Evolutionary Insights into RNA trans-Splicing in Vertebrates

Quan Lei[1], Cong Li[1], Zhixiang Zuo[1], Chunhua Huang[2], Hanhua Cheng[2,*], and Rongjia Zhou[1,*]

[1]Department of Genetics, College of Life Sciences, Wuhan University, P.R. China

[2]Department of Cell Biology, College of Life Sciences, Wuhan University, P.R. China

*Corresponding author: E-mail: rjzhou@whu.edu.cn; hhcheng@whu.edu.cn.

Accepted: February 8, 2016

Abstract

Pre-RNA splicing is an essential step in generating mature mRNA. RNA trans-splicing combines two separate pre-mRNA molecules to form a chimeric non-co-linear RNA, which may exert a function distinct from its original molecules. Trans-spliced RNAs may encode novel proteins or serve as noncoding or regulatory RNAs. These novel RNAs not only increase the complexity of the proteome but also provide new regulatory mechanisms for gene expression. An increasing amount of evidence indicates that trans-splicing occurs frequently in both physiological and pathological processes. In addition, mRNA reprogramming based on trans-splicing has been successfully applied in RNA-based therapies for human genetic diseases. Nevertheless, clarifying the extent and evolution of trans-splicing in vertebrates and developing detection methods for trans-splicing remain challenging. In this review, we summarize previous research, highlight recent advances in trans-splicing, and discuss possible splicing mechanisms and functions from an evolutionary viewpoint.

Key words: trans-splicing, functions, evolution, RNAs, vertebrates.

Introduction

To create fully functional mRNA, pre-mRNA is processed into mature mRNA through three main modifications: 5′-capping, 3′-polyadenylation and RNA splicing. The last modification includes cis- and trans-splicing. Cis-splicing occurs within same pre-mRNA molecule, whereas trans-splicing uses two separate pre-mRNA molecules to form a chimeric non-co-linear RNA, which may encode novel proteins or serve as noncoding or regulatory RNAs. These novel RNAs not only increase the complexity of the proteome but also provide new regulatory mechanisms for gene activities, all of which extend the coding capacity of a genome and shape speciation.

Trans-splicing was first observed in the RNA processes in trypanosomes in which a short leader sequence is transferred to the 5′-end of the pre-mRNA for variant surface glycoprotein (Boothroyd et al. 1982; Van der Ploeg et al. 1982). A 22-nt spliced leader sequence (SL) was also found at the 5′-end of actin mRNA in *Caenorhabditis elegans* (Krause et al. 1987). In SL trans-splicing, a short noncoding exon is spliced to the 5′-end of mRNAs for distinct structural genes, producing mRNAs with a common leading sequence (Van Doren et al. 1988). SL trans-splicing is mediated by the spliceosome, including snRNAs (small nuclear RNAs) U2, U4, U5, and U6, but not U1 (Hannon et al. 1991). In lower eukaryotes, SL trans-splicing plays a pivotal role in mature mRNA processing (Hastings 2005), especially of polycistronic transcription units (Nilsen 1993). In addition, trans-splicing is involved in growth recovery in *C. elegans* (Zaslaver et al. 2011) and in nutrient-dependent translational control in the marine chordate *Oikopleura dioica*. (Danks et al. 2015). Through these functions, SL trans-splicing provides evolutionary advantages to prokaryotes.

Trans-splicing can occur in both prokaryotes and eukaryotes. In some archaea and bacteria, trans-splicing events probably split tRNA genes, implying an evolutionary trace of continuous tRNA genes among different species. In *Drosophila*, the biological significance of trans-splicing in *mod (mdg4)* and *lola* genes has been examined (Dorn et al. 2001; Horiuchi et al. 2003; McManus et al. 2010). Recently, many trans-splicing events have also been detected in vertebrates through high-throughput RNA analysis techniques (Herai et al. 2010; Frenkel-Morgenstern et al. 2013). In addition, in mammals, trans-splicing has been observed in many physiological and pathological processes including cancer (Yu et al. 1999; Li, Wang, et al. 2009). However, the occurrence, extent and implications of trans-splicing could be quite different in vertebrates compared with invertebrates. There are at least two concerns. First, the extent and scope of trans-splicing

© The Author(s) 2016. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

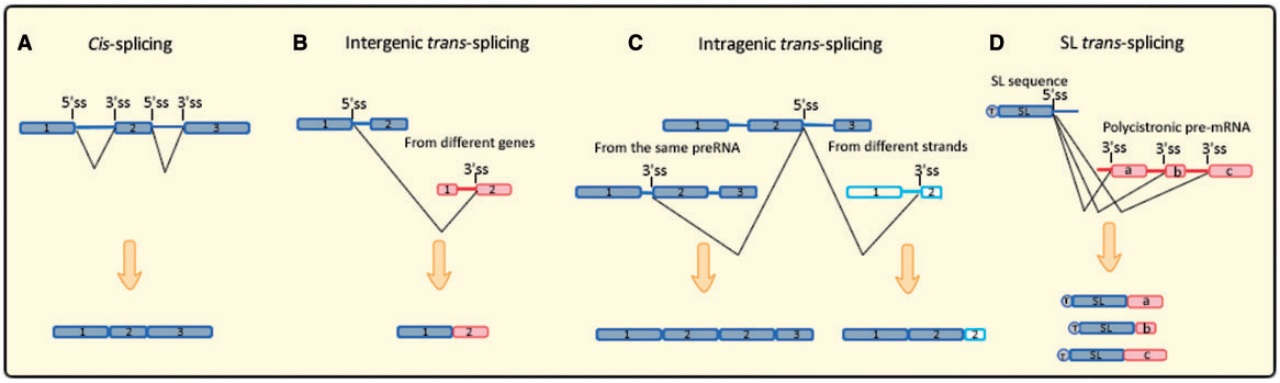

Fig. 1.—Schematic diagram of different types of pre-RNA splicing events. (*A*) Cis-splicing. After excision of introns, exons of the same pre-mRNA are joined together to form a linear molecule. (*B*) Intergenic trans-splicing. Transcripts from different genes or even different chromosomes could be spliced and generate a non-linear chimeric molecule. (*C*) Intragenic trans-splicing. Boxes with vertical line represent exons transcribed from the other strand. In the same gene, splicing reaction occurs between two identical transcripts, alternatively, transcripts from different strands leading to exon-duplication and sense–antisense fusion. (*D*) SL trans-splicing. Red boxes represent structural genes, while T represents for the TMG cap on Spliced-leader (SL) mini-exon. SL exon produced from tandem repeated SL gene cluster, splicing reaction occurs between SL exon and distinct structural genes of a ploycistronic pre-mRNA to generate an array of mature "capped" transcripts.

in vertebrates may be lower compared with invertebrates. Another intriguing question concerns the detection methodologies for trans-splicing, with the key challenge determining whether reverse transcription from RNA to cDNA using reverse transcriptase (RT) may result in artificial chimeras.

Trans-splicing plays important roles in many physiological and pathological processes, although it occurs at a low frequency in humans. The principle has successfully been applied in RNA-based therapy in human genetic diseases (Wally et al. 2012). However, the functions, evolution, and underlying mechanisms of trans-splicing remain unknown. In this review, we systematically discuss trans-splicing with a focus on its extent, functions, and mechanisms in vertebrates from an evolutionary viewpoint.

Pre-RNA Splicing Types: cis- and trans-Splicing

After transcription, the majority of pre-RNAs are processed through splicing to become a mature RNA. There are two types of splicing: cis- and trans-splicing. Trans-splicing involves two pre-RNA molecules, whereas cis-splicing occurs within a single pre-mRNA (fig. 1*A*). There are two types of trans-splicing based on the pre-RNA source: intragenic trans-splicing and intergenic trans-splicing (fig. 1*B* and *C*). In the intragenic subgroup, pre-RNAs are transcribed from the same genome locus, but chimeric RNA is spliced from different strands or exon order. Intragenic trans-splicing may occur through exon repetition, sense-antisense fusion, and exon scrambling. For example, chimeric RNAs of the *mod (mdg4)* and *lola* genes in *Drosophila* originate from intragenic trans-splicing (Dorn et al. 2001; Horiuchi et al. 2003; McManus et al. 2010). In intergenic trans-splicing, exons from diverse genes, even those on different chromosomes, are used to generate a chimeric RNA. For example, in humans, transcripts from the *JAZF1* gene on chromosome 7p15 and the *JJAZ1* gene on chromosome 17q11 can generate a chimeric *JAZF1-JJAZ1* RNA (Li et al. 2008). SL splicing is a special type of trans-splicing that frequently occurs in lower eukaryotes such as nematodes (Nilsen 1993) (fig. 1*D*).

Evolutionary Trends of trans-Splicing

Trans-splicing frequently occurs in lower organisms, such as dinoflagellates (e.g., *Karlodinium micrum*), euglenozoa (e.g., *Trypanosoma brucei*), and some species of nematodes (e.g., *C. elegans*), with more than 70% of genes participating in the process. Trans-splicing even occurs in viruses such as bacteriophage T4, demonstrating an early origin (Galloway Salvo et al. 1990). In archaea, tRNA generation could occur through trans-splicing, for example, in *Thermosphaera aggregans* (Chan et al. 2011) and *Nanoarchaeum equitans* (Randau et al. 2005). Split tRNAs were found in some archaea species (Randau et al. 2005; Fujishima et al. 2009; Chan et al. 2011), and tRNA half homologs were detected in the genomes of archaea, bacteria, and eukaryotes (Zuo et al. 2013). Recent studies indicate that small guide RNA could be involved in tRNA splicing (Randau 2015). Thus, some split tRNAs are proposed to be transcribed from different loci and trans-spliced to generate mature tRNAs. Intriguingly, SL splicing has a much higher frequency, up to approximately 100% compared with other types, including inter/intragenic splicing events, which are observed mainly in dinoflagellates (e.g., 100% in *Amphidinium carterae* and *K. micrum*), euglenozoa (e.g.,

100% in *T. brucei*), nematodes (e.g., 90% in *Ascaris* sp. and 70% in *C. elegans*), and rotifers (e.g., 60% in *Adineta ricciae*). In addition, in a recent mega-data study, a total of 1,627 trans-splicing events involving 2,199 genes were identified in insects, which accounts for 1.58% of the total genes (Kong et al. 2015). This finding, together with many other studies, provides new evidence against the hypothesis that trans-splicing events are merely 'splicing noise.'

Trans-splicing frequency peaks in protozoa, radiate, and protostomia and then decreases, with a dramatic decline in vertebrates (fig. 2). The high percentage of trans-splicing events observed in invertebrates represents SL-type splicing, which can occur in 100% of genes in *A. carterae*, *K. micrum*, and *T. brucei*. The valleys in the percentage of trans-splicing events indicate non-SL species, mainly vertebrates. The analyses imply that along the evolutionary process, trans-splicing has experienced dynamic changes.

Conservation of Splicing Machinery

We believe that trans-splicing shares most characteristics with cis-splicing. Several lines of evidence have shown that trans-splicing utilizes a similar set of splicing machinery to alternative splicing. Trans-splicing has the same splicing signals and factors as alternative splicing. For example, the spliceosome, which contains U1, U2, U4, U5, and U6 snRNAs, catalyzes pre-mRNA in cis-splicing. A recent study demonstrated that U1 snRNP binding may promote *mod* trans-splicing in *Drosophila* (Gao et al. 2015). For SL splicing, the key elements of SL snRNP are very similar to the spliceosomal snRNPs (Bruzik et al. 1988; Van Doren et al. 1988), indicating that SL RNA may originate from a splicing U snRNA in lower organisms with an ancestral cis-splicing mechanism. Additional support comes from the fact that SL trans-splicing exists in some metazoans, including cnidarians, ctenophores, rotifers, flatworms, nematodes, crustaceans, sponges, and chaetognaths. In contrast, plants, fungi, insects, most protists, and vertebrates do not exhibit SL trans-splicing (Douris et al. 2010). However, simultaneous trans-splicing events could take place between SL RNA and inherent transcripts in HeLa cells both in vivo and in vitro (Bruzik et al. 1992). In addition, SL trans-splicing is favored in adenosine-rich 5′ -UTRs in hydrozoans (Derelle et al. 2010). In vertebrates, the 5′ -UTR can be involved in the generation of some trans-spliced mRNA chimeras (Li et al. 1999), which is similar to SL trans-splicing in invertebrates.

Alternative splicing has contributed much more to proteome diversity than trans-splicing. It is noteworthy that multiple protein factors and substantial energy are required in alternative splicing, which is not used by lower organisms, for example, prokaryotes. A recent study indicated that SL trans-splicing provides an evolutionary advantage for species that depend on translational control to regulate early embryogenesis, growth, and oocyte production in response to nutrient levels (Danks et al. 2015). Overall, we suggest that the biological significance of trans-splicing especially SL trans-splicing may be more vital for lower eukaryotes than vertebrates. In higher evolutionary phyla, the more complicated genome structure needs an adaptable regulatory mechanism using preexisting machineries to promote the evolutionary shift from trans-splicing to alternative splicing. In vertebrates, these trans-splicing events focus mainly on some key physiological processes, including gene expression regulation for cell viability and growth. Moreover, dysregulation of trans-splicing could induce pathological events such as cancer (Li et al. 2014).

The spliceosome consists of proteins and U snRNAs complexes that participate in the splicing process. Serine/arginine-rich proteins (SR proteins) and heterogeneous nuclear ribonucleoproteins (hnRNPs) are two of the involved protein families that join in complex A to participate in cross-exon assembly by regulating U1 and U2 snRNP binding to the prespliceosome (Wahl et al. 2015). Both SR proteins and hnRNPs contain RNA-binding domains that can bind to exonic/intronic splicing enhancer (ESE/ISE) sequences and ESS/ISS sequences on pre-mRNAs, respectively (Zhu et al. 2001). Moreover, the SR protein is a key factor for alternative splicing (Gupta et al. 2014); it is intriguing that it can also enhance the efficiency of trans-splicing (Bruzik et al. 1995; Shao et al. 2012). Given these data, we analyzed the evolutionary conservation of spliceosome-associated proteins hnRNPA1, hnRNPI (PTBP1), SRSF1, and SRSF2 from *T. brucei*, *C. elegans*, insects, fish, chicken, and mammals (fig. 3). Although the SL type of trans-splicing is obviously separated from the non-SL type, both cis- and trans-splicing can utilize the same set of splicing factors. Thus, the splicing mechanisms of cis- and trans-splicing seem to be evolutionarily conserved.

From SL1, SL2 to Non-SL trans-Splicing

Although vertebrates and invertebrates exhibit similarities in trans-splicing, there are some distinct features, indicating that trans-splicing is evolutionarily dynamic. In most unicellular organisms and nematodes, SL trans-splicing is an exclusive splicing mode, while no SL trans-splicing occurs in vertebrates. Trans-splicing in vertebrates shows more complexities in magnitude and regulation compared with invertebrates. One example is the donor-acceptor sequence diversity in vertebrates. In mice, the donor-acceptor sequences of the first intron in the *Msh4* β and ϵ pre-mRNAs are "TG-GT" and "TC-CA," respectively, which do not match the consensus sequence of U2-type (GT-AG) or U12-type (AT-AC) (Hirano et al. 2004). In addition, SL trans-splicing does not increase the complexity of proteomes, whereas trans-splicing in vertebrates does, suggesting that this process may generate proteins with functions that differ from those of the parental genes in vertebrates. These observations imply that trans-splicing may have evolved, probably from a broad SL splicing to a precise regulation of

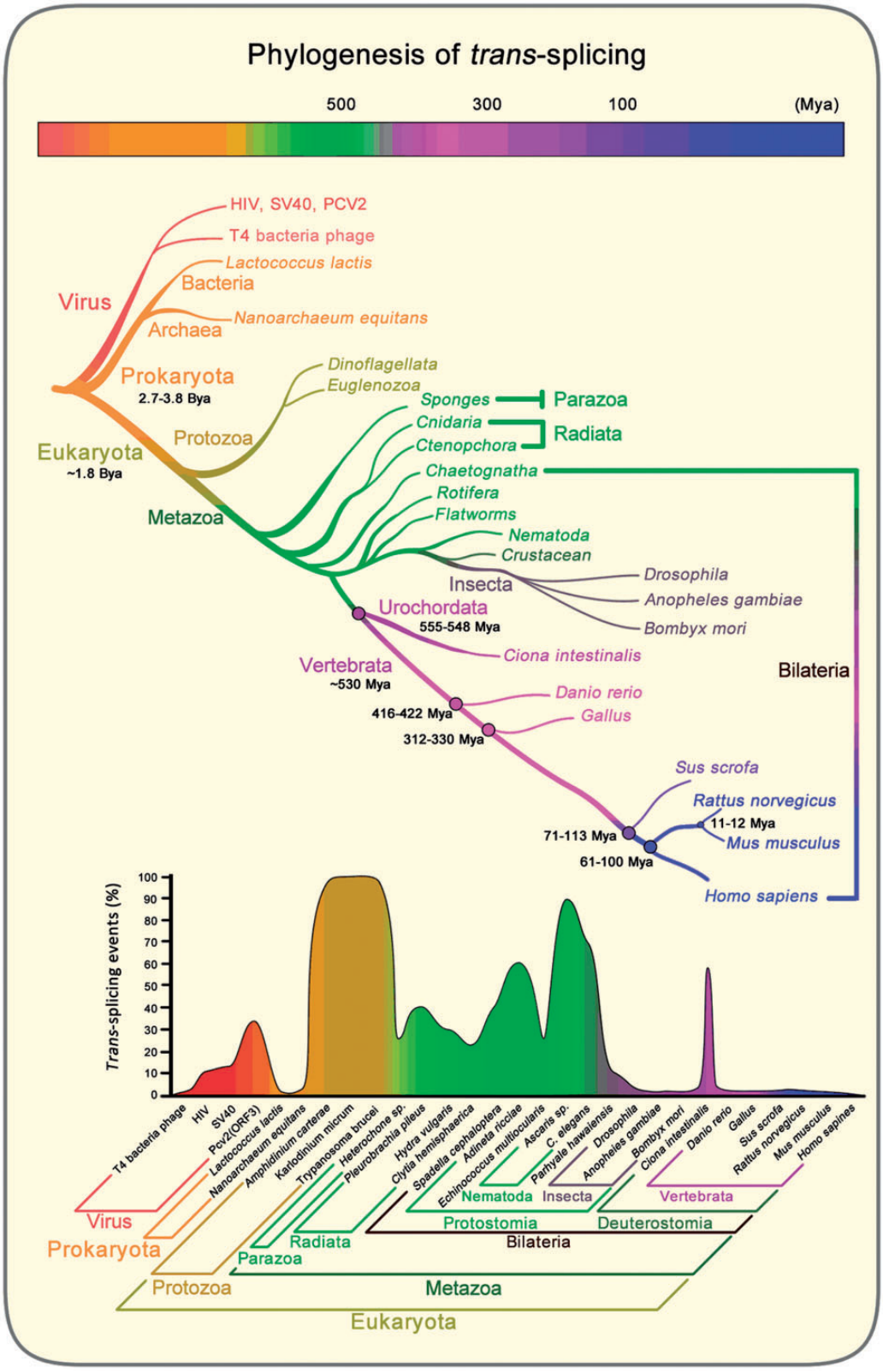

Fig. 2.—Phylogenetic analysis of trans-splicing events. Evolutionary tree and time scale refer to Benton et al. (2007). Ba, billion years ago; Ma, million years ago. In low panel, the percentage of trans-splicing events and trans-spliced gene numbers are relative to the total amount of gene numbers for a species. Trans-splicing data are from published literatures: *Parhyale hawaiensis* (Douris et al. 2010), *Clytia hemisphaerica*

(continued)

non-SL trans-splicing, which matches the complexity of gene regulation in vertebrates.

SL trans-splicing has evolved continually in nematodes. Two types of SL trans-splicing have been identified, namely SL1 and SL2 (Harrison et al. 2010). SL2 RNAs appeared only in the Rhabditina clade of nematodes, including *C. elegans*, which indicates that the SL2 RNAs evolved relatively late in nematode evolution (Guiliano et al. 2006). SL trans-splicing is associated with the evolution of operons (Blumenthal 2005). Operons evolved before SL2-like spliced leaders; nematodes can use trans-splicing to resolve their operonic transcripts into single-gene mRNAs (Guiliano et al. 2006). SL2 in *C. elegans* evolved from SL1; notably, in processing operon pre-mRNAs, SL2 is much more efficient (Blumenthal 2005). Nevertheless, vertebrates lack SL trans-splicing and instead form non-SL type trans-splicing. SL trans-splicing occurs in the organisms that utilize operons, and there are rarely operons in vertebrates. In fact, SL trans-splicing arose after operons. Although the evolutionary mechanisms of trans-splicing remain unknown, we speculate that along with loss of operons and formation of genome complexity, trans-splicing may have shifted to cis-splicing. Thus, trans-splicing is reserved and limited to some essential processes in vertebrates but at a much lower frequency.

Trans-Splicing in Vertebrates

In mammals, the first characterized case of trans-splicing was a novel small T antigen transcript in HeLa cells (Eul et al. 1995). To date, trans-splicing has been observed in various species, including *Danio rerio, Gallus, Rattus norvegicus, Mus musculus, Sus scrofa*, and *Homo sapiens* (Vellard et al. 1991; Caudevilla et al. 1998 ; Hirano et al. 2004; Cadieux et al. 2005; Li et al. 2008; Ma et al. 2012). Trans-splicing occurs in genes involved in some physiological processes (table 1), which expands our understanding of the repertoire of genes and their regulation.

Functions of Chimeric Transcripts in Vertebrates

Chimeric RNA is abundant in both normal and cancer tissues (Romani et al. 2003; Frenkel-Morgenstern et al. 2013). Chimeric RNA may be produced not only by trans-splicing but also by cis-splicing of adjacent genes (Zhang et al. 2012), chromosomal translocation (Mori et al. 2002), and cotranscription across neighboring loci (Magrangeas et al. 1998; Communi et al. 2001). Trans-spliced chimeric RNA was previously ignored as the byproduct of aberrant transcription or "splicing noise" due to its rarity (Maniatis et al. 2002; Tasic et al. 2002), but now it seems to be a "hidden" component of the genome. Evidence suggests that trans-splicing generates an additional layer of genome complexity (Gingeras 2009; Kowarz et al. 2012). These chimeric RNAs are engaged in a versatile range of physiological processes as either protein-coding or noncoding RNAs. Here, we summarize the functions of trans-spliced chimeric RNAs in vertebrates.

Trans-Splicing and Cancer

Although some trans-spliced chimeric RNAs are associated with cancers (Guerra et al. 2008; Kowarz et al. 2011), the causal relationship between trans-splicing and cancer remains unclear. *JAZF1-JJAZ1* in cancer cells is derived from chromosomal translocation. However, it was also detected in normal endometrial stromal cells, indicating that the chimeric RNA is trans-spliced in normal cells (Li et al. 2008; Li,Wang, et al. 2009). The same situation is also found in the chimeric RNA *PAX3-FOXO1* (Yuan, Qin, et al. 2013). In addition, intermolecular recombination events are involved in the tissue-specific expression of the *C-myb* proto-oncogene (Vellard et al. 1991). In human prostate cancer, most partner genes involved in chimeric RNAs have a low expression level (Kannan et al. 2011).

It has been postulated that a trans-spliced RNA molecule may serve as a scaffold to facilitate genomic interactions,

Fig. 2.—Continued

(Derelle et al. 2010), *Echinococcus multilocularis* (Brehm et al. 2000), *Heterochone* sp. (Douris et al. 2010), *Hydra vulgaris* (Stover et al. 2001), *Pleurobrachia pileus* (Derelle et al. 2010), *Spadella cephaloptera* (Marletaz et al. 2008; Marletaz and Le Parco 2008), *Ciona intestinalis* (Vandenberghe et al 2001; Satou et al. 2006; Satou et al. 2008; Matsumoto et al. 2010), *Adineta ricciae* (Pouchkina-Stantcheva et al. 2005), *C. elegans* (Krause et al. 1987; Huang et al. 1989; Zorio et al. 1994), *Ascaris sp.* (Nilsen et al. 1989; Maroney et al. 1995), *Trypanosoma brucei* (Murphy et al. 1986; Sutton et al. 1986; Perry et al. 1987; Liang et al. 2003), *Amphidinium carterae* (Bachvaroff et al. 2008; Zhang et al. 2009) and *Karlodinium micrum* (Zhang et al. 2007). In the other species, the percentage of trans-splicing events are calculated by counting known trans-splicing molecules including T4 bacteria phage (Galloway Salvo et al. 1990), HIV (Caudevilla, Da Silva-Azevedo, et al. 2001), SV40 (Caudevilla, Da Silva-Azevedo, et al. 2001), Pv2 (ORF3) (Gao et al. 2013), *Lactococcus lactis* (Belhocine et al. 2007), *Nanoarchaeum equitans* (Randau et al. 2005), *Drosophila* (Dorn et al. 2001 Horiuch et al. 2003), *Anopheles gambiae* (Robertson et al. 2007), *Bombyx mori* (Shao et al. 2012; Duan et al. 2013), *Danio rerio* (Cadieux et al. 2005), *Gallus* (Vellard et al. 1991), *Sus scrofa* (Ma et al. 2012), *Rattus norvegicus* (Sullivan et al. 1991; Caudevilla et al. 1998; Akopian et al. 1999; Takahara et al. 2002; Zhang et al. 2003; Fitzgerald et al. 2006; Ni et al. 2011), *Mus musculus* (Hirano et al. 2004; Zhang et al. 2010) and *Homo sapiens* (Vellard et al. 1991; Breen et al. 1997; Yu et al. 1999; Chatterjee et al. 2000; Takahara et al. 2000; Finta et al. 2002; Flouriot et al. 2002; Jehan et al. 2007; Guerra et al. 2008; Li et al. 2008; Brooks et al. 2009; Kannan et al. 2011; Kowarz et al. 2011, 2012; Fang et al. 2012; Hu et al. 2013; Kawakami et al. 2013; Yuan, Qin, et al. 2013; Li et al. 2014; Wu et al. 2014). Percentages in insects are consistent with recent mega-data study (Kong et al. 2015).

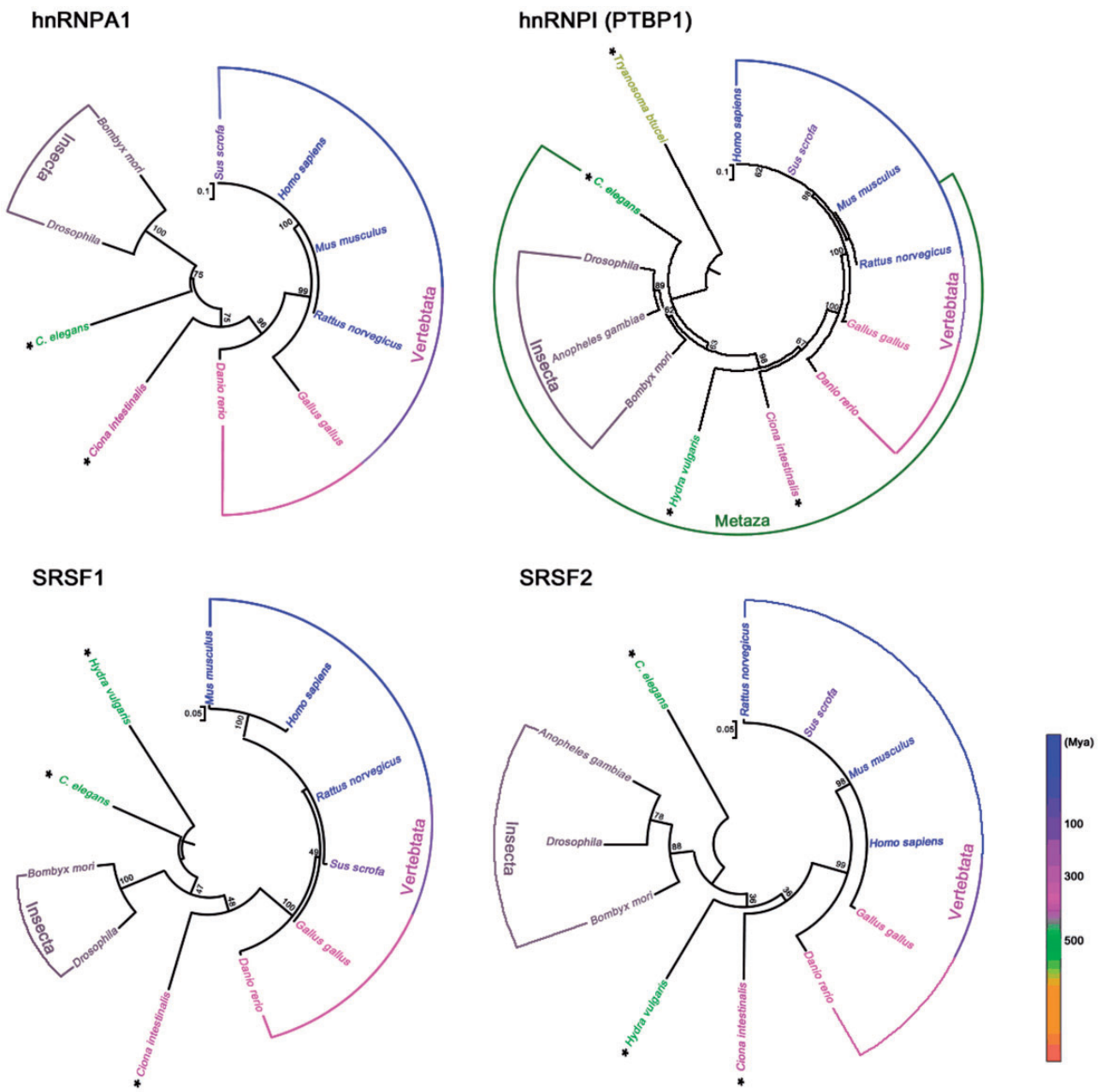

Fig. 3.—Phylogenetic analysis of spliceosome-associated proteins hnRNPA1, hnRNPI (PTBP1), SRSF1, and SRSF2. Ma, million years ago. Species with SL trans-splicing are marked with asterisks. Phylogenetic analysis was performed with MEGA 6 using maximum likelihood method. Numbers on the branches represent the bootstrap values from 1,000 replicates obtained. The scale bar corresponds to the estimated evolutionary distance units. GenBank accession numbers are as follows: *Homo sapiens*, NP_002127.1 (hnRNPA1), NP_002810.1 (PTBP1), NP_001071634.1 (SRSF1), NP_001182356.1 (SRSF2); *Mus musculus*, NP_001034218.1 (hnRNPA1), NP_001070831.1 (PTBP1), NP_001071635.1 (SRSF1); NP_035488.1 (SRSF2); *Rattus norvegicus*, NP_058944.1 (hnRNPA1), NP_001257986.1 (PTBP1), NP_001103022.1, (SRSF1), NP_001009720.1 (SRSF2); *Sus scrofa*, NP_001070686.1 (hnRNPA1), NP_999396.1 (PTBP1), NP_001033096.1 (SRSF1), NP_001070697.1 (SRSF2); *Gallus gallus*, XP_004950342.1 (hnRNPA1), NP_001026106.1 (PTBP1), NP_001107213.1 (SRSF1), NP_001001305.1 (SRSF2); *Danio rerio*, NP_956398.1 (hnRNPA1), NP_001116126.1 (PTBP1), NP_956887.2 (SRSF1), NP_998547.1 (SRSF2); *Drosophila*, NP_001262538.1 (hnRNPA1), NP_001097994.1 (PTBP1), NP_001247139.1 (SRSF1), NP_001188794.1 (SRSF2); *Bombyx mori*, NP_001093319.1 (hnRNPA1), XP_012546585.1 (PTBP1), XP_012548197.1 (SRSF1), NP_001040152.1 (SRSF2); *C. elegans*, NP_500326.2 (hnRNPA1), NP_741041.1 (PTBP1), NP_499649.2 (SRSF1), NP_495013.1 (SRSF2); *Ciona intestinalis*, XP_002128542.1 (hnRNPA1), XP_002127727.3 (PTBP1), XP_002124933.3 (SRSF1), XP_004227013.1 (SRSF2); *Hydra vulgaris*, XP_002156158.1 (PTBP1), XP_002159641.1 (SRSF1), XP_002161458.1 (SRSF2); *Anopheles gambiae*, XP_318405.4 (PTBP1), XP_318826.3 (SRSF2); *Trypanosoma brucei*, XP_827198.1 (PTBP1).

Table 1
Typical trans-Splicing Chimeras

Organisms	Involved Genes or Chimeras	Function Description	References	Experiments Verified
C. elegans	*eri-6/7*	Superfamily I helicase	(Fischer et al. 2008)	RT-PCR; Sequencing
Drosophila (Fruit fly)	*lola*	Transcription factor	(Horiuchi et al. 2003)	RT-PCR; Sequencing
	mod (mdg4)	Transcription factor	(Dorn et al. 2001)	RT-PCR; Sequencing
Anopheles gambiae (Mosquito)	*Bursicon*	Coding bursicon	(Robertson et al. 2007)	Bioinformatics analysis
Bombyx mori (Silk worm)	*mod (mdg4)*	Transcription factor	(Shao et al. 2012)	RT-PCR; Sequencing
	Dsx-dsr2	Sexual development	(Duan et al. 2013)	RT-PCR; Sequencing
Danio rerio (Zebrafish)	*Grn1-Grn2*	Hybrid granulin	(Cadieux et al. 2005)	RT-PCR; Northern blot
Gallus (Chicken)	*C-myb*	Proto-oncogene	(Vellard et al. 1991)	(?)
Rattus norvegicus (Rat)	1038 mRNA	Unknown	(Fitzgerald et al. 2006)	RT-PCR; Northern blot
	ABP-HDC	Unknown	(Sullivan et al. 1991)	RT-PCR; Northern blot
	COT	Gene expression regulation	(Caudevilla et al. 1998)	RT-PCR; Northern blot; in vitro trans-splicing
	HongrE2	Gene expression regulation	(Ni et al. 2011)	RT-PCR; Northern blot (?)
	LAR tyrosine phosphatase receptor	Gene expression regulation	(Zhang et al. 2003)	RT-PCR; RNase protection assay; Northern blot
	SNS-A	Unknownn	(Akopian et al. 1999)	PCR; RNase protection assay; Northern blot
	Sp1	Transcription factor	(Takahara et al. 2002)	RT-PCR; RNase protection assay; Northern, Southern blot
Mus musculus (Mouse)	*Dmrt1-Dmr*	Gene expression regulation	(Zhang et al. 2010)	RT-PCR; Northern blot; Southern blot
	Msh4-Hspa5 Msh4-Pcbp3	Cell death	(Hirano et al. 2004)	RT-PCR; Northern blot
Sus scrofa (Pig)	AK238425, AK351564 and other 667 putative chimeras	Unknownn	(Ma et al. 2012)	Systematic analysis; RT-PCR; RNA-Seq
Homo sapiens (Human)	*ATAC-1-* Exon X_a/X_b	Gene expression regulation	(Yu et al. 1999)	RT-PCR; RNase protection assay; Northern blot
	ATAC-1-Amp^r	Gain antibiotics resistance	(Hu et al. 2013)	RT-PCR; in vitro trans-splicing
	CAMK2G-SRP72	Unknownn	(Breen et al. 1997)	PCR; Genetic mapping; Western blot; (?)
	CDC2L2	Transcriptional regulation	(Jehan et al. 2007)	FISH; RT-PCR
	C-myb	Proto-oncogene	(Vellard et al. 1991)	RT-PCR; Sequencing
	CoAA-RBM4	Regulate stem/progenitor cell differentiation	(Brooks et al. 2009)	RT-PCR; in vitro trans-splicing
	CYCLIND1-TROP2	Cell growth	(Guerra et al. 2008)	RT-PCR; Northern blot; RNase protection assay (?)
	CYP3A4, 5, 7, 43	Catalytic activity	(Finta et al. 2002)	RT-PCR; Northern blot; RNase protection assay

(continued)

Table 1 Continued

Organisms	Involved Genes or Chimeras	Function Description	References	Experiments Verified
	hER alpha	Gene expression regulation	(Flouriot et al. 2002)	RT-PCR; Southern blot
	JAZF1-JJAZ1	Anti-apoptotic protein	(Li et al. 2008)	RT-PCR; Southern blot; in vitro trans-splicing
	PJA2-FER	Cancer biomarker	(Kawakami et al. 2013)	RT-PCR
	PAX3-FOXO1	Cancer biomarker	(Yuan, Qin, et al. 2013)	RT-PCR; FISH
	RGS12	G protein signaling	(Chatterjee et al. 2000)	RT-PCR
	Sp1	Transcription factor	(Takahara et al. 2000)	RT-PCR; Southern blot; RNase protection assay
	ZC3HAV1L-CHMP1A	Genome rearrangement	(Fang et al. 2012)	RT-PCR (?)
	AF4, AF9, ELL, ENL, MLL, ETV6, NUP98, RUNX1, EWSR1	DNA repair and chromosomal translocation	(Kowarz et al. 2012; Kowarz et al. 2011)	RT-PCR (?)
	TMEM79-SMG5	Cancer biomarker	(Kannan et al. 2011)	PCR; Sequencing (?)
	ts*RMST*	Pluripotency maintenance of hESCs	(Wu et al. 2014)	RT-PCR; RNase protection assay
	TSNAX-DISC1	G1/S transition and endometrial carcinoma (EC) development	(Li et al. 2014)	PCR; RNA-Seq (?)

Note.—"?", Probably trans-splicing.

which could lead to chromosomal translocations (Zaphiropoulos 2011). Kowarz et al. observed premature transcriptional termination as a common feature of genome rearrangements, and early terminated RNAs have an "unsaturated" splice donor site that gives rise to trans-splicing events (Kowarz et al. 2011, 2012). In this hypothesis, in case of DNA damage, these chimeric RNAs may direct broken chromosomes to align to the corresponding gene loci and guide chromosomal translocation. It seems that trans-spliced chimeric RNA is a precondition for chromosomal exchange; this may be a good explanation of why some patients have recurrent genetic rearrangements between *AF4* (exon 4) and *MLL* (exon 9) (Kowarz et al. 2011). Because chromosomal translocation is a common event in neoplastic cells (Kowarz et al. 2011), some trans-spliced chimeras may be indicative of tumorigenesis (Guerra et al. 2008; Yuan, Qin, et al. 2013). Indeed, chimeric RNA molecules have been proposed as potential biomarkers for tumor diagnosis (Zhou et al. 2012). For example, the chimeric *TMEM79-SMG5* molecule occurs in approximately 90% of prostate cancer samples, which may enable it to serve as a diagnostic biomarker for that type of cancer (Kannan et al. 2011).

Gene Expression Regulation

As indicated in table 1, trans-spliced chimeric RNAs are involved in the regulation of gene expression. For example, a 4.3-kb mRNA of human Acyl-CoA cholesterol acyltransferase 1 (ACAT-1) is derived from both chromosomes 1 and 7 (Li et al. 1999). The trans-spliced additional exons X_a and X_b, serve as the 5′ -UTR upstream the exon 1 and may account for its unconventional translation initiation. This chimeric RNA encodes a 56-kDa isoform protein with reduced activity (~30%) compared with the common form (Chen et al. 2008). Another case is the epididymis-specific *HongrES2*, composed of exons from different chromosomes, which was found to share a common 3′ -end with the *CES7* gene (Ni et al. 2011). Meanwhile, *HongrES2* can give rise to miRNA-like small RNA (*mil-HongrES2*) that downregulates *CES7* gene expression.

Signal Transduction

Trans-spliced chimeric RNAs are associated with signal transduction. For example, the *SNS-A* transcript comprises a repeat sequence of exons 12, 13, and 14, which encodes four transmembrane regions of domain II (Akopian et al. 1999). Nerve growth factor can induce SNS-A transcript expression. The regulation is probably associated with nervous signal transduction. Similarly, a truncated isoform (γ_{SRP}) of CaM kinase II acquires six amino acids (RNNYKL) from the *SRP72* gene (Breen et al. 1997). Although it has most of the catalytic properties of the holoenzyme, this isoform lacks an association domain, which may change its targeting ability. In addition, the RGS

(regulators of G-protein signaling) protein family has several distinct chimeric transcripts of *RGS12* in COS-7 cells, suggesting that trans-splicing may be a novel mechanism in the regulation of G-protein signaling pathways (Chatterjee et al. 2000).

Cell Viability and Growth

As mentioned above, the antiapoptotic *JAZF1-JJAZ1* protein is associated with aberrant proliferation of neoplastic cells. The chimeric *Msh4* δ variant is generated by trans-splicing between the *Hspa5* and *Msh4* pre-mRNAs, which could induce programed cell death during spermatogenesis (Hirano et al. 2004). In addition, some trans-spliced RNAs play a role in cell growth. For example, a low level of expression of the *CYCLIN D1-TROP2* chimera was shown to be sufficient to induce cell proliferation and to extend the life span of primary culture cells, while high expression of the chimera can induce cell transformation, indicating its role in the regulation of cell growth and cancer (Guerra et al. 2008). A recent study reported a chimeric *TSNAX-DISC1* in human endometrial carcinoma cells, which is regulated by a long intergenic noncoding RNA *lincRNA-NR_034037* (Li et al. 2014). Notably, the regulation of *TSNAX-DISC1* expression is involved in cell transition from G1 to S phase and in tumor growth.

Other Functions

In addition to previously discussed functions, trans-splicing is associated with other biological processes. Some chimeric RNAs in human are tissue-specific and can encode proteins. These proteins may compete with their parental proteins, disturbing protein interaction networks (Frenkel-Morgenstern and Valencia 2012; Frenkel-Morgenstern et al. 2012). In addition, a novel type of trans-splicing has been found in the *ACAT1* transcript, where an exogenous recombinant plasmid-derived Amp^r antisense segment is integrated (Hu et al. 2013). The type of exo-endo trans-splicing is abundant in normal human blood cells. This finding also suggests that exogenous DNA fragments, derived from recombinant plasmids or other sources, may affect cellular gene expression at both RNA and DNA levels.

Putative Mechanisms of trans-Splicing in Vertebrates

Currently, the mechanisms underlying trans-splicing in vertebrates remain largely unknown. Little is known about how the associated partner genes are physically recruited and what factors are involved in the process. Based on previous studies, we summarize several current models and propose new ones to address these issues.

tRNA-Mediated trans-Splicing Model

The tRNA sequence of two partner genes could direct their splicing reaction in a *trans* manner to generate a chimeric molecule in eukaryotic cells (Di Segni et al. 2008) (fig. 4*A*). In this model, the widespread tRNA genes in a genome or a repetitive sequence inside the coding region of an mRNA may be recognized and cleaved by the tRNA splicing endonuclease. Although experiments in vitro have shown that some mammalian mRNAs can be spliced by tRNA splicing endonuclease, tRNA-mediated trans-splicing needs to be explored further (Sidrauski et al. 1996; Deidda et al. 2003). Our recent study suggested that modern tRNA*s* originated from tRNA halves, potentially involving trans-splicing (Zuo et al. 2013).

Transcriptional Slippage Model

The second possible mechanism is the "transcriptional slippage model," which is based on a large-scale screening of chimeric RNAs in yeast, fruit fly, mouse, and human (Li, Zhao, et al. 2009). This model assumes that the transcription machinery "walks" along the primary template strand and dissociates from it in some cases, followed by "misaligning" with certain position of another locus through short homologous sequences (SHS) (fig. 4*B*). Thus, by continuing the transcriptional process on the new template, the chimeric RNA is generated. In this model, chimeric RNAs with classical "GU-AG" junction site only account for a small fraction (<20%), whereas the SHS type accounts for nearly 50%. Distal actively transcribed genes can frequently be corecruited to the same transcription machinery (Osborne et al. 2007), and this may be an environment promoting the occurrence of trans-splicing between two pre-mRNAs. As an example of the model, a 4-bp sequence at the junction site of chimeric *Msh4-Hspa5* molecule can be exactly mapped to each of the two partner genes (Hirano et al. 2004). This homologous region between partner genes may induce transcriptional slippage and further trans-splicing (fig. 4*C*).

Spliceosome-Mediated trans-Splicing Model

The third model is the spliceosome-mediated trans-splicing model. It was assumed that partner genes can be corecruited to the same spliceosome (Osborne et al. 2007) and spliced at canonical "GU-AG" sites (Li, Zhao, et al. 2009) (fig. 4*D*). Several cases of functional trans-splicing molecules with "GU-AG" at splicing sites support this model (Sullivan et al. 1991; Robertson et al. 2007; Fischer et al. 2008). Unsaturated splice donor sites were detected in early terminated transcripts in the human *MLL* gene. These unsaturated splice donor sites can induce a splicing reaction. Early terminated transcripts use cryptic exons to saturate the splice donor sites, which could give rise to trans-splicing events (Kowarz et al. 2011, 2012).

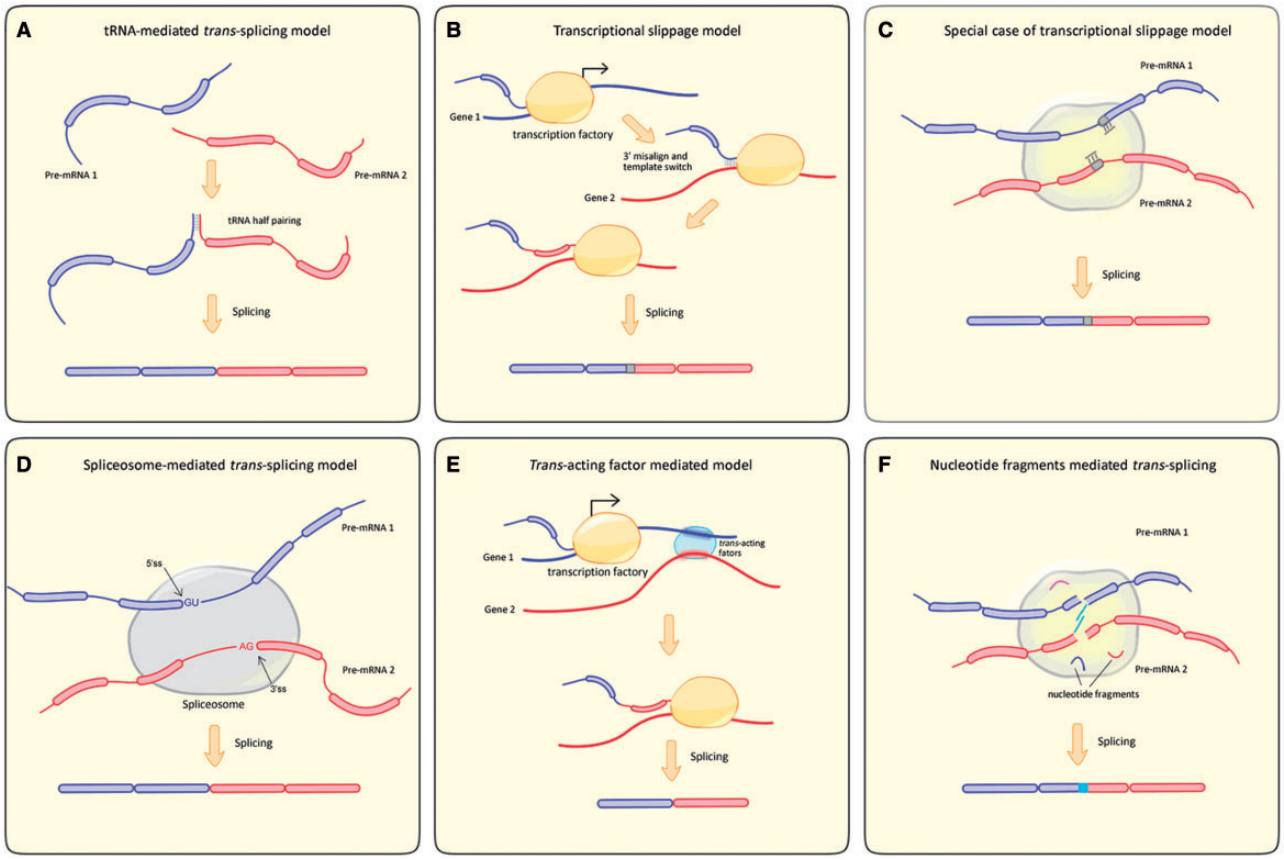

Fig. 4.—Schematic representation of proposed models of trans-splicing mechanisms. (*A*) tRNA-mediated trans-splicing model. Pre-tRNA halve adjacent to pre-mRNA context narrowing two associated molecules through complementary sequences, then the hybrid molecule is cleaved precisely at the sites of the tRNA intron by tRNA splicing endonuclease. (*B*) Transcriptional slippage model. Gray boxes represent pairing of SHSs. A pre-RNA is transcribed from Gene 1 and then misaligns to the DNA template of gene 2 via the SHSs. Transcription machinery keeps on moving on the strand of gene 2, after removal of introns, resulting in the chimeric molecule. (*C*) Special case of transcriptional slippage model. Both partner genes share a forward direction repeat sequence in the junction site of chimeric RNA. (*D*) Spliceosome mediated trans-splicing model. Like canonical cis-splicing, pre-RNA 1 and pre-RNA 2 is precisely spliced at the 5′- and 3′-splicing site and ligated as a non-linear chimeric molecule. (*E*) Trans-acting factor mediated model. Blurry region represent consensus DNA motif in parental gene 1 and gene 2. They can be recognized by trans-acting factor like CTCF and recruited to the shared transcription factory, and then coordinate the transcription by the same or similar transcription machinery. Transcription occurs between the Gene 1 and 2, the chimeric transcript is finally generated after intron removal. (*F*) Nucleotide fragments - mediated trans-splicing model. Short nucleotide fragments could induce transcription or be added into pre-mRNA. Trans-splicing could occur through base paring between two fragments. Through intermolecular splicing, this nucleotide fragments can be introduced into the chimeric molecule.

Trans-Acting Factors-Mediated Model

Compared with the above models, the trans-acting factors-mediated model could be more dynamic and capable of explaining how mRNA precursors are associated with each other before splicing (Ma et al. 2012) (fig. 4*E*). An interesting study has identified 251 chimeric mRNAs in pig, and a considerable fraction of these molecules have the canonical "GU-AG" at junction sites (Ma et al. 2012). The study also observed four consensus DNA sequences in the genomic region of the 5′ and 3′ partner genes, which are similar to the known DNA-binding motifs of the human CCCTC-bind factor (CTCF) binding sites. In this model, it is postulated that some consensus DNA motifs, such as CTCF, that are shared by associated partner genes can be recognized and recruited by CTCF to the same transcriptional machinery. CTCF may bring distal intrachromosomal and interchromosomal regions into proximity, suggesting a role in facilitating trans-splicing events (Ling et al. 2006; Williams et al. 2008). Indeed, while CTCF is silenced in endometrial stromal cells, the trans-spliced *JJAF1-JJAZ1* chimeric RNA was downregulated (Li et al. 2008; Zhang et al. 2012).

Furthermore, in line with the transcriptional slippage model, parental genes can be induced to colocalize to the same transcriptional factory so that they are coordinately transcribed to generate chimeric pre-mRNAs. After the excision of

introns, exons are joined by the spliceosome to generate a mature chimeric molecule. The trans-acting factor model could be universal and sufficiently dynamic to generate trans-splicing molecules.

Nucleotide Fragments-Mediated trans-Splicing Model

Endogenous and random short fragments were observed in cells and could serve as primers for reverse transcription polymerase chain reaction (RT-PCR) without adding extra primers (Yuan, Liu, et al. 2013). These endogenous short fragments can integrate into pre-mRNAs during transcriptional or posttranscriptional processes. Homologous regions in the short fragments could serve as intermediary guides to induce trans-splicing (fig. 4*F*). An example for this model is chimeric *ACAT-1* mRNA. Human *ACAT-1* mRNA is produced from two chromosomes by trans-splicing, but a 10-bp exon Xb could not be mapped to the relevant exons. Thus, an extra nucleotide fragment was inserted into the chimeric molecule (Li et al. 1999). This model could explain the formation of chimeric RNAs without the canonical "GU-AG" junction site, as well as some chimeric molecules with a small insertion that does not exist in the pre-mRNAs.

However, none of the models completely explains the generation of all trans-splicing. Current in silico screening strategies in chimeric RNA analysis rely on the canonical splicing sites "GT/AG." However, in real scenarios, trans-splicing could occur at some infrequent splicing sites (Herai et al. 2010). In addition, some DNA motifs, such as the GAAGAAG box in *COT* gene, can enhance trans-splicing frequency, suggesting a potential regulatory network (Caudevilla, Codony, et al. 2001). We are still far from a comprehensive understanding of trans-splicing mechanisms. Because of the complexity of RNA types in different cell types and different physiological conditions, there may be other mechanisms for the generation of chimeric RNAs that remain to be identified.

Challenges and Perspectives

Methodology Challenges

The identification and elimination of artificial chimeras are major challenges. Current methods utilized in gene expression analysis, such as RT-PCR, transcriptome, and cDNA library construction, typically require transcribing RNA into cDNA with RT. There are several sources of RTs. Lentiviruses (e.g., HIV-1, SIV) (Jamburuthugoda et al. 2011) and oncoretroviruses (e.g., AMV, M-MLV) encode virus RTs. In eukaryotes, both long terminal repeat (LTR) and non-LTR retrotransposable elements can encode RTs (Bibillo et al. 2002). In addition, the telomerase gene also encodes an RT to maintain telomere length. RTs lack 3′-5′ exonuclease activity and proofreading ability, and thus transcribe RNA into DNA with a high error rate (Bakhanashvili et al. 1992). The average error rate is approximately 3×10^{-5} for M-MLV RT and approximately 6×10^{-5} per nucleotide for AMV RT, which is one-tenth of that of HIV-1 RT (Katarzyna Bebenek 1993). The error rate of RTs encoded by LTR retrotransposable elements is similar to that of oncoretroviral RTs. The error rate for human telomerase is much higher, with approximately 2×10^{-3} per nucleotide (Agorio et al. 2003). The error rates of viral RTs with RNA templates are consistent with retroviral mutation rates of 10^{-4} to 10^{-6}. A high error rate results in the rapid evolution of viral genomes, which is essential for the virus to rapidly evade the host. However, it is difficult to avoid the introduction of many biases and artifacts when transcribing RNA into cDNA using RTs. In fact, there is a considerable amount of artificial chimeras in RNA-Seq, transcriptome, and cDNA libraries when using commercial RTs. In addition to substitution errors, it has been shown that the RT process is associated with the generation of artificial sequences due to template switching and fusions (Houseley et al. 2010). Moreover, due to different strategies in adapter ligation and fragmentation, we may generate platform-dependent biased data (Aird et al. 2011; Zheng et al. 2011). These two points may partially explain the inconsistent data generated from different RNA-Seq approaches, as observed in previous studies (Wu et al. 2014). Data retrieving and screening are also difficult. It is not easy to design an effective and reliable algorithm to identify real trans-splicing events from terabytes of data. Currently, there are some programs and databases for screening chimeric transcripts (Li, Zhao, et al. 2009; Kim et al. 2010; Al-Balool et al. 2011; Carrara et al. 2013; Frenkel-Morgenstern et al. 2013; Hoffmann et al. 2014). Further optimization, evaluation, and experimental confirmation are needed.

Even given these disadvantages, the RNA-Seq analysis and bioinformatics pipelines are still the most powerful tools for the analysis of trans-spliced chimeric RNAs. Improvement of cDNA cloning methods, for example, a new 3′-end cloning method (Yuan, Liu, et al. 2013) and other emerging technologies, will enable the discovery of more credible trans-splicing events. A new non-collinear transcript-detecting method was recently developed that can detect trans-spliced, circular, or fusion transcripts (Chuang et al. 2015). In addition, several chimeric RNA databases have been constructed and they reported appealing results (Kim et al. 2010; Abate et al. 2012; Benelli et al. 2012; Frenkel-Morgenstern et al. 2012, 2013; Bruno et al. 2013). For example, the ChiTaRS database includes comprehensive information on more than 16,000 chimeric transcripts from humans, mice, and fruit flies (Frenkel-Morgenstern et al. 2013). It is expected that the use of an optimized algorithm and filtering steps to eliminate false positives will yield more credible candidates using RNA-Seq data. The "TScan" strategy is a good example of a method of screening trans-splicing events in human embryonic stem cells (hESCs) (Wu et al. 2014). This is an integrative transcriptome sequencing technology with multiple experimental validation steps. First, the investigators acquired 0.83 million long

reads (~353.7-bp) and 230.63 million short reads (50-bp) from Roche 454 and SOLiD whole-transcriptome sequencing platforms, respectively. Then, by aligning these long reads with the public human genome database based on 454/Illumina sequencing data, 8,822 preliminary candidates were obtained. Targets validated by short-read information were extracted. The candidate group was then filtered by rules intended to identify non-trans-splicing events including: 1) chimeric junction site with SHS (McManus et al. 2010); 2) sense–antisense fusion containing a noncanonical splicing signal (Houseley et al. 2010); 3) mitochondrial–nuclear fusion events (McManus et al. 2010). Artificial products formed during the reverse-transcription (RT) process were subtracted (Houseley et al. 2010). Finally, Wu and colleagues identified and experimentally confirmed four trans-spliced RNAs (ts*CSNK1G3*, ts*ARHGAP5*, ts*FAT1*, and ts*RMST*) in the hESCs (Wu et al. 2014). These trans-spliced RNAs are all highly expressed in human pluripotent stem cells and differentially expressed during hESC differentiation. ts*RMST* may control pluripotency through repressing lineage-specific genes, involving the pluripotency transcription factor NANOG and PRC2 complex factor SUZ12. This report not only uncovered the importance of trans-splicing as a posttranscriptional event but also established an insightful pipeline to discover trans-splicing events.

New Technologies

Direct RNA Sequencing

Direct RNA sequencing (DRS) (Ozsolak et al. 2009) can profile mRNA transcripts of interest free of interference from RT-based artificial products. Natural RNA molecules can be directly sequenced by DRS without prior conversion to cDNA (Ozsolak et al. 2011), so DRS can detect real chimeric molecules. DRS of single molecules will have practical implications in the real-time monitoring of chimeric transcripts.

Hi-C and Interactome Modeling

Two distant gene loci separated by millions of DNA base pairs can be bridged by enhancers, transcription factors, and insulator proteins, and they can interact to regulate transcription of distant genes. All these activities are carefully orchestrated in the form of the three-dimensional (3D) conformation of chromosomes, which are compartmentalized in the nucleus. Hi-C allows one to probe genome-wide individual chromatin interactions (Lieberman-Aiden et al. 2009). It was reported that active genes can be transcribed and coregulated by the same transcription machinery (Osborne et al. 2007). In this case, both Hi-C and interactome modeling (Fullwood et al. 2009) can be used to characterize the complete repertoire of chromosomal interactions and help us to probe the regulatory activity involved specifically in trans-splicing.

Molecular Labeling Techniques

Nanostring techniques can quantitatively measure the expression levels of RNA transcripts (Geiss et al. 2008). By labeling probes with specific barcode, this method can capture individual RNA transcripts and count the exact copy number with high sensitivity and a digital readout. Another technique is using the tiny molecular beacon LNA/2′-*O*-methyl to mark individual pre-mRNA molecules to trace dynamic mRNA activities in living cells (Catrina et al. 2012). These techniques mean that we may be able to see how the two candidate primary transcripts are recruited together and where they are processed into chimeric molecules.

Proteomic Data Analysis

Based on a comprehensive analysis of 7,424 human chimeric RNAs, Frenkel-Morgenstern et al. (2012) suggested that chimeras potentially contain common and unique domain combinations. In combination with these techniques at the protein level, the accuracy of identified trans-splicing events will be improved tremendously. In addition, multiple experimental validation steps have been shown to be efficient in the validation of trans-splicing variants (Yu et al. 2014). With continued development of the techniques mentioned above and new techniques, we will gain understanding of the nature of trans-splicing.

Conclusions

(1) Trans-splicing is evolutionarily dynamic. The discovery of trans-splicing has updated the definition of genome coding capacity. Trans-splicing may be a mechanism for cells to extend the maximum potential of limited genetic information to adapt to various physiological conditions. In prokaryotes, reprogramming events on the RNA level rely on autocatalytic group II or group I introns and may be a detour from continuous RNAs in eukaryotes (Glanz et al. 2009). Despite a limited understanding of its evolutionary origin, we realize that trans-splicing occurs more frequently in lower species than in higher vertebrates. For example, trans-splicing occurs nearly in all genes in *T. brucei*, while vertebrates are free of SL trans-splicing. There is an evolutionary dynamic that trans-splicing is being replaced by other mechanisms, such as alternative splicing, to adapt to intricate genomic structures through refined regulation systems in vertebrates. Nevertheless, the splicing machinery is evolutionarily conserved between lower eukaryotes and mammals. It has been observed that induced SL RNAs can be accurately trans-spliced in HeLa cells in vivo and in vitro (Bruzik et al. 1992). The SR (Ser/Arg)-rich protein is a key factor for alternative splicing. This protein has also been shown to promote trans-splicing (Bruzik et al. 1995). These data suggest a common evolutionary origin of both cis-splicing and trans-splicing.

(2) Trans-spliced chimeras could contribute to some pathological consequences, such as cancer and apoptosis, and response to external stimuli, given that trans-splicing chimeras are temporally/spatially regulated and have low expression levels in normal cells. Under specific conditions/cell types, such as in cancer cells, they are deregulated and could lead to chromosomal translocation and tumorigenesis (Li et al. 2008; Li, Wang, et al. 2009). Several models for putative mechanisms of trans-splicing in vertebrates have been proposed. Further research will elucidate the underlying mechanisms of trans-splicing and uncover the biological functions and physiological/pathological significance of trans-spliced RNAs. In addition, the development of new trans-splicing RNA technologies and their translations into clinical applications will benefit more patients.

(3) Because a considerable number of RNA chimeras are artificial products generated by RT-based technology, the question of how to identify real trans-splicing molecules remains. A global transcriptome-wide and high-throughput analysis needs to be developed with both high sensitivity and optimized algorithms to detect tissue-specific and low-copy transcripts. DRS analysis with high-throughput, high efficiency, and low cost will be the most promising technique for detection of trans-splicing events in vertebrates.

Acknowledgments

Authors thank Dr Rainer Dorn for his suggestions for the manuscript. This work was supported by the National Natural Science Foundation of China, National Key Technologies R&D Program, Hubei Science & Tech Project and the Chinese 111 Project Grant B06018.

Literature Cited

Abate F, et al. 2012. Bellerophontes: an RNA-Seq data analysis framework for chimeric transcripts discovery based on accurate fusion model. Bioinformatics 28:2114–2121.

Agorio A, Chalar C, Cardozo S, Salinas G. 2003. Alternative mRNAs arising from trans-splicing code for mitochondrial and cytosolic variants of *Echinococcus granulosus* thioredoxin Glutathione reductase. J Biochem. 278:12920–12928.

Aird D, et al. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. 12:R18.

Akopian AN, et al. 1999. Trans-splicing of a voltage-gated sodium channel is regulated by nerve growth factor. FEBS Lett. 445:177–182.

Al-Balool HH, et al. 2011. Post-transcriptional exon shuffling events in humans can be evolutionarily conserved and abundant. Genome Res. 21:1788–1799.

Bachvaroff TR, Place AR. 2008. From stop to start: tandem gene arrangement, copy number and trans-splicing sites in the dinoflagellate *Amphidinium carterae.* PLoS One 3:e2929.

Bakhanashvili M, Hizi A. 1992. Fidelity of the reverse transcriptase of human immunodeficiency virus type 2. FEBS Lett. 306:151–156.

Belhocine K, Mak AB, Cousineau B. 2007. Trans-splicing of the LI.LtrB group II intron in *Lactococcus lactis.* Nucleic Acids Res. 35:2257–2268.

Benelli M, et al. 2012. Discovering chimeric transcripts in paired-end RNA-seq data by using EricScript. Bioinformatics 28:3232–3239.

Benton MJ, Donoghue PC. 2007. Paleontological evidence to date the tree of life. Mol Biol Evol. 24:26–53.

Bibillo A, Eickbush TH. 2002. The reverse transcriptase of the R2 non-LTR retrotransposon: continuous synthesis of cDNA on non-continuous RNA templates. J Mol Biol. 316:459–473.

Blumenthal T. 2005. Trans-splicing and operons. In: WormBook, editors. The C. elegans Research Community, WormBook, doi/10.1895/wormbook.1.5.1. p. 1–9.

Boothroyd JC, Cross GA. 1982. Transcripts coding for variant surface glycoproteins of *Trypanosoma brucei* have a short, identical exon at their 5′ end. Gene 20:281–289.

Breen MA, Ashcroft SJ. 1997. A truncated isoform of Ca2+/calmodulin-dependent protein kinase II expressed in human islets of langerhans may result from trans-splicing. FEBS Lett. 409:375–379.

Brehm K, Jensen K, Frosch M. 2000. mRNA trans-splicing in the human parasitic cestode *Echinococcus multilocularis.* J Biol Chem. 275:38311–38318.

Brooks YS, et al. 2009. Functional pre- mRNA trans-splicing of coactivator CoAA and corepressor RBM4 during stem/progenitor cell differentiation. J Biol Chem. 284:18033–18046.

Bruno AE, Miecznikowski JC, Qin M, Wang J, Liu S. 2013. FUSIM: a software tool for simulating fusion transcripts. BMC Bioinformatics 14:13.

Bruzik JP, Maniatis T. 1992. Spliced leader RNAs from lower eukaryotes are trans-spliced in mammalian cells. Nature 360:692–695.

Bruzik JP, Maniatis T. 1995. Enhancer-dependent interaction between 5′ and 3′ splice sites in trans. Proc Natl Acad Sci U S A. 92:7056–7059.

Bruzik JP, Van Doren K, Hirsh D, Steitz JA. 1988. Trans splicing involves a novel form of small nuclear ribonucleoprotein particles. Nature 335:559–562.

Cadieux B, Chitramuthu BP, Baranowski D, Bennett HP. 2005. The zebrafish progranulin gene family and antisense transcripts. BMC Genomics 6:156.

Carrara M, et al. 2013. State-of-the-art fusion-finder sensitivity and specificity. BioMed Res Int. 2013:340620. doi:10.1155/2013/340620.

Catrina IE, Marras SA, Bratu DP. 2012. Tiny molecular beacons: LNA/2′-O-methyl RNA chimeric probes for imaging dynamic mRNA processes in living cells. ACS Chem Biol. 7:1586–1595.

Caudevilla C, Codony C, et al. 2001. Localization of an exonic splicing enhancer responsible for mammalian natural trans-splicing. Nucleic Acids Res. 29:3108–3115.

Caudevilla C, Da Silva-Azevedo, et al. 2001. Heterologous HIV-nef mRNA trans-splicing: a new principle how mammalian cells generate hybrid mRNA and protein molecules. FEBS Lett. 507:269–279.

Caudevilla C, et al. 1998. Natural trans-splicing in carnitine octanoyltransferase pre-mRNAs in rat liver. Proc Natl Acad Sci U S A. 95:12185–12190.

Chan PP, Cozen AE, Lowe TM. 2011. Discovery of permuted and recently split transfer RNAs in Archaea. Genome Biol. 12:R38.

Chatterjee TK, Fisher RA. 2000. Novel alternative splicing and nuclear localization of human *RGS12* gene products. J Biol Chem. 275:29660–29671.

Chen J, et al. 2008. RNA secondary structures located in the interchromosomal region of human ACAT1 chimeric mRNA are required to produce the 56-kDa isoform. Cell Res. 18:921–936.

Chuang TJ, et al. 2015. NCLscan: accurate identification of non-co-linear transcripts (fusion, trans-splicing and circular RNA) with a good balance between sensitivity and precision. Nucleic Acids Res. 44(3):e29.

Communi D, Suarez-Huerta N, Dussossoy D, Savi P, Boeynaems JM. 2001. Cotranscription and intergenic splicing of human *P2Y11* and *SSF1* genes. J Biol Chem. 276:16561–16566.

Danks GB, et al. 2015. Trans-splicing and operons in metazoans: translational control in maternally regulated development and recovery from growth arrest. Mol Biol Evol. 32:585–599.
Deidda G, Rossi N, Tocchini-Valentini GP. 2003. An archaeal endoribonuclease catalyzes cis- and trans- nonspliceosomal splicing in mouse cells. Nat Biotechnol. 21:1499–1504.
Derelle R, et al. 2010. Convergent origins and rapid evolution of spliced leader trans-splicing in Metazoa: insights from the Ctenophora and Hydrozoa. RNA 16:696–707.
Di Segni G, Gastaldi S, Tocchini-Valentini GP. 2008. Cis- and trans-splicing of mRNAs mediated by tRNA sequences in eukaryotic cells. Proc Natl Acad Sci U S A. 105:6864–6869.
Dorn R, Reuter G, Loewendorf A. 2001. Transgene analysis proves mRNA trans-splicing at the complex *mod(mdg4)* locus in *Drosophila.* Proc Natl Acad Sci U S A. 98:9724–9729.
Douris V, Telford MJ, Averof M. 2010. Evidence for multiple independent origins of trans-splicing in Metazoa. Mol Biol Evol. 27:684–693.
Duan J, et al. 2013. Novel female-specific trans-spliced and alternative splice forms of dsx in the silkworm *Bombyx mori.* Biochem Biophys Res Commun. 431:630–635.
Eul J, Graessmann M, Graessmann A. 1995. Experimental evidence for RNA trans-splicing in mammalian cells. EMBO J. 14:3226–3235.
Fang W, Wei Y, Kang Y, Landweber LF. 2012. Detection of a common chimeric transcript between human chromosomes 7 and 16. Biol Direct. 7:49.
Finta C, Zaphiropoulos PG. 2002. Intergenic mRNA molecules resulting from trans-splicing. J Biol Chem. 277:5882–5890.
Fischer SE, Butler MD, Pan Q, Ruvkun G. 2008. Trans-splicing in *C. elegans* generates the negative RNAi regulator *ERI-6/7*. Nature 455:491–496.
Fitzgerald C, et al. 2006. Mammalian transcription in support of hybrid mRNA and protein synthesis in testis and lung. J Biol Chem. 281:38172–38180.
Flouriot G, Brand H, Seraphin B, Gannon F. 2002. Natural trans-spliced mRNAs are generated from the human estrogen receptor-alpha (*hER alpha*) gene. J Biol Chem. 277:26244–26251.
Frenkel-Morgenstern M, et al. 2012. Chimeras taking shape: potential functions of proteins encoded by chimeric RNA transcripts. Genome Res. 22:1231–1242.
Frenkel-Morgenstern M, et al. 2013. ChiTaRS: a database of human, mouse and fruit fly chimeric transcripts and RNA-sequencing data. Nucleic Acids Res. 41:D142–D151.
Frenkel-Morgenstern M, Valencia A. 2012. Novel domain combinations in proteins encoded by chimeric transcripts. Bioinformatics 28:i67–i74.
Fujishima K, et al. 2009. Tri-split tRNA is a transfer RNA made from 3 transcripts that provides insight into the evolution of fragmented tRNAs in archaea. Proc Natl Acad Sci U S A. 106:2683–2687.
Fullwood MJ, et al. 2009. An oestrogen-receptor-alpha-bound human chromatin interactome. Nature 462:58–64.
Galloway Salvo JL, Coetzee T, Belfort M. 1990. Deletion-tolerance and trans-splicing of the bacteriophage T4 td intron. Analysis of the P6-L6a region. J Mol Biol. 211:537–549.
Gao JL, et al. 2015. A conserved intronic U1 snRNP-binding sequence promotes trans-splicing in *Drosophila.* Genes Dev. 29:760–771.
Gao Z, et al. 2013. Identification and characterization of two novel transcription units of porcine circovirus 2. Virus Genes 47:268–275.
Geiss GK, et al. 2008. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol. 26:317–325.
Gingeras TR. 2009. Implications of chimaeric non-co-linear transcripts. Nature 461:206–211.
Glanz S, Kuck U. 2009. Trans-splicing of organelle introns–a detour to continuous RNAs. Bioessays 31:921–934.
Guerra E, et al. 2008. A bicistronic *CYCLIN D1-TROP2* mRNA chimera demonstrates a novel oncogenic mechanism in human cancer. Cancer Res. 68:8113–8121.
Guiliano DB, Blaxter ML. 2006. Operon conservation and the evolution of trans-splicing in the phylum Nematoda. PLoS Genet. 2:e198.
Gupta SK, et al. 2014. Two splicing factors carrying serine-arginine motifs, TSR1 and TSR1IP, regulate splicing, mRNA stability, and rRNA processing in *Trypanosoma brucei.* RNA Biol. 11:715–731.
Hannon GJ, Maroney PA, Nilsen TW. 1991. U small nuclear ribonucleoprotein requirements for nematode cis- and trans-splicing *in vitro.* J Biol Chem. 266:22792–22795.
Harrison N, Kalbfleisch A, Connolly B, Pettitt J, Muller B. 2010. SL2-like spliced leader RNAs in the basal nematode *Prionchulus punctatus*: new insight into the evolution of nematode SL2 RNAs. RNA 16:1500–1507.
Hastings KE. 2005. SL trans-splicing: easy come or easy go? Trends Genet. 21:240–247.
Herai RH, Yamagishi ME. 2010. Detection of human interchromosomal trans-splicing in sequence databanks. Brief Bioinform. 11:198–209.
Hirano M, Noda T. 2004. Genomic organization of the mouse *Msh4* gene producing bicistronic, chimeric and antisense mRNA. Gene 342:165–177.
Hoffmann S, et al. 2014. A multi-split mapping algorithm for circular RNA, splicing, trans-splicing, and fusion detection. Genome Biol. 15:R34.
Horiuchi T, Giniger E, Aigaki T. 2003. Alternative trans-splicing of constant and variable exons of a *Drosophila* axon guidance gene, *lola.* Genes Dev. 17:2496–2501.
Houseley J, Tollervey D. 2010. Apparent non-canonical trans-splicing is generated by reverse transcriptase *in vitro.* PLoS One 5:e12271.
Hu GJ, et al. 2013. Production of ACAT1 56-kDa isoform in human cells via trans-splicing involving the ampicillin resistance gene. Cell Res. 23:1007–1024.
Huang XY, Hirsh D. 1989. A second trans-spliced RNA leader sequence in the nematode *Caenorhabditis elegans.* Proc Natl Acad Sci U S A. 86:8640–8644.
Jamburuthugoda VK, Eickbush TH. 2011. The reverse transcriptase encoded by the non-LTR retrotransposon R2 is as error-prone as that encoded by HIV-1. J Mol Biol. 407:661–672.
Jehan Z, et al. 2007. Novel noncoding RNA from human Y distal heterochromatic block (Yq12) generates testis-specific chimeric *CDC2L2*. Genome Res. 17:433–440.
Kannan K, et al. 2011. Recurrent chimeric RNAs enriched in human prostate cancer identified by deep sequencing. Proc Natl Acad Sci U S A. 108:9172–9177.
Katarzyna Bebenek TAK. 1993. The fidelity of retroviral reverse transcriptases—chapter 5. In: Skalka AM, Goff Stephen, editors. Reverse transcriptase. New York: Cold Spring Harbor Laboratory Press. p. 85–102.
Kawakami M, et al. 2013. Detection of novel paraja ring finger 2-fer tyrosine kinase mRNA chimeras is associated with poor postoperative prognosis in non-small cell lung cancer. Cancer Sci. 104:1447–1454.
Kim P, et al. 2010. ChimerDB 2.0—a knowledgebase for fusion genes updated. Nucleic Acids Res. 38:D81–D85.
Kong Y, et al. 2015. The evolutionary landscape of intergenic trans-splicing events in insects. Nat Commun. 6:8734.
Kowarz E, Dingermann T, Marschalek R. 2012. Do non-genomically encoded fusion transcripts cause recurrent chromosomal translocations? Cancers 4:1036–1049.
Kowarz E, Merkens J, Karas M, Dingermann T, Marschalek R. 2011. Premature transcript termination, trans-splicing and DNA repair: a vicious path to cancer. Am J Blood Res. 1:1–12.
Krause M, Hirsh D. 1987. A trans-spliced leader sequence on actin mRNA in *C. elegans.* Cell 49:753–761.

Li BL, et al. 1999. Human acyl-CoA:cholesterol acyltransferase-1 (*ACAT-1*) gene organization and evidence that the 4.3-kilobase ACAT-1 mRNA is produced from two different chromosomes. J Biol Chem. 274:11060–11071.
Li H, Wang J, Ma X, Sklar J. 2009. Gene fusions and RNA trans-splicing in normal and neoplastic human cells. Cell Cycle 8:218–222.
Li H, Wang J, Mor G, Sklar J. 2008. A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells. Science 321:1357–1361.
Li N, et al. 2014. Identification of chimeric *TSNAX-DISC1* resulting from intergenic splicing in endometrial carcinoma through high-throughput RNA sequencing. Carcinogenesis 35:2687–2697.
Li X, Zhao L, Jiang H, Wang W. 2009. Short homologous sequences are strongly associated with the generation of chimeric RNAs in eukaryotes. J Mol Evol. 68:56–65.
Liang XH, Haritan A, Uliel S, Michaeli S. 2003. trans and *cis* splicing in trypanosomatids: mechanism, factors, and regulation. Eukaryot Cell. 2:830–840.
Lieberman-Aiden E, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289–293.
Ling JQ, et al. 2006. CTCF mediates interchromosomal colocalization between *Igf2/H19* and *Wsb1/Nf1*. Science 312:269–272.
Ma L, et al. 2012. Identification and analysis of pig chimeric mRNAs using RNA sequencing data. BMC Genomics. 13:429.
Magrangeas F, et al. 1998. Cotranscription and intergenic splicing of human galactose-1-phosphate uridylyltransferase and interleukin-11 receptor alpha-chain genes generate a fusion mRNA in normal cells. Implication for the production of multidomain proteins during evolution. J Biol Chem. 273:16005–16010.
Maniatis T, Tasic B. 2002. Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature 418:236–243.
Marletaz F, et al. 2008. Chaetognath transcriptome reveals ancestral and unique features among bilaterians. Genome Biol. 9:R94.
Marletaz F, Le Parco Y. 2008. Careful with understudied phyla: the case of chaetognath. BMC Evol Biol. 8:251.
Maroney PA, Denker JA, Darzynkiewicz E, Laneve R, Nilsen TW. 1995. Most mRNAs in the nematode Ascaris lumbricoides are trans-spliced: a role for spliced leader addition in translational efficiency. RNA 1:714–723.
Matsumoto J, et al. 2010. High-throughput sequence analysis of *Ciona intestinalis* SL trans-spliced mRNAs: alternative expression modes and gene function correlates. Genome Res. 20:636–645.
McManus CJ, Duff MO, Eipper-Mains J, Graveley BR. 2010. Global analysis of trans-splicing in *Drosophila*. Proc Natl Acad Sci U S A. 107:12975–12979.
Mori H, et al. 2002. Chromosome translocations and covert leukemic clones are generated during normal fetal development. Proc Natl Acad Sci U S A. 99:8242–8247.
Murphy WJ, Watkins KP, Agabian N. 1986. Identification of a novel Y branch structure as an intermediate in *trypanosome* mRNA processing: evidence for trans splicing. Cell 47:517–525.
Ni MJ, et al. 2011. Identification and characterization of a novel non-coding RNA involved in sperm maturation. PLoS One 6:e26053.
Nilsen TW. 1993. Trans-splicing of nematode premessenger RNA. Annu Rev Microbiol. 47:413–440.
Nilsen TW, et al. 1989. Characterization and expression of a spliced leader RNA in the parasitic nematode Ascaris lumbricoides var. suum. Mol Cell Biol. 9:3543–3547.
Osborne CS, et al. 2007. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. PLoS Biol. 5:e192.
Ozsolak F, et al. 2009. Direct RNA sequencing. Nature 461:814–818.
Ozsolak F, Milos PM. 2011. Single-molecule direct RNA sequencing without cDNA synthesis. Wiley Interdiscip Rev RNA. 2:565–570.
Perry KL, Watkins KP, Agabian N. 1987. *Trypanosome* mRNAs have unusual "cap 4" structures acquired by addition of a spliced leader. Proc Natl Acad Sci U S A. 84:8190–8194.
Pouchkina-Stantcheva NN, Tunnacliffe A. 2005. Spliced leader RNA-mediated trans-splicing in phylum Rotifera. Mol Biol Evol. 22:1482–1489.
Randau L. 2015. Evolution of small guide RNA genes in hyperthermophilic archaea. Ann N Y Acad Sci. 1341:188–193.
Randau L, Munch R, Hohn MJ, Jahn D, Soll D. 2005. *Nanoarchaeum equitans* creates functional tRNAs from separate genes for their 5′- and 3′-halves. Nature 433:537–541.
Robertson HM, Navik JA, Walden KKO, Honegger HW. 2007. The bursicon gene in mosquitoes: an unusual example of mRNA trans-splicing. Genetics 176:1351–1353.
Romani A, Guerra E, Trerotola M, Alberti S. 2003. Detection and analysis of spliced chimeric mRNAs in sequence databanks. Nucleic Acids Res. 31:e17
Satou Y, et al. 2008. Improved genome assembly and evidence-based global gene model set for the chordate *Ciona intestinalis*: new insight into intron and operon populations. Genome Biol. 9:R152
Satou Y, Hamaguchi M, Takeuchi K, Hastings KE, Satoh N. 2006. Genomic overview of mRNA 5′-leader trans-splicing in the ascidian *Ciona intestinalis*. Nucleic Acids Res. 34:3378–3388.
Shao W, et al. 2012. Alternative splicing and trans-splicing events revealed by analysis of the *Bombyx mori* transcriptome. RNA 18:1395–1407.
Sidrauski C, Cox JS, Walter P. 1996. tRNA ligase is required for regulated mRNA splicing in the unfolded protein response. Cell 87:405–413.
Stover NA, Steele RE. 2001. Trans-spliced leader addition to mRNAs in a cnidarian. Proc Natl Acad Sci U S A. 98:5693–5698.
Sullivan PM, Petrusz P, Szpirer C, Joseph DR. 1991. Alternative processing of androgen-binding protein RNA transcripts in fetal rat liver. Identification of a transcript formed by trans splicing. J Biol Chem. 266:143–154.
Sutton RE, Boothroyd JC. 1986. Evidence for trans splicing in *trypanosomes*. Cell 47:527–535.
Takahara T, Kanazu SI, Yanagisawa S, Akanuma H. 2000. Heterogeneous Sp1 mRNAs in human HepG2 cells include a product of homotypic trans-splicing. J Biol Chem. 275:38067–38072.
Takahara T, Kasahara D, Mori D, Yanagisawa S, Akanuma H. 2002. The trans-spliced variants of Sp1 mRNA in rat. Biochem Biophys Res Commun. 298:156–162.
Tasic B, et al. 2002. Promoter choice determines splice site selection in protocadherin alpha and gamma pre-mRNA splicing. Mol Cell. 10:21–33.
Van der Ploeg LH, et al. 1982. RNA splicing is required to make the messenger RNA for a variant surface antigen in *trypanosomes*. Nucleic Acids Res. 10:3591–3604.
Van Doren K, Hirsh D. 1988. Trans-spliced leader RNA exists as small nuclear ribonucleoprotein particles in *Caenorhabditis elegans*. Nature 335:556–559.
Vandenberghe AE, Meedel TH, Hastings KEM. 2001. mRNA 5′-leader trans-splicing in the chordates. Genes Dev. 15:294–303.
Vellard M, et al. 1991. *C-myb* proto-oncogene: evidence for intermolecular recombination of coding sequences. Oncogene 6:505–514.
Wahl MC, Luhrmann R. 2015. SnapShot: spliceosome dynamics I. Cell 161:1474–14e1. 471.
Wally V, Murauer EM, Bauer JW. 2012. Spliceosome-mediated trans-splicing: the therapeutic cut and paste. J Invest Dermatol. 132:1959–1966.
Williams A, Flavell RA. 2008. The role of CTCF in regulating nuclear organization. J Exp Med. 205:747–750.

Wu CS, et al. 2014. Integrative transcriptome sequencing identifies trans-splicing events with important roles in human embryonic stem cell pluripotency. Genome Res. 24:25–36.

Yu C, et al. 1999. Human acyl-CoA:cholesterol acyltransferase-1 is a homotetrameric enzyme in intact cells and *in vitro*. J Biol Chem. 274:36139–36145.

Yu CY, Liu HJ, Hung LY, Kuo HC, Chuang TJ. 2014. Is an observed non-co-linear RNA product spliced in trans, in *cis* or just *in vitro*? Nucleic Acids Res. 42:9410–9423.

Yuan CF, Liu YM, Yang M, Liao DJ. 2013. New methods as alternative or corrective measures for the pitfalls and artifacts of reverse transcription and polymerase chain reactions (RT-PCR) in cloning chimeric or anti-sense-accompanied RNA. RNA Biol. 10:958–968.

Yuan H, Qin F, et al. 2013. A chimeric RNA characteristic of rhabdomyosarcoma in normal myogenesis process. Cancer Discov. 3:1394–1403.

Zaphiropoulos PG. 2011. Trans-splicing in Higher Eukaryotes: implications for cancer development? Front Genet. 2:92.

Zaslaver A, Baugh LR, Sternberg PW. 2011. Metazoan operons accelerate recovery from growth-arrested states. Cell 145:981–992.

Zhang C, et al. 2003. A candidate chimeric mammalian mRNA transcript is derived from distinct chromosomes and is associated with nonconsensus splice junction motifs. DNA Cell Biol. 22:303–315.

Zhang H, et al. 2007. Spliced leader RNA trans-splicing in dinoflagellates. Proc Natl Acad Sci U S A. 104:4618–4623.

Zhang H, Lin S. 2009. Retrieval of missing spliced leader in dinoflagellates. PLoS One 4:e4129.

Zhang L, Lu H, Xin D, Cheng H, Zhou R. 2010. A novel ncRNA gene from mouse chromosome 5 trans-splices with Dmrt1 on chromosome 19. Biochem Biophys Res Commun. 400:696–700.

Zhang Y, et al. 2012. Chimeric transcript generated by cis-splicing of adjacent genes regulates prostate cancer cell proliferation. Cancer Discov. 2:598–607.

Zheng W, Chung LM, Zhao H. 2011. Bias detection and correction in RNA-sequencing data. BMC Bioinformatics. 12:290.

Zhou J, Liao J, Zheng X, Shen H. 2012. Chimeric RNAs as potential biomarkers for tumor diagnosis. BMB Rep. 45:133–140.

Zhu J, Mayeda A, Krainer AR. 2001. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. Mol Cell. 8:1351–1361.

Zorio DA, Cheng NN, Blumenthal T, Spieth J. 1994. Operons as a common form of chromosomal organization in *C. elegans*. Nature 372:270–272.

Zuo Z, et al. 2013. Genome-wide analysis reveals origin of transfer RNA genes from tRNA halves. Mol Biol Evol. 30:2087–2098.

Associate editor: Kateryna Makova