

Fast and Furious: insights of back splicing regulation during nascent RNA synthesis

Xue Wei¹, Ma Xu-Kai² and Yang Li²

Citation: [SCIENCE CHINA Life Sciences](#); doi: 10.1007/s11427-020-1881-1

View online: <https://engine.scichina.com/doi/10.1007/s11427-020-1881-1>

Published by the [Science China Press](#)

Articles you may be interested in

[Competition of RNA splicing: line in or circle up](#)

SCIENCE CHINA Life Sciences **57**, 1232 (2014);

[Regulation of mammalian pre-mRNA splicing](#)

Science in China Series C-Life Sciences **52**, 253 (2009);

[Opening new doors for understanding eukaryotic RNA splicing](#)

SCIENCE CHINA Life Sciences **58**, 1171 (2015);

[Landscape and Fruit Developmental Regulation of Alternative Splicing in Tomato by Genome-Wide Analysis](#)

Horticultural Plant Journal **2**, 338 (2016);

[New insights into mTOR structure and regulation](#)

Chinese Science Bulletin **59**, 2927 (2014);

REVIEW

Fast and Furious: insights of back splicing regulation during nascent RNA synthesis

Wei Xue¹, Xu-Kai Ma¹ & Li Yang^{1,2,*}

¹CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, China;

²School of Life Science and Technology, ShanghaiTech University, 393 Middle Huaxia Road, Shanghai 201210, China

*Corresponding author (email: liyang@picb.ac.cn)

Received November 7, 2020; accepted January 7, 2021

Abstract

Alternative splicing of eukaryotic precursor (messenger) RNAs in the nucleus not only increases transcriptomic complexity, but also expands proteomic and functional diversity. In addition to basic types of alternative splicing, recent transcriptome-wide analyses have also suggested other new types of non-canonical splicing, such as back splicing and recursive splicing, and their widespread expression across species. Increasing lines of evidence have suggested mechanisms for back splicing, including insights from analyses of nascent RNA sequencing. In this review, we discuss our current understanding of back splicing regulation, and highlight its distinct characteristics in processing during nascent RNA synthesis by taking advantage of metabolic tagging nascent RNA sequencing. Features of recursive splicing are also discussed in the perspective of nascent RNA sequencing.

Keywords

alternative splicing, back splicing, deep sequencing, nascent RNA-seq, metabolic tagging, 4sUDRB-seq, transcription, recursive splicing

Introduction

One of the fundamental features of eukaryotic genes is that their expressed sequences (exons) are interrupted by stretches of intragenic regions (introns) (Gilbert, 1978). With two transesterification steps of nuclear mRNA splicing that are catalyzed by the spliceosome to link an upstream donor site (5' splice site) with a downstream acceptor site (3' splice site), exons are ligated together and introns are removed from precursor mRNAs (pre-mRNAs), thus providing a versatile layer for gene expression regulation (Black, 2003; Lee and Rio, 2015; Nilsen and Graveley, 2010). In higher eukaryotes, alternations of splice site choice, referred to as alternative splicing, could produce multiple mRNA isoforms from a single pre-mRNA (Black, 2003; Lee and Rio, 2015; Nilsen and Graveley, 2010). Indeed, nearly all human multi-exon genes undergo

alternative splicing (Pan et al., 2008; Wang et al., 2008), which significantly expands the transcriptomic complexity and hence proteomic diversity.

With the application of deep sequencing (RNA-seq) technologies and corresponding bioinformatics tools, complex alternative splicing maps have been generated in different tissues and across species (Barbosa-Morais et al., 2012; Merkin et al., 2012). In addition to well-characterized alternative splicing events (Nilsen and Graveley, 2010), such as cassette-exon inclusion or exclusion, alternative 5' or 3' splice site selection, intron retention and mutually exclusive exons (Figure 1A), other types of non-canonical splicing events have been recently re-identified on a genome-wide scale, including but not limited to circularized exons by back splicing (back-spliced exons, Figure 1B) (Chen, 2016) and recursive splice site selection (Figure 1C) (Georgomanolis et al., 2016). Mounting lines of evidence have suggested that back splicing is more widely expressed than previously thought (Chen, 2020; Zhang et al., 2016a), but unconventionally processed with unique features. In this review, we describe the regulation of how splice sites are chosen for back splicing to produce previously under-appreciated circular RNAs (circRNAs). We start with a brief summary of our current understanding of how back splicing is processed by the same spliceosomal machinery for canonical splicing. After introducing different types of nascent RNA-seq technologies and their pros and cons in the application for back splicing analyses, we then focus on distinct mechanistic processes of back splicing, which have been learned from analyzing metabolic tagging nascent RNA-seq datasets. Features of recursive splicing at the level of nascent RNA synthesis are also discussed.

Recent re-discovery of non-conventional back splicing and recursive splicing

Recent transcriptome-wide analyses of steady-state RNA-seq datasets that profile most mRNAs and long noncoding RNAs (lncRNAs) with 3'-polyadenylated tails have shown the complex and dynamic regulation of alternative splicing (Figure 1A) (Barbosa-Morais et al., 2012; Merkin et al., 2012). Interestingly, with

non-polyadenylated RNA-seq datasets aimed to identify RNAs without 3'-polyadenylated tails (Yang et al., 2011; Yin et al., 2015) and/or nascent RNA-seq datasets (Wissink et al., 2019; Zhang et al., 2016b), previously under-appreciated types of non-canonical splicing events, such as back-spliced exon (Figure 1B) (Jeck and Sharpless, 2014; Zhang et al., 2014) and recursive splice site selection (Figure 1C) (Georgomanolis et al., 2016; Zhang et al., 2018), have also been re-discovered on a genome-wide scale, suggesting an even complex regulation of gene expression.

Back splicing for circRNA biogenesis

Deep-sequencing analyses of non-poly(A)-enriched RNAs (Yang et al., 2011) using specific computational approaches (Gao and Zhao, 2018) to identify fusion splicing junction fragments in a non-sequential order have identified hundreds of thousands of circRNAs from back-spliced exons in animals (Chen, 2016; Li et al., 2018) and plants (Chu et al., 2017; Liu et al., 2020). Different from canonical splicing that ligates an upstream 5' splice site to a downstream 3' splice site, back splicing links a downstream 5' splice site backwards to an upstream 3' splice site (Figure 1B), resulting in the formation of covalently closed circRNAs (Chen, 2020; Li et al., 2018; Wilusz, 2018). Most, if not all, back splice events happen at the same splice sites as canonical splicing for linear RNA, which suggests a direct competition of splice site selections for either canonical splicing or back splicing from the same gene loci (Chen and Yang, 2015). This competition also leads to nearly full sequence overlap between linear and circular RNA formation from the same gene loci, impeding studies of back splicing quantification (Ma et al., 2019) and circRNA function (Li et al., 2018). Nevertheless, increasing lines of evidence have already shown that at least some circRNAs play important roles with distinct modes of action in both physiological and pathogenetically conditions, including but not limited to innate immunity (Li et al., 2017; Liu et al., 2019), cell proliferation and transformation (Chen et al., 2019; Guarnerio et al., 2019), and neuronal function (Piwecka et al., 2017; You et al., 2015). Please see other publications for detailed discussion of the biological significance of

circRNAs (reviewed in (Chen, 2020; Li et al., 2018; Kristensen et al., 2019; Wilusz, 2018)).

Mechanistically, back splicing is catalyzed by the same spliceosomal machinery that processes canonical splicing (Starke et al., 2015; Liang et al., 2017). Meanwhile, back-splicing is regulated by *cis*-elements, mainly intronic complementary sequences (ICSs) flanking circRNA-forming exons, and *trans*-factors that bind to *cis*-elements in flanking introns (Ashwal-Fluss et al., 2014; Conn et al., 2015; Li et al., 2017; Zhang et al., 2014). Specifically, analyses of RNA pairing capacities of different types of ICSs suggested that *SINEs*, especially *Alu* elements in primates, contribute the most for back-splicing to produce circRNAs (Zhang et al., 2014; Dong et al., 2017). In addition, competition of RNA pairs from different sets of ICSs is also correlated with alternative back splice site selections (including alternative 5' or 3' back-splice site selection) (Dong et al., 2017; Zhang et al., 2020), which leads to the formation of multiple circRNAs from the same gene loci. Meanwhile, some proteins, such as Muscleblind (Mbl) in *Drosophila* (Ashwal-Fluss et al., 2014), quaking (QKI) (Conn et al., 2015) and nuclear factor 90/110 (NF90/NF110) (Li et al., 2017) in human, were found to facilitate back-splicing for circRNA biogenesis. Differently, some other RBPs, such as DHX9 (RNA helicase A, RHA) (Aktas et al., 2017) and adenosine deaminase acting on RNA 1 (ADAR1) (Ivanov et al., 2015), were shown to repress back-splicing. Detailed modes of action of how back-splicing is regulated for circRNA formation are awaiting for further investigation.

Noticeably, many of these conclusions on back splicing regulation were largely drawn from studies of individual circRNAs (Salzman et al., 2012; Starke et al., 2015; Liang et al., 2017) and canonical RNA-seq datasets that show steady-state levels of mature (linear and circular) RNAs (Guo et al., 2014; Salzman et al., 2013; Yang et al., 2011; Zhang et al., 2016a). Since steady-state RNA-seq mainly indicates the equilibrium of gene transcription, maturation and degradation (Rabani et al., 2011), it has been unclear whether the pervasive detection of the steady-state levels of circRNAs in a

cell-/tissue-specific manner can reflect their dynamic expression and the endogenous kinetics of back splicing for circRNA production. For instance, if back splicing is unfavorably processed by the same spliceosomal machinery used for canonical splicing, how can a subset of abundant circRNAs from back-spliced exons be processed at higher levels than their cognate linear RNAs (Dong et al., 2017; Rybak-Wolf et al., 2015; You et al., 2015)? This question cannot be addressed solely by analyzing steady-state RNA-seq datasets. Meanwhile, given the fact that circRNAs produced from back splicing are more stable than linear RNAs from canonical (alternative) splicing, it is not possible to directly compare the steady-state levels of circRNAs with their linear RNA cognates to reveal the selection of splice sites for back splicing or canonical splicing. This problem can be, at least partially, solved by analyzing nascent RNA-seq datasets to directly examine how back splicing itself is regulated for circRNA biogenesis and hence their associated functions.

Recursive splicing to remove long introns

Another type of non-conventional splicing, recursive splicing, has been discovered in both *Drosophila melanogaster* (Hatton et al., 1998; Duff et al., 2015) and mammalian cells (Sibley et al., 2015). Mechanistically, an upstream canonical 5' splice site is firstly spliced with a downstream 3' recursive splice site to remove the first part of a long intron, resulting in the formation of a new 5' recursive splice site, which is then subsequently spliced with a downstream canonical 3' splice site to remove the second part of the long intron (Figure 1C). By taking advantage of this multi-step process of recursive splicing, long introns can be efficiently removed (Georgomanolis et al., 2016; Sibley et al., 2015). Current studies showed that recursive splicing can be regulated by *cis*-elements, and has been revealed in both fruitfly and human (Duff et al., 2015; Sibley et al., 2015). When mediated by a variant U1 RNA, recursive splicing can be processed from non-canonical splice sites (Tan et al., 2016). Due to their transient and unstable characteristics, recursive splicing intermediates between 5' splice sites and 3' recursive splice sites or 5' recursive splice sites and 3' splice sites are generally undetectable in canonical RNA-seq datasets for mature RNAs, but could

be detected in nascent RNA-seq datasets (Zhang et al., 2018). Identifying fragments that are mapped to unstable recursive splice sites from nascent RNA-seq datasets is crucial to capture the intermediate steps of recursive splicing and to further depict its regulations during transcription.

Different technologies for nascent RNA-seq and transcription analyses

By purifying nascent RNAs for deep sequencing, a transcriptome-wide view of nascent RNA synthesis and regulation can be obtained (Wissink et al., 2019). Different from canonical RNA-seq datasets that reflect steady-state levels of RNAs in a balance between RNA synthesis, maturation and degradation (Rabani et al., 2011), nascent RNA-seq is applied to detect newly-transcribed RNA molecules with short half-lives and low abundance (Wissink et al., 2019). Various biochemical methods, such as co-immunoprecipitation (Churchman and Weissman, 2011; Mayer et al., 2015; Nojima et al., 2015), RNA polymerase II (Pol II) run-on (Core et al., 2008; Kwak et al., 2013), and metabolic labeling (Fuchs et al., 2014; Rabani et al., 2014; Zhang et al., 2016b), have been used to capture nascent RNAs for subsequent deep sequencing, which uncovers different layers of regulation features for transcription and other co-/post-transcriptional events, such as (alternative) splicing (Wissink et al., 2019).

Immunoprecipitation-based methods

Due to the fact that nascent RNAs are transcribed from DNA templates by RNA polymerase, the most direct and simplest method for nascent RNA enrichment is to capture nascent (messenger and long noncoding) RNAs that are still interacting with chromatin and/or Pol II (Figure 2A) (Wissink et al., 2019). Chromatin-associated RNAs (caRNA) can be isolated for deep sequencing (caRNA-seq) to profile nascent RNAs that are newly transcribed and still associated with chromatin (Bhatt et al., 2012). In addition, Start-seq was developed to profile chromatin-associated capped RNAs, after removing RNAs without the 5' cap protection by enzymatic degradation (Nechaev et al., 2010; Wissink et al., 2019). Deep sequencing of these short (less than 100 nucleotides in length), capped RNAs that are associated with chromatin has

demonstrated initiation, pausing and/or early elongation of transcription, together with providing high-resolution maps of transcription start sites (TSSs).

Alternatively, native elongating transcripts can be isolated by co-immunoprecipitating with Pol II and then applied to deep sequencing analysis (NET-seq) (Mayer et al., 2015). In principle, NET-seq and its application in mammalian cells (mNET-seq) (Nojima et al., 2015) reveal active Pol II positions along the DNA templates at single nucleotide resolution (Figure 2A) (Churchman and Weissman, 2011). Importantly, with antibodies against different phosphorylation states of Pol II's large subunit C-terminal domain (CTD), different signals of nascent RNAs can be determined to show distinct distribution of specific Pol II isoforms during nascent RNA biogenesis (Nojima et al., 2015). For example, most nascent RNAs were found to be accumulated around TSSs by using the 8WG16 antibody against Pol II unphosphorylated CTD for the enrichment. When using the CMA602 antibody against Pol II phosphorylated CTD on Serine 2 (Ser2) for the enrichment, nascent RNAs were found to be widely distributed throughout whole gene bodies, including transcription end sites (TESs). Differently, nascent RNA-seq peaks were enriched at both ends of co-transcriptionally spliced exons when using the antibody CMA603 against Pol II phosphorylated CTD on Serine 5 (Ser5) for the enrichment. In this way, (m)NET-seq with different antibodies against distinct Pol II CTD modifications can be used to reveal characteristic features of nascent RNA biogenesis and process, such as co-transcriptional splicing of protein-coding RNAs (Nojima et al., 2015; Nojima et al., 2018) and processing of long intergenic noncoding RNAs (lincRNAs) (Schlackow et al., 2017).

Despite their broad application in analyses of active transcription, other abundant mature RNAs, such as *XIST* long noncoding RNA (lncRNA) and/or small nuclear RNAs (snRNAs), can also be co-purified with nascent RNAs in these types of chromatin/Pol II immunoprecipitation based methods which allow precise profiling of nascent RNAs. It is also worthwhile noting that RNAs co-immunoprecipitated with

chromatin and/or Pol II are not precisely equivalent to nascent RNAs, since some mature RNA transcripts with 3'-polyadenylated ends are also associated with chromatin fractions (Bhatt et al., 2012).

Run-on based methods

In addition to co-immunoprecipitation-based methods, nascent RNAs can be directly isolated using nuclear run-on (NRO) assays for active transcription analysis (Figure 2B). NRO assays start with chilling cells on ice to stop Pol II from transcribing, followed by permeabilizing cells to isolate nuclei, which also removes endogenous nucleotides. By treating isolated nuclei with labeled nucleoside 5'-triphosphates (NTPs), transcription can be restored *in vitro* for nascent RNA synthesis. By replacing radiolabeled NTPs in the NRO assays with 5-bromouridine 5'-triphosphate (Br-UTP), Br-UTP-labeled nascent RNAs can be purified for subsequent deep sequencing (GRO-seq) to directly profile competent TSSs on a genome-wide scale (Core et al., 2008; Jonkers et al., 2014). Various mechanistic insights into transcription regulation have been achieved by analyzing GRO-seq (Core et al., 2008; Jordan-Pla et al., 2019). These studies showed that only transcription-competent sites could be clearly identified due to the requirement of labeled nucleotide incorporation in NRO based assays, while transcriptionally inactive regions remained undetectable. Meanwhile, the distribution of GRO-seq signals was enriched between ~20 and 50 bases downstream of TSSs, suggesting that promoter-proximal pausing is a key regulatory step during active transcription. With the incorporation of biotin-labeled ribonucleotide triphosphate analogs (biotin-NTP) in NRO assays, precise nuclear run-on sequencing (PRO-seq) has been achieved to map actively-engaged Pol II genome-wide (Kwak et al., 2013). PRO-seq demonstrated a similar conclusion of enriched Pol II accumulation downstream of promoters as with GRO-seq, and further revealed that active Pol II also accumulated at 3' splice sites of intron-exon junctions, thus suggesting a coupling between transcription and splicing (Kwak et al., 2013; Mahat et al., 2016; Wissink et al., 2019).

Despite being useful to detect rules of transcription regulation, such as promoter-proximal pausing in the early stages of elongation, NRO-based nascent RNA-seq methods are limited by their short incubation times for transcription (Core et al., 2008; Kwak et al., 2013; Jonkers et al., 2014; Jordan-Pla et al., 2019; Wissink et al., 2019), which produces short fragments of actively-transcribed RNAs for temporal snapshots of competent transcription sites. In addition, NRO-based nascent RNA synthesis is achieved from isolated nuclei after cell permeabilization, thus indicating the regulation of transcription in an *in vitro*, but not *in vivo*, condition. Direct purification of nascent RNAs from *in vivo* systems is needed to show the biogenesis and processing of nascent RNAs along with transcription elongation in intact cells.

Metabolic labeling based methods

By applying metabolic labeling, such as with 4-thiouridine (4sU), newly-transcribed nascent RNAs can be purified from intact cells for deep sequencing to reveal the whole landscape of RNA transcription, processing and degradation under given conditions (Rabani et al., 2011; Rabani et al., 2014). However, newly-transcribed transcripts purified from this direct labeling method can be mixed with processed/matured RNAs from pre-existing transcription. To circumvent this issue, cells can be pre-treated with 5,6-Dichloro-1- β -D-ribofuranosylbenzimidazole (DRB) to exclude unintended labeling of RNAs from pre-existing transcription, and only nascent RNAs from restored transcription after DRB removal are able to be tagged with 4sU. Theoretically, by inhibiting the P-TEFb-dependent Ser2 phosphorylation of CTD, DRB blocks new entry of initiating Pol II or prevents 5'-proximally paused Pol II from transcription initiation step into productive elongation, but rarely affects active transcription of previously-elongated transcripts by phosphorylated Pol II within gene bodies (Ardehali and Lis, 2009; Singh and Padgett, 2009). After hours of DRB incubation, actively elongated Pol II molecules should be all cleared from gene bodies. Since the inhibition of transcription by DRB is reversible, removal of DRB and simultaneous addition of the uridine analog 4sU could restore transcription elongation that is synchronized from promoter-proximal regions at most gene loci (Fuchs et al.,

2014; Fuchs et al., 2015; Zhang et al., 2016b). After metabolic labeling for designed time periods, 4sU-tagged nascent RNAs are collected and subjected to deep-sequencing (4sUDRB-seq, Figure 2C). Signals of newly-transcribed RNAs with 4sU labeling from TSSs prove the restored transcription from promoter regions after DRB removal and metabolic labeling (Figure 2C). Comparing nascent RNAs from a series of pulse-labeling experiments is useful to determine features of transcription regulation, such as calculating transcription elongation rates (TERs) and identifying the coupling of transcription and splicing (Fuchs et al., 2014; Fuchs et al., 2015; Zhang et al., 2016b). As they are inefficiently produced, fragments mapping to back splice junction (BSJ) sites were barely detected from nascent RNA-seq in PA1 cells with short metabolic labeling times (such as within 30 min), but kept accumulating in samples with prolonged time points up to 240 and 960 min (Zhang et al., 2016b).

Applications of nascent RNA-seq technologies for transcription analyses

Nascent RNA-seq technologies are applicable to determine rules of regulation at various transcription steps. For example, Start-seq and PRO-seq are useful to identify transcription initiation sites (Core et al., 2008; Nechaev et al., 2010; Wissink et al., 2019). GRO-seq and (m)NET-seq are efficient to detect promoter-proximal pausing (Core et al., 2008; Mayer et al., 2015; Nojima et al., 2015; Wissink et al., 2019). PRO-seq and (m)NET-seq can be used to determine transcription termination sites (Nojima et al., 2015; Schlackow et al., 2017; Wissink et al., 2019). Both GRO-seq and 4sUDRB-seq datasets are applicable to calculate transcription elongation rates (Danko et al., 2013; Fuchs et al., 2014; Jonkers et al., 2014; Zhang et al., 2016b). Regulation of other co-/post- transcriptional events, such as splicing, can be also determined by GRO-seq, mNET-seq and 4sUDRB-seq datasets (Jonkers et al., 2014; Nojima et al., 2015; Fuchs et al., 2014; Zhang et al., 2016b).

Which type of these different nascent RNA-seq datasets is more suitable for studying the kinetics studies of back splicing? Compared with immunoprecipitation- and NRO-based methods, 4sUDRB-seq is more appropriate in many ways to be used for back

splicing analysis (Figure 2). First of all, the pre-treatment with DRB in 4sUDRB-seq method guarantees metabolic labeling of freshly transcribed RNAs but not mature RNAs, while in immunoprecipitation-based methods, other non-nascent RNAs could be co-purified together with nascent RNAs. In addition, the pre-treatment with DRB in the 4sUDRB-seq method arrests Pol II at the transcription initiation step and thus synchronizes all transcription events after DRB removal, but chromatin/Pol II immunoprecipitation could pull down (nascent) RNAs at different transcription stages from different locations of gene bodies at the same gene loci. Moreover, with extended labeling times in 4sUDRB-seq, full length metabolic tagging nascent RNAs could be enriched to reveal transcription regulation along the whole gene bodies, thus proving useful to detect unfavorably processed back splice events; however, NRO-based methods only generate short fragments of actively-transcribed RNAs for snapshots of transcription regulation. Finally, the in-cell metabolic labeling of 4sUDRB-seq likely reflects endogenous transcription regulation, while NRO-based methods only restore transcription in an *in vitro* system after cell permeabilization.

Of note, some other types of 4sU labeling methods have also been developed for nascent RNA profiling, such as transient transcriptome sequencing (TT-seq) (Schwalb et al., 2016) and thiol(SH)-linked alkylation for the metabolic sequencing of RNA (SLAM-seq) (Herzog et al., 2017; Muhar et al., 2018). However, these methods are limited for the analysis of back splicing. TT-seq is carried out without DRB treatment and with only 5-min labeling (Schwalb et al., 2016), which is not only insufficient to enrich nascent RNAs without DRB treatment, but also difficult to gain enough signals for back splicing within such a short labeling time. SLAM-seq only enriches metabolically tagged mRNAs with polyadenylated tails (Herzog et al., 2017; Muhar et al., 2018), which excludes circRNAs that lack poly(A) tails for the subsequent analysis. In this case, SLAM-seq is also not suitable for back splicing studies. Taken all these together, 4sUDRB-seq is applicable for studying the kinetics of back splicing.

Regulation of back splicing and recursive splicing revealed by 4sUDRB-seq

To obtain suitable datasets for a global view of back splicing regulation, two modifications have been chosen to prepare 4sUDRB-seq datasets for back splicing analysis. One is to, after DRB removal, treat cells over a wide time course of 4sU labeling, spanning from as short as 10 min to 960 min (about 16 hrs), to maximally detect unfavorably-processed back splicing signals after transcription. The other change is to apply an additional rRNA-depletion step to remove background signals of redundant rRNAs to obtain more fragments mapping to BSJ sites. Collectively, a time series (including 0, 10, 15, 30, 60, 120, 240 and/or 960 min) of 4sU-labeled nascent RNA-seq datasets were generated from human embryonic carcinoma PA1 cells, human embryonic stem cell line H9 cells and H9-differentiated forebrain (FB) cells (Zhang et al., 2016b). Some possible side effects, including the repression of DRB treatment, 4sU labeling and/or subsequent purification steps on back splicing/circRNAs, were excluded by comparing exogenously-expressed nascent circRNAs with or without these treatments, confirming these 4sUDRB-seq datasets were suitable for the subsequent nascent circRNA analysis (Figure 2C). With these comprehensive 4sUDRB-seq datasets from PA1, H9 and FB cells (Zhang et al., 2016b), regulation rules of transcription elongation, splicing, and back splicing were quantitatively measured and compared on a whole genome-wide scale. Characteristic features, including new insights, of back splicing regulation were obtained at a single-nucleotide resolution.

Features of back splicing regulation during nascent RNA synthesis

By comprising 4sUDRB-seq datasets with short 4sU pulse labeling time points (10 and 15 min), transcription elongation rates (TERs) of circRNA-producing or non-circRNA producing genes were obtained (Zhang et al., 2016b). It was interesting to observe that transcription of circRNA-producing genes was generally faster than that of non-circRNA producing ones, with TERs of 2.90 kb/min versus 2.29 kb/min in PA1 cells. This fast transcription of circRNA-producing genes appeared to be a general rule for back splicing regulation. On the one hand, circRNA-producing genes

were transcribed faster than non-circRNA producing ones in both H9 and FB cells, with TERs of 3.05 kb/min versus 2.72 kb/min or 3.34 kb/min versus 2.87 kb/min, respectively. On the other hand, in PA1 cells with artificially-altered transcription rates using Pol II mutants (E1126G and R749H) (Fong et al., 2014), back splicing efficiencies of four examined endogenous circRNAs, individually from *BMP2*, *ZNF148*, *PVT1* and *SPECC1* loci, were positively correlated with Pol II transcription rates (Zhang et al., 2016b). Increased TER with the Pol II E1126G mutant resulted in higher levels of back splicing for these examined circRNAs, and meanwhile, reduced TER with Pol II R749H mutant led to lower levels of back splicing for the same set of circRNAs. The finding that back splicing for circRNA biogenesis is correlated with fast Pol II transcription is compatible with the theory of kinetic coupling between splicing and transcription (Bentley, 2014). Specifically, TERs are coordinated with different selections of splice sites for alternative splicing, and global changes of Pol II elongation rates lead to the alternative-splicing-mediated downregulation of RNA processing (Braunschweig et al., 2013). Fast transcription allows a selection of an upstream splice site to a far downstream splice site for alternative splicing; however, this selection may not preferentially happen when slow transcription occurs, as the far downstream splice site may be not even transcribed yet (Figure 3A). This could be the case for back splicing regulation as well. In the condition of fast Pol II transcription, a far downstream splice donor site is available to be linked with an upstream splice acceptor site for back splicing. Since pairing of complementary sequences across introns anchoring circularized exons facilitates the process of back splicing (Dong et al., 2017; Liang and Wilusz, 2014; Kramer et al., 2015; Zhang et al., 2014), another advantage of fast Pol II transcription is to enhance non-sequential base-pairing of distal complementary sequences for back splicing (Figure 3B).

Comparison of datasets with prolonged labeling time points showed additional insights into back splicing regulation for circRNA biogenesis. In line with what has been learned from individual validated circRNA examples and canonical steady-state RNA-seq datasets (Guo et al., 2014; Liang et al., 2017; Salzman et al., 2013; Starke et

al., 2015), back splicing is more unfavorably processed than canonical splicing at the nascent RNA-seq level (Zhang et al., 2016b). Indeed, at all examined labeling time points, less RNA-seq fragments were mapped to BSJ sites than those to their neighboring canonical splicing junction sites. Despite this, back splicing events kept increasing to the prolonged 960-min 4sU labeling time point, while canonical splicing reached its equilibrium at the 240-min 4sU labeling time point in PA1 cells (Zhang et al., 2016b). Together with the observation that nascent circRNA-producing genes have been transcribed to their termination sites after 60-min labeling (over 90% and 96%, individually at 60-min or 120-min 4sU labeling time points in PA1 cells), these results suggested that back splicing largely occurs in a post-transcriptional manner after transcription has reached to the termination sites of circRNA-producing genes. The post-transcriptional nature of back splicing is in agreement with the theory that alternative splicing of regulated introns often occurs post-transcriptionally, while splicing of constitutive introns is thought to occur co-transcriptionally (Braunschweig et al., 2013).

Extensive comparison of back splicing from nascent RNA-seq datasets in undifferentiated H9 cells and their differentiated FB neuron progenitor cells further uncovered how the abundant and dynamic expression of circRNAs could be achieved upon neuronal differentiation (Zhang et al., 2016b). Given the positive correlation between back splicing and fast Pol II transcription, highly-expressed circRNAs in neurons are coordinated with enhanced transcription of circRNA-producing genes. Although unfavorably processed at the nascent RNA-seq level, back splicing continues to proceed for the production of circRNAs, even after the transcription of nascent circRNA-producing genes has finished. At the same time, circRNAs are more stable than their cognate linear RNAs, evidenced by the longer half-lives of circRNAs than those of linear ones in PA1 cells treated with actinomycin D to inhibit transcription (Zhang et al., 2016b). In this case, the accumulation of nascent circRNAs from inefficiently-processed back splicing also contributes to their high steady-state accumulation in neurons that have slow division rates (Figure 3C). Taken

together, the process of back splicing from fast-transcribed circRNA-producing genes and their accumulation both lead to the remarkable expression of steady-state circRNAs, in terms of their increased number and up-regulated expression, upon neuronal differentiation and in brain tissues.

Regulation of recursive splicing

Other than back splicing, regulation rules of another type of alternative splicing, recursive splicing, have also been revealed by analyzing the same 4sUDRB-seq datasets (Zhang et al., 2018). With a customized computational pipeline, about 342 recursive splice events were identified genome-wide in human PA1, H9 and FB cell lines, suggesting a temporal regulation landscape of recursive splicing along with Pol II transcription elongation. Among them, 108 events were detected in at least two cell lines, confirming wide-spread of recursive splicing. In addition, 16 events displayed a sawtooth feature of step-wide regulation of recursive splicing at the nascent RNA-seq level, which was further validated by RT-PCR and Sanger sequencing experiments. Statistical analysis of these identified recursive splice events suggested that recursive splicing also largely occurred post-transcriptionally and was coordinated with fast-transcribed genes. These findings provide an additional clue to further understand the diversity and dynamics of RNA alternative splicing.

Conclusions and future perspectives

The broad application of next generation sequencing on RNA (RNA-seq) has revealed complex gene expression regulation at the whole transcriptome level. With distinct enrichment of different RNA subpopulations for deep sequencing, various aspects of gene expression at the transcriptome level have been achieved, including alternative splicing. For example, analyzing polyadenylated transcriptomic RNAs suggests a complex alternative splice map, consisting of most well-known alternative splice events; while profiling transcriptomic RNAs without polyadenylated sequences uncovers un-conventionally processed alternative splice events, such as back splicing for circRNA biogenesis (Chen, 2016) and recursive splicing for long intron removal

(Georgomanolis et al., 2016). With well-designed experiments and comprehensive RNA-seq analyses, features of back splicing regulation have been revealed, including but not limited to the competition between back splicing and canonical (alternative) splicing. However, how and to what extent back splicing competes with splicing have remained unclear from solely analyzing most available steady-state RNA-seq datasets. Meanwhile, since splicing is also coupled with transcription, analyses of steady-state RNA-seq datasets are unable to truly show how back splicing itself is regulated during transcription. To address these questions, specific nascent RNA-seq, 4sUDRB-seq, has been modified and applied to study the kinetics of back splicing regulation along with transcription (Zhang et al., 2016b). This study shows that back splicing preferentially occurs in fast-transcribed genes. In addition, despite being less efficiently processed by the same spliceosomes as canonical splicing, back splicing is continuously catalyzed for circRNA synthesis even after Pol II reaches the end of genes. Meanwhile, circRNAs are also very stable since they are covalently-closed structures. All these contribute to up-regulated steady-state expression of circRNAs in neurons. Of note, extensive analyses of the same 4sUDRB-seq datasets also suggest back splicing from downstream transcripts in polycistronic gene loci (Liang et al., 2017) and recursive splice events with featuring sawtooth patterns (Zhang et al., 2018). It is possible that additional layers of nascent RNA regulation await further investigation.

Other advances in deep sequencing technologies and specific RNA enrichments also enhance studies of nascent RNA-seq. Integrating single-cell RNA-seq (scRNA-seq) and nascent RNA-seq can be used to detect dynamics of nascent RNA synthesis at the single cell level. For example, combining 4sU labeling and scRNA-seq, scSLAM-seq and new transcriptome alkylation-dependent single-cell RNA sequencing (NASC-seq) have been developed to enrich both nascent and total RNAs for profiling (Erhard et al., 2019; Hendriks et al., 2019). Similarly, by conjugating single-cell combinatorial indexing and 4sU labeling techniques (sci-fate), total and newly-synthesized RNAs were examined from more than 6,000 cells, simultaneously (Cao et al., 2020).

Long-read sequencing technologies have also been applied in nascent RNA-seq to obtain transcripts of full length (Drexler et al., 2020; Oesterreich et al., 2016), providing direct evidence of different RNA isoforms and their splicing kinetics. Imaging-based techniques, such as intron sequential fluorescence *in situ* hybridization (intron seqFISH) to track nascent RNAs from 10,421 genes (Shah et al., 2018), have uncovered nascent RNAs and their transcription regulation in live cells (Tutucci et al., 2018; Shah et al., 2018), but these fluorescence-labeled nascent RNAs were not enriched for deep sequencing.

Other than these progresses regarding back splicing regulation during nascent RNA synthesis, many questions concerning circRNA biogenesis and functions remain to be further addressed. For example, since protein factors can regulate circRNA formation and their associated functions (Ashwal-Fluss et al., 2014; Conn et al., 2015; Li et al., 2017; Liu et al., 2019), it will be of great interest in the future to identify how these proteins are involved in back splicing regulation for nascent circRNA synthesis and functions. Meanwhile, the sequence overlap between circRNAs from back splicing and their cognate linear RNA isoforms from canonical splicing also impedes in-depth annotation of functional circRNAs. Improvements in technologies, including direct expression comparison of circular and linear RNAs (Ma et al., 2019) and circRNA-specific repression with novel genome editing toolkits to precisely recognize and target the featured BSJ sites (Li et al., 2020), will be the key to fully address how these circular molecules are produced and what they can truly do in cells.

Compliance and ethics

The author(s) declare that they have no conflict of interest.

Acknowledgements

We are grateful to Gordon G. Carmichael and Ling-Ling Chen for critical reading of this article, and Yang laboratory for discussion. Our work is supported by grants 2019YFA0802804 (L.Y.) from National Key R&D Program of China, 31730111

(L.Y.), 31925011 (L.Y.), 91940306 (L.Y.) and 31801073 (W.X.) from National Natural Science Foundation of China, and the Youth Innovation Promotion Association CAS (W.X.) from Chinese Academy of Sciences.

References

- Aktas, T., Avsar Ilik, I., Maticzka, D., Bhardwaj, V., Pessoa Rodrigues, C., Mittler, G., Manke, T., Backofen, R., and Akhtar, A. (2017). DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. *Nature* 544, 115-119.
- Ardehali, M.B., and Lis, J.T. (2009). Tracking rates of transcription and splicing in vivo. *Nat Struct Mol Biol* 16, 1123-1124.
- Ashwal-Fluss, R., Meyer, M., Pamudurti, N.R., Ivanov, A., Bartok, O., Hanan, M., Evantal, N., Memczak, S., Rajewsky, N., and Kadener, S. (2014). circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell* 56, 55-66.
- Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R., Kim, T., Misquitta-Ali, C.M., Wilson, M.D., Kim, P.M., Odom, D.T., Frey, B.J., and Blencowe, B.J. (2012). The evolutionary landscape of alternative splicing in vertebrate species. *Science* 338, 1587-1593.
- Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. *Nat Rev Genet* 15, 163-175.
- Bhatt, D.M., Pandya-Jones, A., Tong, A.J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* 150, 279-290.
- Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 72, 291-336.
- Braunschweig, U., Gueroussov, S., Plocik, A.M., Graveley, B.R., and Blencowe, B.J. (2013). Dynamic integration of splicing within gene regulatory pathways. *Cell* 152, 1252-1269.
- Cao, J., Zhou, W., Steemers, F., Trapnell, C., and Shendure, J. (2020). Sci-fate characterizes the dynamics of gene expression in single cells. *Nat Biotechnol* 38, 980-988.
- Chen, L.L. (2016). The biogenesis and emerging roles of circular RNAs. *Nat Rev Mol Cell Biol* 17, 205-211.
- Chen, L.L. (2020). The expanding regulatory mechanisms and cellular functions of circular RNAs. *Nat Rev Mol Cell Biol* 21, 475-490.
- Chen, L.L., and Yang, L. (2015). Regulation of circRNA biogenesis. *RNA Biol* 12, 381-388.
- Chen, S., Huang, V., Xu, X., Livingstone, J., Soares, F., Jeon, J., Zeng, Y., Hua, J.T., Petricca, J., Guo, H., Wang, M., Yousif, F., Zhang, Y., Donmez, N., Ahmed, M., Volik, S., Lapuk, A., Chua, M.L.K., Heisler, L.E., Foucal, A., Fox, N.S., Fraser,

- M., Bhandari, V., Shiah, Y.J., Guan, J., Li, J., Orain, M., Picard, V., Hovington, H., Bergeron, A., Lacombe, L., Fradet, Y., Tetu, B., Liu, S., Feng, F., Wu, X., Shao, Y.W., Komor, M.A., Sahinalp, C., Collins, C., Hoogstrate, Y., de Jong, M., Fijneman, R.J.A., Fei, T., Jenster, G., van der Kwast, T., Bristow, R.G., Boutros, P.C., and He, H.H. (2019). Widespread and Functional RNA Circularization in Localized Prostate Cancer. *Cell* 176, 831-843 e822.
- Chu, Q., Zhang, X., Zhu, X., Liu, C., Mao, L., Ye, C., Zhu, Q.H., and Fan, L. (2017). PlantcircBase: A Database for Plant Circular RNAs. *Mol Plant* 10, 1126-1128.
- Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469, 368-373.
- Conn, S.J., Pillman, K.A., Toubia, J., Conn, V.M., Salmanidis, M., Phillips, C.A., Roslan, S., Schreiber, A.W., Gregory, P.A., and Goodall, G.J. (2015). The RNA binding protein quaking regulates formation of circRNAs. *Cell* 160, 1125-1134.
- Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322, 1845-1848.
- Danko, C.G., Hah, N., Luo, X., Martins, A.L., Core, L., Lis, J.T., Siepel, A., and Kraus, W.L. (2013). Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Mol Cell* 50, 212-222.
- Dong, R., Ma, X.K., Chen, L.L., and Yang, L. (2017). Increased complexity of circRNA expression during species evolution. *RNA Biol* 14, 1064-1074.
- Drexler, H.L., Choquet, K., and Churchman, L.S. (2020). Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. *Mol Cell* 77, 985-998 e988.
- Duff, M.O., Olson, S., Wei, X., Garrett, S.C., Osman, A., Bolisetty, M., Plocik, A., Celniker, S.E., and Graveley, B.R. (2015). Genome-wide identification of zero nucleotide recursive splicing in *Drosophila*. *Nature* 521, 376-379.
- Erhard, F., Baptista, M.A.P., Krammer, T., Hennig, T., Lange, M., Arampatzi, P., Jorges, C.S., Theis, F.J., Saliba, A.E., and Dolken, L. (2019). scSLAM-seq reveals core features of transcription dynamics in single cells. *Nature* 571, 419-423.
- Fong, N., Kim, H., Zhou, Y., Ji, X., Qiu, J., Saldi, T., Diener, K., Jones, K., Fu, X.D., and Bentley, D.L. (2014). Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev* 28, 2663-2676.
- Fuchs, G., Voichek, Y., Benjamin, S., Gilad, S., Amit, I., and Oren, M. (2014). 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biol* 15, R69.
- Fuchs, G., Voichek, Y., Rabani, M., Benjamin, S., Gilad, S., Amit, I., and Oren, M. (2015). Simultaneous measurement of genome-wide transcription elongation speeds and rates of RNA polymerase II transition into active elongation with 4sUDRB-seq. *Nat Protoc* 10, 605-618.
- Gao, Y., and Zhao, F. (2018). Computational Strategies for Exploring Circular RNAs. *Trends Genet* 34, 389-400.

- Georgomanolis, T., Sofiadis, K., and Papantonis, A. (2016). Cutting a Long Intron Short: Recursive Splicing and Its Implications. *Front Physiol* 7, 598.
- Gilbert, W. (1978). Why genes in pieces? *Nature* 271, 501.
- Guarnerio, J., Zhang, Y., Cheloni, G., Panella, R., Mae Katon, J., Simpson, M., Matsumoto, A., Papa, A., Loretelli, C., Petri, A., Kauppinen, S., Garbutt, C., Petur Nielsen, G., Deshpande, V., Castillo-Martin, M., Cordon-Cardo, C., Dimitrios, S., Clohessy, J.G., Batish, M., and Pandolfi, P.P. (2019). Intragenic antagonistic roles of protein and circRNA in tumorigenesis. *Cell Res* 29, 628-640.
- Guo, J.U., Agarwal, V., Guo, H., and Bartel, D.P. (2014). Expanded identification and characterization of mammalian circular RNAs. *Genome Biol* 15, 409.
- Hatton, A.R., Subramaniam, V., and Lopez, A.J. (1998). Generation of alternative Ultrabithorax isoforms and stepwise removal of a large intron by resplicing at exon-exon junctions. *Mol Cell* 2, 787-796.
- Hendriks, G.J., Jung, L.A., Larsson, A.J.M., Lidschreiber, M., Andersson Forsman, O., Lidschreiber, K., Cramer, P., and Sandberg, R. (2019). NASC-seq monitors RNA synthesis in single cells. *Nat Commun* 10, 3138.
- Herzog, V.A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T.R., Wlotzka, W., von Haeseler, A., Zuber, J., and Ameres, S.L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. *Nat Methods* 14, 1198-1204.
- Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H.T., Orejuela, M.R., Piechotta, M., Levanon, E.Y., Landthaler, M., Dieterich, C., and Rajewsky, N. (2015). Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep* 10, 170-177.
- Jeck, W.R., and Sharpless, N.E. (2014). Detecting and characterizing circular RNAs. *Nat Biotechnol* 32, 453-461.
- Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* 3, e02407.
- Jordan-Pla, A., Perez-Martinez, M.E., and Perez-Ortin, J.E. (2019). Measuring RNA polymerase activity genome-wide with high-resolution run-on-based methods. *Methods* 159-160, 177-182.
- Kramer, M.C., Liang, D., Tatomer, D.C., Gold, B., March, Z.M., Cherry, S., and Wilusz, J.E. (2015). Combinatorial control of *Drosophila* circular RNA expression by intronic repeats, hnRNPs, and SR proteins. *Genes Dev* 29, 2168-2182.
- Kristensen, L.S., Andersen, M.S., Stagsted, L.V.W., Ebbesen, K.K., Hansen, T.B., and Kjems, J. (2019). The biogenesis, biology and characterization of circular RNAs. *Nat Rev Genet* 20, 675-691.
- Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* 339, 950-953.
- Lee, Y., and Rio, D.C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem* 84, 291-323.

- Li, S., Li, X., Xue, W., Zhang, L., Yang, L.Z., Cao, S.M., Lei, Y.N., Liu, C.X., Guo, S.K., Shan, L., Wu, M., Tao, X., Zhang, J.L., Gao, X., Zhang, J., Wei, J., Li, J., Yang, L., and Chen, L.L. (2020). Screening for functional circular RNAs using the CRISPR-Cas13 system. *Nat Methods*.
- Li, X., Liu, C.X., Xue, W., Zhang, Y., Jiang, S., Yin, Q.F., Wei, J., Yao, R.W., Yang, L., and Chen, L.L. (2017). Coordinated circRNA Biogenesis and Function with NF90/NF110 in Viral Infection. *Mol Cell* 67, 214-227 e217.
- Li, X., Yang, L., and Chen, L.L. (2018). The Biogenesis, Functions, and Challenges of Circular RNAs. *Mol Cell* 71, 428-442.
- Liang, D., Tatomer, D.C., Luo, Z., Wu, H., Yang, L., Chen, L.L., Cherry, S., and Wilusz, J.E. (2017). The Output of Protein-Coding Genes Shifts to Circular RNAs When the Pre-mRNA Processing Machinery Is Limiting. *Mol Cell* 68, 940-954 e943.
- Liang, D., and Wilusz, J.E. (2014). Short intronic repeat sequences facilitate circular RNA production. *Genes Dev* 28, 2233-2247.
- Liu, C.X., Li, X., Nan, F., Jiang, S., Gao, X., Guo, S.K., Xue, W., Cui, Y., Dong, K., Ding, H., Qu, B., Zhou, Z., Shen, N., Yang, L., and Chen, L.L. (2019). Structure and Degradation of Circular RNAs Regulate PKR Activation in Innate Immunity. *Cell* 177, 865-880 e821.
- Liu, Y., Su, H., Zhang, J., Liu, Y., Feng, C., and Han, F. (2020). Back-spliced RNA from retrotransposon binds to centromere and regulates centromeric chromatin loops in maize. *PLoS Biol* 18, e3000582.
- Ma, X.K., Wang, M.R., Liu, C.X., Dong, R., Carmichael, G.G., Chen, L.L., and Yang, L. (2019). CIRCexplorer3: A CLEAR Pipeline for Direct Comparison of Circular and Linear RNA Expression. *Genomics Proteomics Bioinformatics* 17, 511-521.
- Mahat, D.B., Kwak, H., Booth, G.T., Jonkers, I.H., Danko, C.G., Patel, R.K., Waters, C.T., Munson, K., Core, L.J., and Lis, J.T. (2016). Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat Protoc* 11, 1455-1476.
- Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., and Churchman, L.S. (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* 161, 541-554.
- Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* 338, 1593-1599.
- Muhar, M., Ebert, A., Neumann, T., Umkehrer, C., Jude, J., Wieshofer, C., Rescheneder, P., Lipp, J.J., Herzog, V.A., Reichholf, B., Cisneros, D.A., Hoffmann, T., Schlapansky, M.F., Bhat, P., von Haeseler, A., Kocher, T., Obenauf, A.C., Popow, J., Ameres, S.L., and Zuber, J. (2018). SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. *Science* 360, 800-805.

- Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science* 327, 335-338.
- Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463, 457-463.
- Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* 161, 526-540.
- Nojima, T., Rebelo, K., Gomes, T., Grosso, A.R., Proudfoot, N.J., and Carmo-Fonseca, M. (2018). RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. *Mol Cell* 72, 369-379 e364.
- Oesterreich, F.C., Herzel, L., Straube, K., Hujer, K., Howard, J., and Neugebauer, K.M. (2016). Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. *Cell* 165, 372-381.
- Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40, 1413-1415.
- Piwecka, M., Glazar, P., Hernandez-Miranda, L.R., Memczak, S., Wolf, S.A., Rybak-Wolf, A., Filipchyk, A., Klironomos, F., Cerda Jara, C.A., Fenske, P., Trimbuch, T., Zywitzka, V., Plass, M., Schreyer, L., Ayoub, S., Kocks, C., Kuhn, R., Rosenmund, C., Birchmeier, C., and Rajewsky, N. (2017). Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. *Science* 357.
- Rabani, M., Levin, J.Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N., Amit, I., and Regev, A. (2011). Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol* 29, 436-442.
- Rabani, M., Raychowdhury, R., Jovanovic, M., Rooney, M., Stumpo, D.J., Pauli, A., Hacohen, N., Schier, A.F., Blackshear, P.J., Friedman, N., Amit, I., and Regev, A. (2014). High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell* 159, 1698-1710.
- Rybak-Wolf, A., Stottmeister, C., Glazar, P., Jens, M., Pino, N., Giusti, S., Hanan, M., Behm, M., Bartok, O., Ashwal-Fluss, R., Herzog, M., Schreyer, L., Papavasileiou, P., Ivanov, A., Ohman, M., Refojo, D., Kadener, S., and Rajewsky, N. (2015). Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Mol Cell* 58, 870-885.
- Salzman, J., Chen, R.E., Olsen, M.N., Wang, P.L., and Brown, P.O. (2013). Cell-type specific features of circular RNA expression. *PLoS Genet* 9, e1003777.
- Salzman, J., Gawad, C., Wang, P.L., Lacayo, N., and Brown, P.O. (2012). Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* 7, e30733.

- Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., and Proudfoot, N.J. (2017). Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Mol Cell* 65, 25-38.
- Schwalb, B., Michel, M., Zacher, B., Fruhauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. *Science* 352, 1225-1228.
- Shah, S., Takei, Y., Zhou, W., Lubeck, E., Yun, J., Eng, C.L., Koulena, N., Cronin, C., Karp, C., Liaw, E.J., Amin, M., and Cai, L. (2018). Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. *Cell* 174, 363-376 e316.
- Sibley, C.R., Emmett, W., Blazquez, L., Faro, A., Haberman, N., Briese, M., Trabzuni, D., Ryten, M., Weale, M.E., Hardy, J., Modic, M., Curk, T., Wilson, S.W., Plagnol, V., and Ule, J. (2015). Recursive splicing in long vertebrate genes. *Nature* 521, 371-375.
- Singh, J., and Padgett, R.A. (2009). Rates of in situ transcription and splicing in large human genes. *Nat Struct Mol Biol* 16, 1128-1133.
- Starke, S., Jost, I., Rossbach, O., Schneider, T., Schreiner, S., Hung, L.H., and Bindereif, A. (2015). Exon circularization requires canonical splice signals. *Cell Rep* 10, 103-111.
- Tan, J., Ho, J.X., Zhong, Z., Luo, S., Chen, G., and Roca, X. (2016). Noncanonical registers and base pairs in human 5' splice-site selection. *Nucleic Acids Res* 44, 3908-3921.
- Tutucci, E., Vera, M., Biswas, J., Garcia, J., Parker, R., and Singer, R.H. (2018). An improved MS2 system for accurate reporting of the mRNA life cycle. *Nat Methods* 15, 81-89.
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470-476.
- Wilusz, J.E. (2018). A 360 degrees view of circular RNAs: From biogenesis to functions. *Wiley Interdiscip Rev RNA* 9, e1478.
- Wissink, E.M., Vihervaara, A., Tippens, N.D., and Lis, J.T. (2019). Nascent RNA analyses: tracking transcription and its regulation. *Nat Rev Genet* 20, 705-723.
- Yang, L., Duff, M.O., Graveley, B.R., Carmichael, G.G., and Chen, L.L. (2011). Genomewide characterization of non-polyadenylated RNAs. *Genome Biol* 12, R16.
- Yin, Q.F., Chen, L.L., and Yang, L. (2015). Fractionation of non-polyadenylated and ribosomal-free RNAs from mammalian cells. *Methods Mol Biol* 1206, 69-80.
- You, X., Vlatkovic, I., Babic, A., Will, T., Epstein, I., Tushev, G., Akbalik, G., Wang, M., Glock, C., Quedenau, C., Wang, X., Hou, J., Liu, H., Sun, W., Sambandan, S., Chen, T., Schuman, E.M., and Chen, W. (2015). Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat Neurosci* 18, 603-610.
- Zhang, P., Zhang, X.O., Jiang, T., Cai, L., Huang, X., Liu, Q., Li, D., Lu, A., Liu, Y., Xue, W., Zhang, P., and Weng, Z. (2020). Comprehensive identification of

- alternative back-splicing in human tissue transcriptomes. *Nucleic Acids Res* 48, 1779-1789.
- Zhang, X.O., Dong, R., Zhang, Y., Zhang, J.L., Luo, Z., Zhang, J., Chen, L.L., and Yang, L. (2016a). Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res* 26, 1277-1287.
- Zhang, X.O., Fu, Y., Mou, H., Xue, W., and Weng, Z. (2018). The temporal landscape of recursive splicing during Pol II transcription elongation in human cells. *PLoS Genet* 14, e1007579.
- Zhang, X.O., Wang, H.B., Zhang, Y., Lu, X., Chen, L.L., and Yang, L. (2014). Complementary sequence-mediated exon circularization. *Cell* 159, 134-147.
- Zhang, Y., Xue, W., Li, X., Zhang, J., Chen, S., Zhang, J.L., Yang, L., and Chen, L.L. (2016b). The Biogenesis of Nascent Circular RNAs. *Cell Rep* 15, 611-624.

Figures and Figure Legends

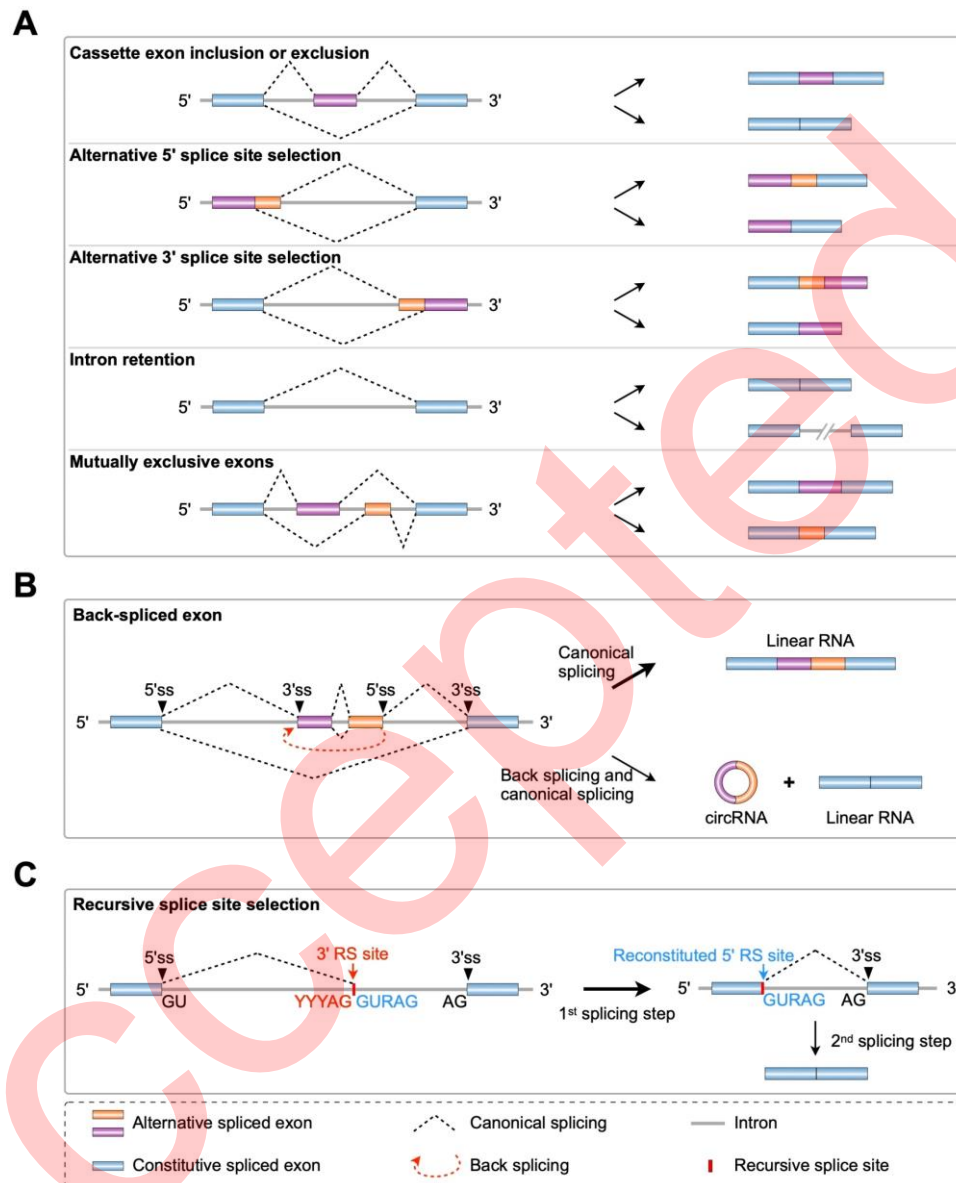


Figure 1 Types of complex alternative splice events.

A, Types of commonly-known alternative splice events. Five types of well-characterized alternative splice events, including cassette exon inclusion or exclusion, alternative 5' splice site selection, alternative 3' splice site selection, intron retention and mutually exclusive exons.

B, Schematic drawing of non-canonical back splicing for circular RNA (circRNA) biogenesis. In addition to canonical splicing to generate a linear RNA (top right), a precursor (m)RNA can also be processed by back splicing to produce a circRNA from back-spliced exon(s) together with an alternatively spliced linear RNA with exon exclusion (bottom right). See main text for details.

C, Schematic drawing of non-canonical recursive splice site selection to remove long introns. Multiple steps are involved in recursive splicing. Basically, an upstream canonical 5' splice site is spliced to a downstream 3' recursive splice (RS) site, which also generates a new 5' RS site. This reconstituted 5' RS site is then spliced to a downstream canonical 3' splice site.

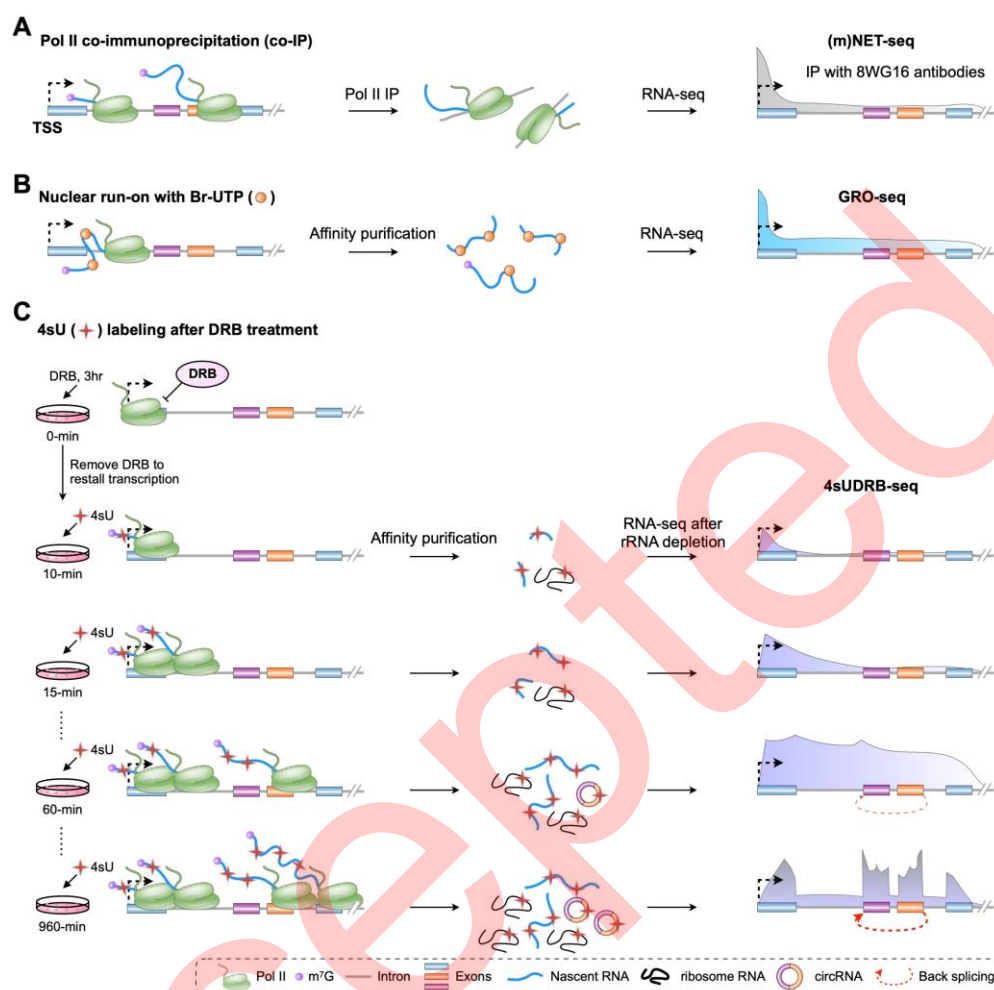


Figure 2 Different nascent RNA-seq technologies.

A, Schematic drawing of immunoprecipitation-based nascent RNA enrichment for deep sequencing. Nascent RNAs that are transcribed from DNA templates can be co-immunoprecipitated with chromatin or Pol II for subsequent RNA-seq, referred to as caRNA-seq and (m)NET-seq, respectively. Different Pol II antibodies can be used in (m)NET-seq to show distinct distribution of specific Pol II for nascent RNA biogenesis. For example, with 8WG16 antibody against unphosphorylated CTD of Pol II, (m)NET-seq signals (grey) are likely enriched around the transcription start sites (TSSs). See main text for details.

B, Schematic drawing of nuclear run-on based nascent RNA enrichment for deep sequencing. In nuclear run-on (NRO) assays, nascent RNA synthesis is processed in isolated nuclei by adding labeled NTP, such as Br-UTP for GRO-seq. After affinity purification, labeled nascent RNAs are enriched for deep sequencing. With short labeling time, GRO-seq signals (light blue) mainly accumulate near TSSs. See main text for details.

C, Schematic drawing of 4sU labeling based nascent RNA enrichment for deep sequencing. Briefly, cells are treated with DRB for 3 hr to block transcription. After DRB removal and the addition of uridine analog 4sU, transcription is restored to produce 4sU-labeled nascent RNAs. Different sets of 4sU-labeled nascent RNAs with a series of labeling time can be purified for deep sequencing. Two major improvements, including prolonged 4sU labeling time (up to 960 min) and an additional rRNA-depletion step, have been applied in preparing 4sUDRB-seq datasets to facilitate back splicing analysis. With short 4sU labeling times, 4sUDRB-seq signals (light purple)

suggest nascent RNA transcription from TSSs. With prolonged 4sU labeling times (such as 960 min, bottom), many more 4sUDRB-seq signals for back splicing can be detected than those with short labeling times (such as 10, 15 and even 60 min, top), suggesting the post-transcriptional regulation of back-splicing. See main text for details.

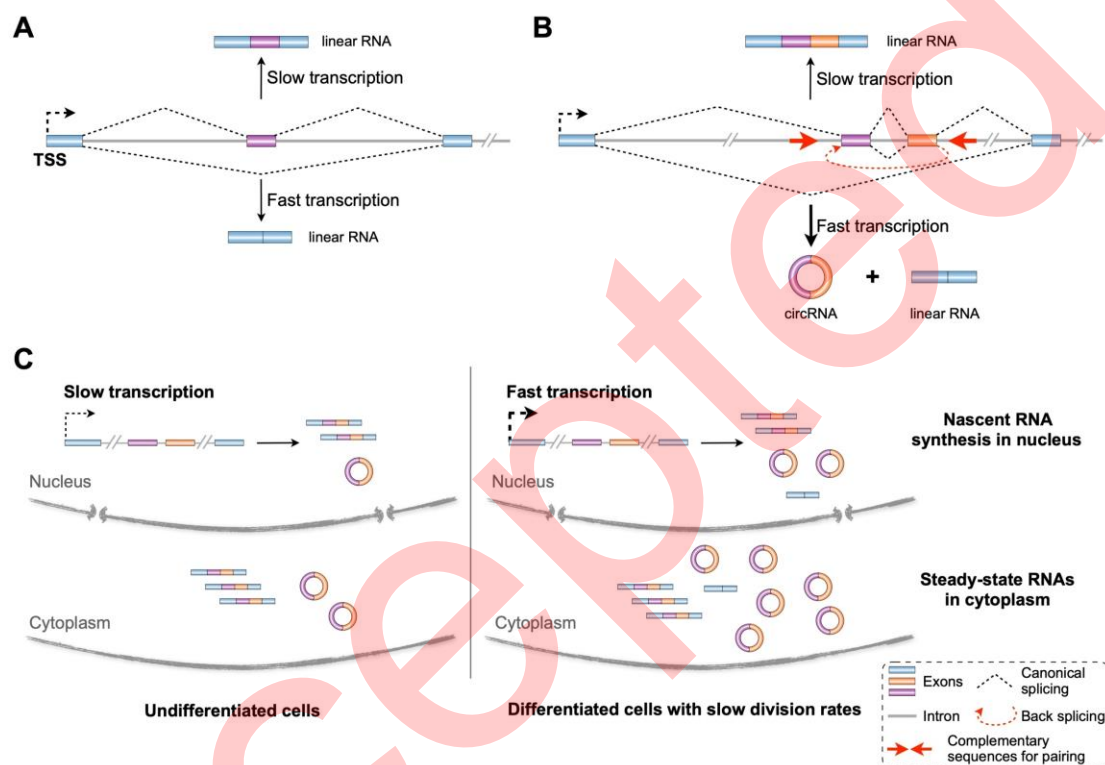


Figure 3 Features of back splicing regulation.

A, Coupling of transcription with alternative splicing. In the condition of slow transcription, pre-RNAs usually undergo splice site selection at nearby exons for cassette exon (purple) inclusion (top). Differently, fast transcription facilitates alternative splicing for cassette-exon exclusion.

B, Coupling of back splicing with fast Pol II transcription. Other than linear RNAs produced by canonical splicing (top), pre-RNA can be processed by back-splicing to produce circRNAs, especially in the condition of fast Pol II transcription (bottom). It is predicted that Pol II transcription might enhance the pairing of distal complementary sequences (red arrows) flanking circularized exons for back-splicing. See main text for details.

C, At the nascent RNA level, fast transcription of circRNA-producing genes facilitates back splicing in the nucleus for nascent circRNA biogenesis (top). Meanwhile, due to their high stability, after synthesis, circRNAs can also accumulate to high levels (bottom), especially in differentiated cells with slow division rates, such as neurons.