Quantitative Biology
DOI 10.1007/s40484-017-0112-7

MINI REVIEW

Multifaceted roles of complementary sequences on circRNA formation

Qin Yang[1], Ying Wang[1] and Li Yang[1,2,*]

[1] CAS Key Laboratory of Computational Biology, Collaborative Innovation Center of Genetics and Development, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences; University of Chinese Academy of Sciences, Shanghai 200031, China
[2] School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China
* Correspondence: liyang@picb.ac.cn

Received March 7, 2017; Revised June 5, 2017; Accepted June 12, 2017

***Background*: Circular RNAs (circRNAs) from back-spliced exon(s) are characterized by the covalently closed loop feature with neither 5′ to 3′ polarity nor polyadenylated tail. By using specific computational approaches that identify reads mapped to back-splice junctions with a reversed genomic orientation, ten thousands of circRNAs have been recently re-identified in various cell lines/tissues and across different species. Increasing lines of evidence suggest that back-splicing is catalyzed by the canonical spliceosomal machinery and modulated by *cis*-elements and *trans*-factors.**
***Results*: In this mini-review, we discuss our current understanding of circRNA biogenesis regulation, mainly focusing on the complex regulation of complementary sequences, especially *Alu*s in human, on circRNA formation.**
***Conclusions*: Back-splicing can be significantly facilitated by RNA pair formed by orientation-opposite complementary sequences that juxtapose flanking introns of circularized exon(s). RNA pair formed within individual introns competes with RNA pair formed across flanking introns in the same gene locus, leading to distinct choices for either canonical splicing or back-splicing. Multiple RNA pairs that bracket different circle-forming exons compete for alternative back-splicing selection, resulting in multiple circRNAs generated in a single gene locus.**

Keywords: circRNA; circRNA biogenesis; back-splicing; *cis*-elements; complementary sequences; *Alu*

INTRODUCTION

A variety of circular RNAs can be produced *via* distinct mechanisms, such as direct single-stranded RNA ligation as circular RNA genome, derived from processed rRNAs as intermediates, processed from self-splicing introns as unstable circular transcripts, and etc. (reviewed in [1–3]). Among them, at least two types of circular RNAs are processed from nuclear (m)RNA precursors through the spliceosomal pathway [4], including circular intronic RNAs (ciRNAs) [5] from excised introns and circular RNAs (circRNAs) [6–10] from back-spliced exon(s) (reviewed in [1,4]). By taking advantage of non-polyadenylated transcriptome enrichment [11–13] and specific computational approaches that identify reads mapped to back-splice junction sites with a reversed genomic orientation (Figure 1A) [10,15,16] (reviewed in [1,2,17]), a large amount of circRNAs have been recently re-discovered from thousands of gene loci in various cell lines/tissues and across different species [6–8,10,18–21]. Increasing lines of evidence have suggested that circRNAs could play important roles in gene expression regulation with different mechanisms of action (reviewed in [1]). These results thus expand our understanding on the complexity and diversity of eukaryotic circular RNAs.

It has been demonstrated that the biogenesis of circRNAs processed from back-splicing requires the canonical spliceosomal machinery [22,23]. Different to canonical splicing, back-splicing ligates a downstream 5′ splice site with an upstream 3′ splice site in a reversed order, which is believed to be inefficiently catalyzed by spliceosome (reviewed in [1,2]). Recent studies aimed to underscore mechanisms of circRNA biogenesis regulation have shown that both *cis*-elements (mainly flanking intronic sequences) and *trans*-factors (mainly RNA binding proteins, RBPs) can promote back-splicing for

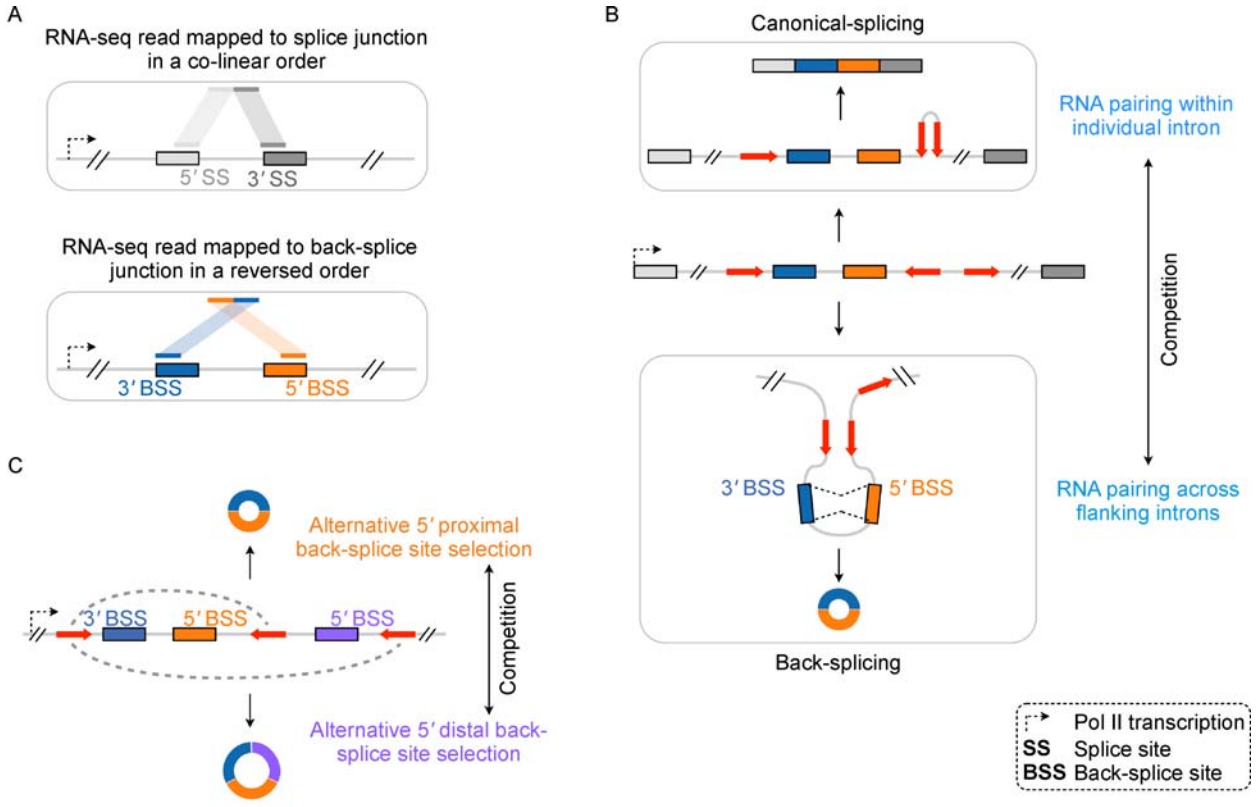

Figure 1. Schematic models of multifaceted roles of complementary sequences on circRNA formation. (A) Identification of back-splice junction sites by specific computational pipelines. Different from canonical splicing with reads mapped to the reference genome in a co-linear order, reads mapped to back-splice junctions exhibit a reversed genomic orientation. (B) Orientation-opposite complementary sequences on circRNA biogenesis. Orientation-opposite complementary sequences (mainly IR*Alu*s in human) across introns flanking back-spliced exons (bottom) enhance circRNA biogenesis, which also compete with RNA pairing formed within an individual intron in the same gene locus (top) for canonical splicing (modified from [2]). (C) The competition of RNA pairing across different pairs of flanking introns results in multiple circRNAs in the same gene locus through alternative back-splicing (modified from [14]). Red arrows, orientation-opposite complementary sequences; dash red arc lines, RNA pairing formed by orientation-opposite complementary sequences; grey bars, canonical spliced exons; colored bars, (alternative) back-spliced exons.

circRNA biogenesis (reviewed in [1,2]). Specifically, genome-wide analyses have revealed the positive correlation of across-intron RNA pairing with circRNA formation [6,10]. Furthermore, back-splicing is efficiently linked to fast RNA polymerase II elongation [24] and largely occurs post-transcriptionally with low efficiency [24,25]. Here, we highlight recent research progress on the regulation of circRNA biogenesis, focusing on our current understanding of the complex regulation of *cis* complementary sequences, especially *Alu*s in human, on circRNA formation.

GENERAL REGULATORY ROLE OF COMPLEMENTARY SEQUENCES ON circRNA FORMATION

Back-splicing is presumably catalyzed by the same canonical spliceosomal machinery as canonical splicing [22,23] (reviewed in [1]), thus back-splicing has been found to compete with canonical splicing [10,22] (reviewed in [2]). Although unfavorably processed in general, circRNA biogenesis can be significantly facilitated by orientation-opposite complementary sequences that juxtapose flanking introns of circularized exon(s) [6,10,25]. These orientation-opposite sequences can be either repetitive sequences, such as *Alu* in human, or non-repetitive but complementary sequences [10,18]. In theory, RNA pairing formed by orientation-opposite complementary sequences, as short as 30 to 40 nucleotides in length [25], can bring the downstream donor and upstream acceptor sites close together to promote spliceosome assembly for back-splicing (Figure 1B, bottom)(reviewed in [2]), however the detailed biochemical evidence is still missing. The large amount of repetitive elements, especially *Alu* sequences in human, contribute the most for enhancing circRNA biogenesis

[10,14,18]. Over one million copies of *Alu* sequences have been found in the human genome, and about half of them are located in intronic regions [10,26]. Due to their sequence similarity, orientation-opposite *Alu*s within certain distance have the potential to form inverted repeated *Alu* sequences (IR*Alu*s) [27], and when located across flanking introns, IR*Alu*s could promote circRNA production [10].

Computational evaluation of pairing capacity of orientation-opposite complementary sequences across circRNA-flanking introns identifies that *SINE*s (short interspersed nuclear repetitive DNA elements), especially *Alu* elements in human, contribute the most for circRNA formation [18]. More specifically, among all types of complementary sequences, 93.3% of circRNA-flanking introns in human exhibiting the strongest RNA pairing capacity are from IR*Alu*s and only a very small portion are from other non-*Alu* repetitive sequences and other non-repetitive but complementary sequences [18]. Despite of higher pairing capacity in both human and mouse, the non-repetitive but complementary sequences are only sparsely present in circRNA-flanking introns, indicating at best a limited role in circRNA formation [18].

In addition to genome-wide annotation that shows the association of circRNA expression with complementary sequences [6,10], direct lines of experimental evidence also demonstrate that orientation-opposite complementary sequences can efficiently enhance circRNA formation. First, using expression vectors, the existence of IR*Alu*s mimicking endogenous conditions was shown to be required for circRNA expression [10]. Short complementary sequences (30 to 40 nucleotides in length) were capable of promoting circRNA biogenesis from expression vectors [25], but stronger pairing capacity with longer sequences could considerably enhance circRNA production [10]. Second, in endogenous condition, disruption of intronic RNA pairing by CRISPR-Cas9 mediated genome editing resulted in the depletion of circRNA expression, while the linear mRNA counterpart remains largely unchanged [24]. Finally, distal intronic sequences from different genes could also be juxtaposed after gene fusion to form RNA pairing that flanks the breakpoint of fusion genes, leading to the formation of aberrant fusion-circRNAs in cancer cells [28].

Other than *cis* intronic complementary sequences, several protein factors have been reported to regulate circRNA biogenesis, either in a positive [22,29] or negative [19,20] manner. It should be noted that *cis*-elements and *trans*-factors can also function in a combinatorial manner to control circRNA biogenesis [30]. Since hundreds of RBPs have been recently identified [31,32], it will be of great interest to identify other *trans*-factors and their combinatorial regulation with *cis*-elements on circRNA biogenesis.

COMPETITION OF BACK-SPLICING AND CANONICAL SPLICING

Back-splicing can compete with canonical splicing [10,22]. Orientation-opposite complementary sequences across two separate introns that flank circle-forming exons are efficient, but may not be sufficient, to boost back-splicing for circRNA formation [10] (reviewed in [2]). In fact, similar RNA pairing could also be formed within individual introns in the same gene locus, which competes with the formation of RNA pairing across flanking introns, leading to distinct choices for either canonical splicing or back-splicing (Figure 1B, top) [10]. By introducing additional RNA pairing within individual intron in expression vector, canonical splicing was observed to compete against back-splicing, resulting in the reduction of circRNA biogenesis [16]. Recently, a quantitative computational method was developed to evaluate RNA pairing capacity of complementary sequences across given circRNA-flanking introns, in which many factors, including sequence pairing strengths and competition ability with other similar complementary sequences, were considered [18].

COMPETITION OF ALTERNATIVE INVERTED REPEATED *Alu* PAIRING ON ALTERNATIVE BACK-SPLICING SELECTION IN HUMAN

Multiple circRNAs could be generated in a single gene locus, through either alternative back-splicing or alternative splicing [10,16]. The alternative back-splice site selection is correlated with the existence of multiple RNA pairs that bracket different circle-forming exons [16]. Specifically, an across-intron RNA pairing that flanks the proximal back-splice sites would lead to proximal back-splice site selection, and meanwhile, an across-intron RNA pairing that flanks the distal back-splice sites would lead to distal back-splice site selection (Figure 1C) [16]. Alternative back-splicing is more common in human than in other examined non-primate species, largely due to the abundance of primate-specific *Alu* sequences in the human genome [14,18]. Genome-wide analysis suggested that over 70% of highly-expressed circRNAs with alternative back-splicing contained alternative RNA pairing across both proximal and distal back-splice sites in human [16]. Importantly, the competition of alternative RNA pairing leading to alternative back-splice site selection could be recapitulated in expression vectors [16].

Although having the same genomic background, i.e., the same *Alu* sequences in introns, expression of alternative back-spliced circRNAs is largely diverse in

various examined human cell lines/tissues [16]. It thus suggests other layers of regulation on alternative back-splicing. Very recently, by using genome-wide siRNA screening and an efficient circRNA expression reporter, we have identified a series of protein factors as key regulators in circRNA biogenesis [33], including those involved in immune responses. Some protein factors regulate circRNA biogenesis in a combinatorial manner by especially associating with intronic *Alu* elements [33,34]. Interestingly, a number of previously-unannotated exons were identified in circRNAs, but not in their linear RNA counterparts [16]. It suggests that the canonical spliceosomal machinery might recognize and catalyze back-splicing in a different manner to splicing. Nevertheless, the finding of widespread alternative back-splicing increases our knowledge on circRNA biogenesis and its regulation.

CONCLUDING REMARKS

By identifying reads mapped to back-splice junction sites, circRNAs have been detected genome-wide in various cells/tissues, and are quantified somewhat by back-splice junction reads. Yet, no direct expression comparison has been established between circRNAs and their correlated linear RNAs bioinformatically, since they share most genomic sequences and that different strategies are applied for their quantification. On the one hand, linear RNAs are determined by normalized reads aligned to full-length genomic regions; on the other hand, circRNAs are determined solely by reads spanning back-splice junction sites (Figure 1A). Nevertheless, biochemical evidence has suggested that the expression of circRNAs is roughly 5%−10% of their linear RNAs by qPCRs analysis of selected cases [8]. In the future, it is of great interest to develop new pipeline(s) to directly compare abundance of circRNAs and their correlated linear RNAs in a genome-wide manner.

Global analyses have also revealed complex roles of *cis* intronic complementary sequences on circRNA biogenesis [10,22]. Among these complementary sequences, *SINE*s (short interspersed nuclear repetitive DNA elements), especially *Alu*s in human, contribute the most for circRNA formation [18]. The existence of primate-specific *Alu*s also results in the complex alternative back-splice selection in human [18]. It should be noted though that some circRNAs are processed without the regulation of complementary sequences, Such as *CDR1as* [7]. Interestingly, *trans*-factors are also suggested to be involved in circRNA formation together with *cis*-complementary sequences [33,34]. Finally, how circRNAs are exported and degraded also remains to be addressed. Fully understanding of the life-cycle of circRNAs and its regulation will provide molecular basis for elucidating their functions.

ACKNOWLEDGEMENTS

We are grateful to L.-L.C. for critical reading of this manuscript. We apologize to authors whose work could not be cited here owing to space/content limitations. Our work is supported by grants 2014CB910601 from MOST and 91540115 and 31471241 from NSFC.

COMPLIANCE WITH ETHICS GUIDELINES

The authors Qin Yang, Ying Wang and Li Yang declare that they have no conflict of interests.

This article is a review article and does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

1. Chen, L. L. (2016) The biogenesis and emerging roles of circular RNAs. Nat. Rev. Mol. Cell Biol., 17, 205–211
2. Chen, L. L. and Yang, L. (2015) Regulation of circRNA biogenesis. RNA Biol., 12, 381–388
3. Lasda, E. and Parker, R. (2014) Circular RNAs: diversity of form and function. RNA, 20, 1829–1842
4. Yang, L. (2015) Splicing noncoding RNAs from the inside out. WIREs RNA, 6, 651–660
5. Zhang, Y., Zhang, X. O., Chen, T., Xiang, J. F., Yin, Q. F., Xing, Y. H., Zhu, S., Yang, L. and Chen, L. L. (2013) Circular intronic long noncoding RNAs. Mol. Cell, 51, 792–806
6. Jeck, W. R., Sorrentino, J. A., Wang, K., Slevin, M. K., Burd, C. E., Liu, J., Marzluff, W. F. and Sharpless, N. E. (2013) Circular RNAs are abundant, conserved, and associated with ALU repeats. RNA, 19, 141–157
7. Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S. D., Gregersen, L. H., Munschauer, M., *et al.* (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. Nature, 495, 333–338
8. 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
9. 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
10. 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
11. 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
12. 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
13. Zhang, Y., Yang, L. and Chen, L. L. (2016) Characterization of Circular RNAs. Methods Mol. Biol., 1402, 215–227
14. Chen, L. L. and Yang, L. (2017) *ALU*ternative regulation for gene expression. Trends Cell Biol., 27, 480–490

15. Hansen, T. B., Veno, M. T., Damgaard, C. K. and Kjems, J. (2016) Comparison of circular RNA prediction tools. Nucleic Acids Res., 44, e58
16. Zhang, X. O., Dong, R., Zhang, Y., Zhang, J. L., Luo, Z., Zhang, J., Chen, L. L. and Yang, L. (2016) Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. Genome Res., 26, 1277–1287
17. Jeck, W. R. and Sharpless, N. E. (2014) Detecting and characterizing circular RNAs. Nat. Biotechnol., 32, 453–461
18. Dong, R., Ma, X. K., Chen, L. L. and Yang, L. (2016) Increased complexity of circRNA expression during species evolution. RNA Biol., 1–11. doi: 10.1080/15476286.2016.1269999
19. Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H. T., Orejuela, M. R., Piechotta, M., Levanon, E. Y., Landthaler, M., Dieterich, C., *et al.* (2015) Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. Cell Rep., 10, 170–177
20. Rybak-Wolf, A., Stottmeister, C., Glazar, P., Jens, M., Pino, N., Giusti, S., Hanan, M., Behm, M., Bartok, O., Ashwal-Fluss, R., *et al.* (2015) Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. Mol. Cell, 58, 870–885
21. Westholm, J. O., Miura, P., Olson, S., Shenker, S., Joseph, B., Sanfilippo, P., Celniker, S. E., Graveley, B. R. and Lai, E. C. (2014) Genome-wide analysis of *Drosophila* circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. Cell Rep., 9, 1966–1980
22. 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
23. 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
24. Zhang, Y., Xue, W., Li, X., Zhang, J., Chen, S., Zhang, J. L., Yang, L. and Chen, L. L. (2016) The biogenesis of nascent circular RNAs. Cell Rep., 15, 611–624
25. Liang, D. and Wilusz, J. E. (2014) Short intronic repeat sequences facilitate circular RNA production. Genes Dev., 28, 2233–2247
26. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) Initial sequencing and analysis of the human genome. Nature, 409, 860–921
27. Chen, L. L., DeCerbo, J. N. and Carmichael, G. G. (2008) *Alu* element — mediated gene silencing. EMBO J., 27, 1694–1705
28. Guarnerio, J., Bezzi, M., Jeong, J. C., Paffenholz, S. V., Berry, K., Naldini, M. M., Lo-Coco, F., Tay, Y., Beck, A. H. and Pandolfi, P. P. (2016) Oncogenic role of fusion-circRNAs derived from cancer-associated chromosomal translocations. Cell, 165, 289–302
29. 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
30. 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
31. Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., Davey, N. E., Humphreys, D. T., Preiss, T., Steinmetz, L. M., *et al.* (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell, 149, 1393–1406
32. He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D. C., Wilusz, J. E., Garcia, B. A. and Bonasio, R. (2016) High-resolution mapping of RNA-binding regions in the nuclear proteome of embryonic stem cells. Mol. Cell, 64, 416–430
33. Li, X., Liu, C. X., Xue, W., Zhang, Z., 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, http://doi.org/10.1016/j.molcel.2017.05.023
34. Aktaş, T., Avşar Ilık, İ., 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

