Cancer Letters 404 (2017) 8–18

Contents lists available at ScienceDirect

Cancer Letters

journal homepage: www.elsevier.com/locate/canlet

Mini-review

Recent progress in circular RNAs in human cancers

Guanqun Huang [a, 1], Shuaihu Li [a, 1], Nuo Yang [b], Yongdong Zou [c], Duo Zheng [a, b, *], Tian Xiao [a, *]

[a] Shenzhen Key Laboratory of Translational Medicine of Tumor, Department of Cell Biology and Genetics, Shenzhen University Health Sciences Center, Shenzhen, Guangdong, China
[b] National Center for International Research of Biological Targeting Diagnosis and Therapy, Guangxi Key Laboratory of Biological Targeting Diagnosis and Therapy Research, Collaborative Innovation Center for Targeting Tumor Diagnosis and Therapy, Guangxi Medical University, Nanning, Guangxi, China
[c] Shenzhen Key Laboratory of Microbial Genetic Engineering, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, Guangdong, China

ARTICLE INFO

Article history:
Received 7 May 2017
Received in revised form
30 June 2017
Accepted 3 July 2017

Keywords:
Circular RNAs
MicroRNA sponge
Transcription regulation
m^{6}A-driven translation
Cancer biomarkers

ABSTRACT

Circular RNAs (circRNAs) are a large class of endogenous RNAs, formed by exon skipping or back-splicing events as covalently closed loops, which are expressed abundantly in mammalian cells. Although their biological functions remain largely unknown, recent studies show that circRNAs have three main functions in mammalian cells. First, circRNAs can regulate transcription and RNA splicing. Second, circRNAs function as microRNA (miRNA) sponges. Third, they can be translated into protein driven by N^6-methyladenosine modification. Taking advantage of RNA sequencing (RNA-seq) technology, the expressions of circRNAs were found to be dysregulated in all types of cancer cell lines, tumor tissues, and even plasma samples from patients, which correlated with certain clinical characteristics, suggesting the potential roles of circRNAs in tumor progression. Considering their conserved sequences and stable structures, circRNAs were deemed to be promising biomarkers for the early diagnosis and prognosis prediction of cancer. In this review, we describe briefly the formation and properties of circRNAs, and focus mainly on recent progress in research into their function, regulation, and clinical relevance in different cancers.

Introduction

Circular RNAs (circRNAs) are a large class of endogenous RNAs that are formed by exon skipping or back-splicing events; however, they attracted little attention until their function in post-transcriptional regulation of gene expression was discovered. Potato spindle tuber viroid (PSTVd) was the first identified circRNA in 1976 when researchers studying potato spindle tuber disease observed that the viroid could infect plants and cause death. Different from viruses, the viroid lacks a protein envelope and the genome is a closed, single-stranded RNA molecule [1]. In 1979, Hsu and Coca-Prados observed the presence of a circular form of RNA in the cytoplasm of several eukaryotic cells using electron microscopy [2].

In the early 1990s, circRNAs in higher eukaryotes were discovered. The first clue to the mechanism of endogenous circRNA generation emerged from studies of the transcripts of the tumor suppressor gene *DCC*. Very low levels of transcripts of the *DCC* gene with exons joined accurately at consensus splice sites were found in normal and neoplastic cells, primarily in the nonpolyadenylated component of cytoplasmic RNA. These results demonstrated that the splicing process does not always pair sequential exons in the order predicted from their positions in the genome, which was called “exon scrambling” [3]. Cocquerelle et al. identified a scrambled transcript of the human *c-ets-1* gene, which is non-polyadenylated and is expressed at much lower levels than the normal transcript [4]. They further determined the structure of these transcripts as circular RNA molecules containing only exons in the genomic order [5]. In adult mouse testis, circular transcripts of the testis-determining gene sex-determining region Y (*Sry*) were detected. These circular RNAs, which represent the most abundant transcript in the testis, are located in the cytoplasm, but are not bound substantially to polysomes [6]. Later, atypical RNA molecules containing an incomplete exon tandem repetition, or having exons

* Corresponding authors. Department of Cell Biology and Genetics, Shenzhen University Health Sciences Center, 3688 Nanhai Ave, Shenzhen, Guangdong Province, 518060, China.
E-mail addresses: dzheng@szu.edu.cn (D. Zheng), txiao@szu.edu.cn (T. Xiao).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.canlet.2017.07.002
0304-3835/

with a different order compared with the corresponding genomic DNA, were identified from the *Drosophila melanogaster muscleblind* (*mbl*) locus. Considering its lack of polyadenylation and downstream splicing events, its small size, and polyacrylamide gel electrophoresis (PAGE) behavior, the non-canonical transcript *mblE2E2′* was deemed to be the first identified circular RNA in invertebrates [7].

Recently, taking advantage of RNA-seq technology and bioinformatic tools, more circRNAs have been discovered and characterized. In 2012, Salzman et al. developed an algorithm to detect scrambled exons in RNA-seq datasets of five bone marrow samples from pediatric acute lymphoblastic leukemia (ALL). They identified more than 1232 genes with evidence of exon scrambling, which were further validated by reverse transcription polymerase chain reaction (RT-PCR). Intriguingly, all examples of exon scrambling were also detected in peripheral blood collected from the same ALL patients and H9 ES cells, providing strong support for the view that a circular RNA isoform resulting from a non-canonical mode of RNA splicing is actually a general feature of the gene expression program in diverse human cells [8]. This group further developed a new bioinformatic approach and investigated circRNA expression using published RNA-Seq data from *Drosophila* brains and a series of cancer and non-cancer cell lines. They showed that circRNA expression was conserved evolutionarily across model organisms. In addition, the expression profiles, the ratio of circular to linear transcripts, and the pattern of splice isoforms of circRNAs of individual genes were cell-type specific. They also estimated that circular RNAs might account for about 1% of poly(A) RNA in humans [9]. Jeck et al. identified over 25,000 distinct RNAs containing non-linear exons from 14.4% of actively transcribed genes in human fibroblasts. Surprisingly, the abundance of certain circRNAs was 10-fold more than that of associated linear mRNAs. Bioinformatic analysis revealed that these circularized exons were always flanked by long introns that contained complementary ALU repeats. Moreover, they found that circRNAs could be degraded by siRNAs, suggesting their potential role as competing endogenous RNAs [10,11]. Combining ribominus sequencing data for HEK293 cells and human leukocyte data, Memczak et al. identified 1950 kinds of human circRNAs from at least two independent junction-spanning reads. They also identified 1903 mouse circRNAs (81 of these mapped to human circRNAs) and 724 circRNAs from various *C. elegans* developmental stages. Further studies focused on a human circRNA, which is antisense to the cerebellar degeneration-related protein 1 transcript (CDR1as), and identified the regulatory potential of circRNA as an miRNA antagonist [12]. Thereafter, they sequenced RNA in human peripheral whole blood and detected thousands of circRNAs reproducibly. Hundreds of circRNAs were expressed at much higher levels than the corresponding linear mRNAs, which were not accessible by classical mRNA specific assays, demonstrating the potential of circRNAs as biomarkers for human disease in an easily accessible body fluid [13].

The biogenesis of CircRNAs

Most eukaryotic genes are split genes, in which exons are interrupted by sections of introns, so the precursor mRNA (pre-mRNA) transcripts must be modified such that non-coding introns are removed and protein coding exons are joined together. In rare cases, the splicing machinery fails to join the 3′ end of one exon to the 5′ end of the next and instead, appears to mis-splice by, for example, connecting the downstream 5′ splice site (5′ ss) to an upstream 3′ splice site (3′ ss), thereby generating a circular transcript [14,15]. Thus, for a long time, circRNAs were thought to be rare errors or 'splicing noise' and few studies were related to the mechanisms of their biogenesis.

Currently, circRNAs in human cells are mainly derived from single or several exons, the so-called exonic circRNAs (ecircRNAs). However, a number of studies also showed that the splicing mechanisms of circRNAs are complicated and they can be generated from a variety of gene structures. Furthermore, the same position of a gene can produce different types of circRNAs [16,17]. According to the constituent sections, three other types of circRNAs have been defined. One is termed circular intronic RNAs (ciRNAs), which contain introns only. Another is termed exon-intron circRNAs (EIciRNAs), which contain both exons and introns. Finally, tRNA introns can form stable circRNA *via* pre-tRNA splicing, which are called tRNA intronic circRNAs (tricRNAs) (Fig. 1).

CircRNAs are derived from pre-mRNAs; therefore, it is believed that the regulation of circRNA biogenesis also requires the canonical spliceosomal machinery. Recently, Jeck et al. proposed two models of circRNAs formation. Model 1 is termed lariat-driven circularization or exon skipping. A partially folded pre-mRNA transcript brings the original non-adjacent exons close to each other, and then exon skipping occurs, resulting in a crossed region that forms a lariat intermediate containing the exons. Next, the introns in the lariat are removed, generating ecircRNAs [10]. Model 2 is termed intron-pairing driven circularization, or direct backsplicing. Circular structures are formed *via* base-pairing of ALU complementarity, or other RNA secondary structures across flanking introns, resulting in the downstream splice donor being connected to an upstream splice acceptor [14]. The introns are then removed or retained to form ecircRNAs or EIciRNAs [10,18]. Another model of circRNAs biogenesis, which is independent of the flanking sequences, was also reported. Interactions between RNA binding proteins (RBPs) form a bridge between the flanking introns, which brings the splice donor and acceptor close to each other, thereby promoting the circularization of exons [19,20].

CiRNAs were first discovered by Zhang and colleagues in human cells and are derived only from introns. CiRNAs biogenesis depends on a consensus motif containing a 7-nt GU-rich element near the 5′ splice site and an 11-nt C-rich element near the branchpoint site. Interestingly, ciRNAs are localized mainly in the nucleus and have little enrichment for miRNA target sites. In addition, ciRNA can associate with the elongation RNA polymerase II (Pol II) machinery to regulate Pol II transcription positively, suggesting that the function of ciRNAs may be different from cytoplasmic circRNAs [21]. TricRNAs are a special type of intronic circRNAs that are generated during pre-tRNA splicing and have been discovered in Archaea [22,23] and *Drosophila* [24]. The biogenesis of tricRNAs requires anciently conserved tRNA sequence motifs and processing enzymes. Excised tRNA introns are removed by tRNA splicing enzymes and then circular RNAs are formed by a 3′, 5'-phosphodiester linkage between the two termini of the introns derived from their pre-tRNAs.

The biological function of CircRNAs

Although thousands of circRNAs have been identified using RNA-seq in diverse cell types from several model organisms, and circular transcripts are the predominant isoforms of hundreds of human genes, the biological functions of most circRNAs remain unknown. Recent studies have demonstrated that circRNAs have important physiological functions *via* binding to RBPs or other proteins, neutralizing endogenous miRNAs, or even translation into proteins, implying that circRNAs may regulate gene expression at multiple levels (Fig. 2).