

Biomedical Impact of Splicing Mutations Revealed through Exome Sequencing

Bahar Taneri,^{1,2} Esra Asilmaz,³ and Terry Gaasterland⁴

¹Institute of Public Health Genomics, Department of Genetics and Cell Biology, Research Institutes CAPHRI and GROW, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, the Netherlands; ²Department of Biological Sciences, Faculty of Arts and Sciences, Eastern Mediterranean University, Famagusta, North Cyprus; ³King's College London, Guy's and St. Thomas' Hospital NHS Foundation Trust, London, United Kingdom; and ⁴University of California San Diego, Scripps Genome Center, Scripps Institution of Oceanography, La Jolla, California, United States of America

Splicing is a cellular mechanism, which dictates eukaryotic gene expression by removing the noncoding introns and ligating the coding exons in the form of a messenger RNA molecule. Alternative splicing (AS) adds a major level of complexity to this mechanism and thus to the regulation of gene expression. This widespread cellular phenomenon generates multiple messenger RNA isoforms from a single gene, by utilizing alternative splice sites and promoting different exon-intron inclusions and exclusions. AS greatly increases the coding potential of eukaryotic genomes and hence contributes to the diversity of eukaryotic proteomes. Mutations that lead to disruptions of either constitutive splicing or AS cause several diseases, among which are myotonic dystrophy and cystic fibrosis. Aberrant splicing is also well established in cancer states. Identification of rare novel mutations associated with splice-site recognition, and splicing regulation in general, could provide further insight into genetic mechanisms of rare diseases. Here, disease relevance of aberrant splicing is reviewed, and the new methodological approach of starting from disease phenotype, employing exome sequencing and identifying rare mutations affecting splicing regulation is described. Exome sequencing has emerged as a reliable method for finding sequence variations associated with various disease states. To date, genetic studies using exome sequencing to find disease-causing mutations have focused on the discovery of nonsynonymous single nucleotide polymorphisms that alter amino acids or introduce early stop codons, or on the use of exome sequencing as a means to genotype known single nucleotide polymorphisms. The involvement of splicing mutations in inherited diseases has received little attention and thus likely occurs more frequently than currently estimated. Studies of exome sequencing followed by molecular and bioinformatic analyses have great potential to reveal the high impact of splicing mutations underlying human disease.

Online address: <http://www.molmed.org>
doi: 10.2119/molmed.2011.00126

INTRODUCTION

Splicing and alternative splicing (AS) have been a major area of focus in the post-genome era. These important cellular mechanisms are not only considered to be significantly involved in evolution of phenotypic complexity in mammals, but also have prominent roles in the physiology and disease of organisms. Splicing is the cellular mechanism for re-

moval of introns and ligation of exons during the expression of multiexon eukaryotic genes. Constitutive splicing involves removal of an intron and ligation of its flanking exons in every mature transcript for a given gene. AS, on the other hand, processes a given intron in more than one way and leads to multiple transcript isoforms from the same genomic region, by making use of alterna-

tive splice sites. The majority of human genes have multiple exons, and more than 90% have been observed to exhibit AS (1). AS provides a means for differential messenger RNA (mRNA) generation from the same gene, via varying exon-intron exclusion and inclusion events. Major versions of AS include skipped exons, retained introns, mutually exclusive exons, alternative 5' splice sites, and alternative 3' splice sites (2).

Address correspondence to Bahar Taneri, Department of Biological Sciences, Faculty of Arts and Sciences, Eastern Mediterranean University, Famagusta, North Cyprus, Mersin 10, Turkey. Phone: +90 392 630 2477; Fax: +90 392 365 1604; E-mail: bahar.taneri@emu.edu.tr. Submitted April 4, 2011; Accepted for publication November 29, 2011; Epub (www.molmed.org) ahead of print December 1, 2011.

MECHANISM AND REGULATION OF SPLICING

The splicing process is carried out by a dynamic RNA-protein complex known as the spliceosome. Recognition of exon-intron boundaries by the spliceosome depends on four major signals in the precursor mRNA (pre-mRNA) sequence. These signals are the 5' and 3' splice junc-

tions of the intron, the branch site and the polyrimidine tract located close to the 3' splice site of the intron (3). Nonetheless, for exon and intron definition, these four splice signals alone are not sufficient. There are several other *cis*-acting sequence elements within the exons and the introns that signal for binding of specific *trans*-acting splice factors, which influence the selection of splice junctions. Most RNA-binding domains recognize short *cis*-acting sequences, and they are guided to the RNA molecule via clusters of such short sequences (4). These regulatory sequences, which could be enhancers or silencers of splicing, recruit specific *trans*-acting splicing factors, such as serine/arginine-rich protein (SR) proteins (2) or neuron-specific splicing factor (NOVA) (4) and guide the localization of the spliceosome. The complex network of regulatory *cis*- and *trans*-acting factors are involved both in constitutive splicing and in AS.

A further level of complexity of AS regulation lies in its tissue specificity. It has been documented that different splice variants transcribed from the same gene exhibit tissue specificity (5). In addition, even though the majority of RNA-binding proteins (RBPs) are expressed ubiquitously, a certain number of them are tissue specific and regulate AS in this manner (4, 6). Tissue-specific expression of mouse transcription factor (TF) splice forms is a good example of both the presence of tissue-specific mRNA isoforms and of the added level of gene expression regulation via alternatively spliced TFs (5). Neuron-specific splicing regulatory protein NOVA provides a specific example, by which presence of tissue-specific RBPs provide a mechanism for generating tissue specific isoforms. Identified NOVA regulated target genes by far have been those that encode proteins working at the synapse, further indicating the tissue-specific function of AS (4).

AS is sensitive also to developmental stage. Many isoforms are expressed in a developmental-stage-dependent manner and function differently across development. Tissue-specific expression of different splice forms mediated temporally

regulates many developmental processes. Sex determination is a well-established example of splicing-mediated developmental process in *Drosophila*. RBPs alternating particular mRNA isoforms regulate this developmental process via regulation of splicing (7).

It has also been documented that AS changes with physiological conditions. The switch between acetylcholinesterase (AChE) enzyme's two splice forms, *AChE-R* and *AChE-S* isoforms, under the influence of stress provides an interesting example (8). Under acute stress, expression of *AChE* shifts to the *AChE-R* isoform, which is not abundant in neuronal cells under normal conditions. It is suggested that *AChE-R* isoform, which unlike *AChE-S* isoform excludes exon 6 and retains an intron, has a neuronal protective role during stress (8).

Epigenetic mechanisms are also at play in AS regulation. There is accumulating evidence that histone modifications and chromatin remodeling are involved in splicing regulation (9). As reviewed by Luco *et al.*, alternative exon inclusion–exclusion has been shown to be affected by the chromatin structure (9). Nucleosomes have been shown to be enriched in exon–intron junctions, and hence, are thought to be implicated in exon definition. Further evidence for chromatin structure involvement in AS comes from the documentation that nucleosomes are more densely found around alternatively spliced exons (9). In addition, histone modifications are thought to function as splicing regulators (9). It has been shown that histone acetyltransferases and methyltransferases interact with certain small nuclear ribonucleoproteins (snRNPs), implying their roles in spliceosome assembly (9).

As discussed above, regulation of splicing and AS have many complex layers and involves a network of factors. Since splicing and AS affect the majority of human genes and therefore tissues and development, proper regulation of splicing and AS is essential to healthy physiology.

SPlicing AND DISEASE

AS accounts for 95% of multiexon human genes (10) and 75% of alternatively spliced exons encode protein parts (11). Therefore, AS significantly alters the coding potential of the mRNAs. In addition, AS can alter 3' untranslated regions (UTRs) in ways that influence gene expression via nonsense-mediated decay (NMD). On one hand, if AS introduces an early stop codon, it thus alters which part of the mRNA transcript remains as 3' UTR. A 3' UTR that contains an exon splice junction further than 50 nucleotides downstream of the stop codon or an excessively long 3' UTR even without splice junction may be subjected to NMD, rendering the mRNA unavailable for translation. On the other hand, if AS removes an optional intron within the 3' UTR, this will introduce an exon splice junction after the stop codon and lead to NMD (12). Therefore, it is essential to consider the disease relevance of AS that modifies 3' UTR, as well as AS that modifies coding regions. Defects in splicing cover a wide spectrum of genes, and therefore cause a wide range of common and rare diseases. In fact, there is higher rate of mutation accumulation in alternative regions of genes compared with constitutive regions. The nonsynonymous substitution rate has been reported to be higher in alternative versus constitutive exons in mammalian genes (13). Conversely, synonymous substitution rate is lower in alternative exons compared with constitutive exons (14). Similar results showing more rapid accumulation of nucleotide substitutions in alternative exons are reported in other organisms as well (15). Aberrant splicing of constitutive and alternative exons has been reported in severe human diseases such as cystic fibrosis and spinal muscular atrophy. As summarized by Caceres and Kornblihtt, various genes from which aberrant forms are spliced are implicated in different disease states including *cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7) (CFTR)* in cystic fibrosis, *microtubule-associated protein tau (tau)* in

frontotemporal dementia, *coagulation factor VIII (factor VIII)* in hemophilia, *fibriillin 1* in Marfan syndrome, *neurofibromin 1 (NF1)* in neurofibromatosis, *survival of motor neuron 1 (SMN1)* and *survival of motor neuron 2 (SMN2)* in spinal muscular atrophy, and *Fanconi anemia complementation group G (FANCG)* in Fanconi anemia (16).

It has been established that in cancer, AS is fundamentally affected. In cancer cells, not only the levels of different alternative splice forms change, but also some splice forms are documented to be cancer tissue specific (17). A cancer-specific exon in *misshaped-like kinase 1 (MINK1)* transcripts implicated in ovarian cancer is an example for this case (17). Aberrant splicing has also been implicated in *breast cancer 1, early onset (BRCA1)* and *breast cancer 2, early onset (BRCA2)* genes, which are implicated in breast and ovarian cancers (18). Involvement of altered spliced patterns in tumor development has been documented for many cancer states including colon cancer, small cell lung cancer, gastric, cervical and thyroid tumors (11). Furthermore, it has been shown that the splicing machinery plays an integral role in oncogenesis. For example, in various human cancers SF2/ASF, a splicing factor, has been shown to be overexpressed (19).

Interestingly, altered spliced patterns due to single nucleotide polymorphisms (SNPs) could also generate sex-specific and developmental-stage specific disease cases. For example, a particular exon skipping in *low density lipoprotein receptor (LDLR)* gene due to presence of a SNP generates a splicing pattern, which leads to hypercholesterolemia, in premenopausal women only (11).

Several extensive reviews are available on AS and disease states, which detail mutations in *cis*-acting and *trans*-acting splicing factors, and discuss targeted therapeutic approaches in the field (11,20,21).

EXOME SEQUENCING TECHNOLOGY

A recent and novel methodological approach in genetic diagnosis of rare

Mendelian disease states uses exome sequencing technology (22). Starting from the disease phenotypes, identification of novel mutations through exome sequencing provides clues into the genetic basis of certain diseases. In fact, exome sequencing has been suggested as a reliable method for finding sequence variations associated with various disease states (23).

The success of exome sequencing or targeted exome capture comes from the fact that this method is both a cost-effective and a time-efficient alternative to the whole genome sequencing (22,24,25). As discussed in detail in the following section, in recent years this technology yielded the identification of various novel mutations causing rare Mendelian conditions. Ng *et al.* provide a detailed overview of the type of sequence variation interrogated by exome sequencing, which go beyond common coding SNPs and cover rare variants. Variations investigated by exome sequencing include nonsense mutations, insertions-deletions and splice-site disruptions (26).

In addition to identification of rare variants in rare diseases, exome sequencing could also be of use in identification of rare variants in common diseases, in large, adequately powered cohorts of case and control individuals (27). Some examples where exome sequencing proved useful in identifying mutations as causes of common disorders include high myopia (28) and prostate cancers (29). In their study, covering seven different common diseases including Crohn's disease, bipolar disorder and Type 1 diabetes, Lehne *et al.* were able to identify genetic variations and suggest exome sequencing as a feasible and cost-effective technology in identifying variants relevant to common complex diseases (23).

Exome sequencing is not without limitations. Firstly, new statistical measures are needed to assess the statistical significance of the identified rare mutations (22,30). Secondly, and more importantly for splicing purposes, exome sequencing is limited in that with this technology

only the exons of the genome are sequenced (31). Because only 1% of the whole genome is translated, variants in the untranslated regions, which might be functionally significant would be missed by this method. For example, untranslated regions, which flank the exons are usually not included in exome sequencing. With this approach, only the coding region of the genomes are sequenced even though the uncoding regions are likely to be functionally significant as regulatory sequences (31). Ku *et al.* offer a comprehensive review on exome sequencing technology and its limitations (31).

The RNA-Seq technology, which has been shown to be a fast and a cost-effective alternative approach in gene variant identification (32), could be used to complement the exome sequencing in detecting splice variations. RNA-Seq has been suggested as an alternative method for quantification of alternative and abundant transcripts, as well as for the detection of variants that can potentially affect mature transcript structure. A recent study by Montgomery *et al.*, which is one of the first studies employing second generation sequencing to investigate genetic variation on the transcriptome level, provides a good example (33). In this study, Montgomery *et al.* used second generation sequencing in a Caucasian population where they sequenced mRNA fraction of the transcriptome of 60 individuals to quantify allele specific isoforms (33). RNA-Seq not only provides a good quantitation of gene and isoform expression, but also an estimate of coding sequence variation and allele-specific expression. The authors investigated AS in detail and detected a significant number of variants that influence the structure of mature transcripts (33). They discovered a range of genetic factors affecting isoform abundance and showed heterogeneity in the transcript distribution due to alternative internal exons and alternative 3' ends. In addition, they quantified AS events and were able to assess the effect of genetic variants that contributed to alternative iso-

forms. The authors showed that the majority of alternative isoforms were due to exon skipping. They also showed that alternate acceptors, alternate donors, mutually exclusive exons and retained introns contributed to a lesser degree to the alternatively spliced isoforms detected (33). This study by Montgomery *et al.* provides a direction in that employing RNA-Seq as a complement to exome sequencing could prove of significant use for AS studies.

EXOME SEQUENCING AND IDENTIFIED NOVEL MUTATIONS

In recent years, genome-wide sequencing of all protein-coding regions with the goal of mutation identification has been employed by various groups. One of the very first such studies published in 2009 by Choi *et al.* identified a rare homozygous missense mutation in *solute carrier family 26, member 3 (SLC26A3)*, yielding to genetic diagnosis of a congenital chloride diarrhea condition in a patient (34). In the last 2 years, a rapid increase has been seen in the biomedical literature covering exome sequencing, mutation identification and genetic diagnosis. To provide some examples, in recent studies, exome sequencing was used to identify underlying mutations for Charcot-Marie-Tooth disease (35), autosomal recessive nonsyndromic mental retardation (36), inflammatory bowel disease (37), familial amyotrophic lateral sclerosis (ALS) (38), spinocerebral ataxias (39), familial combined hypolipidemia (40), Brown-Vialetto-van Laere syndrome (41), Hajdu-Cheney syndrome (42) and osteogenesis imperfecta (43). In the next section, we provide examples of how exome sequencing mediated unraveling of novel mutations, leading to disrupted splicing and disease states.

EXOME SEQUENCING AND SPLICING MUTATIONS

Mutations that affect mRNA splicing patterns alter gene expression and hence could result in disease phenotypes (11,20,21). It is documented that many SNPs are located in splicing regulatory

sequence regions. These cover not only the canonical splice sites but also exonic and intronic splicing enhancer sequences (44). Rare variations in the genome could affect splice-site usage, and therefore, could alter expression of a specific gene with possible implications in disease states. Single base pair (bp) substitutions at splice junctions account for a minimum of 10% inherited disease causing mutations in humans (45). Krawczak *et al.* estimate 1.6% of such substitutions to have an impact on the splicing regulation of the transcripts coded by that gene (45).

On the basis of the prevalence of AS within the genome, the impact of alternative exons on protein coding, and the estimates of many SNPs affecting the splicing regulatory sequences, exome sequencing stands out as a methodology with potential to reveal novel splicing mutations as underlying causes of both rare and common diseases. By utilizing exome sequencing, rare mutations affecting splice-site usage can be identified. Such mutations affect mRNA architecture and gene expression, and could result in disease states. Two recent studies reveal this potential. Byun *et al.* applied exome sequencing and described a splice-site mutation in *stromal interaction molecule 1 (STIM1)* gene, which leads to Kaposi sarcoma (46). Similarly, a splice-site mutation in *WD repeat domain 35 (WDR35)* gene identified in a recent study by Gilissen *et al.* has been implicated in Sensenbrenner syndrome (47).

Byun *et al.* applied whole-exome sequencing for the genetic diagnosis of a particular Kaposi sarcoma case. In this study, the authors were able to identify the single base substitution mutation in *STIM1* gene through massively parallel sequencing of the exome from a single affected individual. In healthy cells, *STIM1* gene codes for an ER-resident transmembrane protein, which acts as a sensor for calcium stores in the endoplasmic reticulum (46). Through exome sequencing, Byun *et al.* identified a rare splice-site mutation at a consensus splice acceptor site at the intron–exon junction of exon 8 of *STIM1*. Specifically, a G to A substitu-

tion was observed at –1 position 5' of the exon. Investigation of this base change in control subjects, SNP database (dbSNP) records, the 1000 Genomes Project, as well as the authors' own in-house exome database revealed this mutation to be a rare allele (46). As a result of this splice-site mutation, transcripts of *STIM1* containing consecutive exons 7, 8 and 9 were completely absent from this affected individual's transcriptome, resulting in the absence of *STIM1* protein in affected cells. Instead, several different aberrant splice variants of this gene were detected in affected cells. The most abundant aberrant splice variants included a truncated version of exon 8 (missing 64 bp), intron retention upstream of exon 8 (addition of 101 bp, 99 bp, or 88 bp), and exon 8 skipping. However, none of these transcripts yielded in translation of the wild-type *STIM1* protein in the patient's cells (46). Identification of the autosomal recessive *STIM1* deficiency caused by a splice-site mutation from a single case (46) reveals the power of whole-exome sequencing, when supplemented by molecular assays and bioinformatic analyses, in identification of rare and novel splicing mutations.

In another recent study, Gilissen *et al.* identified a novel splice-site mutation implicated in Sensenbrenner syndrome, an autosomal recessive disease also known as cranioectodermal dysplasia (CED) (47). Phenotype of this syndrome could vary from individual to individual but generally includes skeletal, facial and ectodermal abnormalities (47). Working with the exome sequence data from two unrelated Sensenbrenner syndrome patients, Gilissen *et al.* identified a rare splice-site mutation in exon 2 of the *WDR35* gene in one of the affected individuals. This gene is homologous to *tubby like protein 4 (TULP4)*, which is an intraflagellar transport component implicated in ciliary function (47). The authors further showed that the splice-site mutation yielded an mRNA product with a 58 bp insertion, which contained a premature stop codon. In addition to this specific splice-site mutation, they identified two other missense mutations in *WDR35* gene as well as a

single nucleotide deletion, which leads to a frameshift mutation and premature termination. In a syndrome with a highly heterogeneous phenotype, such as Sensenbrenner syndrome, it is not surprising that they did not observe any causative mutations in *WDR35* gene in 6 additional cases with CED (47). Similar to the Byun *et al.* study, this study shows that exome sequencing is still useful in identification of causative genes even in a heterogeneous syndrome in the presence of as few as two affected individuals with very similar phenotypes. Both studies show the implication of mutated splice sites in rare disorders and their effective identification through exome sequencing (46,47).

DISCUSSION

Whole-exome sequencing is a recent innovation and has gained traction in the past year as a means to discover new mutations underlying disease. Exome sequencing is a novel and alternative approach to identify genetic causes for rare human genetic diseases. Instead of using known SNPs to search statistically for genome regions harboring disease-causing changes and then sequencing those regions in many subjects, exome sequencing allows a directed search for deleterious mutations. First publications in the field appearing in 2009, and the subsequent literature suggest the utility of this technology in identification of rare mutations. Here, we look at a subset of the identified mutations and focus on the importance of exome sequencing in identifying rare mutations affecting splice sites. Rare mutations that alter splice-site usage or splicing regulation in general change the mRNA exon-intron architecture and hence have a strong potential to be implicated in rare disease states (46,47).

It should be noted that natural variation identified across exome sequencing studies is likely to yield more and more information about splice-site disruption and splicing modifying sequence variation. For example, Ng and colleagues discovered a number of new splice-site

disrupting variants in their parallel sequencing of exomes from 12 humans (26). Ng *et al.* performed targeted capture and massively parallel sequencing of these 12 exomes, eight of which were HapMap participants and four of which were unrelated individuals with Freeman-Sheldon syndrome, a rare autosomal dominant disorder. They covered 300 Mb of protein-coding regions and uncovered both rare and common sequence variants. Novel coding SNPs were identified, which were not included in dbSNP; 49 of which were novel splice-site disrupting SNPs, indicating that uncommon sequence variations are likely to have impacts on splicing regulation (26).

Involvement of splicing mutations in inherited diseases is likely to be more prevalent than observed in studies to date. Studies of exome sequencing followed by molecular and bioinformatic analyses will soon reveal the high impact of splicing mutations in such disease states. Starting from disease phenotypes, identification of novel rare mutations leading to splicing pattern changes via exome sequencing will no doubt reveal more insight into defective splicing in disease states. Since many SNPs are located within splicing regulatory regions and as AS affects almost all human genes, it is important to pay attention to the potential of splicing mutation discovery through exome sequencing within the context of genetic diagnosis of rare diseases.

DISCLOSURE

The authors declare that they have no competing interests as defined by *Molecular Medicine*, or other interests that might be perceived to influence the results and discussion reported in this paper.

REFERENCES

1. Wang ET, *et al.* (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*. 456:470–6.
2. Cartegni L, Chew SL, Krainer AR. (2002) Listening to silence and understanding nonsense: Exonic mutations that affect splicing. *Nat. Rev. Genet.* 3:285–98.

3. Kim E, Goren A, Ast G. (2007) Alternative splicing: current perspectives. *BioEssays*. 30:38–47.
4. Darnell RB. (2007) Developing global insight into RNA regulation. Cold Spring Harbor Symposia on Quantitative Biology. LXXI:1–7.
5. Taneri B, Snyder B, Novoradovsky A, Gaasterland T. (2004) Alternative splicing of mouse transcription factors affects their DNA-binding domain architecture and is tissue specific. *Gen. Biol.* 5:R75.
6. Nilsen TW, Graveley BR. (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature*. 463:457–462.
7. Black DL. (2003) Mechanisms of alternative pre-mRNA splicing. *Annu. Rev. Biochem.* 72:291–336.
8. Marden JH. (2006) Quantitative and evolutionary biology of alternative splicing: how changing the mix of alternative transcripts affects phenotypic plasticity and reaction norms. *Heredity*. 100:111–20.
9. Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T. (2011) Epigenetics in alternative pre-mRNA splicing. *Cell*. 144:16–26.
10. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40:1413–15.
11. Tazi J, Bakkour N, Stamm S. (2008) Alternative splicing and disease. *Biochim. Biophys. Acta*. 1792:14–25.
12. McGlincy NJ, Smith CWJ. (2008) Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends Biochem. Sci.* 33:385–93.
13. Iida K, Akashi H. (2000) A test of translational selection at 'silent' sites in the human genome: base composition comparisons in alternatively spliced genes. *Gene*. 261:93–105.
14. Parmley JL, Chamary JV, Hurst LD. (2006) Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. *Mol. Biol. Evol.* 23:301–9.
15. Ermakova EO, Mal'ko DB, Gelfand MS. (2006) Evolutionary differences between alternative and constitutive protein-coding regions of alternatively spliced genes of *Drosophila*. *Mol. Biophys.* 51:515–22.
16. Caceres JF, Kornblihtt AR. (2002). Alternative splicing: Multiple control mechanisms and involvement in human disease. *Trends Gene.* 18:186–93.
17. Venables JP, *et al.* (2009) Cancer-associated regulation of alternative splicing. *Nature Struct. Mol. Biol.* 16:670–6.
18. Sanz DJ, *et al.* (2010) A high proportion of DNA variant of BRCA1 and BRCA2 is associated with aberrant splicing in breast/ovarian cancer patients. *Clin. Cancer Res.* 16:1957–67.
19. Karni R, *et al.* (2010) The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nature Struct. Mol. Bio.* 14:185–93.
20. Garcia-Blanco MA, Baraniak AP, Lasda EL. (2004) Alterantive splicing in disease and therapy. *Nat. Biotechnol.* 22:535–46.
21. Wang G-S, Cooper TA. (2007) Splicing in disease:

- disruption of the splicing code and the decoding machinery. *Nature Rev. Genet.* 8:749–61.
22. Biesecker LG. (2010) Exome sequencing makes medical genomics a reality. *Nature Gen.* 42:13–4.
 23. Lehne B, Lewis CM, Schlitt T. (2011) Exome localization of complex disease association signals. *BMC Genom.* 12:92.
 24. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J. (2011) Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Rev. Gen.* 12:745–55.
 25. Singleton AB. (2011) Exome sequencing: a transformative technology. *Lancet Neurol.* 10:942–6.
 26. Ng SB, et al. (2009) Targeted capture and massively parallel sequencing of twelve human exons. *Nature.* 461:272–6.
 27. Kryukov GV, Shpunt A, Stamatoyannopoulos JA, Sunyaev SR. (2009). Power of deep, all-exon resequencing for discovery of human trait genes. *PNAS.* 106:3871–6.
 28. Shi Y, et al. (2011). Exome sequencing identifies ZNF644 mutations in high myopia. *PLOS Gen.* 7:e1002084.
 29. Kumar A, et al. (2011). Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. *PNAS.* 108:17087–92.
 30. Ng SB, et al. (2010) Exome sequencing identifies the cause of a mendelian disorder. *Nature Gen.* 42:30–5.
 31. Ku CS, Naidoo N, Pawitan Y. (2011) Revisiting Mendelian disorders through exome sequencing. *Human Gen.* 129:351–70.
 32. Cirulli ET, et al. (2010) Screening the human exome: a comparison of whole genome and whole transcriptome sequencing. *Gen. Biol.* 11:R57.
 33. Montgomery SB, et al. (2010) Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature.* 464:773–9.
 34. Choi M, et al. (2009) Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *PNAS.* 106:19096–101.
 35. Montenegro G, et al. (2011) Exome sequencing allows for rapid gene identification in a Charcot-Marie-Tooth family. *Ann. Neurol.* 69:464–70.
 36. Caliskan M, et al. (2011) Exome sequencing reveals a novel mutation for autosomal recessive nonsyndromic mental retardation in the *TECR* gene on chromosome 19p13. *Hum. Mol. Genet.* 20:1285–9.
 37. Worthey EA, et al. (2011) Making a definitive diagnosis: Successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet. Med.* 13:255–62.
 38. Johnson JO, et al. (2010). Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron.* 68:857–64.
 39. Wang JL, et al. (2010). TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. *Brain.* 133:3510–8.
 40. Musunuru K, et al. (2010) Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N. Eng. J. Med.* 363:2220–7.
 41. Johnson JO, Gibbs JR, van Maldergen L, Houlden H, Singleton AB. (2010) Exome sequencing in Brown-Vialetto-van Laere Syndrome. *Am. J. Hum. Genet.* 87:567–9.
 42. Simpson MA, et al. (2011) Mutations in *NOTCH2* cause Hajdu-Cheney Syndrome, a disorder of severe and progressive bone loss. *Nat. Genet.* 43:303–5.
 43. Becker J, et al. (2011). Exome sequencing identifies truncating mutations in human *SERPINF1* in autosomal-recessive Osteogenesis Imperfecta. *Am. J. Hum. Genet.* 88:362–71.
 44. ElSharawy A, et al. (2009) Systematic evaluation of the effect of common SNPs on pre-mRNA splicing. *Hum. Mutat.* 30:625–32.
 45. Krawczak M, et al. (2007) Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Human Mutat.* 28:150–8.
 46. Byun M, et al. (2010) Whole exome sequencing-based discovery of *STIM1* deficiency in a child with fatal classic Kaposi sarcoma. *J. Exp. Med.* 207:2307–12.
 47. Gilissen C, et al. (2010) Exome sequencing identifies *WDR35* variants involved in Sensenbrenner syndrome. *Am. J. Hum. Genet.* 87:418–23.