouck, E.J. Sodergren, K.C. Worley, C.M. Rives, J.H. Gorrell, M.L. Metzker, S.L. Naylor, R.S. Kucherlapati, D.L. Nelson, G.M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D.R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H.M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R.W. Davis, N.A. Federspiel, A.P. Abola, M.J. Proctor, R.M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D.R. Cox, M.V. Olson, R. Kaul, N. Shimizu, K. Kawasaki, S. Minoshima, G.A. Evans, M. Athanasiou, R. Schultz, B.A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W.R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J.A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D.G. Brown, C.B. Burge, L. Cerutti, H.C. Chen, D. Church, M. Clamp, R.R. Copley, T. Doerks, S.R. Eddy, E.E. Eichler, T.S. Furey, J. Galagan, J.G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L.S. Johnson, T.A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W.J. Kent, P. Kitts, E.V. Koonin, I. Korf, D. Kulp, D. Lancet, T.M. Lowe, A. McLysaght, T. Mikkelsen, J.V. Moran, N. Mulder, V.J. Pollara, C.P. Ponting, G. Schuler, J. Schultz, G. Slater, A.F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y.I. Wolf, K.H. Wolfe, S.P. Yang, R.F. Yeh, F. Collins, M.S. Guyer, J. Peterson, A. Felsenfeld, K.A. Wetterstrand, A. Patrinos, M.J. Morgan, P. de Jong, J.J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y.J. Chen, Initial sequencing and analysis of the human genome, Nature, 409 (**2001**) 860-921.

[84]   J. Leipzig, P. Pevzner, S. Heber, The Alternative Splicing Gallery (ASG): bridging the gap between genome and transcriptome, Nucleic Acids Res, 32 (**2004**) 3977-3983.

[85]   D. Holste, G. Huo, V. Tung, C.B. Burge, HOLLYWOOD: a comparative relational database of alternative splicing, Nucleic Acids Res, 34 (**2006**) D56-62.

[86]   S. Stamm, J.J. Riethoven, V. Le Texier, C. Gopalakrishnan, V. Kumanduri, Y. Tang, N.L. Barbosa-Morais, T.A. Thanaraj, ASD: a bioinformatics resource on alternative splicing, Nucleic Acids Res, 34 (**2006**) D46-55.

[87]   H. Ji, Q. Zhou, F. Wen, H. Xia, X. Lu, Y. Li, AsMamDB: an alternative splice database of mammals, Nucleic Acids Res, 29 (**2001**) 260-263.

[88]   N. Kim, S. Shin, S. Lee, ASmodeler: gene modeling of alternative splicing from genomic alignment of mRNA, EST and protein sequences, Nucleic Acids Res, 32 (**2004**) W181-186.

[89]   G. Koscielny, V. Le Texier, C. Gopalakrishnan, V. Kumanduri, J.J. Riethoven, F. Nardone, E. Stanley, C. Fallsehr, O. Hofmann, M. Kull, E. Harrington, S. Boue, E. Eyras, M. Plass, F. Lopez, W. Ritchie, V. Moucadel, T. Ara, H. Pospisil, A. Herrmann, G.R. J, R. Guigo, P. Bork, M.K. Doeberitz, J. Vilo, W. Hide, R. Apweiler, T.A. Thanaraj, D. Gautheret, ASTD: The Alternative Splicing and Transcript Diversity database, Genomics, 93 (**2009**) 213-220.

[90]   H.D. Huang, J.T. Horng, C.C. Lee, B.J. Liu, ProSplicer: a database of putative alternative splicing information derived from protein, mRNA and expressed sequence tag sequence data, Genome Biol, 4 (**2003**) R29.

[91]   H. Nagasaki, M. Arita, T. Nishizawa, M. Suwa, O. Gotoh, Automated classification of alternative splicing and transcriptional initiation and construction of visual database of classified patterns, Bioinformatics, 22 (**2006**) 1211-1216.

[92]   H.D. Huang, J.T. Horng, F.M. Lin, Y.C. Chang, C.C. Huang, SpliceInfo: an information repository for mRNA alternative splicing in human genome, Nucleic Acids Res, 33 (**2005**) D80-85.

[93]   A. Krause, S.A. Haas, E. Coward, M. Vingron, SYSTERS, GeneNest, SpliceNest: exploring sequence space from genome to protein, Nucleic Acids Res, 30 (**2002**) 299-300.

[94]   Y.H. Huang, Y.T. Chen, J.J. Lai, S.T. Yang, U.C. Yang, PALS db: Putative Alternative Splicing database, Nucleic Acids Res, 30 (**2002**) 186-190.

[95]   J. Takeda, Y. Suzuki, M. Nakao, T. Kuroda, S. Sugano, T. Gojobori, T. Imanishi, H-DBAS: alternative splicing database of completely sequenced and manually annotated full-length cDNAs based on H-Invitational, Nucleic Acids Res, 35 (**2007**) D104-109.

[96]   Y. Lee, B. Kim, Y. Shin, S. Nam, P. Kim, N. Kim, W.H. Chung, J. Kim, S. Lee, ECgene: an alternative splicing database update, Nucleic Acids Res, 35 (**2007**) D99-103.

[97]   J. Takeda, C. Yamasaki, K. Murakami, Y. Nagai, M. Sera, Y. Hara, N. Obi, T. Habara, T. Gojobori, T. Imanishi, H-InvDB in 2013: an omics study platform for human functional gene and transcript discovery, Nucleic Acids Res, 41 (**2013**) D915-919.

[98]   S. Foissac, M. Sammeth, ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets, Nucleic Acids Res, 35 (**2007**) W297-299.

[99]   T. Castrignano, M. D'Antonio, A. Anselmo, D. Carrabino, A. D'Onorio De Meo, A.M. D'Erchia, F. Licciulli, M. Mangiulli, F. Mignone, G. Pavesi, E. Picardi, A. Riva, R. Rizzi, P. Bonizzoni, G. Pesole, ASPicDB: a database resource for alternative splicing analysis, Bioinformatics, 24 (**2008**) 1300-1304.

[100] E. Buratti, M. Chivers, G. Hwang, I. Vorechovsky, DBASS3 and DBASS5: databases of aberrant 3'- and 5'-splice sites, Nucleic Acids Res, 39 (**2011**) D86-91.

[101] N. Kim, D. Lim, S. Lee, H. Kim, ASePCR: alternative splicing electronic RT-PCR in multiple tissues and organs, Nucleic Acids Res, 33 (**2005**) W681-685.

[102] A. ElSharawy, C. Manaster, M. Teuber, P. Rosenstiel, R. Kwiatkowski, K. Huse, M. Platzer, A. Becker, P. Nurnberg, S. Schreiber, J. Hampe, SNPSplicer: systematic analysis of SNP-dependent splicing in genotyped cDNAs, Hum Mutat, 27 (**2006**) 1129-1134.

[103] Q. Liu, C. Chen, E. Shen, F. Zhao, Z. Sun, J. Wu, Detection, annotation and visualization of alternative splicing from RNA-Seq data with SplicingViewer, Genomics, 99 (**2012**) 178-182.

[104] M.F. Rogers, J. Thomas, A.S. Reddy, A. Ben-Hur, SpliceGrapher: detecting patterns of alternative splicing from RNA-Seq data in the context of gene models and EST data, Genome Biol, 13 (**2012**) R4.

[105] C. Trapnell, L. Pachter, S.L. Salzberg, TopHat: discovering splice junctions with RNA-Seq, Bioinformatics, 25 (**2009**) 1105-1111.

[106] S. Shen, J.W. Park, J. Huang, K.A. Dittmar, Z.X. Lu, Q. Zhou, R.P. Carstens, Y. Xing, MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data, Nucleic Acids Res, 40 (**2012**) e61.

[107] Y. Zhang, E.W. Lameijer, P.A. t Hoen, Z. Ning, P.E. Slagboom, K. Ye, PASSion: a pattern growth algorithm-based pipeline for splice junction detection in paired-end RNA-Seq data, Bioinformatics, 28 (**2012**) 479-486.

[108] J. Wu, M. Akerman, S. Sun, W.R. McCombie, A.R. Krainer, M.Q. Zhang, SpliceTrap: a method to quantify alternative splicing under single cellular conditions, Bioinformatics, 27 (**2011**) 3010-3016.

[109] D. Emig, N. Salomonis, J. Baumbach, T. Lengauer, B.R. Conklin, M. Albrecht, AltAnalyze and DomainGraph: analyzing and visualizing exon expression data, Nucleic Acids Res, 38 (**2010**) W755-762.

[110] J. Seok, W. Xu, H. Gao, R.W. Davis, W. Xiao, JETTA: junction and exon toolkits for transcriptome analysis, Bioinformatics, 28 (**2012**) 1274-1275.

[111] N.A. Twine, K. Janitz, M.R. Wilkins, M. Janitz, Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease, PLoS One, 6 (**2011**) e16266.

[112] M. Wang, A. Marin, Characterization and prediction of alternative splice sites, Gene, 366 (**2006**) 219-227.

[113] M. Pertea, X. Lin, S.L. Salzberg, GeneSplicer: a new computational method for splice site prediction, Nucleic Acids Res, 29 (**2001**) 1185-1190.

[114] K. Wang, D. Singh, Z. Zeng, S.J. Coleman, Y. Huang, G.L. Savich, X. He, P. Mieczkowski, S.A. Grimm, C.M. Perou, J.N. MacLeod, D.Y. Chiang, J.F. Prins, J. Liu, MapSplice: accurate mapping of RNA-seq reads for splice junction discovery, Nucleic Acids Res,

38 (**2010**) e178.

[115]   Y. Katz, E.T. Wang, E.M. Airoldi, C.B. Burge, Analysis and design of RNA sequencing experiments for identifying isoform regulation, Nat Methods, 7 (**2010**) 1009-1015.

[116]   B.J. Haas, A.L. Delcher, S.M. Mount, J.R. Wortman, R.K. Smith, Jr., L.I. Hannick, R. Maiti, C.M. Ronning, D.B. Rusch, C.D. Town, S.L. Salzberg, O. White, Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies, Nucleic Acids Res, 31 (**2003**) 5654-5666.

[117]   M.A. Campbell, B.J. Haas, J.P. Hamilton, S.M. Mount, C.R. Buell, Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis, BMC Genomics, 7 (**2006**) 327.

[118]   S. Anders, A. Reyes, W. Huber, Detecting differential usage of exons from RNA-seq data, Genome Res, 22 (**2012**) 2008-2017.

[119]   M.T. Dimon, K. Sorber, J.L. DeRisi, HMMSplicer: a tool for effi-

cient and sensitive discovery of known and novel splice junctions in RNA-Seq data, PLoS One, 5 (**2010**) e13875.

[120]   S. Stamm, Signals and their transduction pathways regulating alternative splicing: a new dimension of the human genome, Hum Mol Genet, 11 (**2002**) 2409-2416.

[121]   Q. Pan, M.A. Bakowski, Q. Morris, W. Zhang, B.J. Frey, T.R. Hughes, B.J. Blencowe, Alternative splicing of conserved exons is frequently species-specific in human and mouse, Trends Genet, 21 (**2005**) 73-77.

[122]   P. Wang, B. Yan, J.T. Guo, C. Hicks, Y. Xu, Structural genomics analysis of alternative splicing and application to isoform structure modeling, Proc Natl Acad Sci U S A, 102 (**2005**) 18920-18925.

[123]   Y. Xing, C. Lee, Alternative splicing and RNA selection pressure--evolutionary consequences for eukaryotic genomes, Nat Rev Genet, 7 (**2006**) 499-509.