

Review



**Cite this article:** Pacini C, Koziol MJ. 2018 Bioinformatics challenges and perspectives when studying the effect of epigenetic modifications on alternative splicing. *Phil. Trans. R. Soc. B* **373**: 20170073. <http://dx.doi.org/10.1098/rstb.2017.0073>

Accepted: 14 September 2017

One contribution of 18 to a discussion meeting issue 'Frontiers in epigenetic chemical biology'.

**Subject Areas:**

bioinformatics, genomics, theoretical biology

**Keywords:**

epigenetics, splicing, RNA-SEQ, bioinformatics, modelling, networks

**Author for correspondence:**

Magdalena J. Koziol  
e-mail: [mjk39@cam.ac.uk](mailto:mjk39@cam.ac.uk)

# Bioinformatics challenges and perspectives when studying the effect of epigenetic modifications on alternative splicing

Clare Pacini<sup>1,2</sup> and Magdalena J. Koziol<sup>1,2</sup>

<sup>1</sup>Wellcome Trust Cancer Research UK Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QN, UK

<sup>2</sup>Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, UK

CP, 0000-0001-7791-0940; MJK, 0000-0002-4123-4343

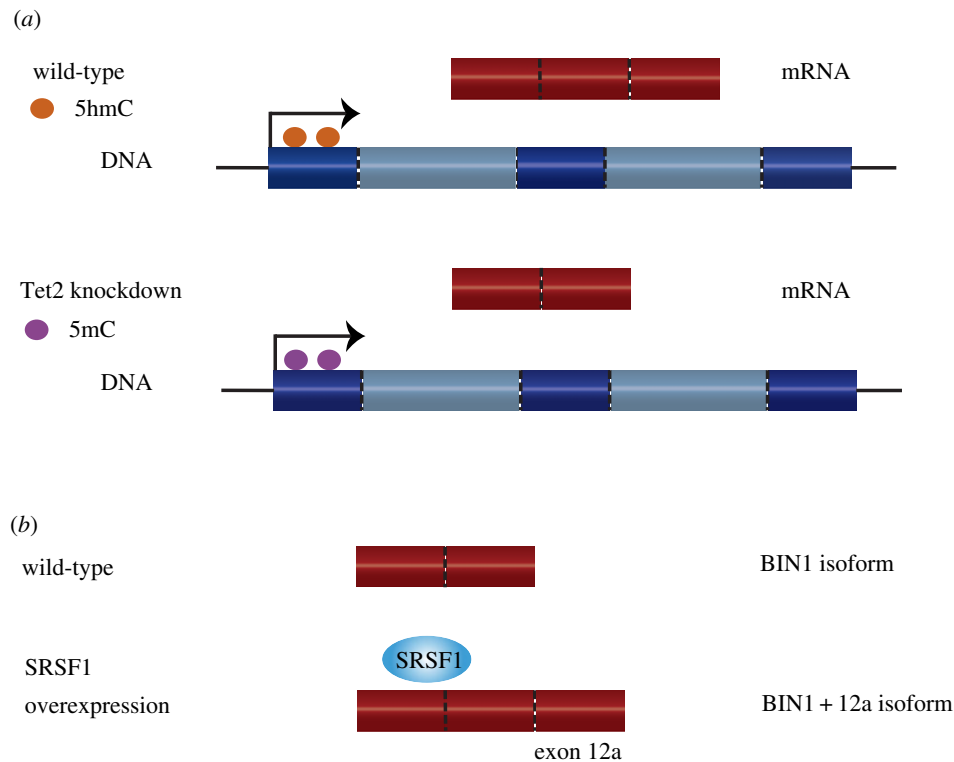
It is widely known that epigenetic modifications are important in regulating transcription, but several have also been reported in alternative splicing. The regulation of pre-mRNA splicing is important to explain proteomic diversity and the misregulation of splicing has been implicated in many diseases. Here, we give a brief overview of the role of epigenetics in alternative splicing and disease. We then discuss the bioinformatics methods that can be used to model interactions between epigenetic marks and regulators of splicing. These models can be used to identify alternative splicing and epigenetic changes across different phenotypes.

This article is part of a discussion meeting issue 'Frontiers in epigenetic chemical biology'.

## 1. Introduction

Epigenetic modifications change the regulation of the genome without changing the DNA sequence. As such, our increased knowledge of epigenetics significantly improves our understanding of gene regulation and biological pathways that are regulated epigenetically. Epigenetic modifications can occur directly on DNA, and on histones that bind to DNA. For example, chemical modifications such as methylation, acetylation and phosphorylation are known to alter histone function. Each of these modifications is reversible, and this has increased the prospects of epigenetic markers as therapeutic drug targets [1,2]. In the case of eukaryotic DNA, mostly DNA methylation affecting cytosine residues (methylcytosine) has been extensively studied. Along with methylcytosine, other modified cytosine states have recently been identified, such as hydroxymethylcytosine, carboxylcytosine and formylcytosine [1]. Moreover, adenosine methylation has been found in the genome of various eukaryotes [3–7] and is not well understood. Further experimental and bioinformatics analysis of these modifications is required to understand their function. While DNA cytosine modifications are known to influence transcription, there is now also growing evidence for its role in the regulation of alternative splicing [8–10].

Alternative splicing of pre-mRNA from a single gene results in multiple isoforms that contribute greatly to the RNA and protein diversity in eukaryotes. Alternative splicing can include intron retention, alternative splice site selection and exon skipping. During splicing, the spliceosome, aided by splicing factors, recognizes and binds to specific sequences in the pre-mRNA. These sequences are at exon/intron boundaries and once the spliceosome binds to the sequence it excises the introns, which results in the mature mRNA [11]. The recruitment of the spliceosome is influenced by both *cis*-acting RNA elements and *trans*-acting splicing factors. *Cis*-acting RNA elements include enhancer or repressor sequences within the pre-mRNA that facilitate or inhibit the binding of the spliceosome.



**Figure 1.** (a) Example of epigenetic markers influencing alternative splicing. Knockdown of the oxidase Tet2 has been shown to reduce levels of cytosine hydroxymethylation (5hmC) (orange circles). For some genomic regions with a reduction in 5hmC, a corresponding increase in cytosine methylation (5mC) was observed (blue circles). 5hmC was shown to be prevalent in exons, particularly at the boundaries, and the loss of 5hmC resulted in differential inclusion of exons. (b) The overexpression of splicing factor SRSF1 is known to cause differential splicing of BIN1. Overexpression of SRSF1 results in an alternative isoform BIN1+12a. This isoform is unable to bind the tumour repressor cMYC and therefore promotes tumorigenesis.

*Trans*-acting splicing factors can bind to the spliceosome or to bound or open *cis*-elements to direct binding of the spliceosome to the pre-mRNA and consequently alternative splicing.

The role of epigenetic modifications in alternative splicing has been extensively reviewed in detail [12–16]. Here, we only give a brief overview of the role of epigenetic modifications in alternative splicing. In this review, we focus on bioinformatics methods that have been developed to understand splicing regulation. This includes the regulatory relationships that govern splicing and differences in isoform expression relating to disease. We highlight the potential to integrate epigenetic markers in these analyses to elucidate the interaction between epigenetics and splicing. We provide some, though not exhaustive, examples on each topic while also directing the reader to comprehensive reviews.

## 2. Epigenetics and alternative splicing

In the classic view of splicing, transcription and splicing are two distinct processes, with transcription occurring first followed by splicing. However, splicing can occur co-transcriptionally [17]: introns can be spliced from pre-mRNA as the transcription of DNA occurs [18]. As a result, epigenetic modifications, such as histone and DNA modifications, are able to regulate splicing and transcription [8,13].

Chromatin structure can control splicing through access to and the recruitment of splicing factors. In yeast, the histone acetyltransferase Gcn5 was shown to interact with two U2 snRNP-encoding genes facilitating its recruitment to splice sites, which results in co-transcriptional splicing [19]. The compactness of chromatin structure also influences splicing

through modifying the elongation rate of Polymerase II; a slower transcription rate has been shown to lead to differential splice site selection as weaker splice sites can be detected [20]. Weaker splice sites are *cis*-regulatory sequences at the exon–intron boundaries that are not as easily detected by the spliceosome. The subunit Brm of the chromatin remodeller, SWI/SNF, has been shown to regulate splicing. Brm causes an accumulation of RNA Polymerase II and pauses transcription. This seems to promote inclusion of exons with weaker splice sites, resulting in alternative splicing [21].

Histone modifications such as H3K36me<sub>3</sub>, which is significantly enriched in exons, have been implicated in the regulation of alternative splicing [22]. The H3K4me<sub>3</sub> binding protein CHD1 was shown to facilitate splicing through the recruitment of the spliceosome component U2 snRNP. Knockdown of CHD1 resulted in reduced efficiency of splicing, while the combinatorial knockdown of CHD1 and H3K4me<sub>3</sub> further reduced U2 snRNP recruitment [23].

Finally, there is growing evidence of DNA modifications regulating splicing [13]. Increased DNA cytosine methylation in exons may regulate splicing by aiding the splicing machinery in the detection of exon/intron boundaries [24]. DNA cytosine methylation was significantly different between exons and introns, even when taking into consideration nucleosome and CpG density [25]. Moreover, knockdown of Tet2 decreases the level of cytosine hydroxymethylation and results in differential exon usage [26] (figure 1a). This further suggests a role of DNA modifications in alternative splicing.

Although we have focused here on examples of epigenetic modifications regulating splicing, splicing factors themselves can influence chromatin structure and histone modifications. This suggests a complex feedback model of regulation [11].

This complexity in the splicing process gives rise to multiple ways in which splicing can be misregulated and it is this misregulation that can lead to disease.

### 3. Alternative splicing and disease

It is known that certain mutations in splice sites, *cis*-regulatory elements, splicing factors and the spliceosome machinery can lead to disease. These diseases include chronic lymphocytic leukaemia, myelodysplastic syndrome and spinomuscular atrophy [12]. Furthermore, splicing factors have been shown to act as oncoproteins and tumour suppressors [27]. For example, the splicing factor SRSF1 is overexpressed in many tumours and found to influence processes leading to tumorigenesis [28]. Specific to alternative splicing, SRSF1 overexpression produces an alternative isoform of the protein BIN1 that is then unable to bind and repress the activity of the proto-oncogene cMYC [29] (figure 1*b*).

From an epigenetic perspective, the analysis of transcriptional networks of late onset Alzheimer's disease found that differentially expressed genes also had differential DNA cytosine methylation levels. This coincided with differential expression at the exon level and reduced splicing [30]. Another example that highlights the interplay between epigenetic markers, splicing and disease, is the protein MeCP2. Mutations of MeCP2 are known to cause neurodevelopmental disorders. MeCP2 can bind methylated cytosines and is involved in splicing. Exons bound by MeCP2 are enriched with specific epigenetic markers; histone modifications H3K4me3 and H3K36me3 are enriched in transcriptionally inactive and active exons, respectively. These regulated exons are involved in mRNA splicing and synaptic function within the brain. The model suggests that epigenetic markers facilitate splicing and the loss of this regulation can influence disease [31]. Despite these examples, the role of epigenetic modifications in alternative splicing is still not well understood. Therefore, identifying and analysing epigenetic drivers of alternative splicing may provide new insights into disease progression and treatment.

To elucidate the drivers of alternative splicing and disease progression, bioinformatics techniques need to be developed to model splicing regulation. To understand the effect of epigenetic modifications on splicing, observing and measuring the effects at the RNA transcript level are required [27]. To measure alternative splicing on a genome-wide scale, splice-aware or exon microarrays have previously been used [32]. RNA-seq is now the method of choice, as it has greater dynamic range and the ability to detect novel and known splice junctions. The analysis of RNA-seq data for alternative splicing requires computational models that are specifically designed for the analysis of differential transcripts. We discuss some of the methods that can be used with high-throughput data to model the relationships between splicing regulation, epigenetics and disease.

### 4. RNA-seq analysis of differential alternative splicing

There are multiple computational analysis tools designed for the analysis of alternative splicing using RNA-seq data. The most recent software packages estimate both biological and


technical variation when testing for differential alternative splicing across conditions [33]. In differential alternative splicing, sequencing reads are first aligned to a reference genome. Given the aligned reads, one approach is to model directly observable exon counts (e.g. DEXSeq) and splice junction counts (e.g. JunctionSeq). The number of reads assigned to an exon or splice junction gives a measure of its expression. While DEXSeq models the observed exon counts, JunctionSeq uses both the observed exon counts and the counts at splice junctions. Therefore, JunctionSeq can identify differential splicing patterns of an exon even when the exon has consistent expression over all conditions. Alternatively, methods such as Cuffdiff 2, Tuxedo 2 and MISO assign reads to isoform transcripts and therefore model isoform as opposed to exon expression (figure 2). While this additional level of complexity in the modelling can be more error prone, Cuffdiff 2 has been shown to outperform differential exon usage analysis when the annotation is largely incomplete [34]. In selecting the correct tool for analysing splicing, it is important to consider whether replicates need to be modelled and the experimental conditions to be analysed. Two group (with replicates) and two sample (without replicates) comparisons can be done using tools such as DiffSplice and MATS, respectively. Experimental designs containing multiple groups or confounding factors would require a tool that can model design matrices such as JunctionSeq or Tuxedo 2. An overview of the available tools for differential alternative splicing analysis is shown in table 1. RNA-seq analysis can generate hypotheses of novel and known splice variants and it is advisable that any important alternative splicing results from RNA-seq analyses are verified. This can be done by either RT-PCR or sequencing capable of generating long full-length transcripts such as 454 Sequencing [44], SMRT sequencing [45] and Nanopore [35].

Results from differential alternative splicing analysis are usually interpreted via enrichment of biological function annotations provided by gene ontologies [36]. This enrichment analysis provides a good starting point to understanding results of differential alternative splicing. RNA-seq data give information on the observed changes in splicing at a transcript level. However, they provide little insight into the mechanisms that caused the observed changes in splicing. To understand interactions between elements, such as epigenetic marks, splicing factors and binding motifs, bioinformatics modelling is required. Bioinformatics models can integrate multiple data sources to reverse engineer and identify the effect of epigenetic landscapes on alternative splicing.

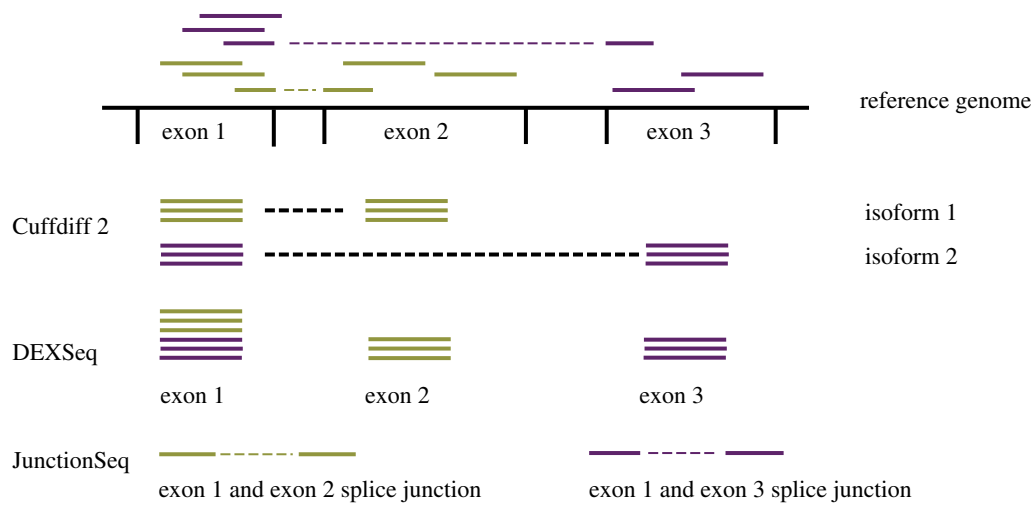
### 5. Bioinformatics modelling in epigenetics and splicing

In bioinformatics data analysis and modelling, there is a trade-off between the scale of the model and its complexity [37,38]. When only a few factors are being modelled, models can be built with detail. For tens of proteins, it is feasible to perturb them individually and in combination to resolve causal relationships between them [39]. By contrast, bioinformatics analysis of high-throughput data at an 'omics' scale is affected by the curse of dimensionality [40]. This refers to the comparatively large number of measured entities, such as genes, compared to the small number of replicates or experimental conditions. Because of the small sample size constraints, simplifying assumptions about the biological system are used to

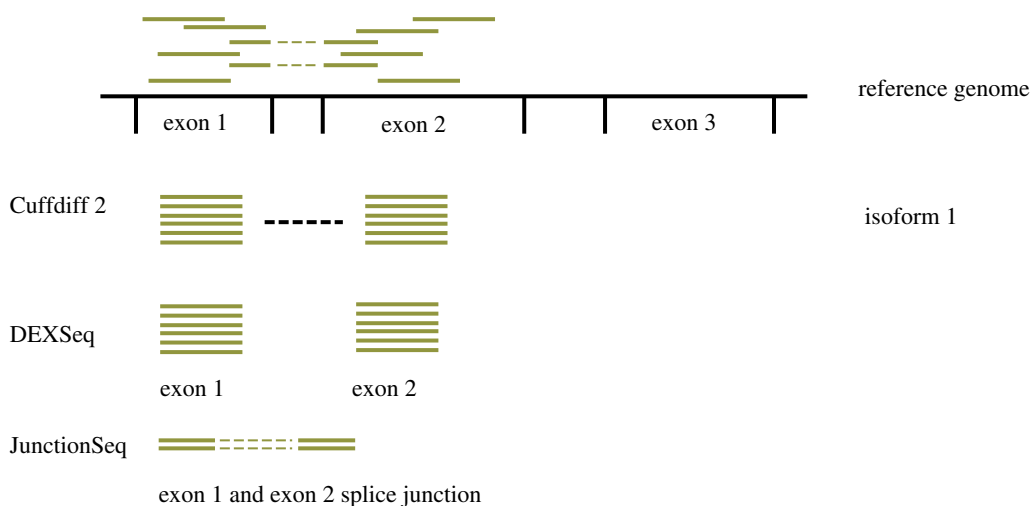
the gene has two isoforms: isoform 1 contains exon 1 and exon 2;  
isoform 2 comprises exon 1 and exon 3

reads from RNA-seq: 

**condition 1**, two isoforms are produced



**condition 2**, only one isoform is produced



**Figure 2.** Sequencing reads from RNA-seq (shown as solid purple or green lines) are aligned to a reference genome. Three methods for analysing differential alternative splicing are shown: Cuffdiff 2, DEXSeq and JunctionSeq. The differences in how sequencing reads are assigned to give either isoform counts (Cuffdiff 2), exon counts (DEXSeq) or splice junction counts (JunctionSeq) are illustrated. For Cuffdiff 2, the difficulty is in correctly assigning the RNA-seq reads at exon 1 to isoform 1 (green reads) and isoform 2 (purple reads). For each method, the respective observed counts are used to detect differences in alternative splicing across two different conditions. The difference in these counts is shown; in the first condition, two isoforms of a gene are expressed and in the second condition only one isoform is expressed. For DEXSeq, no change in the expression of exon 1 will be detected. By contrast, JunctionSeq can detect a change in splicing involving exon 1 as there is an observed change in the number of reads at the exon 1 and exon 2 splice junctions.

enable the analysis of 'omics' scale data. For example, the differential expression of genes between two conditions may be calculated independently for each gene, even though genes are not independent of each other [41,42]. Despite these simplifications, these analyses can be useful when little is known about the function or the genes targeted by the splicing factors. An alternative approach that has been used extensively is meta-analysis. Meta-analysis is the combination of samples from multiple experiments into one analysis to increase the sample size. In all cases, useful bioinformatics methods applied to high-throughput data will generate hypotheses at a smaller scale, for example the interaction between a few proteins, that can be experimentally validated [43,46].

We view the modelling of splicing regulation and epigenetics from two, separate but related networks: splicing regulatory networks [47,48] and functional isoform networks [49]. Interactions between the elements responsible for the splicing process are referred to as splicing regulation networks. This includes the *trans*-acting splicing factors and the *cis*-acting binding motifs (figure 3a). Epigenetic marks may be viewed in these networks as *cis*-acting motifs that influence the activity of splicing factors [50]. The second type of network is a functional isoform network. In these networks, interactions between different isoforms across genes are modelled. This is used to infer functional units according to the co-expression or shared regulation of isoforms (figure 3b). These functional

**Table 1.** Comparison of tools available for differential alternative splicing analysis. The tools calculate alternative splicing at different levels; isoform, exon or using splice junction counts. The tools differ in their ability to use biological replicates and the complexity of the experimental design. Tools that can only compare between two conditions are either two samples (where replicates are not used) or two groups (where replicates are used). Methods that use design matrices can incorporate more complex designs including multiple group comparison and confounding factors.

	biological replicates	model	experimental design	reference
JunctionSeq	yes	exon & junction	design matrix	[35]
Tuxedo 2	yes	isoform, exon & junction	design matrix	[36]
DEXSeq	yes	exon	design matrix	[37]
MATS	no	exon & junction	two sample	[38]
MISO	no	isoform	two sample	[39]
Cuffdiff 2	yes	isoform & exon	two groups	[40]
DSGseq	yes	exon	two groups	[41]
DiffSplice	yes	exon & junction	two groups	[42]
ARH-seq	yes	exon & junction	two sample	[43]

isoform networks are usually specific to either disease status, cell or tissue type [51]. Given the role of epigenetic marks in splicing and disease, epigenetic information can be used to aid the analysis or inference of these networks. Below we discuss examples of resources and models that can be used to perform the analysis of epigenetic and splicing data both at a small scale and using high-throughput data.

## 6. Splicing regulation networks

Splicing regulation networks (SRNs) are used to represent the interactions within and between *trans*-acting RNA-binding proteins (RBPs) or splicing factors and *cis*-regulatory targets or epigenetic marks. In network analysis, the splicing factors and their targets are represented as nodes, while edges between nodes represent that an interaction exists between them. To infer networks, computational methods select the set of nodes and interactions between them that best explain the observed data [52,53]. SRNs can be built for specific disease or tissue types to identify splicing factors that contribute to phenotype. Differential splicing networks can be inferred, reflecting changes in splicing networks under different conditions. Below we provide examples of SRNs that have been inferred and discuss some, though not exhaustive, models from regulatory and signalling network inference that, we believe, can be adapted to model splicing and epigenetics.

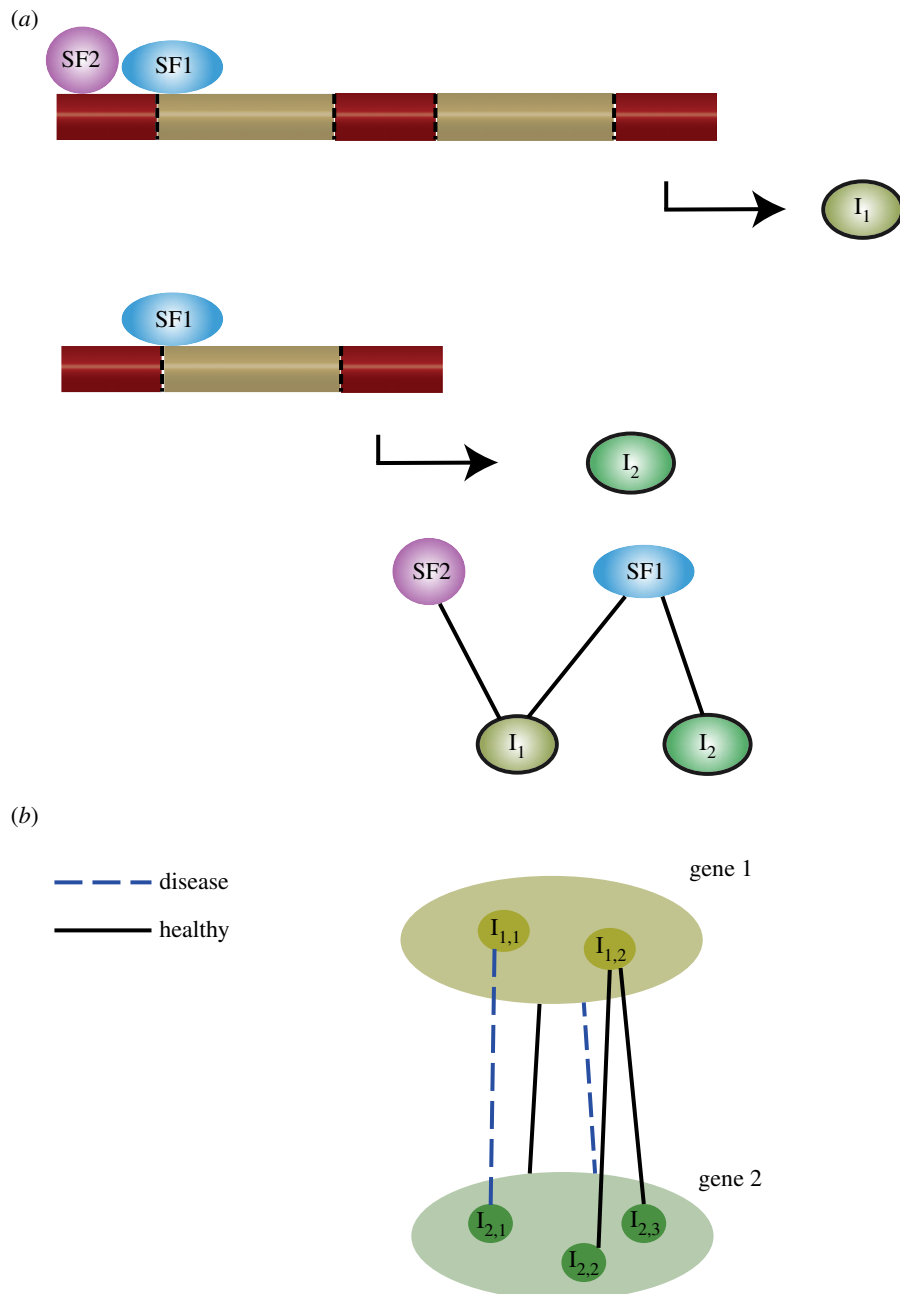
A splicing regulatory network for the tissue-specific splicing factor FOX-1 and its paralogue FOX-2 (Fox-1/2) was identified using genome-wide searching for their *cis*-binding motif. The motif for FOX-1/2 was found in exons and intronic flanking regions across the genome. False discovery rates were controlled by using the conservation of motif location in 28 additional vertebrate genomes to categorize FOX-binding targets. RT-PCR was used to validate exon splicing patterns [48]. The work highlighted the combinatorial nature of splicing motifs; differential splicing patterns could not be entirely explained by the FOX-1/2 binding motifs, suggesting a combinatorial regulation in conjunction with additional splicing factors. This opens interesting avenues for future research. Logic models can be used for modelling combinatorial relationship such as OR, AND and NOT between regulators of splicing [37]. For example, if the separate overexpression of two

splicing factors led to alternative splicing of the same gene, this would indicate an OR relationship of these two splicing factors and the gene, where either splicing factor can regulate the genes spliced.

To model splicing regulatory networks between *trans*-acting factors using logical relations, the most direct method would be to knock down individual or combinations of splicing factors and measure the expression level of each splicing factor [54,55]. Complete perturbation of the SRN allows for the topology of the network to be inferred. By contrast, any unobserved combinations of splicing factors can lead to uncertainties in resolving the relationships between them. Therefore, this type of analysis requires a known splicing factor or factors to have their expression altered to elucidate targets and function.

To overcome the lack of targeted perturbation data, methods, including Cell Net Optimizer (CellNOptR), have been developed that use prior knowledge networks to facilitate the inference of the logic-based signalling networks. In signalling networks, prior knowledge networks have been used to build initial models of signalling regulation. These initial models are refined using phosphoproteomic data that measure activity in the network following stimulation of the signalling pathway [56,57]. These models can be used to infer SRNs provided the proteins to be included in the model have either been perturbed or their abundance measured across perturbations. In practice, it has been shown that alternative splicing is regulated by signalling pathways, such as MAPK and AKT, through phosphorylation of splicing factors [17]. Therefore, activation of these pathways may be used to provide experimental data for inference. Alternatively, altering histone modifications may provide an indirect perturbation of splicing due to the recognition of histone marks by splicing factors that can be used for modelling. The difficulty with methods such as CellNOptR is having sufficient data to select between different models. Consequently, these methods are often used to resolve interactions between splicing factors but not to identify novel factors contributing to splicing regulation.

RNA-seq data have been used to infer splicing modules and the splicing factors that regulate them. RNA-seq data from multiple experiments in humans were combined to build co-expression networks at an exon level [58]. In this work, a splicing module was defined as a set of co-expressed

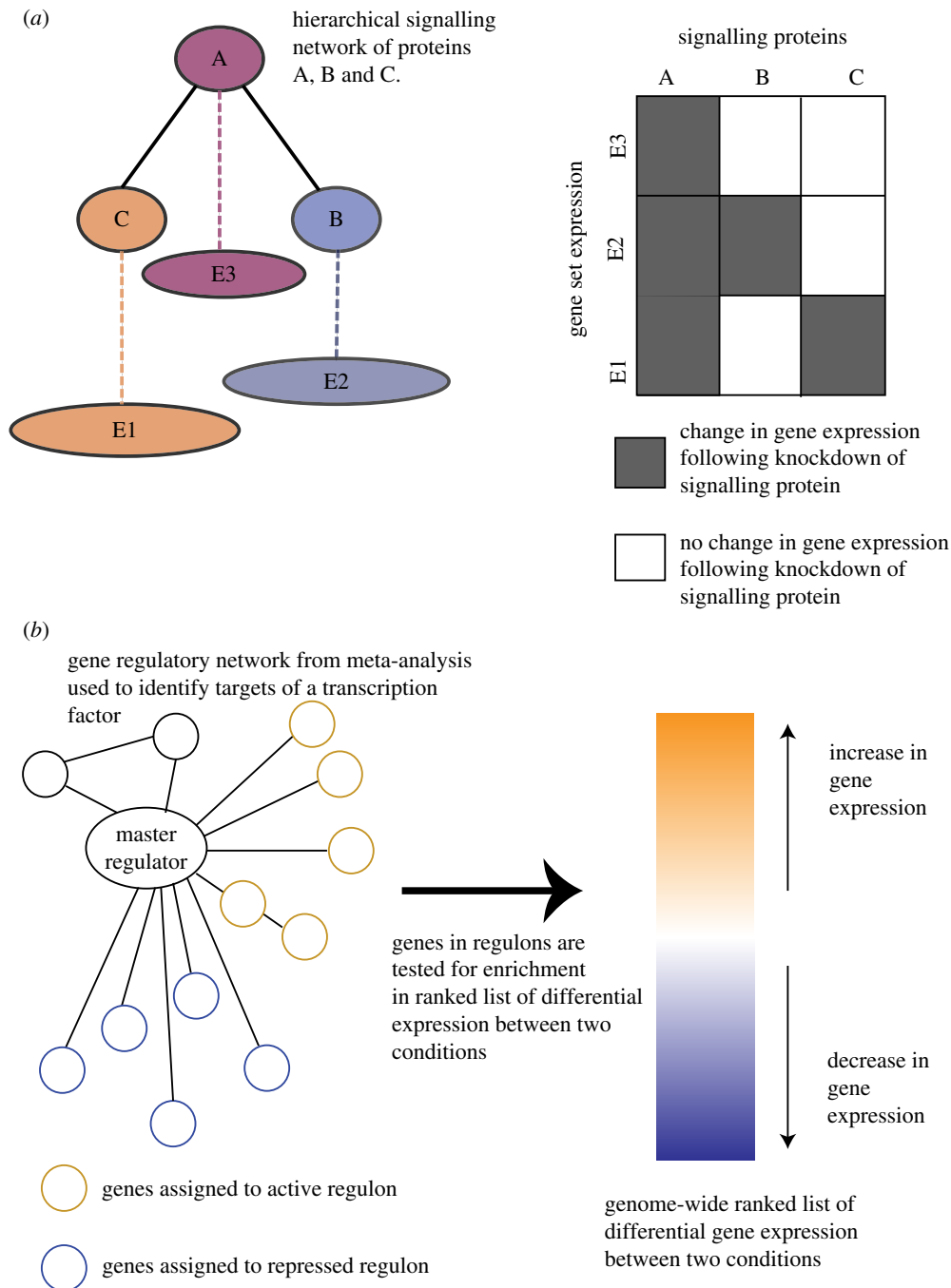


**Figure 3.** (a) Two different isoforms are shown. Isoform 1 ( $I_1$ ) is regulated by two splicing factors (SF1 and SF2). By contrast, isoform 2 ( $I_2$ ) is only regulated by SF1. The network view of this relationship is shown underneath. (b) This example shows a toy network between two genes and their isoforms. At the gene level, a connection between the two genes is shown under both healthy and disease states. However, at the isoform level, differences can be seen in the interaction between isoforms in the healthy and disease states.

exons that were assumed to be co-regulated. The analysis of known splicing factor motifs was used to assign splicing factor(s) that regulated each splicing module [58]. These results also showed a combinatorial pattern of splicing factors; multiple splicing modules were enriched for two splicing factor motifs *Tra2 $\alpha$*  and *SRp30c*. Bioinformatics analysis combined multiple experiments that together provided enough data to infer splicing modules. Moreover, ENCODE datasets of transcription factor binding and epigenetic markers were analysed for enrichment within the network, the results of which supported a model of co-transcriptional splicing [59]. While this method identified splicing factors that regulate a set of exons, it did not infer a corresponding network between splicing factors. To do this, methods that have been developed for signalling networks can be adapted to model interactions between splicing factors.

Nested effects models (NEMs) are an alternative method for inferring signalling networks that can be applied to the analysis of SRNs [60–62]. In contrast to CellNOptR, NEMs use indirect observations of gene expression to infer the signalling network. Using indirect gene expression has the advantage that the data are easier to analyse and a wider repository of publicly available transcriptomic data exists in comparison to proteomic data. The model assumes that the hierarchy of the signalling network can be inferred by using the subset of expression observed at the gene level, following knockdown of the signalling proteins (figure 4a). For example, an NEM framework was used to model five genes involved in the invasion pathway in human colon cancer, identifying novel genes involved in invasiveness that were validated experimentally [63].

As with the logic models, the knockdown of splicing factors can be used to infer the hierarchy of effects between a known



**Figure 4.** (a) The nested effects model uses the effect of knockdowns of signalling proteins (A, B, C), measured using genome-wide expression data, to infer a hierarchical signalling network. In this example, protein A is above B and C in the hierarchy. Knockdown of protein B results in differential gene expression of all genes in gene set E2. Similarly, knockdown of protein C causes differential gene expression of genes in gene set E1. Knockdown of protein A causes differential gene expression in gene sets E1, E2 and E3, as knockdown of protein A changes the expression of proteins B and C. The differences in gene expression following perturbation of the signalling proteins allow a hierarchical network to be inferred. (b) In master regulator analysis, first a gene regulatory network (GRN) is inferred based on the co-expression of genes over multiple conditions (meta-analysis). Using the GRN, a transcription factor is assigned two regulons, one containing active genes (brown circles) and the other repressed genes (blue circles). Second, a ranked list of changes in genome-wide expression between two conditions is calculated. For each transcription factor in the GRN, the two regulons are tested for enrichment in the ranked list of genome-wide differential expression. A transcription factor that has the most significantly differentially expressed target genes between the two conditions is hypothesized to be a master regulator of that phenotype.

set of splicing factors using NEMs. However, if the set of splicing factors to be modelled is unknown, an interesting approach would be to use indirect perturbations such as activation of signalling pathways. The effect on splicing of perturbing signalling networks and target splicing factors can be measured by differential exon or isoform expression using RNA-seq data. In the NEM, each node in the network will have a set of differentially spliced genes or differentially expressed exons associated with it. These sets of genes or

exons can be searched for common regulatory elements, such as splicing factor motifs and epigenetic marks. Common motifs can also be used to generate hypotheses of elements regulating each of the gene sets [64]. As well as regulatory motifs, epigenetic marks, such as histone modifications, can be integrated with RNA-seq data to elucidate the interactions between epigenetics and splicing.

In humans, integration of multiple datasets has been used to study the relationship between splicing, histone modifications

[65] and cytosine methylation [66]. Tissue-independent patterns of DNA cytosine methylation were identified that are related to alternative splicing [66]. By contrast, cell-specific changes in histone modifications were associated with alternative splicing through comparison of RNA-seq, exon arrays and ChIP-seq profiles. Exons displaying cell-specific splicing were found to have less of the H3K36me3 and H3K79me1 histone marks [65]. Thus, bioinformatics analysis of integrated datasets has been successful in identifying relationships between alternative splicing and epigenetic modifications. Master regulators of transcription have also been identified using multiple datasets or metaanalysis [67].

Master regulator analysis (MRA) is designed to discriminate between all regulatory factors to identify a master regulator of a phenotype. The term ‘master regulator’ has been used to describe both transcription factors and splicing factors. Master regulators are defined by their initiation of transcription or splicing networks [68]. In splicing, it is known that RBPs can self-regulate and regulate each other, resulting in single or multiple stable states of the cell [69]. A potential master regulator in splicing, RbFOX2, was identified using iCLIP. The RBP RbFOX2 regulates the autoregulation of several RBPs in mouse embryonic stem cells through alternative splicing-coupled nonsense-mediated decay of the target RBPs [70]. To identify master regulators of transcription, combined multiple experiments (meta-analysis) have been widely used [67].

Meta-analysis relies on the availability of multiple experiments perturbing in different ways a cell line or tissue. Using these combined datasets from multiple experiments, networks of gene regulation patterns were inferred that allowed master regulators of transcription to be identified in prostate cancer [71], breast cancer [72], Parkinson’s disease [73] and human B-cells [74]. The networks are inferred by calculating measures of pairwise co-regulation between gene expression; genes involved in the same biological process will show similar patterns of expression over different experimental conditions [74]. Within this gene regulatory network (GRN), transcription factor activity is determined by a two-stage process. First, a transcription factor is assigned a set of active and repressed regulated targets (regulons), based on the genes it is connected to in the GRN. Second, the activity of these targets is used to determine the influence of the transcription factor in the phenotype. Genome-wide differential expression between two conditions is calculated and each gene is ranked according to its differential expression. Each regulon is tested for the enrichment of genes at the top of this ranked list, thus relating the activity of a transcription factor to the contribution of a phenotype via the differential expression of its targets (figure 4b) [67]. MRA requires large datasets that perturb the system to generate a large dynamic range of effects in the targets, to successfully identify regulons.

From a splicing perspective, it is possible to perturb the splicing networks and generate high-throughput data that can be used in MRA. Under- and overexpression of RBPs have been used to perturb the splicing system to infer targets and function of alternative splicing [15]. Similarly, drug compounds have been used to reduce methylation to elucidate the impact of epigenetic markers on splicing regulation and disease [75]. This gives mechanisms that can be used to perturb the system that can impact splicing regulation. Previous analysis of GRNs has identified transcription factors rather than splicing factors as central to the network; this is most likely

because the underlying network is based on expression at the gene level. Therefore, exon regulatory networks could be used analogously to the GRN to identify master regulators of splicing. Regulatory networks at the exon level have been inferred using RNA-seq datasets, for example for schizophrenia and bipolar disease. This analysis inferred regulatory ‘hubs’ centred around several genes that had specific allelic differential expression between disease and control patients, including CpG islands and splicing enhancers. This analysis required the additional level of information contained within exon expression as the gene-level analysis identified random networks without any gene hubs [76]. Alternatively, splicing motifs can be used to identify splicing factor regulons within the GRN. The differential gene expression profile used for transcription factor MRA can easily be adapted to a differential alternative splicing signature for splicing factor MRA. Furthermore, since epigenetic markers can be assigned to either genes or exons in these networks, this will enable the identification of regulons with shared epigenetic and splicing regulation.

Successful meta-analysis relies on the co-ordination of experiments and outcomes from multiple sources. As a result, databases of experimental data and prior knowledge are required to maximize the data that can be analysed, such as ArrayExpress [77]. The Expression Atlas provides a map of gene expression over different tissues and cell types based on data in ArrayExpress [78,79]. Recently, the ISOexpresso database has been released that gives analogous information but at an isoform level [80]. This includes isoform coverage for different tissue types or disease status. MiasDB contains interactions between splicing factors, epigenetic marks and *cis*-regulatory elements as well as the interaction between *trans*-acting proteins. This resource is currently available only for human interactions and coverage is still incomplete [81]. In future work, there are plans to integrate MiasDB with established signalling pathway database KEGG [82]. It is hoped that this will be extended to allow for ontologies based on splicing regulation. The creation of isoform-specific ontologies will also aid the interpretation of isoform-specific networks discussed in the next section.

## 7. Functional isoform networks

As well as SRNs, networks have also been inferred for the interaction between isoform transcripts [83]. Epigenetic data can be integrated into isoform networks to identify the role of epigenetics in splicing. Understanding the function of changes to transcript expression is fundamental to understanding how different phenotypes materialize. The assumptions underlying these models are that isoforms showing similar patterns of expression share regulators and biological function [84]. The assumption of shared expression equating to shared function has been used extensively in modelling gene regulatory networks. Network views of isoforms provide a more accurate representation than lists of differentially spliced transcripts, where each transcript is treated independently.

SpliceNet built a disease-specific isoform network for lung adenocarcinoma using RNA-seq data. This method identified differences in isoform interactions between *Siva1*, *Cflar* and *Bcl-x* in normal and disease samples not identifiable at a gene level [85]. The gene-level network showed the same dependencies between these three genes in normal and cancer tissues. By contrast, the isoform network showed a cancer-specific



relationship between the *Bcl-xL* isoform and a *Siva1* isoform. Literature evidence supports the hypothesis that the binding of SIVA1 to BCL-XL inhibits its anti-apoptotic function [85]. The isoform network, therefore, generated hypotheses of interactions that differ between normal and cancer cells that can be experimentally tested where the gene level analysis could not.

More recently, integration of multiple genomic and proteomic data sources was used to infer a network of functional interactions at the isoform level in mouse [84]. These results found different isoform connections for the two isoforms of *Anxa6* that predicted different functional roles. The connections of one isoform predicted a role in vesicle and organelle fusion, while the second one was connected to isoforms enriched for genes related to the regulation of cell shape. These predictions were supported by experimental evidence in the literature. As the authors note, this is a first step towards building isoform-level networks and future work will focus on building phenotype-, cell- or tissue-specific networks. Moreover, this method integrates heterogeneous data sources including RNA-seq and isoform-docking data, therefore opening the possibility of including epigenetic information in the data used to infer the networks. In this way, interactions between isoforms and epigenetic marks can be identified that are predicted to share a common functional role.

From an epigenetic perspective, the functional isoform networks outlined above can be used as a prior knowledge network and mined using epigenetic data. This is analogous to the work of West *et al.* [86], where age-related gene methylation data were mapped onto a protein–protein interaction network. Modules of age-enriched methylation activity were identified that were associated with stem cell differentiation pathways. In future work, epigenetic markers that are differential with respect to a disease can analogously be mapped to the isoform-specific networks. This would enable identification of isoform-specific functions, not observed at the gene level, that are altered in disease due to differential epigenetic profiles. As well as mining existing signalling or regulatory networks for enriched epigenetic marks, epigenetic marks may also be used to infer the networks.

Epigenetic markers can be used for resolving targets of splicing factors assuming they recognize, or work in combination with, epigenetic marks. As an example, knockdown of PARP1 in *Drosophila* revealed PARP1 to be a ‘splicing hub’ [87]. Nuc-ChIP high-throughput assay was used to find nucleosome targets of PARP1 and alternative splicing was measured through RNA-seq. Experimental validation using ChIP-qPCR showed a reduction of occupancy of PARP1 following knockdown at PARP1-targeted exons, while overall nucleosome density remained constant. A concordant reduction of H3K4me3 was seen at these locations but not after histone H1 knockdown, indicating a specific association between PARP1 and H3K4me3. PAR-CLIP [88] showed that PARP1 facilitates the RNA binding to chromatin, while ChIP pulldown revealed that PARP1 also recruits the splicing factor SF3B1. The multiple mechanisms by which PARP1 influences splicing conferred its role as a splicing hub. The specific interaction between PARP1 and H3K4me3 supports the idea of using epigenetic marks to predict splicing factor binding and regulation. From a modelling perspective, epigenetic marks can be treated as *cis*-regulatory elements to facilitate their integration into isoform networks.

*Cis*-regulatory motif analysis has been used in regulatory networks to identify direct targets of a transcription factor and to distinguish them from downstream effects [55].

Similarly, in co-splicing networks based on the correlation of inclusion of exons, *cis*-regulatory motifs of splicing factors were used to identify splicing factors regulating sets of strongly co-expressed exons [64]. In addition to splicing motifs, integrating epigenetic marks, such as histone modifications and methylation data, can be used to understand the interaction between epigenetic factors and splicing. In the example above, knowledge of the H3K4me3 mark can be used to separate direct targets of PARP1 from indirect effects.

## 8. Current limitations and challenges

The bioinformatics analysis of epigenetics and splicing has so far relied heavily on the meta-analysis of multiple data sources [65,66]. This meta-analysis requires sufficient numbers of experiments from either the same cell type or disease condition that can be combined [85,86]. This restricts the experimental conditions under which epigenetics and splicing can be analysed [76,85]. Furthermore, the annotation of results requires prior knowledge [59,79] but databases on prior knowledge are not complete and exist only for specific organisms [80,81].

Meta-analysis has been successful in identifying both splicing factors and their targets and a combinatorial pattern of splicing factors [64]. However, resolving the interactions between splicing factors has so far been limited. Methods that have been used to infer signalling networks could be used to infer splicing networks [56,60,67]. To understand combinatorial patterns or networks between splicing and *trans*-acting factors, specific experiments will need to be designed [37]. This would involve single and combinatorial knockdowns to investigate the interplay between splicing factors. Similarly, as the tools for differential alternative splicing analysis can now model replicates and complicated experimental designs, targeted splicing factors can be investigated to unravel splicing regulation [89–91]. These experiments should also consider the combined analysis of RNA-seq and ChIP-seq datasets designed to investigate splicing and epigenetic markers.

Although splicing can occur co-transcriptionally, results from gene expression networks indicate that correlation or information-based networks are dominated by transcription as opposed to splicing [67,76]. Correlation of splicing factors and alternative splicing will not be identified by these networks unless splicing results in a concordant decrease in gene expression. Therefore, the analysis of exon co-expression networks [76,85] is expected to increase our understanding of co-transcriptional splicing and give a more complete view of regulatory networks.

Integrating epigenetic data into these exon co-expression networks is a current challenge for bioinformatics. For example, when integrating DNA methylation with gene expression it is common to use the cytosine methylation at promoter regions assuming a negative relationship between DNA cytosine methylation and gene expression [92,93]. However, the relationship between DNA cytosine methylation and splicing is less clear [85,94,95]. The analysis of epigenetic profiles has found a positive association between methylation density and the inclusion of intragenic exons [93]. Therefore, representing epigenetic marks for splicing factor targets that can be used in modelling is a current challenge in bioinformatics [66]. We anticipate future work will model DNA modifications for a gene as several quantities split across different genomic

features, such as promoter, exon and intron. The analysis of nonlinear patterns of methylation levels across a gene may provide insight into predictive signatures of phenotypes based on methylation levels.

## 9. Conclusion

Here, we have reviewed recent progress in models that can be used for understanding epigenetic and splicing regulation. Multiple data sources, including gene and protein expression, histone modifications and DNA methylation, can be integrated to give an overall picture of epigenetics, splicing and disease. However, there remain several challenges to fully understanding epigenetics and splicing regulation. This includes selection of appropriate datasets and methods to model epigenetics and splicing. There are multiple methods from modelling signalling or regulatory networks that can be adapted to study and model the effect of epigenetics on splicing. We anticipate that future work will establish how to use genome-wide DNA methylation profiles to create a 'splicing signature' that can relate DNA methylation levels or histone modifications to alternative splicing outcomes. While many analyses of single

splicing factors have been successful in identifying splicing targets and modes of regulation, we expect future models of splicing to allow for combinatorial mechanisms between splicing factors. Furthermore, we anticipate that recently discovered epigenetic marks, such as adenosine methylation, will be investigated for a potential role in splicing. Identifying the mechanisms by which isoform expression is regulated and its differential expression in diseases will aid understanding of isoform function and provide potential therapeutic targets.

**Data Accessibility.** This article has no additional data.

**Authors' contributions.** C.P. wrote the article, while M.J.K. assisted with editing.

**Competing interests.** We have no competing interests.

**Funding.** The work was sponsored by the Biotechnology and Biological Sciences Research Council grant no. BB/M022994/1 (C.P., M.J.K. and John Bertrand Gurdon). The Gurdon laboratory is funded by the grant 101050/Z/13/Z (John Bertrand Gurdon) from the Wellcome Trust. The Gurdon Institute is supported by core grants, namely by the Wellcome Trust Core grant no. (203144) and by the Cancer Research UK Grant (C6946/A24843).

**Acknowledgments.** The authors are grateful to John Gurdon, Charles Bradshaw and Jerome Jullien for reading the manuscript, useful comments and discussions on the topic.

## References

- Allis CD, Jenuwein T. 2016 The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17**, 487–500. (doi:10.1038/nrg.2016.59)
- Wu H, Zhang Y. 2014 Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* **156**, 45–68. (doi:10.1016/j.cell.2013.12.019)
- Greer EL *et al.* 2015 DNA Methylation on N6-Adenine in *C. elegans*. *Cell* **161**, 868–878. (doi:10.1016/j.cell.2015.04.005)
- Fu Y *et al.* 2015 N<sup>6</sup>-methyldeoxyadenosine marks active transcription start sites in *Chlamydomonas*. *Cell* **161**, 879–892. (doi:10.1016/j.cell.2015.04.010)
- Kozioł MJ, Bradshaw CR, Allen GE, Costa ASH, Frezza C, Gurdon JB. 2015 Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nat. Struct. Mol. Biol.* **23**, 24–30. (doi:10.1038/nsmb.3145)
- Wu TP *et al.* 2016 DNA methylation on N<sup>6</sup>-adenine in mammalian embryonic stem cells. *Nature* **532**, 329–333. (doi:10.1038/nature17640)
- Zhang G *et al.* 2015 N<sup>6</sup>-methyladenine DNA modification in *Drosophila*. *Cell* **161**, 893–906. (doi:10.1016/j.cell.2015.04.018)
- Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T. 2011 Epigenetics in alternative Pre-mRNA splicing. *Cell* **144**, 16–26. (doi:10.1016/j.cell.2010.11.056)
- Gutierrez-Arcelus M *et al.* 2015 Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. *PLoS Genet.* **11**, 1–25. (doi:10.1371/journal.pgen.1004958)
- Narayanan SP, Singh S, Shukla S. 2017 A saga of cancer epigenetics: linking epigenetics to alternative splicing. *Biochem. J.* **474**, 885–896. (doi:10.1042/BCJ20161047)
- de Almeida SF, Carmo-Fonseca M. 2012 Design principles of interconnections between chromatin and pre-mRNA splicing. *Trends Biochem. Sci.* **37**, 248–253. (doi:10.1016/j.tibs.2012.02.002)
- Cieply B, Carstens RP. 2015 Functional roles of alternative splicing factors in human disease. *Wiley Interdiscip. Rev. RNA* **6**, 311–326. (doi:10.1002/wrna.1276)
- Khan DH, Jahan S, Davie JR. 2012 Pre-mRNA splicing: role of epigenetics and implications in disease. *Adv. Biol. Regul.* **52**, 377–388. (doi:10.1016/j.jbior.2012.04.003)
- Scotti MM, Swanson MS. 2015 RNA mis-splicing in disease. *Nat. Rev. Genet.* **17**, 19–32. (doi:10.1038/nrg.2015.3)
- Baralle FE, Giudice J. 2017 Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* **18**, 437–451. (doi:10.1038/nrm.2017.27)
- House AE, Lynch KW. 2008 Regulation of alternative splicing: more than just the ABCs. *J. Biol. Chem.* **283**, 1217–1221. (doi:10.1074/jbc.R700031200)
- Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ. 2013 Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* **14**, 153–165. (doi:10.1038/nrm3525)
- Kornblihtt AR, de la Mata M, Fededa JP, Muñoz MJ, Nogués G. 2004 Multiple links between transcription and splicing. *RNA* **10**, 1489–1498. (doi:10.1261/rna.7100104)
- Gunderson FQ, Johnson TL. 2009 Acetylation by the transcriptional coactivator Gcn5 plays a novel role in co-transcriptional spliceosome assembly. *PLoS Genet.* **5**, e1000682. (doi:10.1371/journal.pgen.1000682)
- Allo M, Schor IE, Muñoz MJ, de la Mata M, Agirre E, Valcarcel J, Eyras E, Kornblihtt AR. 2010 Chromatin and alternative splicing. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 103–111. (doi:10.1101/sqb.2010.75.023)
- Batsché E, Yaniv M, Muchardt C. 2006 The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat. Struct. Mol. Biol.* **13**, 22–29. (doi:10.1038/nsmb1030)
- Spies N, Nielsen CB, Padgett RA, Burge CB. 2009 Biased chromatin signatures around polyadenylation sites and exons. *Mol. Cell* **36**, 245–254. (doi:10.1016/j.molcel.2009.10.008)
- Sims RJ, Millhouse S, Chen CF, Lewis BA, Erdjument-Bromage H, Tempst P, Manley JL, Reinberg D. 2007 Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Mol. Cell* **28**, 665–676. (doi:10.1016/j.molcel.2007.11.010)
- Ehrlich M, Ehrlich KC. 2014 DNA cytosine methylation and hydroxymethylation at the borders. *Epigenomics* **6**, 563–566. (doi:10.2217/epi.14.48)
- Gelfman S, Cohen N, Yearim A, Ast G. 2013 DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon–intron structure. *Genome Res.* **23**, 789–799. (doi:10.1101/gr.143503.112)
- Huang Y *et al.* 2014 Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA* **111**, 1361–1366. (doi:10.1073/pnas.1322921111)
- Dvinge H, Kim E, Abdel-Wahab O, Bradley RK. 2016 RNA splicing factors as oncoproteins and tumour suppressors. *Nat. Rev. Cancer* **16**, 413–430. (doi:10.1038/nrc.2016.51)
- Karni R, de Stanchina E, Lowe SW, Sinha R, Mu D, Krainer AR. 2007 The gene encoding the splicing

- factor SF2/ASF is a proto-oncogene. *Nat. Struct. Mol. Biol.* **14**, 185–193. (doi:10.1038/nsmb1209)
29. Anczuków O, Rosenberg AZ, Akerman M, Das S, Zhan L, Karni R, Muthuswamy SK, Krainer AR. 2012 The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nat. Struct. Mol. Biol.* **19**, 220–228. (doi:10.1038/nsmb.2207)
  30. Humphries CE, Kohli MA, Nathanson L, Whitehead P, Beecham G, Martin E, Mash DC, Pericak-Vance MA, Gilbert J. 2015 Integrated whole transcriptome and DNA methylation analysis identifies gene networks specific to late-onset Alzheimer's disease. *J. Alzheimer's Dis.* **44**, 977–987.
  31. Cheng T, Chen J, Wan H, Tang B, Tian W, Liao L, Qiu Z. 2017 Regulation of mRNA splicing by MeCP2 via epigenetic modifications in the brain. *Sci. Rep.* **7**, 42790. (doi:10.1038/srep42790)
  32. Zimmermann K, Jentsch M, Rasche A, Hummel M, Leser U. 2015 Algorithms for differential splicing detection using exon arrays: a comparative assessment. *BMC Genomics* **16**, 136. (doi:10.1186/s12864-015-1322-x)
  33. Hooper JE. 2014 A survey of software for genome-wide discovery of differential splicing in RNA-Seq data. *Hum. Genomics* **8**, 3. (doi:10.1186/1479-7364-8-3)
  34. Liu R, Loraine AE, Dickerson JA. 2014 Comparisons of computational methods for differential alternative splicing detection using RNA-seq in plant systems. *BMC Bioinformatics* **15**, 364. (doi:10.1186/s12859-014-0364-4)
  35. Deamer D, Akeson M, Branton D. 2016 Three decades of nanopore sequencing. *Nat. Biotechnol.* **34**, 518–524. (doi:10.1038/nbt.3423)
  36. Blake JA *et al.* 2015 Gene ontology consortium: going forward. *Nucleic Acids Res.* **43**, D1049–D1056. (doi:10.1093/nar/gku1179)
  37. Le Novère N. 2015 Quantitative and logic modelling of molecular and gene networks. *Nat. Rev. Genet.* **16**, 146–158. (doi:10.1038/nrg3885)
  38. Fisher J, Henzinger TA. 2007 Executable cell biology. *Nat. Biotechnol.* **25**, 1239–1249. (doi:10.1038/nbt1356)
  39. Markowetz F. 2010 How to understand the cell by breaking it: network analysis of gene perturbation screens. *PLoS Comput. Biol.* **6**, e1000655. (doi:10.1371/journal.pcbi.1000655)
  40. Bellman RE. 1957 *Dynamic programming*. Princeton, NJ: Princeton University Press.
  41. Robinson MD, McCarthy DJ, Smyth GK. 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140. (doi:10.1093/bioinformatics/btp616)
  42. Gao D, Kim J, Kim H, Phang TL, Selby H, Tan AC, Tong T. 2010 A survey of statistical software for analysing RNA-seq data. *Hum. Genomics* **5**, 56–60. (doi:10.1186/1479-7364-5-1-56)
  43. Kitano H. 2002 Systems biology: a brief overview. *Science* **295**, 1662–1664. (doi:10.1126/science.1069492)
  44. Margulies M *et al.* 2006 Corrigendum: genome sequencing in microfabricated high-density picolitre reactors. *Nature* **441**, 120. (doi:10.1038/nature04726)
  45. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. 2010 Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **7**, 461–465. (doi:10.1038/nmeth.1459)
  46. Butcher EC, Berg EL, Kunkel EJ. 2004 Systems biology in drug discovery. *Nat. Biotechnol.* **22**, 1253–1259. (doi:10.1038/nbt1017)
  47. Zhang C *et al.* 2010 Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. *Science* **329**, 439–443. (doi:10.1126/science.1191150)
  48. Zhang C, Zhang Z, Castle J, Sun S, Johnson J, Krainer AR, Zhang MQ. 2008 Defining the splicing regulatory network of tissue-specific splicing factors Fox-1/2. *Genes Dev.* **22**, 2550–2563. (doi:10.1101/gad.1703108)
  49. Li H-D, Menon R, Omenn GS, Guan Y. 2014 The emerging era of genomic data integration for analyzing splice isoform function. *Trends Genet.* **30**, 340–347. (doi:10.1016/j.tig.2014.05.005)
  50. Worsley-Hunt R, Bernard V, Wasserman WW. 2011 Identification of cis-regulatory sequence variations in individual genome sequences. *Genome Med.* **3**, 65. (doi:10.1186/gm281)
  51. Calarco JA, Zhen M, Blencowe BJ. 2011 Networking in a global world: establishing functional connections between neural splicing regulators and their target transcripts. *RNA* **17**, 775–791. (doi:10.1261/rna.2603911)
  52. Lee W-P, Tzou W-S. 2009 Computational methods for discovering gene networks from expression data. *Brief. Bioinform.* **10**, 408–423.
  53. Markowetz F, Spang R. 2007 Inferring cellular networks—a review. *BMC Bioinformatics* **8**, 55. (doi:10.1186/1471-2105-8-56-55)
  54. Karlebach G, Shamir R. 2008 Modelling and analysis of gene regulatory networks. *Nat. Rev. Mol. Cell Biol.* **9**, 770–780. (doi:10.1038/nrm2503)
  55. Davidson EH *et al.* 2002 A genomic regulatory network for development. *Science* **295**, 1669–1678. (doi:10.1126/science.1069883)
  56. Terfve C, Cokelaer T, Henriques D, MacNamara A, Goncalves E, Morris MK, van Iersel M, Lauffenburger DA, Saez-Rodriguez J. 2012 CellNOptR: a flexible toolkit to train protein signaling networks to data using multiple logic formalisms. *BMC Syst. Biol.* **6**, 133. (doi:10.1186/1752-0509-6-133)
  57. Saez-Rodriguez J, Alexopoulos LG, Epperlein J, Samaga R, Lauffenburger DA, Klamt S, Sorger PK. 2009 Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol. Syst. Biol.* **5**, 331. (doi:10.1038/msb.2009.87)
  58. Dai C, Li W, Liu J, Zhou X. 2012 Integrating many co-splicing networks to reconstruct splicing regulatory modules. *BMC Syst. Biol.* **6**, S17. (doi:10.1186/1752-0509-6-S1-517)
  59. The ENCODE Project Consortium. 2012 An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74. (doi:10.1038/nature11247)
  60. Markowetz F, Kostka D, Troyanskaya OG, Spang R. 2007 Nested effects models for high-dimensional phenotyping screens. *Bioinformatics* **23**, i305–i312. (doi:10.1093/bioinformatics/btm178)
  61. Anchang B, Sadeh MJ, Jacob J, Tresch A, Vlad MO, Oefner PJ, Spang R. 2009 Modeling the temporal interplay of molecular signaling and gene expression by using dynamic nested effects models. *Proc. Natl Acad. Sci. USA* **106**, 6447–6452. (doi:10.1073/pnas.0809822106)
  62. Pirkl M, Diekmann M, van der Wees M, Beerenwinkel N, Fröhlich H, Markowetz F. 2017 Inferring modulators of genetic interactions with epistatic nested effects models. *PLoS Comput. Biol.* **13**, e1005496. (doi:10.1371/journal.pcbi.1005496)
  63. Vaske CJ, House C, Luu T, Frank B, Yeang CH, Lee NH, Stuart JM. 2009 A factor graph nested effects model to identify networks from genetic perturbations. *PLoS Comput. Biol.* **5**, e1000274. (doi:10.1371/journal.pcbi.1000274)
  64. Li W, Dai C, Kang S, Zhou XJ. 2014 Integrative analysis of many RNA-seq datasets to study alternative splicing. *Methods* **67**, 313–324. (doi:10.1016/j.ymeth.2014.02.024)
  65. Shindo Y, Nozaki T, Saito R, Tomita M. 2013 Computational analysis of associations between alternative splicing and histone modifications. *FEBS Lett.* **587**, 516–521. (doi:10.1016/j.febslet.2013.01.032)
  66. Zheng Z, Wei X, Hildebrandt A, Schmidt B. 2016 A computational method for studying the relation between alternative splicing and DNA methylation. *Nucleic Acids Res.* **44**, e19. (doi:10.1093/nar/gkv906)
  67. Lefebvre C *et al.* 2010 A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol. Syst. Biol.* **6**, 377. (doi:10.1038/msb.2010.31)
  68. Kin Chan SS. 2013 What is a master regulator? *J. Stem Cell Res. Ther.* **3**, 10–13.
  69. Jangi M, Sharp PA. 2014 Building robust transcriptomes with master splicing factors. *Cell* **159**, 487–498. (doi:10.1016/j.cell.2014.09.054)
  70. Jangi M, Boutz PL, Paul P, Sharp PA. 2014 Rbfox2 controls autoregulation in RNA-binding protein networks. *Genes Dev.* **28**, 637–651. (doi:10.1101/gad.235770.113)
  71. Aytes A *et al.* 2014 Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell* **25**, 638–651. (doi:10.1016/j.ccr.2014.03.017)
  72. Tovar H, García-Herrera R, Espinal-Enríquez J, Hernández-Lemus E. 2015 Transcriptional master regulator analysis in breast cancer genetic networks. *Comput. Biol. Chem.* **59**, 67–77. (doi:10.1016/j.compbiolchem.2015.08.007)
  73. Brichta L *et al.* 2015 Identification of neurodegenerative factors using translome—regulatory network analysis. *Nat. Neurosci.* **18**, 1325–1333. (doi:10.1038/nn.4070)

74. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A. 2006 ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7**(Suppl 1), S7. (doi:10.1186/1471-2105-7-S1-S7)
75. Ding X-L, Yang X, Liang G, Wang K. 2016 Isoform switching and exon skipping induced by the DNA methylation inhibitor 5-Aza-2'-deoxycytidine. *Sci. Rep.* **6**, 24545. (doi:10.1038/srep24545)
76. Hong S, Chen X, Jin L, Xiong M. 2013 Canonical correlation analysis for RNA-seq co-expression networks. *Nucleic Acids Res.* **41**, 1–15. (doi:10.1093/nar/gkt145)
77. Rustici G *et al.* 2013 ArrayExpress update—trends in database growth and links to data analysis tools. *Nucleic Acids Res.* **41**, D987–D990. (doi:10.1093/nar/gks1174)
78. Petryszak R *et al.* 2014 Expression Atlas update—a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. *Nucleic Acids Res.* **42**, D926–D932. (doi:10.1093/nar/gkt1270)
79. Kapushesky M *et al.* 2012 Gene expression atlas update—a value-added database of microarray and sequencing-based functional genomics experiments. *Nucleic Acids Res.* **40**, D1077–D1081. (doi:10.1093/nar/gkr913)
80. Yang IS, Son H, Kim S, Kim S. 2016 ISOexpresso: a web-based platform for isoform-level expression analysis in human cancer. *BMC Genomics* **17**, 631. (doi:10.1186/s12864-016-2852-6)
81. Xing Y *et al.* 2016 MiasDB: a database of molecular interactions associated with alternative splicing of human pre-mRNAs. *PLoS ONE* **11**, 1–9. (doi:10.1371/journal.pone.0155443)
82. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016 KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462. (doi:10.1093/nar/gkv1070)
83. van Dam S, Vösa U, van der Graaf A, Franke L, de Magalhães JP. 2017 Gene co-expression analysis for functional classification and gene–disease predictions. *Brief. Bioinform.* 1–18. (doi:10.1093/bib/bbw139)
84. Li H-D, Menon R, Eksi R, Guerler A, Zhang Y, Omenn GS, Guan Y. 2016 A network of splice isoforms for the mouse. *Sci. Rep.* **6**, 24507. (doi:10.1038/srep24507)
85. Yalamanchili HK, Li Z, Wang P, Wong MP, Yao J, Wang J. 2014 SpliceNet: recovering splicing isoform-specific differential gene networks from RNA-Seq data of normal and diseased samples. *Nucleic Acids Res.* **42**, e121. (doi:10.1093/nar/gku577)
86. West J, Beck S, Wang X, Teschendorff AE. 2013 An integrative network algorithm identifies age-associated differential methylation interactome hotspots targeting stem-cell differentiation pathways. *Sci. Rep.* **3**, 1630. (doi:10.1038/srep01630)
87. Matveeva E *et al.* 2016 Involvement of PARP1 in the regulation of alternative splicing. *Cell Discov.* **2**, 15046. (doi:10.1038/celldisc.2015.46)
88. Hafner M *et al.* 2010 Transcriptome-wide identification of RNA-binding protein and MicroRNA target sites by PAR-CLIP. *Cell* **141**, 129–141. (doi:10.1016/j.cell.2010.03.009)
89. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016 Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667. (doi:10.1038/nprot.2016.095)
90. Hartley SW, Mullikin JC. 2016 Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq. *Nucleic Acids Res.* **44**, e127. (doi:10.1093/nar/gkw501)
91. Anders S, Reyes A, Huber W. 2012 Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22**, 2008–2017. (doi:10.1101/gr.133744.111)
92. Iorio F *et al.* 2016 A landscape of pharmacogenomic interactions in cancer. *Cell* **166**, 740–754. (doi:10.1016/j.cell.2016.06.017)
93. Li S, Zhang J, Huang S, He X. 2017 Genome-wide analysis reveals that exon methylation facilitates its selective usage in the human transcriptome. *Brief. Bioinform.* 1–11. (doi:10.1093/bib/bbx019)
94. Ma X, Liu Z, Zhang Z, Huang X, Tang W. 2017 Multiple network algorithm for epigenetic modules via the integration of genome-wide DNA methylation and gene expression data. *BMC Bioinformatics* **18**, 72. (doi:10.1186/s12859-017-1490-6)
95. Wagner JR, Busche S, Ge B, Kwan T, Pastinen T, Blanchette M. 2014 The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. *Genome Biol.* **15**, R37. (doi:10.1186/gb-2014-15-2-r37)