Genome Biology

OPINION

Open Access

CrossMark

Splicing heterogeneity: separating signal from noise

Yihan Wan and Daniel R. Larson[*]

Abstract

Single-cell analyses have revealed a tremendous variety among cells in the abundance and chemical composition of RNA. Much of this heterogeneity is due to alternative splicing by the spliceosome. Little is known about how many of the resulting isoforms are biologically functional or just provide noise with little to no impact. The dynamic nature of the spliceosome provides numerous opportunities for regulation but is also the source of stochastic fluctuations. We discuss possible origins of splicing stochasticity, the experimental approaches for studying heterogeneity in isoforms, and the potential biological significance of noisy splicing in development and disease.

Introduction

In recent years, there has been substantial progress in the development of methodologies to interrogate gene expression in single cells. Single-cell imaging has historically been the workhorse technology for such studies, but applications such as single-cell sequencing have rapidly advanced, with recent publications drawing conclusions from tens of thousands of individual cells [1–4]. The picture that emerges from these studies is that gene expression varies from cell to cell. These differences can be both genetic and non-genetic, and they can be stable or dynamic. Differences can arise from programmed specialization during development or through random processes that occur in the cell. Even at the mRNA level, abundance, sequence, and chemical modifications can vary among transcripts that are produced from the same sequence of DNA. Making sense of this variation has become an immense experimental and theoretical challenge.

The process of RNA synthesis leads to variation in mRNA abundance, which has been studied extensively [5]. However, RNA processing, specifically pre-mRNA splicing, has the potential to be an equally important source of variability in gene expression. Since the first discovery of splicing 40 years ago [6–8], accumulating knowledge about the spliceosome's assembly and enzymatic mechanism, about the process of splice site selection, and on the coupling with transcription depicts a complex, multi-step, dynamic model involving a massive molecular machine. Each of these steps in splicing is subject to regulation, leading to the amazing diversity of alternatively spliced transcripts in virtually every organism in which RNA splicing is present. Each of these steps is, however, also subject to random fluctuations. Like all reactions that occur at the molecular level and rely on small numbers of molecules, stochastic (i.e., random) effects are the rule rather than the exception. This phenomenon was evident in the earliest observations of alternative splicing using chromatin spreads of the *Drosophila* chorion gene. In the same transcription unit, two alternative splicing isoforms were observed at the single-molecule level [9]. Since then, the proportion of transcripts that show alternative splicing has asymptotically approached 100%. From 'a list of genes' in the 1980s [10], to '74% of all genes' inferred from expressed sequence tag (EST)–genomic alignments and microarrays [11, 12], to '98–100% of multi-exonic genes' in the next-generation sequencing era [13–16]. Single-cell sequencing has now revealed that splicing variability exists among tissues and between individuals [17–20].

Which transcripts are functional? How do we detect meaningful changes not only in alternative splicing but also in RNA editing or alternative poly-adenylation? And what experimental and conceptual advances will be needed for the next stage of research? In this special issue, new techniques and datasets are presented that are at the forefront of RNA biology. Here, we focus on the current understanding of variability in RNA processing, mostly on splicing. We hope to frame the following

* Correspondence: dan.larson@nih.gov
Laboratory of Receptor Biology and Gene Expression, Center for Cancer Research, National Cancer Institute, Bethesda, MD 20892, USA

questions. 1) Where does splicing stochasticity come from? 2) How do we measure splicing variability? 3) What is the biological significance of splicing heterogeneity?

Noise in splicing: where does it come from?

To understand the source of stochastic variability, a comparison with transcription is illuminating. One major source of variability in mRNA abundance is the time-varying activity of RNA synthesis called transcriptional 'bursting'. Periods of active RNA synthesis are punctuated by long inactive periods [21–23]. The properties of the bursts are determined by *cis*-acting elements such as enhancers [24–27] and promoters [28–30], and by *trans*-acting activators [31, 32] and chromatin remodelers [33–35]. In particular, the initiation of RNA synthesis is the major source of variability, with downstream processes such as elongation, cleavage, release, and termination contributing little. Notably, enhanceosomes and pre-initiation complexes assemble and dissemble within a timescale of seconds [36, 37], and a 'successful' event results in the production of a transcript with low efficiency (from about 10% of complexes formed) [38–40]. Similarly, splicing is also a dynamic process that relies on both *cis*-acting elements and *trans*-acting modulators [41]. The assembly and disassembly of the spliceosome E complex occurs at a timescale of seconds to minutes [42]. The spliceosome is also a single turnover enzyme that disassembles after the completion of each splicing event (Fig. 1). Thus, the spliceosome would need to assemble and disassemble dozens of times (or more) during the production of any one transcript. The assembly of a spliceosome is determined by information residing in the consensus branch point and 5′, 3′ splice sites, but it can be affected by multiple levels of regulation, such as activities of silencer or enhancer sequences, the binding of SR proteins or heterogeneous nuclear ribonucleoproteins (hnRNPs), transcriptional kinetics, nucleosome positioning, and DNA template or chromatin modifications [15, 43]. When attempting to understand splicing noise, we can begin by looking at the composition and kinetics of the splicing machinery (Fig. 1).

For each nascent RNA molecule generated from transcription, the spliceosome needs to first recognize the correct splice sites, then assemble to complete intron removal and exon ligation, and then disassemble. Intron and exon definition is the key step in the initiation of spliceosome assembly. The 5′ splice site consensus AG|GURAGU is present at the exon–intron junction. The 15-nucleotide 3′ splice site $Y_{10}NCAG|G$ is present at the intron–exon junction. At a variable distance upstream of the 3′ splice site (10–50 nucleotides (nt) for human transcripts) is the branch point consensus YNYURAY [44–47]. The dinucleotide pair GU-AG is present in over 98% of all intron sequences that are removed by the spliceosome, but variations are found in neighboring bases [48, 49] (Fig. 1c).

Randomness is generated by several aspects of this splice-site-recognition step. First, the sequence information from the nascent RNA transcripts is ambiguous and highly degenerate, especially in mammals. The intron or exon definition step requires spliceosomes to read the information from more than ~ 30 bases accurately [50]. This recognition mostly relies on the base-pairing between U1 and U2 small nuclear RNAs (snRNAs) and the nascent RNA, but RNA modifications and bulged nucleotides make this base-pairing highly flexible [49, 51]. Sequence alone is not sufficient allow the accurate identification of splicing boundaries, even for short introns (≤ 134 bp) in human transcripts [52]. Moreover, many sequences in the mammalian genome match the consensus but are not recognized as real splice sites and the mechanism behind this discrimination is poorly understood. Second, mutations and single nucleotide variants (SNVs) in the template sequence generate moving targets for the spliceosome. Millions of genetic variants in the human genome have been uncovered through the 1000 Genomes project [53]. Multiple methods, such as machine learning [54], splicing quantitative trait loci (QTL) [55], and integrative genome-wide association studies (iGWAS) [56] have revealed that SNVs are associated with alternative splicing. These SNVs could change the splice sites directly or could alter a splicing regulatory sequence. Furthermore, the long introns in human transcripts also provide ample mutational opportunities for the creation of new or weak splice sites and for the generation of new exons (exonization) [20, 57]. Third, this 'reading and recognition' process is coordinated by splicing enhancer and silencer sequences through recruitment of SR proteins and hnRNPs [58]. Binding motifs for SR proteins and hnRNPs can be found in the majority of exons and introns [59, 60]. The role of RNA-binding proteins (RBP) can be synergic or competitive. The output of a splicing event will be affected by the motif sequence of the pre-mRNA and the array of RBP concentrations in the cell.

The complexity in the template pre-mRNA brings a primary source of stochasticity, even before considering the assembly of the spliceosome itself. The spliceosome consists of hundreds of proteins and multiple snRNAs. Initially, splicing 'commitment' was thought to occur once the intron–exon boundary has been defined [61]. Recent studies have revealed, however, that the spliceosome is a highly flexible and reversible enzyme. Spliceosome assembly can be initiated by either a U1- or a U2-first pathway [62]. After assembly initiation, the spliceosome can switch between different catalytic conformations that favor forward or reverse progress [41]. The

Fig. 1 (See legend on next page.)

(See figure on previous page.)
Fig. 1 Stochastic events in splicing. The spliceosome is a single-turnover enzyme that assembles and disassembles for each splicing event. Splicing consists of a complex sequence of steps, and each step represents several biochemical reactions. These reactions involve binding and dissociation events, which include random variables at the molecular level. **a** Schematic representation of the steps associated with mRNA production: transition of the promoter between a repressed and an active state, transcription, co-transcriptional or post-transcriptional splicing to create heterogenous isoforms and mRNA degradation. **b** Kinetic scheme for co-transcriptional spliceosome assembly. The formation of the catalytically competent spliceosome starts with splice site recognition, which is a highly dynamic process. Although the in vivo measurements of snRNP kinetics are still missing, in vitro experiments provide evidence for the reversible binding of almost all of the major subcomplexes to the nascent RNA (e.g., the pairing between U1 and 5'ss, U2 and branchpoint, the binding of tri-snRNP and NTC are in a kinetic range of $k = 0.13–0.35$ min^{-1}). The binding dynamics between the U2AF complex and poly-pyrimidine tract and 3'ss are still poorly understood. The binding of heterogeneous nuclear ribonucleoproteins (hnRNPs) and SR proteins also regulates splicing dynamics. Their kinetics need to be further explored. **c** Variability of splice sites in the human genome. (i) Consensus motifs of the U2-type 3' splice sites with AG at the border. (ii) Non-canonical motifs of U2-type 3' splice sites with dinucleotides other than AG at the border. (iii) Consensus motifs of the U1-type 5' splice sites with GT at the border. (iv) Non-canonical motifs of U1-type 5' splice sites with dinucleotides other than GT at the border. *NTC* NineTeen complex, *Pol II* RNA polymerase II, *snRNP* small nuclear ribonucleoprotein

splicing catalytic process is iso-energetic and driven by numerous ATPases, resulting in two transesterification processes that are both reversible in the proper ionic environment [63, 64]. Recent single-molecule research on spliceosome assembly has revealed that almost all of the steps in splicing are reversible [65, 66]. In the context of a highly flexible and reversible spliceosome assembly process, the alternative splicing decision may be the result of kinetic competition between different spliceosome assembly pathways.

Years of work on transcription have solidified the view that heterogeneity is a dynamic phenomenon: a gene may appear to be 'off', only to be expressed again minutes or hours later. Likewise, understanding splicing stochasticity requires an understanding of splicing dynamics. Splicing can be viewed as a process that is affected by multiple kinetic variables: 1) transcription kinetics affecting the generation of nascent RNAs; 2) the diffusion kinetics and assembly dynamics of the macromolecules involved in recognizing splice sites; 3) the spliceosome catalytic dynamics. Determining these kinetic parameters in vivo and how they work in concert are important for understanding splicing stochasticity. To address this question, a simple starting point is the time-lapse measurement of the splicing output—the generated mRNA isoforms. Population-level measurements that are based on time-resolved nascent-RNA sequencing and quantitative real-time PCR (RT-PCR) elucidate average splicing times ranging from 5 to 14 min in mammalian cells and of less than 5 min in yeast [67–69]. Nevertheless, the average measurement of a cell population may not reflect the behaviors of individual cells. Live-cell fluorescence microscopy based on a set of reporter genes labeled by MS2 and/or PP7 stem loops (Fig. 2a) probes the splicing kinetics at the single-cell level and reveals variable timescales (for example, from 20 s to many minutes) [70–72]. Single-molecule intron tracking (SMIT) combined with long-read sequencing showed that splicing can take place before the RNA polymerase transcribes even a
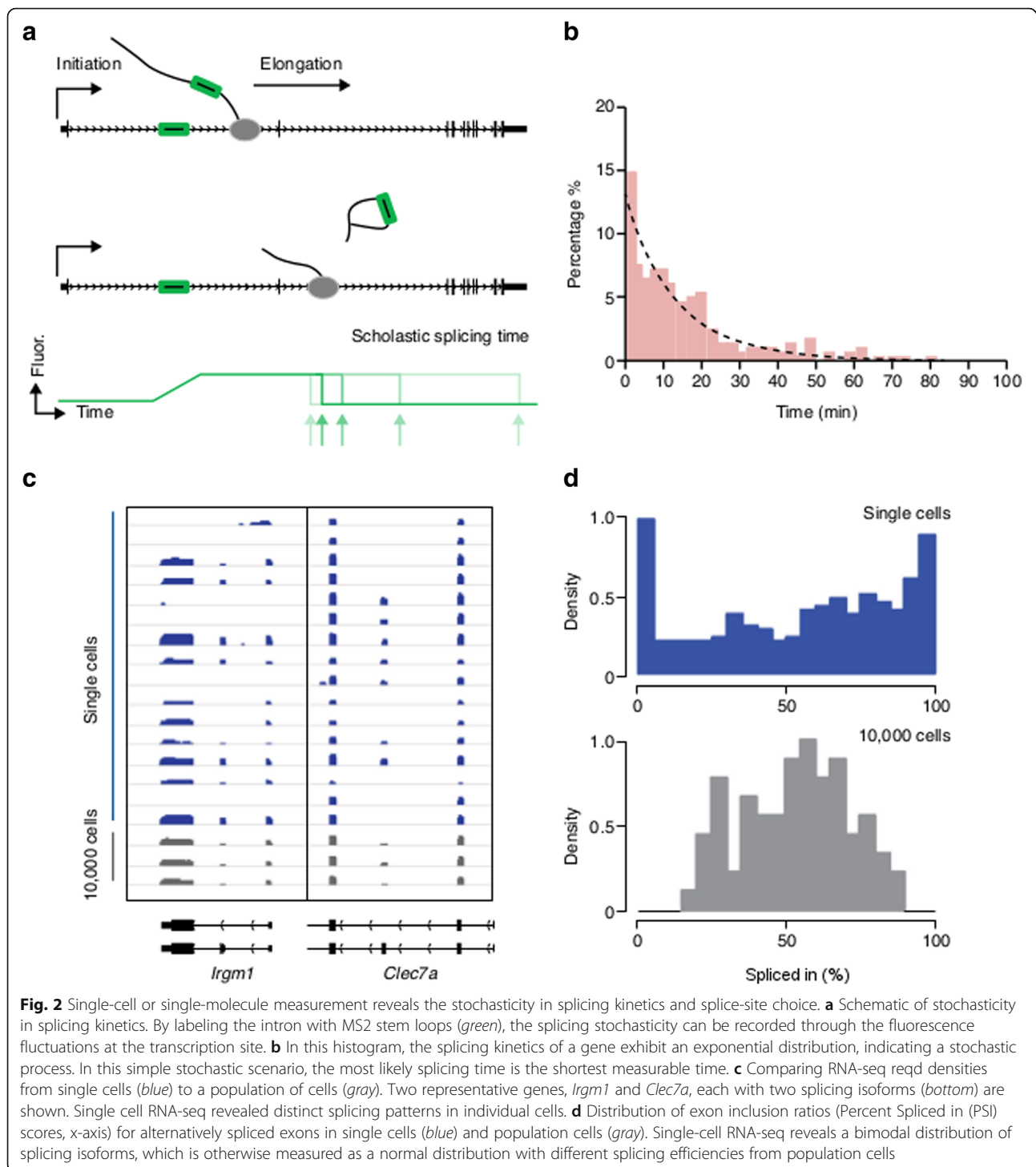
few dozen nucleotides downstream of the 3′ splice site (i.e., a few seconds given the polymerase transcribing rate) in yeast [73].

One explanation for the lack of consensus across the various methods used might be the difficulties in detecting the whole dynamic range of splicing. For example, if splicing time is exponentially distributed over a broad range (Fig. 2a, b), the measured time will depend on the time resolution of the method. Imaging or pulse-chase methods might overestimate the duration of very short events or might underestimate the duration of extremely long events. Likewise, for steady-state biochemical methods, the inferred dynamic parameters rely on the assumption that all intermediates are identified and analyzed, whether they are on chromatin or in the nucleoplasm.

Above all, the stochasticity of splicing could result in variability in both splice site selection and splicing kinetics. How do splicing kinetics associate with splice site selection? Does alternative splicing exhibit different kinetics? Evidence is emerging that, at least for certain genes, alternative splicing occurs mostly post-transcriptionally, whereas constitutive exons are spliced co-transcriptionally [74]. In addition, changing the nucleotides next to GU at the 5′ splice site can alter both the kinetics of spliceosome remodeling and splicing efficiency [75]. How the spliceosome makes a choice amongst splice sites during the kinetic competition between splicing and transcription [70, 71, 74, 76] is still an unanswered question that requires further investigation.

Can we measure the extent of stochastic RNA processing experimentally?
The initial concept of 'splicing noise' comes from the analysis of EST sequences and microarray-based mRNA abundance measurements [77]. These data suggested a positive correlation between the number of alternative isoforms and the number of splicing reactions (i.e., the number of introns per gene and the level of gene expression). A more precise evaluation was provided by the de novo identification of splice junctions based on

Fig. 2 Single-cell or single-molecule measurement reveals the stochasticity in splicing kinetics and splice-site choice. **a** Schematic of stochasticity in splicing kinetics. By labeling the intron with MS2 stem loops (*green*), the splicing stochasticity can be recorded through the fluorescence fluctuations at the transcription site. **b** In this histogram, the splicing kinetics of a gene exhibit an exponential distribution, indicating a stochastic process. In this simple stochastic scenario, the most likely splicing time is the shortest measurable time. **c** Comparing RNA-seq reqd densities from single cells (*blue*) to a population of cells (*gray*). Two representative genes, *Irgm1* and *Clec7a*, each with two splicing isoforms (*bottom*) are shown. Single cell RNA-seq revealed distinct splicing patterns in individual cells. **d** Distribution of exon inclusion ratios (Percent Spliced in (PSI) scores, x-axis) for alternatively spliced exons in single cells (*blue*) and population cells (*gray*). Single-cell RNA-seq reveals a bimodal distribution of splicing isoforms, which is otherwise measured as a normal distribution with different splicing efficiencies from population cells

RNA-seq data. Such studies revealed the existence of a large class of low-abundance isoforms [78]. Most of these isoforms contain the GU-AG dinucleotides, which indicate that they are generated from a random splice site choice. When thousands of independent RNA-seq datasets were combined, a significant number of previously unannotated splice junctions became evident across different tissues and cell types [79]. Although the current focus when analyzing these data is still on major alternative splicing events, a more comprehensive analysis across all splicing junctions would be beneficial for elucidating the distribution of isoform frequencies. Interestingly, a simple two-parameter Weibull distribution can be used to explain the statistical distribution of the

isoforms of all transcribed genes, indicating a possible general model of stochastic splicing [80].

Ideally, measurement of the stochasticity in splicing requires capturing each individual event in a population. Single-cell RNA-seq [3, 81–87] provides a promising avenue, but there are two major challenges: the first comes from the single-molecule capture efficiency. Using a spike-in assisted evaluation, Wold and colleagues [88] were able to provide an estimate of single-molecule capture efficiency of around 0.1, meaning that rare events are not represented in the single-cell sequencing library. The second challenge is to distinguish the biological stochasticity from technical noise, which is an enduring issue in single-cell analysis. Careful evaluation of the technical noise with quantitative statistical methods is necessary. Two recent studies carried out splicing analysis at the single-cell level [89, 90]. One unexpected discovery is that about 20% of the genes exhibit a bimodal distribution of certain splicing isoforms (Fig. 2c, d). These bimodal genes are related to differentiation and cell-type determination. After excluding technical artifacts caused by a low capture rate, there are two possible explanations for the bimodal distribution. First, the distribution may be due to extrinsic noise. For example, heterogeneity in the concentration of splicing regulators in different cells might result in the same pre-mRNA being processed differently. Second, the bi-modality might be caused by intrinsic noise. For example, in transcription, slow promoter kinetics will result in a bimodal distribution of gene expression [91]. Similarly, a slow transition parameter in isoform processing could also generate a bimodal distribution of isoforms in a cell population.

Single molecule long-read sequencing (Pacbio RNA-seq, iso-seq) [92, 93] is another promising technique for surveying isoform diversity. It can provide confident high-quality reads for transcripts over 20 kb, and over 10% of novel splice junctions have been identified through this strategy. The drawbacks are low throughput (i.e., limited reads per SMRT cell) and the potential for relatively high error rates in long reads.

Single-cell sequencing is comprehensive but suffers from low sensitivity and the potential for the introduction of error during library preparation and analysis. Single-molecule imaging is a complementary method. Single-molecule fluorescent in situ hybridization (smFISH) [94, 95] is a powerful way to quantify the absolute abundance of endogenous RNA transcripts in individual cells. Alternative splicing can be visualized by detecting the unique sequences of the different isoforms. The major advantage of this method compared to single-cell RNA-seq is that it provides both spatial information and sequence-specific information. For example, by probing the introns undergoing alternative splicing in the genes

Sxl and nPTB, Vargas et al. [74] showed that alternatively processed introns have delayed kinetics and are more frequently detected in the nucleoplasm. Waks et al. [96] probed the alternative spliced exons in genes CAPRIN1 and MKNK2, and examined the cell-to-cell variability by measuring the fraction of isoform abundance. Notably, they found that the distribution of isoform ratio could be explained by a theoretical stochastic model [96]. Nevertheless, standard smFISH requires the targeting of a single transcript with probes of approximately 48 oligonucleotides, each spanning about 17–22 nt and labeled at their 3′ end with one fluorophore. For the large majority of alternatively spliced isoforms, which only have slight differences in their mRNA sequences, a more sensitive approach such as the recently developed inosine fluorescence in situ hybridization (inoFISH) [97] is necessary.

Both smFISH and inoFISH require killing cells, and neither addresses the dynamic nature of splicing. To explore the stochasticity in splicing, it is necessary to record splicing kinetics in living cells. Taking advantage of the bacteriophage MS2 stem-loop and fluorescence-labeled coat proteins, researchers now can record RNA dynamics at the single-molecule level. Initially, fluorescence recovery after photobleaching (FRAP) together with MS2 stem-loop-labeled genes were used to monitor splicing and transcription kinetics [98]. The improvement in the imaging and analysis of RNAs at the single-molecule level enabled the direct observation of nascent RNAs at the gene locus. The fluctuation of intron and exon signals was recorded, and transcription and splicing kinetics were extracted through the cross-correlation function [70, 72]. With the advance of genome editing [99, 100], it is now possible to label single molecules of RNA produced from endogenous loci, which will allow tracing of the nascent RNA synthesized under physiological conditions. The information provided by live imaging of splicing of endogenous genes will extend our understanding of the stochasticity in splicing kinetics, including the impact of signaling networks and the chromatin environment.

Tremendous progress in the single-cell sequencing and real-time measurement of single-molecule fluorescence has accelerated our understanding of splicing stochasticity. An integrated method that combines the 'bird's-eye view' provided by high-throughput sequencing and the detailed information from time-lapse single-molecule microscopy will facilitate further advancements.

Understanding the physiological role of noise in RNA processing

To understand a potential functional role for variability (stochastic or otherwise) in RNA sequence, a potential starting point is the assessment of the protein products. The proposition of 'one gene, multiple proteins' is rooted in the early days soon after the discovery of

alternative splicing. Yet, there is debate on the extent to which alternative splicing can change the protein reservoir. Of course, there are numerous examples showing that functionally distinct proteins are generated from alternative splicing isoforms. More recently, using ribosome profiling, it has been shown that more than 75% of medium-to-high abundance alternative cassette exons are occupied by ribosomes [101]. Over 60% of these cassette exons preserve the reading frame, in agreement with the observation that short, frame-preserving cassette exons are more evolutionarily favored [102]. An opposing view is that although thousands of alternative splicing isoforms are identified through RNA-seq, only a small portion of them are identified by large-scale mass spectrometry [103]. In the early days of GENCODE, Tress et al. [104] examined the limited number of reported alternative splicing events. They concluded that many alternative spliced transcripts, if translated, would drastically change the structure and function of the protein products. Nevertheless, it is hard to predict the protein structure that would result from some isoforms, or whether the sequence would result in an unstable folding status [104]. The follow-up study, based on a large-scale human proteomics database analysis, suggests that most highly expressed genes have one dominant isoform [105]. Nevertheless, owing to the limited sensitivity of mass spectrometry-based proteomics, we still do not know what proportion of alternative splicing isoforms will result in functional proteins.

Did biological systems evolve to suppress splicing noise? Alternatively, has the system evolved to exploit this noise? The most common noise-reducing regulatory mechanism is negative feedback. RNA quality control systems, such as nonsense-mediated decay (NMD), nonstop decay (NSD), and no-go decay (NGD), have evolved to mitigate errors in RNA processing [106]. In addition to negative feedback, kinetic proofreading also plays a role in dampening splicing noise [107, 108]. On the other hand, noisy splicing has been proposed to give rise to population heterogeneity and may be essential in neurogenesis [109, 110], innate immunity [111], and evolution [112, 113]. Notably, recent work has also demonstrated a global alteration in splicing in cancers that involve mutations in core spliceosomal subunits such as U2AF1 and SF3B1 [114]. Intensive sequencing efforts from patients' samples argued that the splicing changes in these patients are minor and highly variable [115–117]. To date, it has been difficult to attribute either the cancer phenotype or the prognosis to isoform changes affecting a specific set of genes. Cancer is an evolutionary disease and these spliceosomal mutations often occur at an early stage [118–120]. One possibility might be that the mutations in spliceosomal proteins function as an amplifier of splicing noise, as has been suggested for splicing alterations in other disease states [121].

Low-abundance isoforms that are generated through splicing noise may allow the new variant to be evolutionarily tested and could benefit tumor progression in a heterogeneous way.

Current limitations and outlook

Splicing has been studied intensively, but it is only one of the processes that determine the chemical composition of mRNA. The roles of RNA editing and RNA modifications are now coming into focus as additional potential sources of heterogeneity. Transcriptome profiling techniques are powerful because of the exquisite detail they provide, and imaging allows researchers to follow cells over time. Future efforts to combine these advantages in order to generate longitudinal studies of transcription and splicing are promising but in the early stages [122]. In the meantime, the problem of interpreting the phenotypic consequences of variability remains a considerable challenge.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Published online: 09 July 2018

References
1. Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. Science. 2017;357:661–7.
2. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA. The technology and biology of single-cell RNA sequencing. Mol Cell. 2015;58:610–20.
3. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods. 2009;6:377–82.
4. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. Nat Rev Immunol. 2018;18:35–45.
5. Lenstra TL, Rodriguez J, Chen H, Larson DR. Transcription dynamics in living cells. Annu Rev Biophys. 2016;45:25–47.
6. Berget SM, Moore C, Sharp PA. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci U S A. 1977;74:3171–5.
7. Chow LT, Roberts JM, Lewis JB, Broker TR. A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids. Cell. 1977;11:819–36.
8. Kitchingman GR, Lai SP, Westphal H. Loop structures in hybrids of early RNA and the separated strands of adenovirus DNA. Proc Natl Acad Sci U S A. 1977;74:4392–5.

9. Beyer AL, Osheim YN. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev. 1988;2:754–65.
10. Breitbart RE, Andreadis A, Nadal-Ginard B. Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. Annu Rev Biochem. 1987;56:467–95.
11. Yeo G, Holste D, Kreiman G, Burge CB. Variation in alternative splicing across human tissues. Genome Biol. 2004;5:R74.
12. Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, et al. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science. 2003;302:2141–4.
13. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008; 456:470–6.
14. Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, et al. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science. 2008;321:956–60.
15. Blencowe BJ. Alternative splicing: new insights from global analyses. Cell. 2006;126:37–47.
16. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet. 2008;40:1413–5.
17. Consortium GTE. Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348:648–60.
18. Gibson G. Human genetics. GTEx detects genetic effects. Science. 2015;348: 640–1.
19. Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. Human genomics. The human transcriptome across tissues and individuals. Science. 2015;348:660–5.
20. Rivas MA, Pirinen M, Conrad DF, Lek M, Tsang EK, Karczewski KJ, et al. Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. Science. 2015;348:666–9.
21. Harper CV, Finkenstadt B, Woodcock DJ, Friedrichsen S, Semprini S, Ashall L, et al. Dynamic analysis of stochastic transcription cycles. PLoS Biol. 2011;9: e1000607.
22. Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. Mammalian genes are transcribed with widely different bursting kinetics. Science. 2011; 332:472–4.
23. Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. Cell. 2005;123:1025–36.
24. Fiering S, Whitelaw E, Martin DI. To be or not to be active: the stochastic nature of enhancer action. BioEssays. 2000;22:381–7.
25. Fukaya T, Lim B, Levine M. Enhancer control of transcriptional bursting. Cell. 2016;166:358–68.
26. Bartman CR, Hsu SC, Hsiung CC, Raj A, Blobel GA. Enhancer regulation of transcriptional bursting parameters revealed by forced chromatin looping. Mol Cell. 2016;62:237–47.
27. Chubb JR, Trcek T, Shenoy SM, Singer RH. Transcriptional pulsing of a developmental gene. Curr Biol. 2006;16:1018–25.
28. Tantale K, Mueller F, Kozulic-Pirher A, Lesne A, Victor JM, Robert MC, et al. A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. Nat Commun. 2016;7:12248.
29. Raser JM, O'Shea EK. Control of stochasticity in eukaryotic gene expression. Science. 2004;304:1811–4.
30. Hendy O, Campbell J Jr, Weissman JD, Larson DR, Singer DS. Differential context-specific impact of individual core promoter elements on transcriptional dynamics. Mol Biol Cell. 2017;28:3360–70.
31. Paszek P, Jackson DA, White MR. Oscillatory control of signalling molecules. Curr Opin Genet Dev. 2010;20:670–6.
32. White MD, Angiolini JF, Alvarez YD, Kaur G, Zhao ZW, Mocskos E, et al. Long-lived binding of Sox2 to DNA predicts cell fate in the four-cell mouse embryo. Cell. 2016;165:75–87.
33. Sato N, Nakayama M, Arai K. Fluctuation of chromatin unfolding associated with variation in the level of gene expression. Genes Cells. 2004;9:619–30.
34. Vinuelas J, Kaneko G, Coulon A, Vallin E, Morin V, Mejia-Pous C, et al. Quantifying the contribution of chromatin dynamics to stochastic gene expression reveals long, locus-dependent periods between transcriptional bursts. BMC Biol. 2013;11:15.
35. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. PLoS Biol. 2006;4:e309.
36. Swinstead EE, Paakinaho V, Presman DM, Hager GL. Pioneer factors and ATP-dependent chromatin remodeling factors interact dynamically: a new perspective: multiple transcription factors can effect chromatin pioneer functions through dynamic interactions with ATP-dependent chromatin remodeling factors. BioEssays. 2016;38:1150–7.
37. Sprouse RO, Karpova TS, Mueller F, Dasgupta A, McNally JG, Auble DT. Regulation of TATA-binding protein dynamics in living yeast cells. Proc Natl Acad Sci U S A. 2008;105:13304–8.
38. Cisse II, Izeddin I, Causse SZ, Boudarene L, Senecal A, Muresan L, et al. Real-time dynamics of RNA polymerase II clustering in live human cells. Science. 2013;341:664–7.
39. Darzacq X, Shav-Tal Y, de Turris V, Brody Y, Shenoy SM, Phair RD, Singer RH. In vivo dynamics of RNA polymerase II transcription. Nat Struct Mol Biol. 2007;14:796–806.
40. Cho WK, Jayanth N, English BP, Inoue T, Andrews JO, Conway W, et al. RNA polymerase II cluster dynamics predict mRNA output in living cells. elife. 2016;5:e13617. https://doi.org/10.7554/eLife.13617.
41. Hoskins AA, Moore MJ. The spliceosome: a flexible, reversible macromolecular machine. Trends Biochem Sci. 2012;37:179–88.
42. Larson JD, Hoskins AA. Dynamics and consequences of spliceosome E complex formation. elife. 2017;6:e27592. https://doi.org/10.7554/eLife.27592.
43. Lee Y, Rio DC. Mechanisms and regulation of alternative pre-mRNA splicing. Annu Rev Biochem. 2015;84:291–323.
44. Mount SM. A catalogue of splice junction sequences. Nucleic Acids Res. 1982;10:459–72.
45. Senapathy P, Shapiro MB, Harris NL. Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. Methods Enzymol. 1990;183:252–78.
46. Krawczak M, Reiss J, Cooper DN. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum Genet. 1992;90:41–54.
47. Gao K, Masuda A, Matsuura T, Ohno K. Human branch point consensus sequence is yUnAy. Nucleic Acids Res. 2008;36:2257–67.
48. Sibley CR, Blazquez L, Ule J. Lessons from non-canonical splicing. Nat Rev Genet. 2016;17:407–21.
49. Taggart AJ, Lin CL, Shrestha B, Heintzelman C, Kim S, Fairbrother WG. Large-scale analysis of branchpoint usage across species and cell lines. Genome Res. 2017;27:639–49.
50. Lynch M. Evolution of the mutation rate. Trends Genet. 2010;26:345–52.
51. Roca X, Akerman M, Gaus H, Berdeja A, Bennett CF, Krainer AR. Widespread recognition of 5′ splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. Genes Dev. 2012;26:1098–109.
52. Lim LP, Burge CB. A computational analysis of sequence features involved in recognition of short introns. Proc Natl Acad Sci U S A. 2001;98:11193–8.
53. Lappalainen T, Sammeth M, Friedlander MR, 't Hoen PA, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature. 2013;501:506–11.
54. Xiong HY, Alipanahi B, Lee LJ, Bretschneider H, Merico D, Yuen RK, et al. RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. Science. 2015;347:1254806.
55. Monlong J, Calvo M, Ferreira PG, Guigo R. Identification of genetic variants associated with alternative splicing using sQTLseekeR. Nat Commun. 2014;5:4698.
56. Hsiao YH, Bahn JH, Lin X, Chan TM, Wang R, Xiao X. Alternative splicing modulated by genetic variants demonstrates accelerated evolution regulated by highly conserved proteins. Genome Res. 2016;26:440–50.
57. Sterne-Weiler T, Sanford JR. Exon identity crisis: disease-causing mutations that disrupt the splicing code. Genome Biol. 2014;15:201.
58. Zhang XH, Chasin LA. Computational definition of sequence motifs governing constitutive exon splicing. Genes Dev. 2004;18:1241–50.
59. Fu XD, Ares M Jr. Context-dependent control of alternative splicing by RNA-binding proteins. Nat Rev Genet. 2014;15:689–701.
60. Wang J, Smith PJ, Krainer AR, Zhang MQ. Distribution of SR protein exonic splicing enhancer motifs in human protein-coding genes. Nucleic Acids Res. 2005;33:5053–62.
61. Lim SR, Hertel KJ. Commitment to splice site pairing coincides with a complex formation. Mol Cell. 2004;15:477–83.
62. Shcherbakova I, Hoskins AA, Friedman LJ, Serebrov V, Correa IR Jr, Xu MQ, et al. Alternative spliceosome assembly pathways revealed by single-molecule fluorescence microscopy. Cell Rep. 2013;5:151–65.
63. Tseng CK, Cheng SC. The spliceosome catalyzes debranching in competition with reverse of the first chemical reaction. RNA. 2013;19:971–81.
64. Tseng CK, Cheng SC. Both catalytic steps of nuclear pre-mRNA splicing are reversible. Science. 2008;320:1782–4.

65. Hoskins AA, Rodgers ML, Friedman LJ, Gelles J, Moore MJ. Single molecule analysis reveals reversible and irreversible steps during spliceosome activation. elife. 2016;5:e14166. https://doi.org/10.7554/eLife.14166.
66. Hoskins AA, Friedman LJ, Gallagher SS, Crawford DJ, Anderson EG, Wombacher R, et al. Ordered and dynamic assembly of single spliceosomes. Science. 2011;331:1289–95.
67. Rabani M, Raychowdhury R, Jovanovic M, Rooney M, Stumpo DJ, Pauli A, et al. High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. Cell. 2014;159:1698–710.
68. Singh J, Padgett RA. Rates of in situ transcription and splicing in large human genes. Nat Struct Mol Biol. 2009;16:1128–33.
69. Barrass JD, Reid JE, Huang Y, Hector RD, Sanguinetti G, Beggs JD, Granneman S. Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling. Genome Biol. 2015;16:282.
70. Coulon A, Ferguson ML, de Turris V, Palangat M, Chow CC, Larson DR. Kinetic competition during the transcription cycle results in stochastic RNA processing. elife. 2014;3 https://doi.org/10.7554/eLife.03939.
71. Schmidt U, Basyuk E, Robert MC, Yoshida M, Villemin JP, Auboeuf D, et al. Real-time imaging of cotranscriptional splicing reveals a kinetic model that reduces noise: implications for alternative splicing regulation. J Cell Biol. 2011;193:819–29.
72. Martin RM, Rino J, Carvalho C, Kirchhausen T, Carmo-Fonseca M. Live-cell visualization of pre-mRNA splicing with single-molecule sensitivity. Cell Rep. 2013;4:1144–55.
73. Carrillo Oesterreich F, Herzel L, Straube K, Hujer K, Howard J, Neugebauer KM. Splicing of nascent RNA coincides with intron exit from RNA polymerase II. Cell. 2016;165:372–81.
74. Vargas DY, Shah K, Batish M, Levandoski M, Sinha S, Marras SA, et al. Single-molecule imaging of transcriptionally coupled and uncoupled splicing. Cell. 2011;147:1054–65.
75. Query CC, Konarska MM. Splicing fidelity revisited. Nat Struct Mol Biol. 2006; 13:472–4.
76. Fong N, Kim H, Zhou Y, Ji X, Qiu J, Saldi T, et al. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. Genes Dev. 2014;28:2663–76.
77. Melamud E, Moult J. Stochastic noise in splicing machinery. Nucleic Acids Res. 2009;37:4873–86.
78. Pickrell JK, Pai AA, Gilad Y, Pritchard JK. Noisy splicing drives mRNA isoform diversity in human cells. PLoS Genet. 2010;6:e1001236.
79. Tapial J, Ha KCH, Sterne-Weiler T, Gohr A, Braunschweig U, Hermoso-Pulido A, et al. An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. Genome Res. 2017;27: 1759–68.
80. Hu J, Boritz E, Wylie W, Douek DC. Stochastic principles governing alternative splicing of RNA. PLoS Comput Biol. 2017;13:e1005761.
81. Tang F, Barbacioru C, Bao S, Lee C, Nordman E, Wang X, et al. Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. Cell Stem Cell. 2010;6:468–78.
82. Tang F, Lao K, Surani MA. Development and applications of single-cell transcriptome analysis. Nat Methods. 2011;8:S6–11.
83. Ozsolak F, Ting DT, Wittner BS, Brannigan BW, Paul S, Bardeesy N, et al. Amplification-free digital gene expression profiling from minute cell quantities. Nat Methods. 2010;7:619–21.
84. Islam S, Kjallquist U, Moliner A, Zajac P, Fan JB, Lonnerberg P, Linnarsson S. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. 2011;21:1160–7.
85. Brouilette S, Kuersten S, Mein C, Bozek M, Terry A, Dias KR, et al. A simple and novel method for RNA-seq library preparation of single cell cDNA analysis by hyperactive Tn5 transposase. Dev Dyn. 2012;241:1584–90.
86. Hashimshony T, Wagner F, Sher N, Yanai I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep. 2012;2:666–73.
87. Ramskold D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol. 2012;30:777–82.
88. Marinov GK, Williams BA, McCue K, Schroth GP, Gertz J, Myers RM, Wold BJ. From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. Genome Res. 2014;24:496–510.
89. Song Y, Botvinnik OB, Lovci MT, Kakaradov B, Liu P, Xu JL, Yeo GW. Single-cell alternative splicing analysis with expedition reveals splicing dynamics during neuron differentiation. Mol Cell. 2017;67:148–61. e145

90. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature. 2013;498:236–40.
91. Iyer-Biswas S, Hayot F, Jayaprakash C. Stochasticity of gene products from transcriptional pulsing. Phys Rev E Stat Nonlinear Soft Matter Phys. 2009;79:031911.
92. Sharon D, Tilgner H, Grubert F, Snyder M. A single-molecule long-read survey of the human transcriptome. Nat Biotechnol. 2013;31:1009–14.
93. Rhoads A, Au KF. PacBio sequencing and its applications. Genomics Proteomics Bioinformatics. 2015;13:278–89.
94. Femino AM, Fay FS, Fogarty K, Singer RH. Visualization of single RNA transcripts in situ. Science. 1998;280:585–90.
95. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods. 2008;5:877–9.
96. Waks Z, Klein AM, Silver PA. Cell-to-cell variability of alternative RNA splicing. Mol Syst Biol. 2011;7:506.
97. Mellis IA, Gupte R, Raj A, Rouhanifard SH. Visualizing adenosine-to-inosine RNA editing in single mammalian cells. Nat Methods. 2017;14:801–4.
98. Brody Y, Neufeld N, Bieberstein N, Causse SZ, Bohnlein EM, Neugebauer KM, et al. The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. PLoS Biol. 2011;9:e1000573.
99. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell. 2014;157:1262–78.
100. Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science. 2014;346:1258096.
101. Weatheritt RJ, Sterne-Weiler T, Blencowe BJ. The ribosome-engaged landscape of alternative splicing. Nat Struct Mol Biol. 2016;23:1117–23.
102. Zhang C, Krainer AR, Zhang MQ. Evolutionary impact of limited splicing fidelity in mammalian genes. Trends Genet. 2007;23:484–8.
103. Tress ML, Abascal F, Valencia A. Alternative splicing may not be the key to proteome complexity. Trends Biochem Sci. 2017;42:98–110.
104. Tress ML, Martelli PL, Frankish A, Reeves GA, Wesselink JJ, Yeats C, et al. The implications of alternative splicing in the ENCODE protein complement. Proc Natl Acad Sci U S A. 2007;104:5495–500.
105. Ezkurdia I, Rodriguez JM, Carrillo-de Santa Pau E, Vazquez J, Valencia A, Tress ML. Most highly expressed protein-coding genes have a single dominant isoform. J Proteome Res. 2015;14:1880–7.
106. Doma MK, Parker R. RNA quality control in eukaryotes. Cell. 2007;131:660–8.
107. Egecioglu DE, Chanfreau G. Proofreading and spellchecking: a two-tier strategy for pre-mRNA splicing quality control. RNA. 2011;17:383–9.
108. Koodathingal P, Staley JP. Splicing fidelity: DEAD/H-box ATPases as molecular clocks. RNA Biol. 2013;10:1073–9.
109. Hattori D, Chen Y, Matthews BJ, Salwinski L, Sabatti C, Grueber WB, Zipursky SL. Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms. Nature. 2009;461:644–8.
110. Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, et al. *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell. 2000;101:671–84.
111. Dong Y, Cirimotich CM, Pike A, Chandra R, Dimopoulos G. *Anopheles* NF-kappaB-regulated splicing factors direct pathogen-specific repertoires of the hypervariable pattern recognition receptor AgDscam. Cell Host Microbe. 2012;12:521–30.
112. Ast G. How did alternative splicing evolve? Nat Rev Genet. 2004;5:773–82.
113. Sorek R. The birth of new exons: mechanisms and evolutionary consequences. RNA. 2007;13:1603–8.
114. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature. 2011;478:64–9.
115. Ilagan JO, Ramakrishnan A, Hayes B, Murphy ME, Zebari AS, Bradley P, Bradley RK. U2AF1 mutations alter splice site recognition in hematological malignancies. Genome Res. 2015;25:14–26.
116. Kim E, Ilagan JO, Liang Y, Daubner GM, Lee SC, Ramakrishnan A, et al. SRSF2 mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. Cancer Cell. 2015;27:617–30.
117. Shirai CL, Ley JN, White BS, Kim S, Tibbitts J, Shao J, et al. Mutant U2AF1 expression alters hematopoiesis and pre-mRNA splicing in vivo. Cancer Cell. 2015;27:631–43.
118. Nadeu F, Delgado J, Royo C, Baumann T, Stankovic T, Pinyol M, et al. Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia. Blood. 2016;127:2122–30.
119. Mian SA, Rouault-Pierre K, Smith AE, Seidl T, Pizzitola I, Kizilors A, et al. SF3B1 mutant MDS-initiating cells may arise from the haematopoietic stem cell compartment. Nat Commun. 2015;6:10004.

120. Makishima H, Yoshizato T, Yoshida K, Sekeres MA, Radivoyevitch T, Suzuki H, et al. Dynamics of clonal evolution in myelodysplastic syndromes. Nat Genet. 2017;49:204–12.
121. Fox-Walsh KL, Hertel KJ. Splice-site pairing is an intrinsically high fidelity process. Proc Natl Acad Sci U S A. 2009;106:1766–71.
122. Single Cell Analysis Challenge. https://commonfund.nih.gov/singlecell/challenge. Accessed 30 Sep 2017.