

Missed threads

The impact of pre-mRNA splicing defects on clinical practice

Diana Baralle, Anneke Lucassen & Emanuele Buratti

Over the past two decades, molecular genetic research has helped to elucidate the genetic component of many diseases and conditions. Patients can now be tested for a wide variety of disease-causing genes to improve diagnoses and risk predictions for patients and their families (www.eddnl.com). This, in turn, has allowed physicians to provide advice on lifestyle or other preventative measures and on surveillance, prophylactic and therapeutic possibilities. Moreover, the identification of a genetic component of a disease can act as a starting point for further research to unravel disease mechanisms and for the development of therapeutic interventions.

It is [...] important to develop new methods to interpret the clinical impact of genetic variations that could cause splicing defects

Most disease-causing genes have been identified through mutations that introduce an early stop codon, which leads to a truncated protein product or prevents protein synthesis altogether. Yet, during the past two decades, molecular genetic research has also elucidated several other mechanisms that alter gene function, such as epigenetic factors, distant regulatory elements and small interfering RNAs, to name but a few. Among these, pre-messenger RNA (pre-mRNA) splicing defects are likely to have an impact on clinical practice as these seem to have a role in almost all known diseases with a genetic aetiology (Tazi *et al*, 2009; Wang & Cooper, 2007). Table 1 lists the frequency of pre-mRNA

splicing defects in some common genetic diseases in humans. Although the frequency of splicing mutations varies considerably between individual genes, it is likely that approximately 15% of pathogenic mutations cause disease through the defect that they introduce in the splicing mechanisms. This is a rather conservative estimate, as research has only recently begun to assess routinely for splicing abnormalities, and there is evidence that many unclassified genetic variants might turn out to result in splicing abnormalities.

The reason that the role of splicing mutations has been realized relatively late is because it has been difficult to show a clear correlation between the suspected mutation and disease. This is in contrast with other types of genetic mutation, in which this connection is relatively straightforward. For example, Fig 1A shows how defects of all four α -globin genes cause hydrops fetalis, a condition resulting from severely impaired haemoglobin production. If three or fewer α -globin genes are mutated, the clinical picture is less severe. By contrast, it is much more difficult to interpret the clinical significance of a gene sequence variant that does not lead to an altered protein or a major sequence change. Even if such a variant is identified in a gene that is suspected of causing a particular clinical feature, it often remains possible that it is a benign polymorphism and that the real disease-causing mutation has not yet been detected.

However, as genetic testing improves and becomes more widely requested by patients and doctors, medical geneticists are increasingly called upon to comment on the clinical significance of such unclassified sequence variants. It is therefore important to develop new methods to interpret

the clinical impact of genetic variations that could cause splicing defects. This involves both 'wet biology'—the development and improvement of methods to characterize the effects of mutations on pre-mRNA splicing—and 'dry biology'—the development of software tools to predict whether a given variation would interfere with mRNA splicing. Another important task is the quality assurance of such tests and improving their reliability in order to be useful for genetic counselling in the clinic.

The normal splicing process removes the unnecessary regions (introns) from the gene transcript and splices the remaining sections (exons) back together. The result is a mature mRNA that is exported to the cytoplasm and translated by the ribosomal machinery into protein (Black, 2003). Several factors determine the fidelity of this process. The first factors to be identified were the strength of the splicing signals—a correct acceptor and donor site at the intron, and a branch point. Other more recently identified factors are splicing regulatory elements (SREs), which are important in addition to a plethora of other factors, including the surrounding genomic architecture, RNA polymerase II, extracellular signalling, nonsense-mediated decay (NMD) and RNA secondary structure, all of which can modify the processing of the pre-mRNA. When evaluating the potential consequences of an unclassified sequence variant on splicing, it is important to take into account all of these factors and not only the acceptor and donor sequences (Fig 1B). Not surprisingly, it is difficult to identify sequence variants that affect gene function, and in particular the splicing spoilers, from a background of harmless

polymorphisms (Baralle & Baralle, 2005; Pagani & Baralle, 2004). As research has progressed, it has become clear that almost any genomic variation, even if found within intronic regions, should be considered as a potential disease-causing mutation through aberrant splicing. Yet, in most cases it remains difficult, if not impossible, to find a clear causal relationship between a splicing mutation and a disease.

Nonetheless, our evolving understanding of splicing mechanisms owes a great deal to research that aims to correlate patient pathology with splicing alterations. This process is similar to the elucidation of biochemical pathways in the 1970s through the study of metabolic disease. These and other research on splicing systems have yielded a large amount of data about splicing regulatory sequences, regulatory proteins, the importance of the local sequence context in which they occur, and about how these sequences and factors interact with each other (Cartegni *et al*, 2002). As our knowledge of splicing mechanisms progresses, we are improving methods to predict whether a nucleotide substitution is a possible splicing spoiler.

Several *in silico* approaches have been developed to assess the effects of sequence variants on splicing (Fig 2). These focus on the basic *cis*-acting elements within the pre-mRNA that define an exon—5'-sequence, 3'-sequence and branch-point—or search for potential splicing mutations within SREs. The main problem that remains for any algorithmic solution is the degree of certainty of these predictions in the identification of splicing anomalies (Hartmann *et al*, 2008; Houdayer *et al*, 2008).

As donor and acceptor elements are reasonably conserved in humans, programs that focus on these tend to be more accurate than those that analyse less conserved SREs. In fact, our current lack of knowledge about the effect of local sequence context makes variations in SREs difficult to assess. Despite these drawbacks, a combination of various predictive algorithms is, so far, the best chance of identifying putative splicing mutations, whether these lie in conserved or less conserved regions (Houdayer *et al*, 2008).

In addition to *cis*-acting sequences and SREs, other factors such as RNA secondary structure also affect the regulation of splicing. Although several programs are now

Table 1 | Frequency of splicing defects in common human genetic disorders

Gene	Disease/phenotype	Splicing/total mutation* (%)
<i>ATM</i>	Ataxia telangiectasia	18
<i>BRCA1</i>	Breast cancer predisposition	9
<i>CADM</i>	Medium chain acyl CoA dehydrogenase deficiency	10
<i>CFTR</i>	Cystic fibrosis	14
<i>DMD</i>	Duchenne muscular dystrophy	9
<i>HBA1/2</i>	Blood disorders (thalassaemias, anaemia etc)	3
<i>HBB</i>	Blood disorders (thalassaemias, anaemia etc)	10
<i>HPRT</i>	Hypoxanthine-guanine phosphoribosyltransferase 1 deficiency	15
<i>IKBKAP</i>	Dysautonomia, familial	33
<i>MAPT</i>	Frontotemporal dementia and Parkinsonism	33
<i>MLH1</i>	Colorectal cancer	18
<i>MSH2</i>	Colorectal cancer	9
<i>NF1</i>	Neurofibromatosis type 1	19
<i>NF2</i>	Neurofibromatosis type 2	22
<i>RHO</i>	Retinitis pigmentosa	3
<i>SMN1/2</i>	Spinal muscular atrophy	4
<i>WT1</i>	Wilms tumour	11

*Data calculated from the public Human Gene Mutation Database (24/11/2008).

available to assess secondary structures and their effects on splicing—one main approach uses energy minimization based on *in vitro* RNA folding studies, whereas another uses an evolutionary algorithm on conserved sequences (Baird *et al*, 2006; Major & Griffey, 2001)—these are not yet at a stage where they can be used for routine clinical testing (Buratti & Baralle, 2004). It is also worth highlighting the development of databases that are partly or fully dedicated to collecting mRNA splicing data. The Human Gene Mutation Database (HGMD), a repository of pathological gene mutations for many diseases, is one such example, as is the Alternative Splicing Mutation Database (ASMD); in addition, there are a growing number of locus-specific databases that collect data on particular genes (Paalman *et al*, 2000). Given that these databases are regularly updated with newly discovered sequence variations, these are an easy source of reference for clinicians and researchers, and can help to save a lot of work.

Notwithstanding their growing usefulness, all these programs still show a considerable degree of unreliability depending on the genes analysed. Consequently, even when

in silico approaches predict splicing pathology, these findings need to be confirmed by using 'wet-laboratory' experiments.

For obvious reasons, direct analysis of the processed transcripts from a patient is the most reliable method of establishing with certainty whether a particular DNA substitution affects splicing (Fig 3A). The availability of genomic sequences for all human genes means that it is a straightforward task to design appropriate primers and perform routine reverse-transcription (RT)-PCR screening assays for potential mutations. Furthermore, recent advances in sequencing technology allow rapid and cheap sequencing of whole genes in order to look for potential mutations. In practice, the direct analysis of potential splicing defects should therefore be the method of choice for a full analysis of splicing mutations. However, this direct approach is not always possible. First, the patient might not be available. Second, most samples for clinical diagnosis are leukocytes, which are normally discarded once the DNA has been extracted, thus preventing further analysis of the RNA. Third, even when RNA samples from

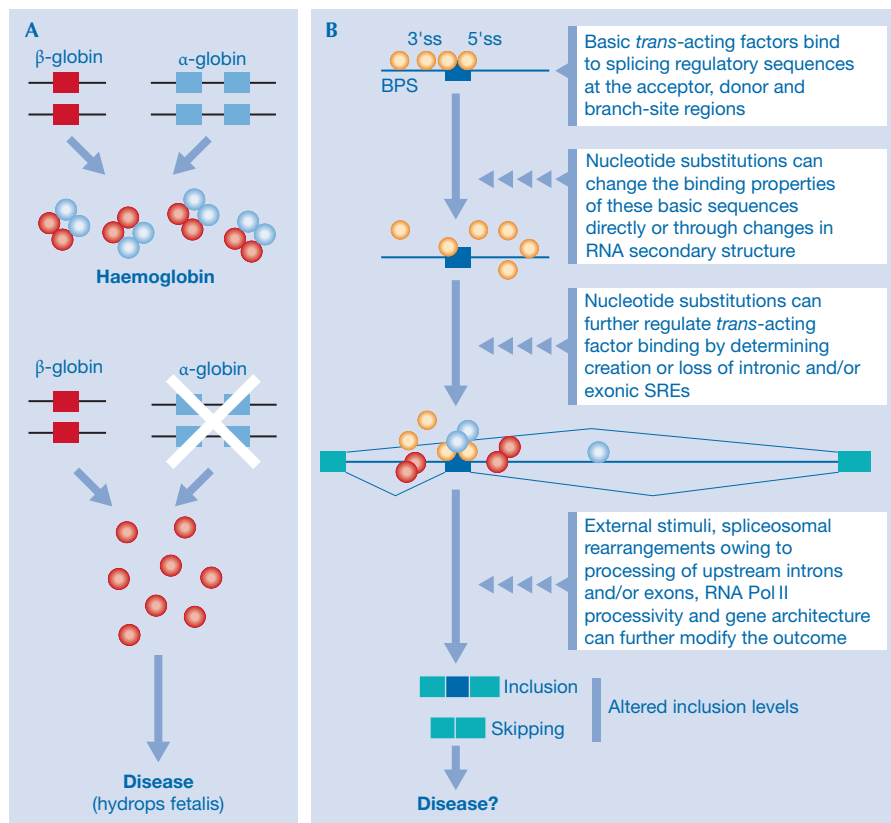


Fig 1 | Comparison between a classical genetic disease and a splicing-related alteration. (A) Schematic representation of the effects of gene deletion in hydrops fetalis. Normally, haemoglobin is made up of two α -subunits and two β -subunits that are transcribed by one β -gene on chromosome 11 (two copies per individual) and two α -genes on chromosome 16 (four copies per individual). People of Asian ancestry often have two α -globin genes deleted on the same chromosome 16, which results in mild thalassaemia. However, an offspring that inherits no α -genes from two heterozygous parents will develop hydrops fetalis, an often fatal disease *in utero*. In these cases, early diagnosis can be easily performed. (B) Mutations in the pre-mRNA splicing process, however, can affect several crucial steps in the recognition of introns and/or exons and can be further modified by many factors outside the mutation location. These context effects make it hard to determine the final outcome, eventual disease severity, and thus genetic counselling. BPS, branch-point site; SRE, splicing regulatory element.

leukocytes are available, these are of little use when the mutation affects tissue-specific gene expression, for example, in the brain or heart.

Even if an RNA sample from the affected tissue is available, the detection of a particular splicing product is not straightforward. Aberrant transcripts can be degraded quickly by NMD, a regulatory mechanism that controls the post-transcriptional quality of the mRNA and selectively degrades transcripts that contain premature termination codons (Maquat, 2004). NMD would therefore mean that such aberrant transcripts are undetectable in normal RNA analysis. This problem of degradation can be solved in

some cases by establishing stable cell lines from the patient's lymphoblasts and treating them with antibiotics such as anisomycin or gentamycin to block NMD. However, this represents a considerable investment of time and resources, and would be difficult to use in a routine, high-throughput manner; in addition, the interpretation of the *in vivo* clinical effects would remain elusive. Finally, a patient's genotype should be evaluated carefully in order to unambiguously assign the aberrant products to a specific mutation so as to rule out the presence of different splicing mutations on different alleles.

To overcome these uncertainties, researchers now use more experimentally controlled

methods. So far, the methods that are most widely used to analyse splicing aberrations are *in vitro* splicing and minigene systems. The *in vitro* splicing assay uses bacterial T7/SP6 polymerases to synthesize radioactively labelled pre-mRNA molecules that are incubated in the presence of nuclear extracts; the resulting spliced products are resolved on a polyacrylamide denaturing gel (Fig 3B). This assay allows the evaluation of splicing kinetics and intermediates of the splicing reactions such as lariat formation. Another advantage of this method is that, once the template is ready, the results of an *in vitro* splicing assay can be observed within hours and it does not require cell culture facilities. *In vitro* splicing has therefore had an important role in the unravelling of many of the basic aspects of the splicing mechanism (Hicks *et al*, 2005).

However, *in vitro* splicing systems have only limited use in clinical and/or diagnostic applications. First, the *in vitro* transcription of pre-mRNA puts an upper limit on the length of sequence that can be transcribed—approximately 1,000–2,000 nucleotides—which makes *in vitro* systems generally unable to mimic complex genomic regions. However, there are some exceptions; for example, cryptic splice-site activation in the β -globin gene was one of the first medically relevant splicing defects described by using an *in vitro* splicing method (Treisman *et al*, 1983; Wieringa *et al*, 1983).

Investigating the pathogenicity of a splicing mutation has recently been supplanted by 'minigene' system assays, which are a refinement of minigene-based technologies for alternative splicing analysis that was initially described more than 20 years ago (Vibe-Pedersen *et al*, 1984). The hybrid minigene splicing assay represents a relatively fast approach for identifying splicing spoilers and the study of their underlying functional mechanisms (Fig 3C; Baralle *et al*, 2003; Cooper, 2005; Kishore *et al*, 2008). Any genomic region of interest—for example, an exon with part of its intronic flanking regions—can be amplified from normal or affected individuals and cloned into the minigene. The resultant plasmid is then transiently transfected in an appropriate cell line where it will be transcribed by RNA polymerase II before the resulting pre-mRNA is processed to mature mRNA. The mRNA splicing pattern can be analysed by RT-PCR with primers designed specifically to amplify the minigene-processed transcripts and to avoid the amplification

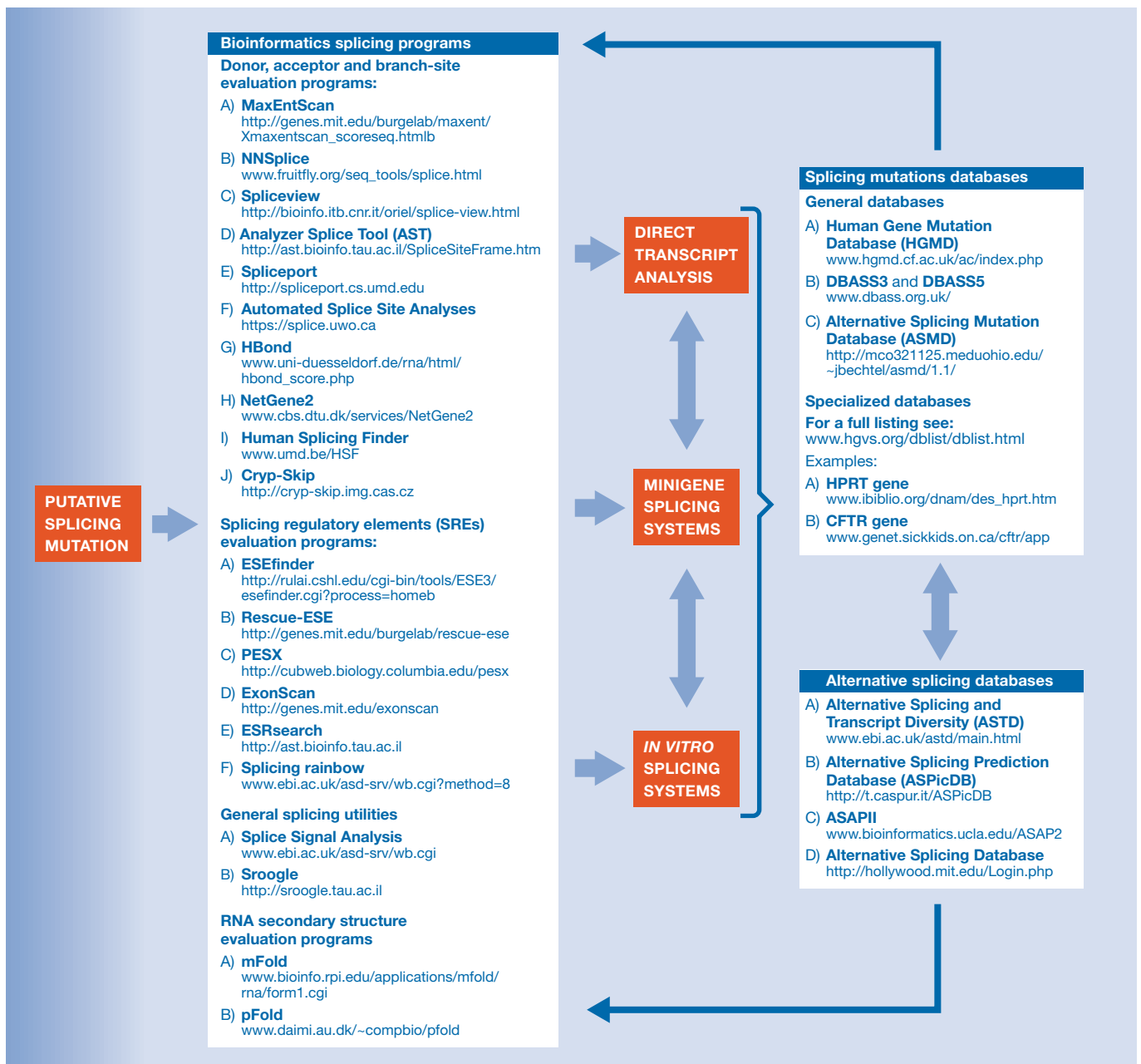


Fig 2 | Putative splicing mutations can be analysed using various publicly available bioinformatic tools that provide predictions on potential disruption of basic splicing sequences (acceptor, donor and branch-point sites), regulatory elements, and other features such as RNA secondary structure and protein binding sites. Splicing mutations are then collected in various databases, and the data stored can be used to improve the predictive analysis of mutation-detecting software.

of endogenous plasmid mRNAs. In many cases, the genomic sequence will contain only the exon that is suspected to be affected by the splicing mutation and a portion of its flanking intronic regions. Yet, it is also possible to analyse more complex exon–intron compositions. In these cases, the part of the gene that needs to be inserted into the plasmid has to be selected carefully to mimic

the hypothesized *in vivo* mutational effects, such as multiple exon skipping (Baralle *et al*, 2006).

However, the minigene system can only yield information about unprocessed and mature mRNA species; splicing intermediates are difficult to detect as the end-point is normally obtained by RT-PCR. Moreover, unless real-time RT-PCR is used, it is difficult to

quantify each processed transcript, although the relative proportion of each splicing variant can often be deduced. Fortunately, for clinical or diagnostic use, it is usually sufficient to determine whether there is a splicing abnormality, so the above limitations have little practical impact. However, if the read-out is a change in the proportion of the splicing products, rather than a clear exon skip,

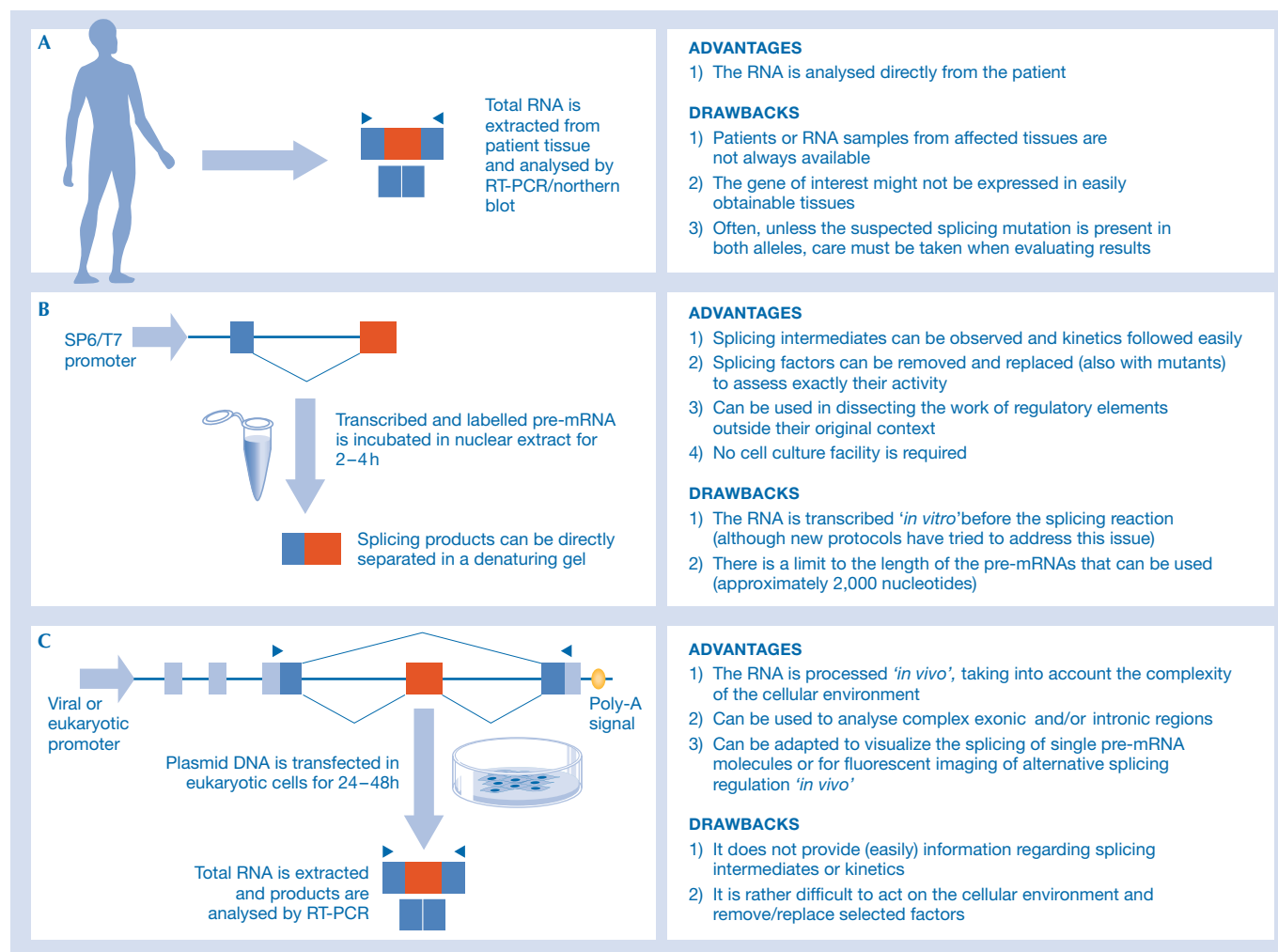


Fig 3 | Advantages and drawbacks of the diagnostic splicing tools commonly used in the laboratory. mRNA, messenger RNA; RT-PCR, reverse-transcription PCR.

the limitations of the minigene method make the assessment of the clinical significance more difficult.

At present, the best manner in which to identify splicing mutations that have a role in disease is to use a combination of these 'wet' and 'dry' techniques (Fig 4). As more data are collected, however, it will generate a greater evidence basis about the risks of disease caused by splicing alterations. This, in turn, might allow the development of more appropriate, effective and safer therapies.

In the meantime, more research is needed in the basic science of splicing mechanisms, and its application in diagnostic and therapeutic research. Although still in their infancy, various strategies have already been developed to explore therapeutic approaches (Table 2; Sumanasekera

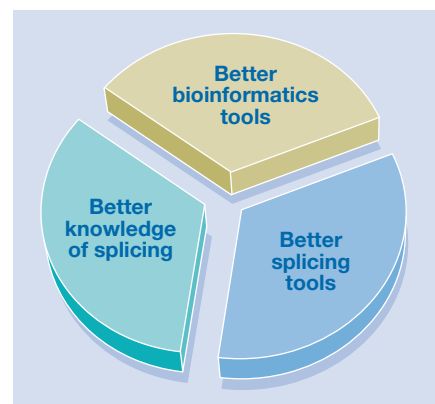
et al, 2008; Wang & Cooper, 2007; Wood et al, 2007). Yet, before human application becomes a reality, several challenges have to be overcome, including optimizing specificity for each gene and/or exon of interest, the development of appropriate carrier systems, the avoidance of immune responses, and optimization of recurrent-administration protocols as none of these methods will cause a permanent correction of mRNA splicing defects. Notwithstanding these difficulties, oligonucleotide antisense technology (Aartsma-Rus & van Ommen, 2007; Garcia-Blanco, 2005) has already entered the clinical trial stage to treat diseases such as Duchenne muscular dystrophy or siRNA for the treatment of neovascularization associated with age-related macular degenerations (<http://www.clinicaltrials.gov>). In particular, antisense

oligonucleotide strategies use base-pairing to target specific sequences in RNAs. Antisense oligonucleotides are thought to modulate splicing by steric hindrance of the recruitment of the splicing factors to the targeted *cis*-elements, thus forcing the machinery to use the natural sites. This is a rapidly evolving field.

The increasing knowledge about how splicing defects cause or contribute to disease also has implications that go beyond clinical diagnosis and drug development, and affect the private lives of patients and their families. Predicting serious diseases, for which there are no treatments, many years before the probable onset of the disease remains a controversial matter. Individuals might wish to know such information while their relatives, who are possibly at the same risk, might decide that they

Table 2 | Potential therapeutic approaches that target splicing abnormalities

Therapeutic approach	Mode of action	Genes targeted
Small molecules	For many small molecules the mechanism of action is unknown. Some of them act by selectively modifying the activity of splicing regulatory proteins through altered cellular distribution or changing phosphorylation states.	<i>SMN2, MAPT, IKBKAP</i>
Antisense oligonucleotides	Antisense oligonucleotides target splicing controlling regions of the selected gene to block their use by the normal splicing machinery. They can be used to inhibit the inclusion of unwanted exons and/or promote the production of a truncated but functional protein.	<i>DMD, CFTR, BRCA1, HBB, ATM</i>
Bifunctional oligonucleotides	Bifunctional oligonucleotides contain a complementary targeting region and an effector region that can recruit or mimic splicing factor activities.	<i>SMN2</i>
Spliceosome-mediated RNA <i>trans</i> -splicing (SMaRT)	In this approach, an exogenous RNA is expressed and through the mechanism of <i>trans</i> -splicing can replace an aberrant transcript with a wild-type sequence.	<i>SMN2, HBB</i>
Isoform-specific RNA interference	This approach involves the use of exon-specific RNA interference to specifically knock down aberrant splicing isoforms.	<i>VEGF</i>
Modified U1 snRNPs	Modified U1 snRNPs (small nuclear ribonucleoproteins) can be used either to block aberrant splice-site sequences (that is, acting as antisense oligonucleotides) or to reverse mis-splicing by carrying compensatory mutations in the 5' end of their U1 snRNA sequence.	<i>HBB, RHO</i>

**Fig 4** | Recipe for optimizing the identification of splicing mutations.

do not want to know. So far, such debates have been largely limited to rare, high-risk, single-gene disorders. But, as we gain more knowledge about the genetic components of many common diseases, 'genetic medicine' is becoming more important and at the same time more complicated. Although many scientists have already abandoned

the one-gene, one-disease concept, this idea still has to find its way into the clinic and into the public realm, as many people still have a rather simplistic idea of genes and their role in disease (Carver *et al*, 2008).

The study of splicing mechanisms in disease adds another level of complexity and complications. The identification of a potential splicing mutation does not mean that there will necessarily be a close correlation between it and any clinical manifestations. In order to be able to advise patients correctly about their health risks, we first need to evaluate the expressivity and penetrance of a splicing mutation. This could be done by studying segregation of the mutation in patients' families or through larger population studies. As already mentioned, laypeople and the media often hold a deterministic view of genetics, and patients sometimes attribute a greater risk of disease from a genetic finding than actually exists. This is particularly likely for certain splicing abnormalities, some of which—in isolation—may have only a minor effect, if at all; however, if these come with a particular genetic or environmental background,

they might have a larger role in the pathogenicity of a disease. These are complex counselling issues and our knowledge will need to expand from its relatively primitive state over the next decade. It can be difficult to convince patients that the exciting research developments and attendant promises of improved diagnosis and therapeutics that they have heard or read about, are still some years from being applicable to them or their family. Present progress is still hampered by technical and scientific limitations, but we must ensure that the societal implications are debated and practical guidelines developed for the time that these limitations are resolved.

ACKNOWLEDGEMENTS

E.B. is supported by Telethon Onlus Foundation (Italy; grant number GGP06147) and by a European community grant (EURASNET-LSHG-CT-2005-518238). D.B. is supported by Action Medical Research grant number SP4175 and EURASNET. We also wish to thank Professor F.E. Baralle for helpful discussions and suggestions.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

- Aartsma-Rus A, van Ommen GJ (2007) Antisense-mediated exon skipping: a versatile tool with therapeutic and research applications. *RNA* **13**: 1609–1624
- Baird SD, Turcotte M, Korneluk RG, Holcik M (2006) Searching for IRES. *RNA* **12**: 1755–1785
- Baralle D, Baralle M (2005) Splicing in action: assessing disease causing sequence changes. *J Med Genet* **42**: 737–748
- Baralle M, Baralle D, De Conti L, Mattocks C, Whittaker J, Knezevich A, French-Constant C, Baralle FE (2003) Identification of a mutation that perturbs NF1, a gene splicing using genomic DNA samples and a minigene assay. *J Med Genet* **40**: 220–222
- Baralle M, Skoko N, Knezevich A, De Conti L, Motti D, Bhuvanagiri M, Baralle D, Buratti E, Baralle FE (2006) NF1 mRNA biogenesis: effect of the genomic milieu in splicing regulation of the NF1 exon 37 region. *FEBS Lett* **580**: 4449–4456
- Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72**: 291–336
- Buratti E, Baralle FE (2004) Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol Cell Biol* **24**: 10505–10514
- Cartegni L, Chew SL, Krainer AR (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* **3**: 285–298
- Carver R, Waldahl R, Breivik J (2008) Frame that gene. A tool for analysing and classifying the communication of genetics to the public. *EMBO Rep* **9**: 943–947

- Cooper TA (2005) Use of minigene systems to dissect alternative splicing elements. *Methods* **37**: 331–340
- Garcia-Blanco MA (2005) Making antisense of splicing. *Curr Opin Mol Ther* **7**: 476–482
- Hartmann L, Theiss S, Niederacher D, Schaal H (2008) Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? *Front Biosci* **13**: 3252–3272
- Hicks MJ, Lam BJ, Hertel KJ (2005) Analyzing mechanisms of alternative pre-mRNA splicing using in vitro splicing assays. *Methods* **37**: 306–313
- Houdayer C *et al* (2008) Evaluation of *in silico* splice tools for decision-making in molecular diagnosis. *Hum Mutat* **29**: 975–982
- Kishore S, Khanna A, Stamm S (2008) Rapid generation of splicing reporters with pSpliceExpress. *Gene* **427**: 104–110
- Major F, Griffey R (2001) Computational methods for RNA structure determination. *Curr Opin Struct Biol* **11**: 282–286
- Maquat LE (2004) Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat Rev Mol Cell Biol* **5**: 89–99
- Paalman MH, Cotton RG, Kazazian HH Jr (2000) VARIATION, DATABASES, and DISEASE: new directions for human mutation. *Hum Mutat* **16**: 97–98
- Pagani F, Baralle FE (2004) Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet* **5**: 389–396
- Sumanasekera C, Watt DS, Stamm S (2008) Substances that can change alternative splice-site selection. *Biochem Soc Trans* **36**: 483–490
- Tazi J, Bakkour N, Stamm S (2009) Alternative splicing and disease. *Biochim Biophys Acta* **1792**: 14–26
- Treisman R, Orkin SH, Maniatis T (1983) Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. *Nature* **302**: 591–596
- Vibe-Pedersen K, Kornblihtt AR, Baralle FE (1984) Expression of a human alpha-globin/fibronectin gene hybrid generates two mRNAs by alternative splicing. *EMBO J* **3**: 2511–2516
- Wang GS, Cooper TA (2007) Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* **8**: 749–761
- Wieringa B, Meyer F, Reiser J, Weissmann C (1983) Unusual splice sites revealed by mutagenic inactivation of an authentic splice site of the rabbit beta-globin gene. *Nature* **301**: 38–43
- Wood M, Yin H, McClorey G (2007) Modulating the expression of disease genes with RNA-based therapy. *PLoS Genet* **3**: e109



Diana Baralle⁺ (top left) is at the Academic Unit of Genetic Medicine, Human Genetics Division, University of Southampton School of Medicine, UK. Anneke Lucassen (top right) is at the Academic Unit of Genetic Medicine, Cancer Sciences division, University of Southampton School of Medicine, UK. Emanuele Buratti (bottom left) is at the Department of Molecular Pathology, ICGEB, Trieste, Italy. ⁺Corresponding author. Tel: +44 2380 796162; Fax: +44 2380 714346; E-mail: d.baralle@soton.ac.uk doi:10.1038/embor.2009.170