

Review

Translating RNA Splicing Analysis into Diagnosis and Therapy

Andrew G. L. Douglas^{1,2,*}, Diana Baralle^{1,2,*}

1. Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK; E-Mails: a.g.douglas@soton.ac.uk; d.baralle@soton.ac.uk
2. Wessex Clinical Genetics Service, University Hospital Southampton NHS Foundation Trust, Southampton, UK.

* **Correspondence:** Diana Baralle and Andrew G. L. Douglas; E-Mails: d.baralle@soton.ac.uk; a.g.douglas@soton.ac.uk

Academic Editor: Michael R. Ladomery

Special Issue: [Alternative Splicing: A Key Process in Development and Disease](#)

OBM Genetics

2021, volume 5, issue 1

doi:10.21926/obm.genet.2101125

Received: December 31, 2020

Accepted: February 26, 2021

Published: March 08, 2021

Abstract

A large proportion of rare disease patients remain undiagnosed and the vast majority of such conditions remain untreatable whether diagnosed or not. RNA splicing analysis is able to increase the diagnostic rate in rare disease by identifying cryptic splicing mutations and can help in interpreting the pathogenicity of genomic variants. Whilst targeted RT-PCR analysis remains a highly sensitive tool for assessing the splicing effects of known variants, RNA-seq can provide a more comprehensive transcriptome-wide analysis of splicing. Appropriate care should be taken in RNA-seq experimental design since sample quality, processing, choice of library preparation and sequencing parameters all introduce variability. Many bioinformatic tools exist to aid both in the prediction of splicing effects from DNA sequence and in the handling of RNA-seq data for splicing analysis. Once identified, splicing abnormalities may be amenable to correction using antisense oligonucleotide compounds by masking cryptic splice sites or blocking key splice regulatory elements, or by use of alternative corrective technologies such as *trans*-splicing. A growing number of such drugs have started to enter clinical use, most notably nusinersen for the treatment of spinal



© 2021 by the author. This is an open access article distributed under the conditions of the [Creative Commons by Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium or format, provided the original work is correctly cited.

muscular atrophy. By bringing together the fields of RNA diagnostics and antisense therapeutics, it is becoming feasible to envisage the development of a truly personalised medicine pipeline. This has already been shown to be possible in the case of milasen, an n=1 bespoke antisense drug, and the growth and convergence of these technologies means that similar therapeutic opportunities should arise in the near future.

Keywords

RT-PCR; RNA-seq; splicing; splicing prediction; bioinformatic tools; antisense oligonucleotides

1. Introduction

Rare diseases affect between 3.5-5.9% of the global population (260-450 million people) and around 72% of these are genetic in origin [1]. However, although rapid advances in next-generation sequencing (NGS) technology in recent years have led to great improvements in diagnostic yield with trio whole genome sequencing (WGS) achieving diagnostic rates of up to 42%, the majority of such individuals still remain undiagnosed [2]. Furthermore, although over 6000 rare diseases are currently known to exist, only some 6% of them have any specific treatments and less than 1% of these can be considered curative [3]. A wide translational gap therefore exists between our increasing ability to diagnose genetic disorders and our relative inability to treat individuals affected by these conditions.

One particular area of genomic medicine that has only recently started to gain widespread traction in rare disease diagnostics is RNA-based testing and in particular RNA splicing analysis [4-7]. Whilst DNA sequencing can consistently and accurately detect germline variants in any given genomic region, interpretation of their effects on gene function is heavily reliant upon predictions of how we expect cellular molecular machinery to work. Given our limited knowledge of macromolecular structures and their functional interactions, together with our generally poor understanding of how such complexes are regulated, it is not surprising that these predictions often turn out to be wrong [8-10]. This holds true not only for protein-level predictions but also for predictions relating to splicing. However, by directly assessing RNA it becomes possible to provide an objective window into the earliest steps of gene function (i.e. transcription and pre-mRNA splicing). RNA analysis can therefore help to remove at least one level of functional effect prediction when it comes to variant interpretation.

As well as its diagnostic potential, RNA also represents a unique therapeutic target that sits halfway between DNA sequence information and protein structure and function. Being a more accessible and modifiable cellular molecule than DNA but still retaining its nucleic acid sequence specificity, RNA therapeutic manipulation is now a well-established field of research with multiple clinical applications [11]. However, these two areas of genomic medicine, genomic diagnostics and genome-based therapeutics, in many ways still remain largely disconnected in everyday clinical practice. In this review, we will illustrate how splicing diagnostics and splicing therapeutics can be brought together into a coherent pipeline for the development of personalised medicines.

2. Diagnosis of Splicing Mutations in Clinical Practice

2.1 RT-PCR Analysis

For many years, the mainstay of RNA-based splicing analysis for variant interpretation has been reverse transcription polymerase chain reaction (RT-PCR) [12]. A variety of reverse transcriptase enzymes are commercially available and these can be utilised to synthesise cDNA through the use of random hexamer or oligo(dT) primers, depending on whether total RNA or just polyadenylated transcripts are required [13]. Gene-specific primers can also be used for reverse transcription if greater specificity is needed or if a one-step RT-PCR protocol is to be employed. Following reverse transcription, primers sited in exons flanking a specific variant can be used to amplify the cDNA region of interest. Straightforward gel electrophoresis and Sanger sequencing of PCR products will often then be able to detect abnormal splicing events such as exon skipping. Molecular cloning of PCR products may sometimes be required to aid in identifying specific alternative splicing products, especially where the RT-PCR reaction yields multiple products. However, when compared against control samples, the splicing effect of a given variant can usually be determined via this method (see Figure 1). Once identified, gel densitometry can be used as a semi-quantitative method for different splice isoforms but if more accurate relative quantification is needed then quantitative PCR (RT-qPCR) can be performed on cDNA templates, while digital PCR (dPCR) can also potentially be employed for the purposes of relative or absolute quantification [14-18].

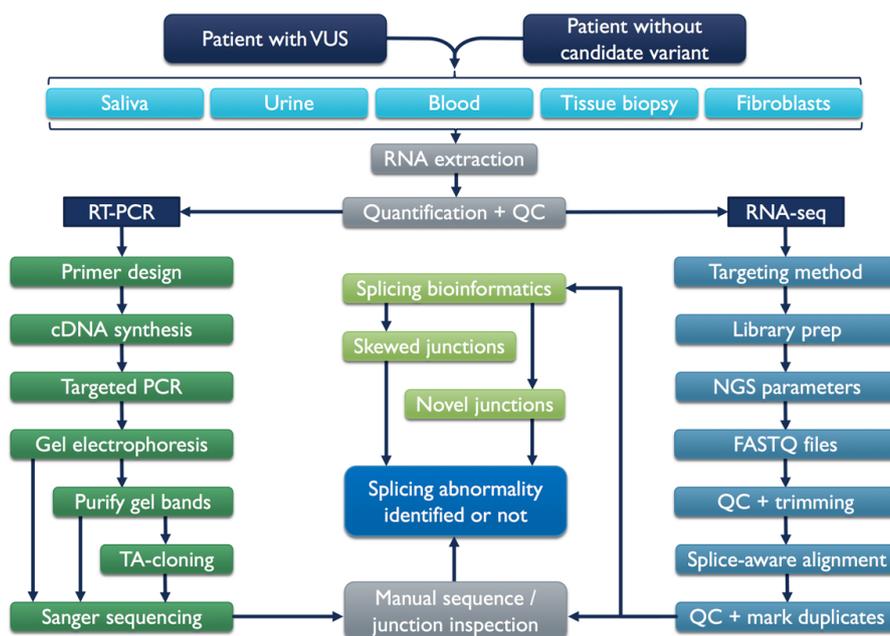


Figure 1 RNA splicing analysis for rare disease diagnostics. Patients with or without candidate variants of uncertain significance (VUSs) can have RNA sampled from a variety of sources. The RT-PCR analysis pipeline is most often applicable to targeted VUS interpretation. RNA-seq analysis can be used for detection of abnormal splicing whether or not a candidate VUS is present. Quality control (QC) of materials and data remains relevant at all stages of the process.

Whilst RT-PCR remains a powerful and highly sensitive technique for targeted RNA analysis, it is limited by several factors. Principal among these is the requirement for the gene of interest to be expressed in a clinically available tissue (most often blood). Although blood has been shown to express at least 80% of human coding sequences at a detectable level, a significant proportion of human disease genes are still not expressed well enough for reliable analysis of splicing [19, 20]. A reasonable estimate of whether a gene is likely to be detectable in blood can be made by reference to the Genotype-Tissue Expression (GTEx) project's freely available data (accessible either via data download or via the GTEx online portal - <https://www.gtexportal.org/home/>) [21]. Analysis of the GTEx data shows that 57% (32,056/56,200) of named human genes have a median transcript per million (TPM) value of zero in whole blood RNA and these are therefore unlikely to be suitable candidates for splicing analysis in blood. Furthermore, 66% (37,111/56,200) have a median TPM under 0.1 and these genes are also unlikely to be reliably detectable in blood by RT-PCR. However, looking solely at disease-associated genes in comparison (in this case referring to genes listed on Genomics England's PanelApp resource), only 10% (561/5516) have median TPM values of zero and 25% (1399/5516) have a TPM value of less than 0.1 (see Figure 2) [22]. One may therefore expect a potentially detectable level of coverage of the remaining 75% of disease-associated genes with respect to blood splicing analysis by RT-PCR.

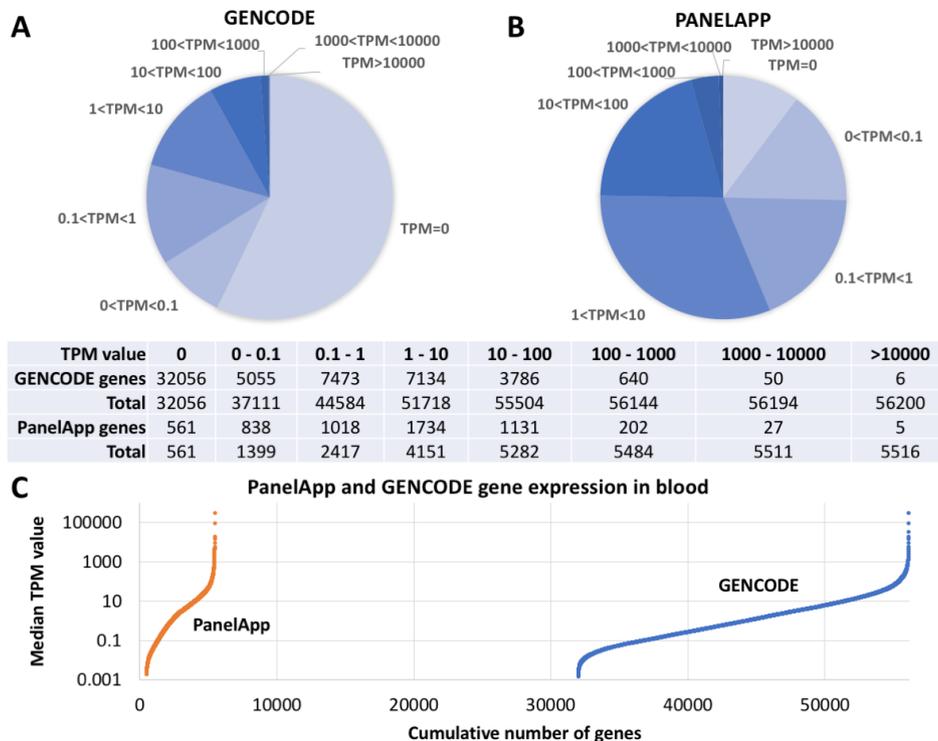


Figure 2 Median transcripts per million (TPM) values in whole blood. A. A chart including all GENCODE listed genes (56200 in total) demonstrates that the majority have low TPM values. B. A chart of clinically relevant genes listed on PanelApp shows that the majority have TPM values above 1. C. Expression values (logarithmic scale) of all GENCODE and PanelApp genes arranged in order of increasing TPM value. Note that genes with TPM values of zero cannot be displayed.

For genes that are not expressed in whole blood, alternative sources of RNA may include (see Figure 1): cultured fibroblasts obtained via skin biopsy, cultured lymphocytes or lymphoblastoid cell lines, other types of tissue biopsy such as skeletal muscle or biofluids such as urine or saliva (or potentially more usefully a buccal swab of cheek epithelial cells since saliva cellular material is largely of leukocytic origin) [23]. The availability of cultured cells in particular provides an opportunity to examine samples for splice isoforms subject to nonsense-mediated decay (NMD). Through the application of NMD inhibitors such as cycloheximide or anisomycin to such cultures, the otherwise degraded splicing products of pathogenic splicing mutations can subsequently be detected and quantified, as has been demonstrated in both fibroblasts and lymphocytes [24, 25].

Another important limitation of RT-PCR analysis is that an abnormal splicing event may yield a product that cannot readily be amplified by the predetermined primer set. This may either be because the resulting amplicon is too large (e.g. long intron retention) or else because a multi-exon skipping event may encompass one or other of the primer binding sites. In some cases, transcript-wide RT-PCR assays can be accomplished by setting up overlapping PCR amplicons spanning contiguous exon regions. This can work to some extent for small genes or where high sample throughput justifies assay development (as has been done in some clinical laboratories for *NF1* analysis and historically was also demonstrated for *DMD* mutation scanning)[26, 27]. However, for most genes the time and effort involved in setting up and validating this type of assay is unlikely to prove viable on a clinical diagnostic basis. Hence, the very nature of targeted RT-PCR that lends strength to its specificity and sensitivity in terms of its lower limit of detection, also conversely gives rise to an inherent lack of sensitivity when it comes to detecting unexpected events.

2.2 RNA Sequencing

NGS technologies have allowed RNA splicing analysis to progress beyond the limitations of RT-PCR. In particular, transcriptome-wide RNA sequencing (RNA-seq) can provide a relatively comprehensive assessment of RNA splicing, potentially allowing detection of unexpected mis-splicing events that may be missed by RT-PCR [28]. The sequence-level mapping employed in RNA-seq alignments also lends itself ideally to the identification of both large-scale and fine-level splicing alterations without the need for PCR product purification, cloning and/or Sanger sequencing. Whilst still reliant on the tissue-specificity of an individual gene's expression, RNA-seq can therefore be used relatively easily to look for abnormal splicing events related to variants of uncertain significance (VUSs) of interest (see Figure 1).

RNA-seq data generation can be achieved via multiple routes and any laboratory embarking on such work must carefully consider its choice of library preparation method and sequencing parameters, since these will largely influence the suitability of the output data for subsequent analyses. RNA quality is distinctly important in this regard, since long intact transcripts are preferable for adequate analysis of splicing. The RNA integrity number (RIN) that can be generated from Agilent Bioanalyzer/Tapestation assays provides a measure of RNA sample degradation on a scale from 10 (no degradation) to 1 (total degradation) [29]. High-quality RNA is especially important if a poly(A) library prep method is employed. This is because using a degraded sample can lead to pronounced skewing of coverage towards the 3' end of transcripts and this can in turn

severely limit the assay's ability to capture and analyse splice junction reads. Quantification can also be affected since different transcripts can be degraded at different rates [30].

A common clinical starting point is a patient blood sample and if this is the case then a frequently used technique is globin depletion, which employs probe-based removal or inhibition of haemoglobin-related transcripts. This greatly increases the relative number of reads that will be generated from non-globin RNA, since globin transcripts comprise between 50-80% of blood mRNA.[31-33] Removal of ribosomal RNA through ribodepletion is another commonly used approach to increase relevant read coverage as rRNA can account for some 75-90% of total cellular RNA in blood [34, 35]. This type of preparation allows retention of RNA species that may lack polyadenylation, such as many non-coding RNAs [36]. Alternatively, poly(A)-selection may be preferred if mRNAs are the sole species of interest. Importantly, most commonly used poly(A) and total RNA library prep methods include a size-selection step, which effectively excludes short RNAs and so this must be considered if, for example, miRNAs and/or similarly sized RNA species are to be studied.

Illumina-style short-read sequencing platforms can generate relatively consistent outputs in terms of numbers and lengths of sequence reads per flowcell. However, the maximum read length available and the total sequencing capacity per flowcell are instrument-dependent. Using longer read lengths increases the likelihood of individual reads capturing splice events and employing paired-end sequencing increases this still further by sequencing the first and second reads from the opposite ends of the inserted DNA fragments within a library. The choice of how many reads to sequence per sample largely depends on the needs of the downstream analysis. Since splice isoforms can exist at variable abundance and are often subject to RNA degradation, the expression levels of the relevant target genes of interest need to be taken into account. As such, there is no set standard for the minimum required read count per sample when it comes to transcriptome-wide splicing analysis and in practical terms it is cost that becomes the ultimate limiting factor. It must also be emphasised that adequate quality control is essential at every step of the RNA-seq process, not only relating to the quality of starting RNA material but also to the quality of the sequencing output and the quality of subsequent alignment steps [37].

2.3 Detecting Splicing Mutations

Once sequenced, RNA-seq data in the form of .fastq files must be aligned to the reference genome (unless *de novo* transcriptome assembly is attempted) using a splice-aware mapping program to produce .bam files. One of the most widely-used aligners is STAR, which has the benefit of being very fast (usually providing alignments within a couple of hours) but with a disadvantage of the user needing access to a high-performance computing (HPC) cluster owing to its high memory requirements [38]. If HPC access is not available, similar alignments can be produced by a program such as HISAT2 running on a personal computer [39]. However, it should be noted that alignments do vary depending on what aligner is used and employing different command options and settings can significantly affect the resulting output. Aligned .bam files can be subsequently sorted and marked for duplicate reads if appropriate. Marking of duplicates is a common QC procedure in DNA-based NGS owing to the possibility of PCR duplicates introduced during library amplification, which can potentially lead to a bias in read counting. However, there is some debate as to whether duplicate marking is always appropriate in RNA-seq [40-42].

Perhaps the most difficult and rapidly evolving part of RNA-seq splicing analysis comes next, in the form of identifying abnormal splicing events in relevant genes. Where a known VUS exists in a patient's DNA, the process is fairly straightforward since the spliced reads that are mapped to any given locus can be inspected visually using software such as the Integrative Genomics Viewer (IGV) and splice junction usage can be highlighted using Sashimi plots [43]. By comparing such visualisations in a patient's data against that of similar batched controls (e.g. other patient samples), a specific splicing alteration can become immediately apparent. However, in situations where no candidate variants are known, the problem of performing a 'comprehensive' analysis of splicing becomes less tractable. The issue is somewhat akin to undertaking whole-genome analysis, where there is no such thing as a 'complete' analysis; one can only ever perform limited sets of analyses looking at the data in certain ways and using specified parameters. Indeed, transcriptome-wide splicing analysis is in some ways conceptually more complex than genome analysis. This is because it encompasses additional variables such as technical variation in RNA handling, preparation and sequencing, relative isoform usage levels, the dynamic effects of post-transcriptional RNA regulation and a much larger potential space for unannotated splice variants.

In the setting of a genomic sequence variant that creates an entirely novel splice junction, detection of the event can potentially be achieved through a process of splice junction filtering. In its most basic form, this relies on the novel junction not being present in any of the control samples against which the sample is being filtered. However, this approach suffers from two significant problems. Firstly, unannotated sample-specific splicing events are surprisingly abundant in RNA-seq data (see Figure 3). This means that a substantial number of batched control samples (e.g. samples from other patients) may be needed if the numbers of unique filtered junctions are to be reduced to a manageably short and manually curatable candidate list. Utilising publicly available RNA-seq datasets, such as that provided through the Genotype-Tissue Expression (GTEx) project, may prove helpful in terms of boosting control numbers [21]. However, it remains to be seen whether such datasets, whose samples are invariably processed and sequenced under diverse conditions and with different parameters, can be reliably used in this way. Secondly, it is not uncommon for a pathogenic cryptic splice junction to be present at low levels in at least some control samples. Blanket filtering out of shared junctions across samples therefore risks removing and thus overlooking such splice variants. One possibility to help address this second issue might be to pre-filter control data to remove low-level splice junctions prior to their use in filtering. This could help ensure that only higher-quality *bona fide* splice junctions are used for subsequent filtering steps.

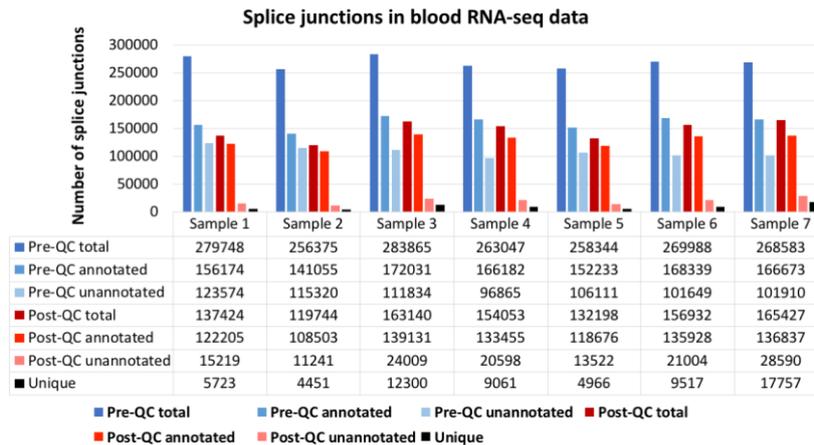


Figure 3 Example of splice junction filtering among a batch of seven blood RNA-seq samples. PAXgene blood RNA samples underwent globin and rRNA depletion with stranded total RNA library prep and 70M 150bp paired-end read sequencing per sample. Data were mapped to GRCh37 using STAR and GENCODE v19 annotations. STAR splice junctions were quality-control (QC) filtered to exclude those with fewer than 3 spliced reads and those with apparently artefactual "intron lengths" of 1bp. Filtering out junctions shared between samples still results in several thousand unique sample-specific junctions being retained.

Filtering for the presence of unique splice junctions will not generally detect intron retention and neither will it detect differential alternative splicing between existing annotated or otherwise shared splice junctions. Alternative splicing can usually be categorised into a set number of possible types or modes: constitutive splicing (CS), mutually exclusive exons (MXE), cassette alternative exon (CAE), alternative 5' splice site (A5SS), alternative 3' splice site (A3SS), and intron retention (IR) [44, 45]. Assessing differential alternative splicing between samples requires a measure of relative usage, such as the commonly used percent-spliced-in (PSI) value [46]. When properly calculated, the PSI value for a splice event takes into account both the sequencing read length and the length of the alternatively included or excluded feature (such as a skipped exon). PSI therefore cannot be calculated from splice junction count data alone but requires read-level coverage data from across the entire interval spanning the splice event of interest. This is especially relevant in the case of intron retention, where the event may be completely missed if relying on analysis of splice junction counts alone.

Several recent studies have demonstrated how RNA-seq can be used to identify splicing mutations in a rare disease diagnostic setting [47-52]. Cummings et al. analysed muscle RNA-seq data from a cohort of patients with undiagnosed neuromuscular conditions and looked primarily for unique splicing abnormalities compared to 184 selected control samples from the GTEx project, yielding an overall diagnostic rate of 35% [47]. In order to allow more valid comparison to GTEx data, sequencing was performed using similar parameters of non-strand-specific poly(A) library preparation and 76-bp paired-end reads with 50 or 100 million reads per sample. Kremer et al. performed RNA-seq on cultured fibroblasts from 48 patients with undiagnosed mitochondrial disorders and looked at aberrant expression, splicing and monoallelic expression, yielding a diagnosis in 10% of cases [48]. Non-strand-specific poly(A) selection was used in library

preparation and sequencing was performed with 100-bp paired-end reads. Abnormal splicing was investigated using LeafCutter software with individual samples being compared to the others in the cohort as internal controls [53]. Fresard et al. performed whole blood RNA-seq on 94 rare disease patients compared to 49 unaffected relatives with additional comparison to existing datasets from 1594 controls [49]. By looking at outlier expression of candidate genes in patient samples as likely evidence for a loss-of-function variant, and by looking at outlier splice junction usage in a similar way, the authors successfully identified a causal variant in 7.5% and highlighted a candidate gene in 16.7% of patients. Globin depletion and poly(A) selection were used and sequencing was performed at around 50 million reads per sample with a mixture of 75-bp and 150-bp paired end reads. Hamanaka et al. performed a focussed study on six undiagnosed cases of nemaline myopathy and undertook RNA-seq on muscle biopsies, fibroblasts and lymphoblastoid cell lines using poly(A) selection and stranded library preparation with 92-bp paired-end reads [50]. By analysing splicing across 161 muscle disorder genes and using LeafCutter, four out of six cases were found to have *NEB* splicing mutations in their second alleles. Gonorazky et al. again looked at neuromuscular conditions and performed RNA-seq on 25 undiagnosed patients and four positive control patients with known disorders, utilising GTEx control samples for comparison [51]. Samples were taken either from skeletal muscle, cultured fibroblasts or from myotubes transdifferentiated from fibroblasts. Library preparation used poly(A) selection (or ribodepletion in one family) and sequencing employed 50-100 million 126-bp paired-end reads per sample. Splice junction filtering was carried out based upon the method of Cummings et al. and the overall diagnostic rate in this study was 36% using combined analysis of splicing, allelic imbalance and gene expression outliers. Finally, in our own study, we analysed 257 VUSs in rare disease patients by RT-PCR of whole blood RNA and in 17 cases also performed RNA-seq using ribodepletion and globin depletion with stranded library preparation and 70 million 150-bp paired-end reads per sample [10]. In four cases the RNA-seq analysis confirmed abnormal splicing seen by RT-PCR but in one case RNA-seq revealed a splice mutation previously undetected by RT-PCR, whilst in another case the abnormal RT-PCR event had insufficient read support in the RNA-seq data to reliably report.

3. Bioinformatic Tools in Splicing Analysis

A growing plethora of bioinformatic tools are available for analysis of splicing. These can be broadly divided into those aiming to predict the occurrence of splicing based on DNA sequence data and those that seek to identify changes in normal splicing within RNA-seq data. Prediction of splicing from DNA has long been something of a 'holy grail' in molecular biology and much has been written in search of a 'splicing code' [54-57]. However, to date a comprehensive code remains elusive. This should perhaps not be especially surprising, given the complexity of the splicing system and the many influences it receives from both *cis*- and *trans*-acting elements whose effects are context-dependent and which are themselves subject to differential regulation from tissue to tissue and from cell to cell.

From the clinical perspective of variant interpretation, several splice prediction programs are in common usage, most of which were first developed over at least a decade ago. SpliceSiteFinder-like computes donor and acceptor splice site scores based on a sequence scoring algorithm first published in 1987 [58]. NNSplice (1997) uses a neural network approach to predict donor and

acceptor splice sites by analysis of dinucleotide frequencies [59]. GeneSplicer (2001) uses maximal dependence decomposition enhanced with Markov modelling to predict splice sites from sequences focussing on a 16-nt region around the putative donor site and a 29-nt region around the putative acceptor site but also incorporating information from up to 80 nt flanking the predicted sites [60]. Another commonly used and reliably performing algorithm is MaxEntScan (2004), which relies on maximum entropy modelling to score 9-nt sequence motifs as splice donor sites and 23-nt sequence motifs as splice acceptor sites [61]. Human Splicing Finder (2009), incorporates a range of different splice prediction tools but principally uses position weight matrices to predict the strengths of donor (9-mer matrix) and acceptor (14-mer matrix) splice sites [62]. More recently, SpliceAI has been developed using a deep learning neural network approach to predict splice donor and acceptor sites from within the context of 10,000 nt of flanking sequence [63].

The splicing predictions of these tools have been compared against the results of experimentally determined splicing effects and sensitivities and specificities of between 70-95% are variously reported [8-10]. The machine learning approach of SpliceAI in particular has shown itself to frequently outperform other algorithms in this regard. However, the accuracy of all such predictions does somewhat depend on user-defined criteria of what scores to accept as significant. There is also some variability between 5' and 3' splice site predictions and a general decrease in accuracy with increasing distance from canonical splice regions. Furthermore, limitations in our understanding of splicing mutations mean that GT>GC 5' splice donor site variants, which can quite often retain the ability to splice correctly, are often misinterpreted in predictions [64]. Interpretation of variants affecting putative splice regulatory elements is another area of uncertainty and currently in most cases lies outside the scope of clinical application. However, a number of predictive tools exist that try to identify such regulatory elements, although again several of these commonly used tools were developed over 15 years ago and it may be that more modern machine learning techniques will prove helpful in future when applied to these problems. ESEfinder searches for putative exonic splicing enhancers (ESEs) in query sequences using SELEX-determined 6-8-nt motifs that bind the serine/arginine-rich (SR) proteins SF2/ASF (SRSF1), SC35 (SRSF2), SRp40 (SRSF5) and SRp55 (SRSF6) [65]. RESCUE-ESE is a computational method that looks for putative ESE hexamer sequences that are enriched in exons compared to introns and that are more frequent in exons with non-consensus splice sites [66]. Sequences forming exonic splicing silencers (ESSs) have also been investigated experimentally and these can be searched for in sequences using tools such as FAS-ESS [67]. Computational predictive methods have also been developed to try to identify ESS sequences by looking at motif enrichment within pseudoexons [68, 69]. The prediction of RNA-binding protein (RBP) interactions with RNA targets is intrinsically linked to the identification of enhancer and silencer elements. Databases of experimentally determined RBP motifs can be used to query sequences for potential splice factor binding sites via tools such as SpliceAid 2 and RBPmap [70, 71]. Deep learning has also recently been applied to predictions of RBP binding sites and changes in RNA-protein interactions based upon sequence changes [72].

Beyond the prediction of splicing, an even larger and ever-growing cohort of tools have been developed to try to detect alternative splicing from RNA-seq data. Cufflinks was one of the first such programs to attempt transcript isoform quantification using a probabilistic method [73]. MISO (mixture-of-isoforms) is a model that statistically estimates the expression of alternatively

spliced exons and their isoforms [74]. Insert length information is incorporated into the probabilistic assignment of read pairs to specific isoforms, which appears to increase the accuracy of PSI estimates. DEXSeq statistically tests for differential exon usage via the fitting of negative binomial generalised linear models [75]. This is a computationally intense process and also relies on the transcript inventory being predefined. rMATS (replicate multivariate analysis of transcript splicing) employs statistical modelling to detect differential alternative splicing events between groups of replicate samples with RNA-seq data [76]. It uses a hierarchical framework to model variability among replicates as well as modelling the estimation uncertainty of isoform proportions within each replicate. MAJIQ (Modeling Alternative Junction Inclusion Quantification) uses GFF3 transcript annotations and also identifies unannotated exons from sample .bam files to characterise and quantify local splicing variations in terms of PSI values and changes in PSI [20]. LeafCutter analyses mapped split reads to identify and quantify alternative splicing without requiring isoform inference [53]. It is based upon intron excision events and consequently does not detect intron retention. However, it is memory efficient in terms of processing and is therefore computationally fast.

4. Antisense Oligonucleotide Correction of Splicing Mutations

Since splice site selection is heavily reliant on the recognition of sequence motifs by the spliceosome and by splicing factors, masking of such motifs within specific pre-mRNA molecules can prove an effective way to manipulate specific splice events. This idea forms the basis for the growing number of splice-switching antisense oligonucleotide (ASO) compounds that are undergoing drug development or in some cases are now in clinical use. ASOs are chemical analogues of nucleic acids that retain the ability to perform Watson-Crick base pairing with their complementary RNA targets but which usually have chemical modifications of their backbone structure both to enhance stability and resist nuclease degradation and also to help direct their mechanism of action based upon their chemistry [77]. Commonly used modifications in currently available ASO drugs include 2'-O-methyl (2'OMe) and 2'-O-methoxy-ethyl (2'MOE) ribose sugar modifications in combination with phosphorothioate (PS) linkages in place of phosphate, and phosphorodiamidate morpholino (PMO) compounds, which employ a morpholine ring configuration instead of a sugar [78, 79].

Importantly, the chemical design of an ASO will determine the cellular pathway by which it acts [80]. A significant proportion of ASO drugs currently in development and/or in clinical use target dominantly inherited diseases such as Huntington disease (IONIS-HTT_{Rx} now known as RG6042), hereditary transthyretin-related amyloidosis (inotersen), *SOD1*-related amyotrophic lateral sclerosis (tofersen) and others, where a toxic accumulation of aberrant protein products is linked to disease pathology [81-83]. Non-splice-switching ASOs of this type typically utilise a "gapmer" design, whereby the flanking nucleotides employ nuclease-resistant modifications such as 2'MOE-PS, while the internal nucleotides retain a more natural DNA-like structure (for example only utilising PS linkages) so as to retain the ability to engage RNase H enzymes (primarily RNase H1) when bound as a heteroduplex to their target RNA, inducing its cleavage [84]. However, for splice-switching ASOs, the aim is not to induce RNase H-mediated cleavage but simply to act as a steric blocker and so their chemical design tends to utilise nuclease-resistant modifications throughout. An additional factor to consider in the design of PS-modified ASOs is stereoisomerism, since the

use of PS linkages introduces chirality around the bridging phosphorus atom of the backbone [85]. This can effectively result in such drugs comprising highly heterogeneous mixtures of stereoisomers with differing physicochemical and pharmacological properties. On account of this, methods have now been developed that allow production of stereopure ASOs and indeed control of stereochemistry has been shown to significantly improve ASO stability and efficacy [86].

To date, at least 10 different ASO drugs have been licensed for clinical use across the world and of these, four involve manipulation of splicing (see Table 1) [11]. The most dramatically effective of these drugs so far has been nusinersen, a 2'MOE phosphorothioate compound targeting an intronic splicing silencer element (ISS-N1) located in intron 7 of the *SMN2* gene [87]. Children with spinal muscular atrophy (SMA) have biallelic *SMN1* gene mutations causing motor neurone degeneration and death in infancy [88]. The highly homologous duplicated gene *SMN2* can potentially compensate for *SMN1* loss but usually skips exon 7 leading to an unstable protein [89]. However, when given intrathecally to infants with SMA, the nusinersen ASO sterically blocks the ISS-N1 silencer and promotes exon 7 inclusion within *SMN2* transcripts [90]. This treatment leads to dramatically improved motor function in affected children and has changed the natural history of SMA from a lethal disease of infancy to one where the condition appears to be treatable and manageable with motor milestones of unaided sitting, standing and walking being achieved [91-93]. Later-onset milder forms of SMA have also been found to demonstrate improvement following ASO treatment [94]. Furthermore, when treatment is started pre-symptomatically in early infancy, current trial evidence suggests that motor milestones can actually be rescued to within the normal range in the majority of cases [95].

Table 1 Clinically licensed splice-switching ASO drugs. Golodirsen and viltolarsen both have the same PMO chemistry and target the same *DMD* exon but have slightly differing sequences. 2'MOE, 2'O-methoxyethyl phosphorothioate; PMO, phosphorodiamidate morpholino.

ASO drug	Chemistry	Gene target	Mechanism of action	Year approved	first
Eteplirsen	PMO	<i>DMD</i>	Binds exon 51 inducing exon skipping to restore reading frame	2016 [96]	
Nusinersen	2'MOE	<i>SMN2</i>	Binds ISS in intron 7 to promote exon inclusion	2016 [97]	
Golodirsen	PMO	<i>DMD</i>	Binds exon 53 inducing exon skipping to restore reading frame	2019 [98]	
Viltolarsen	PMO	<i>DMD</i>	Binds exon 53 inducing exon skipping to restore reading frame	2020 [99]	

Although the ASO drugs licensed so far have been for SMA and for Duchenne muscular dystrophy (DMD), neither of which are typically caused by splicing mutations *per se*, ASO-based approaches do naturally lend themselves to the therapeutic silencing of cryptic splice sites.

However, this brings with it a difficulty of scale, since most such mutations are novel or so-called 'private' mutations and are not widely shared amongst cohorts of individuals affected by rare diseases. Nevertheless, the sequence specificity of ASO design means that these compounds, perhaps above and beyond any other pharmacological modality, have the potential to be used as truly personalised medicines. One notable example of this has been the development of milasen, a 22-mer 2' MOE ASO that was designed solely for the treatment of a specific individual, a child named Mila with a diagnosis of *CLN7*-related Batten disease [100]. Milasen targets and silences a cryptic splice site introduced by insertion of a transposable element within intron 6 of the *CLN7* gene. This 2kb retrotransposition event was undetectable by initial exome sequencing but was identified by whole genome analysis. Remarkably, the time that elapsed between confirming the genetic diagnosis in this case and delivering the first intrathecal injection of the drug was less than one year.

5. *Trans*-Splicing Therapy

Whilst ASO compounds represent an easily adaptable and intuitive means by which to therapeutically manipulate splicing, they are not the only way in which to do so. One alternative approach is to employ the phenomenon of *trans*-splicing [101-103]. This is where splicing occurs across two separate RNA molecules using the splice donor site from one and the splice acceptor site from the other. The process was originally identified in trypanosomes but has subsequently been found to be a widespread feature of natural mRNA processing across viruses, prokaryotes and higher eukaryotes including humans [104-112]. Despite the occurrence of *trans*-splicing being much lower in vertebrates compared to protozoa and its physiological role being for the most part poorly understood, its potential for application as a therapeutic strategy of splicing correction has been demonstrated for a number of diseases, including cystic fibrosis, haemophilia, Duchenne muscular dystrophy and also correction of mutated *TP53* in hepatocellular carcinoma [113-116]. This can be achieved through substitution of part of a mutated pre-mRNA sequence with a corrected coding sequence. The most widely described version of this approach is spliceosome-mediated RNA *trans*-splicing (SMaRT), where a pre-mRNA *trans*-splicing molecule (PTM) can be designed that contains the following features: a binding domain sequence complementary to the target intron, an artificial intronic sequence region including polypyridine tract and branch point and a coding sequence flanked by the appropriate splice site (either 5' or 3' depending on the position of the desired splicing replacement). By including strong splice sites within the PTM, the replacement sequence is able to compete against the native molecule's splice sites and achieve *trans*-splicing [117].

Despite *trans*-splicing representing a promising therapeutic approach, its use has thus far been limited by several factors. These include frequently low rates of *trans*-splicing efficiency, issues of adequate PTM delivery to target cells, potential for off-target *trans*-splicing to affect other genes and the potential for aberrant *cis*-splicing of the PTM itself and unintentional PTM translation [101, 118]. Nevertheless, continued development and refinement of *trans*-splicing technology will likely prove beneficial, not only in terms of understanding its biology but also by offering a potential therapeutic solution for genomic variants unamenable to ASO-mediated therapy. Whilst alternative approaches such as clustered regularly interspaced short palindromic repeat/CRISPR associated protein 9 (CRISPR/Cas9) gene editing do of course exist for the targeted correction of

almost any given genomic variant, RNA-based therapies benefit from their pharmacological titratability, their relative ease of manufacture and in most cases the need to only deliver a single therapeutic compound rather than a combination.

6. Conclusion

We are now able to predict and detect clinically relevant splicing abnormalities more accurately and more easily than ever before. In some cases we are also now learning how to correct the abnormal splicing and to treat the resulting disease. This parallel advancement and convergence of technologies means that we are in effect gradually accumulating all the prerequisite knowledge and expertise needed for the development of a personalised medicine pipeline of splice-modulating therapeutics (see Figure 4). As the detection of splicing mutations becomes easier and more widely implemented in a clinical setting, the next main focus of investigation that will likely need much greater research effort and investment is in the understanding of splicing regulation. Whilst regulatory elements can be predicted bioinformatically to a degree, there remains no substitute for wet-lab-based experimental work in this regard. Tools such as minigene assays and CRISPR/Cas9 genome editing screens facilitate the investigation of splicing effects in response to sequence element changes, whilst molecular biological confirmations of predicted macromolecular interactions will always be needed [119, 120]. In determining the individual regulatory elements of specific mis-splicing events, it should in many cases become feasible to design bespoke splice-switching ASOs and other compounds to help shift the balance of splicing back towards normality. Better understanding of molecular pathogenesis pathways should also bring to light alternative therapeutic targets, not only for correction of abnormal splicing *per se* but also for up- and down-regulation of relevant target genes, for example through destructive splice-switching [121]. Thus, notwithstanding the considerable challenges inherent in RNA-targeted drug development, such as ensuring adequate tissue drug delivery, the future looks bright for splice-switching therapeutics, as evidenced by the multibillion dollar industry that ASO pharmaceuticals have become [122].

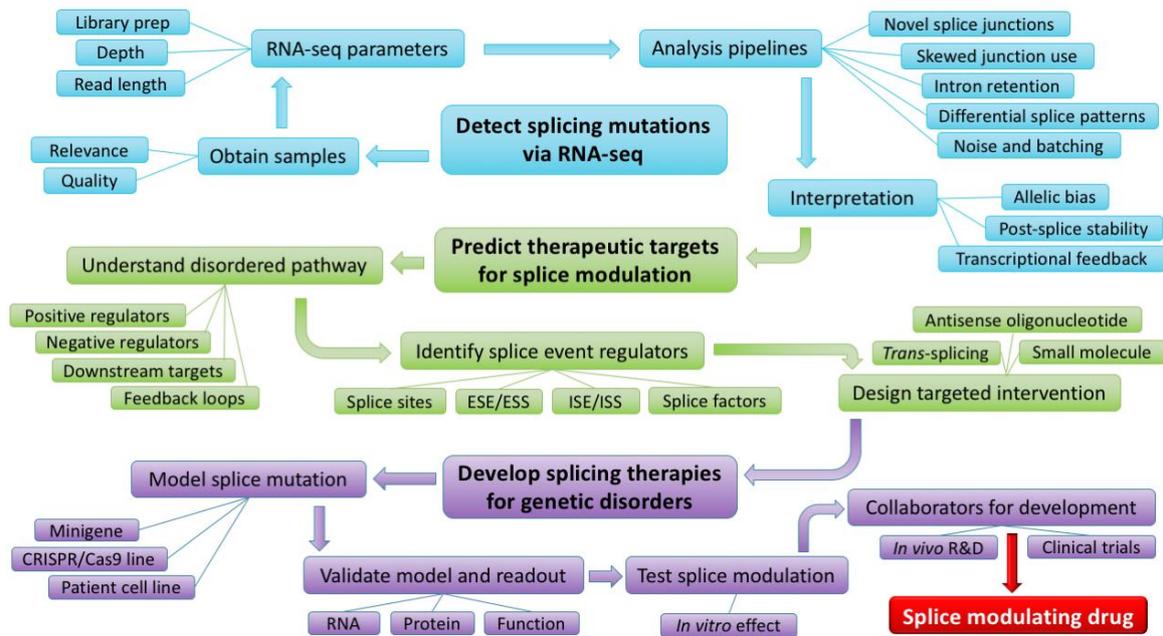


Figure 4 From RNA splicing analysis to personalised splice-modulating therapies. Detecting splicing mutations from RNA-seq data requires not only appropriate samples and sequencing parameters but also comprehensive analysis and interpretation. Designing therapeutically effective splice-switching compounds requires an understanding of splicing regulation and knowledge of a disease's molecular pathogenesis, since targeting other genes in a pathway may be an alternative route to achieving therapeutic benefit. Adequate modelling of abnormal splice events and accurate validation of their correction is a prerequisite for developing a splice modulating drug. Later stage research and development (R&D) trials generally require pharmaceutical industry collaboration.

Having said this, a number of key issues still need to be addressed if we are to bring to reality the dream of an RNA diagnostics to RNA therapeutics pipeline. To begin with, RNA-seq will need to be brought from the research laboratory setting into routine clinical diagnostic practice for rare disease, along with the necessary standard operating procedures and accreditations. Aside from the technical aspects of how to control for variable batch effects in sequencing and how to deal with tissue-specific splicing and splicing artefacts apparent in read mapping, a critical part of this will be the development of clinical guidelines relating to how splicing abnormalities should be interpreted in terms of their pathogenicity in variant classification. Initial attempts at such guidelines have been made in relation to cancer susceptibility genes but it is likely that a much more nuanced and perhaps experimentally evidenced approach will be needed in order to try to take account of the complexity of RNA metabolism and splice isoform regulation [123]. Beyond diagnostics, funding of translational research into therapeutic splicing manipulation will be key. Few rare disease families have access to the philanthropy and crowd-sourced funding that made milasen's rapid development possible. Going forward, it will be important for all relevant stakeholders from family support groups and charities through to researchers, research funders and drug companies, together with clinicians, medicines regulators and wider society at large to discuss and consider how these novel technologies should best be used and how they can be

utilised in a fair and equitable way for all those in need. Only then can we hope to bridge the translational gap in personalised medicine, completing the circle from RNA diagnostics to personalised splicing therapeutics.

Author Contributions

Both AD and DB were involved in the design and composition of this manuscript. The manuscript was drafted by AD and reviewed and edited by DB.

Funding

AD and DB are funded by a NIHR Research Professorship grant awarded to DB (RP-2016-07-011).

Competing Interests

The authors have declared that no competing interests exist.

References

1. Wakap SN, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, et al. Estimating cumulative point prevalence of rare diseases: Analysis of the orphanet database. *Eur J Hum Genet.* 2020; 28: 165-173.
2. Wright CF, FitzPatrick DR, Firth HV. Paediatric genomics: Diagnosing rare disease in children. *Nat Rev Genet.* 2018; 19: 253-268.
3. Dawkins HJ, Draghia-Akli R, Lasko P, Lau LP, Jonker AH, Cutillo CM, et al. Progress in rare diseases research 2010-2016: An IRDiRC perspective. *Clin Transl Sci.* 2018; 11: 11-20.
4. Wai H, Douglas AG, Baralle D. RNA splicing analysis in genomic medicine. *Int J Biochem Cell Biol.* 2019; 108: 61-71.
5. Marco-Puche G, Lois S, Benítez J, Trivino JC. RNA-Seq perspectives to improve clinical diagnosis. *Front Genet.* 2019; 10: 1152.
6. Murdock DR. Enhancing diagnosis through RNA sequencing. *Clin Lab Med.* 2020; 40: 113-119.
7. Stenton SL, Prokisch H. The clinical application of RNA sequencing in genetic diagnosis of Mendelian disorders. *Clin Lab Med.* 2020; 40: 121-133.
8. Houdayer C, Caux-Moncoutier V, Krieger S, Barrois M, Bonnet F, Bourdon V, et al. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum Mutat.* 2012; 33: 1228-1238.
9. Tang R, Prosser DO, Love DR. Evaluation of bioinformatic programmes for the analysis of variants within splice site consensus regions. *Adv Bioinformatics.* 2016; 2016: 5614058.
10. Wai HA, Lord J, Lyon M, Gunning A, Kelly H, Cibir P, et al. Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance. *Genet Med.* 2020; 22: 1005-1014.
11. Kuijper EC, Bergsma AJ, Pijnappel WP, Aartsma-Rus A. Opportunities and challenges for antisense oligonucleotide therapies. *J Inher Metab Dis.* 2021; 44: 72-87.
12. Farrell RE. RT-PCR: A science and an art form. In *RNA Methodologies*. 4th ed. London: Academic Press; 2010. pp. 385-448.

13. Martín-Alonso S, Frutos-Beltrán E, Menéndez-Arias L. Reverse transcriptase: From transcriptomics to genome editing. *Trends Biotechnol.* 2021; 39: 194-210.
14. Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol.* 2000; 25: 169-193.
15. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 2001; 29: e45.
16. Van Heetvelde M, Van Loocke W, Trypsteen W, Baert A, Vanderheyden K, Crombez B, et al. Evaluation of relative quantification of alternatively spliced transcripts using droplet digital PCR. *Biomol Detect Quantif.* 2017; 13: 40-48.
17. Sun B, Tao L, Zheng YL. Simultaneous quantification of alternatively spliced transcripts in a single droplet digital PCR reaction. *Biotechniques.* 2014; 56: 319-325.
18. Schindler S, Heiner M, Platzer M, Szafranski K. Comparison of methods for quantification of subtle splice variants. *Electrophoresis.* 2009; 30: 3674-3681.
19. Liew CC, Ma J, Tang HC, Zheng R, Dempsey AA. The peripheral blood transcriptome dynamically reflects system wide biology: A potential diagnostic tool. *J Lab Clin Med.* 2006; 147: 126-132.
20. Aicher JK, Jewell P, Vaquero-Garcia J, Barash Y, Bhoj EJ. Mapping RNA splicing variations in clinically accessible and nonaccessible tissues to facilitate Mendelian disease diagnosis using RNA-seq. *Genet Med.* 2020; 22: 1181-1190.
21. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science.* 2015; 348: 648-660.
22. Martin AR, Williams E, Foulger RE, Leigh S, Daugherty LC, Niblock O, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet.* 2019; 51: 1560-1565.
23. Theda C, Hwang SH, Czajko A, Loke YJ, Leong P, Craig JM. Quantitation of the cellular content of saliva and buccal swab samples. *Sci Rep.* 2018; 8: 6944.
24. Bergsma AJ, In't Groen SL, Catalano F, Yamanaka M, Takahashi S, Okumiya T, et al. A generic assay for the identification of splicing variants that induce nonsense-mediated decay in Pompe disease. *Eur J Hum Genet.* 2020. Doi: 10.1038/s41431-020-00751-3.
25. Häuser F, Gökce S, Werner G, Danckwardt S, Sollfrank S, Neukirch C, et al. A non-invasive diagnostic assay for rapid detection and characterization of aberrant mRNA-splicing by nonsense mediated decay inhibition. *Mol Genet Metab.* 2020; 130: 27-35.
26. Evans DG, Bowers N, Burkitt-Wright E, Miles E, Garg S, Scott-Kitching V, et al. Comprehensive RNA analysis of the NF1 gene in classically affected NF1 affected individuals meeting NIH criteria has high sensitivity and mutation negative testing is reassuring in isolated cases with pigmentary features only. *EBioMedicine.* 2016; 7: 212-220.
27. Roberts RG, Barby TF, Manners E, Bobrow M, Bentley DR. Direct detection of dystrophin gene rearrangements by analysis of dystrophin mRNA in peripheral blood lymphocytes. *Am J Hum Genet.* 1991; 49: 298-310.
28. Hrdlickova R, Toloue M, Tian B. RNA-Seq methods for transcriptome analysis. *Wiley Interdiscip Rev RNA.* 2017; 8: e1364.
29. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, et al. The RIN: An RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol.* 2006; 7: 3.

30. Romero IG, Pai AA, Tung J, Gilad Y. RNA-seq: Impact of RNA degradation on transcript quantification. *BMC Biol.* 2014; 12: 42.
31. Mastrokolas A, den Dunnen JT, van Ommen GB, AC't Hoen P, van Roon-Mom WM. Increased sensitivity of next generation sequencing-based expression profiling after globin reduction in human blood RNA. *BMC Genom.* 2012; 13: 28.
32. Shin H, Shannon CP, Fishbane N, Ruan J, Zhou M, Balshaw R, et al. Variation in RNA-Seq transcriptome profiles of peripheral whole blood from healthy individuals with and without globin depletion. *PLoS ONE.* 2014; 9: e91041.
33. Krjutškov K, Koel M, Roost AM, Katayama S, Einarsdottir E, Jouhilahti EM, et al. Globin mRNA reduction for whole-blood transcriptome sequencing. *Sci Rep.* 2016; 6: 31584.
34. O'Neil D, Glowatz H, Schlumpberger M. Ribosomal RNA depletion for efficient use of RNA-seq capacity. *Curr Protoc Mol Biol.* 2013; 103: 4.19.1-4.19.8.
35. Chomczynski P, Wilfinger WW, Eghbalnia HR, Kennedy A, Rymaszewski M, Mackey K. Inter-individual differences in RNA levels in human peripheral blood. *PLoS ONE.* 2016; 11: e0148260.
36. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* 2012; 22: 1775-1789.
37. Sheng Q, Vickers K, Zhao S, Wang J, Samuels DC, Koues O, et al. Multi-perspective quality control of Illumina RNA sequencing data analysis. *Brief Funct Genomics.* 2017; 16: 194-204.
38. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013; 29: 15-21.
39. Kim D, Langmead B, Salzberg SL. HISAT: A fast spliced aligner with low memory requirements. *Nat Methods.* 2015; 12: 357-360.
40. Dozmorov MG, Adrianto I, Giles CB, Glass E, Glenn SB, Montgomery C, et al. Detrimental effects of duplicate reads and low complexity regions on RNA-and ChIP-seq data. *BMC Bioinform.* 2015; 16: S10.
41. Parekh S, Ziegenhain C, Vieth B, Enard W, Hellmann I. The impact of amplification on differential expression analyses by RNA-seq. *Sci Rep.* 2016; 6: 25533.
42. Fu Y, Wu PH, Beane T, Zamore PD, Weng Z. Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. *BMC Genom.* 2018; 19: 531.
43. Katz Y, Wang ET, Silterra J, Schwartz S, Wong B, Thorvaldsdóttir H, et al. Quantitative visualization of alternative exon expression from RNA-seq data. *Bioinformatics.* 2015; 31: 2400-2402.
44. Wang Y, Liu J, Huang BO, Xu YM, Li J, Huang LF, et al. Mechanism of alternative splicing and its regulation. *Biomed Rep.* 2015; 3: 152-158.
45. Bhadra M, Howell P, Dutta S, Heintz C, Mair WB. Alternative splicing in aging and longevity. *Hum Genet.* 2020; 139: 357-369.
46. Schafer S, Miao K, Benson CC, Heinig M, Cook SA, Hubner N. Alternative splicing signatures in RNA-seq data: Percent spliced in (PSI). *Curr Protoc Hum Genet.* 2015; 87: 11.16.1-11.16.14.
47. Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med.* 2017; 9: eaal5209.
48. Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, Iuso A, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun.* 2017; 8: 15824.

49. Frésard L, Smail C, Ferraro NM, Teran NA, Li X, Smith KS, et al. Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat Med.* 2019; 25: 911-919.
50. Hamanaka K, Miyatake S, Koshimizu E, Tsurusaki Y, Mitsuhashi S, Iwama K, et al. RNA sequencing solved the most common but unrecognized NEB pathogenic variant in Japanese nemaline myopathy. *Genet Med.* 2019; 21: 1629-1638.
51. Gonorazky HD, Naumenko S, Ramani AK, Nelakuditi V, Mashouri P, Wang P, et al. Expanding the boundaries of RNA sequencing as a diagnostic tool for rare Mendelian disease. *Am J Hum Genet.* 2019; 104: 466-483.
52. Wai HA, Lord J, Lyon M, Gunning A, Kelly H, Cibir P, et al. Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance. *Genet Med.* 2020; 22: 1005-1014.
53. Li YI, Knowles DA, Humphrey J, Barbeira AN, Dickinson SP, Im HK, et al. Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet.* 2018; 50: 151-158.
54. Wang Z, Burge CB. Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA.* 2008; 14: 802-813.
55. Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, et al. Deciphering the splicing code. *Nature.* 2010; 465: 53-59.
56. Xiong HY, Alipanahi B, Lee LJ, Bretschneider H, Merico D, Yuen RK, et al. The human splicing code reveals new insights into the genetic determinants of disease. *Science.* 2015; 347: 1254806.
57. Baralle M, Baralle FE. The splicing code. *Biosystems.* 2018; 164: 39-48.
58. Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: Sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* 1987; 15: 7155-7174.
59. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol.* 1997; 4: 311-323.
60. Perteira M, Lin X, Salzberg SL. GeneSplicer: A new computational method for splice site prediction. *Nucleic Acids Res.* 2001; 29: 1185-1190.
61. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol.* 2004; 11: 377-394.
62. Desmet FO, Hamroun D, Lalande M, Collod-Bérout G, Claustres M, Bérout C. Human Splicing Finder: An online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* 2009; 37: e67.
63. Jaganathan K, Panagiotopoulou SK, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting splicing from primary sequence with deep learning. *Cell.* 2019; 176: 535-548.
64. Chen JM, Lin JH, Masson E, Liao Z, Férec C, Cooper DN, et al. The experimentally obtained functional impact assessments of 5'splice site GT > GC variants differ markedly from those predicted. *Curr Genomics.* 2020; 21: 56-66.
65. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* 2003; 31: 3568-3571.
66. Fairbrother WG, Yeo GW, Yeh R, Goldstein P, Mawson M, Sharp PA, et al. RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. *Nucleic Acids Res.* 2004; 32: W187-W190.

67. Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB. Systematic identification and analysis of exonic splicing silencers. *Cell*. 2004; 119: 831-845.
68. Sironi M, Menozzi G, Riva L, Cagliani R, Comi GP, Bresolin N, et al. Silencer elements as possible inhibitors of pseudoexon splicing. *Nucleic Acids Res*. 2004; 32: 1783-1791.
69. Zhang XH, Chasin LA. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev*. 2004; 18: 1241-1250.
70. Piva F, Giulietti M, Burini AB, Principato G. SpliceAid 2: A database of human splicing factors expression data and RNA target motifs. *Hum Mutat*. 2012; 33: 81-85.
71. Paz I, Kosti I, Ares Jr M, Cline M, Mandel-Gutfreund Y. RBPmap: A web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res*. 2014; 42: W361-W367.
72. Grønning AG, Doktor TK, Larsen SJ, Petersen US, Holm LL, Bruun GH, et al. DeepCLIP: Predicting the effect of mutations on protein–RNA binding with deep learning. *Nucleic Acids Res*. 2020; 48: 7099-7118.
73. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010; 28: 511-515.
74. Katz Y, Wang ET, Airoidi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods*. 2010; 7: 1009-1015.
75. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. *Genome Res*. 2012; 22: 2008-2017.
76. Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, et al. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A*. 2014; 111: E5593-E5601.
77. Khvorova A, Watts JK. The chemical evolution of oligonucleotide therapies of clinical utility. *Nat Biotechnol*. 2017; 35: 238-248.
78. Manoharan M. 2'-Carbohydrate modifications in antisense oligonucleotide therapy: Importance of conformation, configuration and conjugation. *Biochim Biophys Acta Gene Struct Expr*. 1999; 1489: 117-130.
79. Summerton JE. Morpholino, siRNA, and S-DNA compared: Impact of structure and mechanism of action on off-target effects and sequence specificity. *Curr Top Med Chem*. 2007; 7: 651-660.
80. Lundin KE, Gissberg O, Smith CE. Oligonucleotide therapies: The past and the present. *Hum Gene Ther*. 2015; 26: 475-485.
81. Tabrizi SJ, Leavitt BR, Landwehrmeyer GB, Wild EJ, Saft C, Barker RA, et al. Targeting huntingtin expression in patients with Huntington's disease. *N Engl J Med*. 2019; 380: 2307-2316.
82. Benson MD, Waddington-Cruz M, Berk JL, Polydefkis M, Dyck PJ, Wang AK, et al. Inotersen treatment for patients with hereditary transthyretin amyloidosis. *N Engl J Med*. 2018; 379: 22-31.
83. Miller T, Cudkowicz M, Shaw PJ, Andersen PM, Atassi N, Bucelli RC, et al. Phase 1-2 trial of antisense oligonucleotide tofersen for SOD1 ALS. *N Engl J Med*. 2020; 383: 109-119.
84. Wu H, Lima WF, Zhang H, Fan A, Sun H, Crooke ST. Determination of the role of the human RNase H1 in the pharmacology of DNA-like antisense drugs. *J Biol Chem*. 2004; 279: 17181-17189.
85. Eckstein F. Developments in RNA chemistry, a personal view. *Biochimie*. 2002; 84: 841-848.

86. Iwamoto N, Butler DC, Svrzikapa N, Mohapatra S, Zlatev I, Sah DW, et al. Control of phosphorothioate stereochemistry substantially increases the efficacy of antisense oligonucleotides. *Nat Biotechnol.* 2017; 35: 845-851.
87. Corey DR. Nusinersen, an antisense oligonucleotide drug for spinal muscular atrophy. *Nat Neurosci.* 2017; 20: 497-499.
88. Lunn MR, Wang CH. Spinal muscular atrophy. *Lancet.* 2008; 371: 2120-2133.
89. Lorson CL, Hahnen E, Androphy EJ, Wirth B. A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc Natl Acad Sci U S A.* 1999; 96: 6307-6311.
90. Groen EJ, Talbot K, Gillingwater TH. Advances in therapy for spinal muscular atrophy: Promises and challenges. *Nat Rev Neurol.* 2018; 14: 214-224.
91. Finkel RS, Chiriboga CA, Vajsaar J, Day JW, Montes J, De Vivo DC, et al. Treatment of infantile-onset spinal muscular atrophy with nusinersen: A phase 2, open-label, dose-escalation study. *Lancet.* 2016; 388: 3017-3026.
92. Mercuri E, Darras BT, Chiriboga CA, Day JW, Campbell C, Connolly AM, et al. Nusinersen versus sham control in later-onset spinal muscular atrophy. *N Engl J Med.* 2018; 378: 625-635.
93. Tizzano EF, Finkel RS. Spinal muscular atrophy: A changing phenotype beyond the clinical trials. *Neuromuscul Disord.* 2017; 27: 883-889.
94. Darras BT, Chiriboga CA, Iannaccone ST, Swoboda KJ, Montes J, Mignon L, et al. Nusinersen in later-onset spinal muscular atrophy: Long-term results from the phase 1/2 studies. *Neurology.* 2019; 92: e2492-e2506.
95. Darryl C, Bertini E, Swoboda KJ, Hwu WL, Crawford TO, Finkel RS, et al. Nusinersen initiated in infants during the presymptomatic stage of spinal muscular atrophy: Interim efficacy and safety results from the Phase 2 NURTURE study. *Neuromuscul Disord.* 2019; 29: 842-856.
96. Syed YY. Eteplirsen: First global approval. *Drugs.* 2016; 76: 1699-1704.
97. Hoy SM. Nusinersen: First global approval. *Drugs.* 2017; 77: 473-479.
98. Heo YA. Golodirsen: First approval. *Drugs.* 2020; 80: 329-333.
99. Dhillon S. Viltolarsen: First approval. *Drugs.* 2020; 80: 1027-1031.
100. Kim J, Hu C, Moufawad El Achkar C, Black LE, Douville J, Larson A, et al. Patient-customized oligonucleotide therapy for a rare genetic disease. *N Engl J Med.* 2019; 381: 1644-1652.
101. Berger A, Maire S, Gaillard MC, Sahel JA, Hantraye P, Bemelmans AP. mRNA trans-splicing in gene therapy for genetic diseases. *Wiley Interdiscip Rev RNA.* 2016; 7: 487-498.
102. Lei Q, Li C, Zuo Z, Huang C, Cheng H, Zhou R. Evolutionary insights into RNA trans-splicing in vertebrates. *Genome Biol Evol.* 2016; 8: 562-577.
103. Hong EM, Ingemarsdotter CK, Lever AM. Therapeutic applications of trans-splicing. *Br Med Bull.* 2020; 136: 4-20.
104. Murphy WJ, Watkins KP, Agabian N. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: Evidence for trans splicing. *Cell.* 1986; 47: 517-525.
105. Sutton RE, Boothroyd JC. Evidence for trans splicing in trypanosomes. *Cell.* 1986; 47: 527-535.
106. Salvo JL, Coetzee T, Belfort M. Deletion-tolerance and trans-splicing of the bacteriophage T4 td intron: Analysis of the P6-L6a region. *J Mol Biol.* 1990; 211: 537-549.
107. Eul J, Patzel V. Homologous SV40 RNA trans-splicing: A new mechanism for diversification of viral sequences and phenotypes. *RNA Biol.* 2013; 10: 1689-1699.

108. Randau L, Münch R, Hohn MJ, Jahn D, Söll D. Nanoarchaeum equitans creates functional tRNAs from separate genes for their 5'-and 3'-halves. *Nature*. 2005; 433: 537-541.
109. Belhocine K, Mak AB, Cousineau B. Trans-splicing of the Ll. LtrB group II intron in *Lactococcus lactis*. *Nucleic Acids Res*. 2007; 35: 2257-2268.
110. Flouriot G, Brand H, Seraphin B, Gannon F. Natural trans-spliced mRNAs are generated from the human estrogen receptor- α (hER α) gene. *J Biol Chem*. 2002; 277: 26244-26251.
111. Romani A, Guerra E, Trerotola M, Alberti S. Detection and analysis of spliced chimeric mRNAs in sequence databanks. *Nucleic Acids Res*. 2003; 31: e17.
112. Wu CS, Yu CY, Chuang CY, Hsiao M, Kao CF, Kuo HC, et al. Integrative transcriptome sequencing identifies trans-splicing events with important roles in human embryonic stem cell pluripotency. *Genome Res*. 2014; 24: 25-36.
113. Mansfield SG, Clark RH, Puttaraju M, Kole J, Cohn JA, Mitchell LG, et al. 5' exon replacement and repair by spliceosome-mediated RNA trans-splicing. *RNA*. 2003; 9: 1290-1297.
114. Koo T, Popplewell L, Athanasopoulos T, Dickson G. Triple trans-splicing adeno-associated virus vectors capable of transferring the coding sequence for full-length dystrophin protein into dystrophic mice. *Hum Gene Ther*. 2014; 25: 98-108.
115. Chao H, Mansfield SG, Bartel RC, Hiriyantha S, Mitchell LG, Garcia-Blanco MA, et al. Phenotype correction of hemophilia A mice by spliceosome-mediated RNA trans-splicing. *Nat Med*. 2003; 9: 1015-1019.
116. He X, Liu F, Yan J, Zhang Y, Yan J, Shang H, et al. Trans-splicing repair of mutant p53 suppresses the growth of hepatocellular carcinoma cells in vitro and in vivo. *Sci Rep*. 2015; 5: 8705.
117. Philippi S, Lorain S, Beley C, Peccate C, Précigout G, Spuler S, et al. Dysferlin rescue by spliceosome-mediated pre-mRNA trans-splicing targeting introns harbouring weakly defined 3' splice sites. *Hum Mol Genet*. 2015; 24: 4049-4060.
118. Monjaret F, Bourg N, Suel L, Roudaut C, Le Roy F, Richard I, et al. Cis-splicing and translation of the pre-trans-splicing molecule combine with efficiency in spliceosome-mediated RNA trans-splicing. *Mol Ther*. 2014; 22: 1176-1187.
119. Gaildrat P, Killian A, Martins A, Tournier I, Frébourg T, Tosi M. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. *Methods Mol Biol*. 2010; 653: 249-257.
120. Liu Y, Cao Z, Wang Y, Guo Y, Xu P, Yuan P, et al. Genome-wide screening for functional long noncoding RNAs in human cells by Cas9 targeting of splice sites. *Nat Biotechnol*. 2018; 36: 1203-1210.
121. Lu-Nguyen N, Malerba A, Popplewell L, Schnell F, Hanson G, Dickson G. Systemic antisense therapeutics for dystrophin and myostatin exon splice modulation improve muscle pathology of adult mdx mice. *Mol Ther Nucleic Acids*. 2017; 6: 15-28.
122. Wang F, Zuroske T, Watts JK. RNA therapeutics on the rise. *Nat Rev Drug Discov*. 2020; 19: 441-442.
123. Garrett A, Callaway A, Durkie M, Cubuk C, Alikian M, Burghel GJ, et al. Cancer Variant Interpretation Group UK (CanVIG-UK): An exemplar national subspecialty multidisciplinary network. *J Med Genet*. 2020; 57: 829-834.



Enjoy *OBM Genetics* by:

1. [Submitting a manuscript](#)
2. [Joining in volunteer reviewer bank](#)
3. [Joining Editorial Board](#)
4. [Guest editing a special issue](#)

For more details, please visit:

<http://www.lidsen.com/journals/genetics>