nature biotechnology

REVIEW

Detecting and characterizing circular RNAs

William R Jeck[1] & Norman E Sharpless[1–3]

Circular RNA transcripts were first identified in the early 1990s but knowledge of these species has remained limited, as their study through traditional methods of RNA analysis has been difficult. Now, novel bioinformatic approaches coupled with biochemical enrichment strategies and deep sequencing have allowed comprehensive studies of circular RNA species. Recent studies have revealed thousands of endogenous circular RNAs in mammalian cells, some of which are highly abundant and evolutionarily conserved. Evidence is emerging that some circRNAs might regulate microRNA (miRNA) function, and roles in transcriptional control have also been suggested. Therefore, study of this class of noncoding RNAs has potential implications for therapeutic and research applications. We believe the key future challenge for the field will be to understand the regulation and function of these unusual molecules.

Circular RNAs (circRNAs) are a recent addition to the growing list of types of noncoding RNA. Although biologists have known of the existence of circular transcripts for at least 20 years[1], such molecules were long considered molecular flukes—artifacts of aberrant RNA splicing[2] or specific to a few pathogens, such as the hepatitis δ virus[3] and some plant viroids[4]. However, recent work has revealed large numbers of circRNAs that are endogenous to mammalian cells, and many of these are abundant and stable. CircRNAs can arise from exons (exonic circRNA) or introns (intronic circRNA); these are distinct species with independent modes of generation. Evidence of potential functions in the regulation of gene expression is emerging for both exonic and intronic circRNAs[5–7].

Most circRNAs have eluded identification until recently for several reasons. CircRNAs, unlike miRNAs and other small RNAs, are not easily separated from other RNA species by size or electrophoretic mobility. Commonly used molecular techniques that require amplification and/or fragmentation destroy circularity, and because circRNAs have no free 3′ or 5′ end, they cannot be found by molecular techniques that rely on a polyadenylated free RNA end (such as rapid amplification of cDNA ends (RACE), or poly(A) enrichment of samples for RNA-seq studies). Furthermore, a key feature of circRNAs, an out-of-order arrangement of exons known as a 'backsplice' (described below), is not unique to circRNAs, and early RNA-seq mapping algorithms filtered out such sequences. These problems have recently been addressed through the development of exonuclease-based enrichment approaches, novel bioinformatic tools, sequencing with longer reads and higher throughput, and sequencing of ribosomal RNA (rRNA)-depleted RNA libraries (rather than polyA-enriched libraries).

The first hint of endogenously produced circRNAs emerged in the early 1990s from studies of the *DCC* transcript in human cells[1]. The authors of that study described transcripts with exons out of the expected order: 5′ exons were 'shuffled' downstream of 3′ exons. Despite the noncanonical ordering, the exons were complete and used the usual splice donor and acceptor sites. This arrangement was referred to as 'exon shuffling' (distinct from the evolutionary process described by Gilbert[8]). The observed shuffled transcripts were less abundant than the expected transcripts by several orders of magnitude and were nonpolyadenylated, predominantly cytoplasmic, and expressed in human and rat tissues. The authors speculated that such a product might emerge from intramolecular (*cis*) splicing, which would result in an exonic circRNA. A site at which the 3′ 'tail' of an expected downstream exon within the gene is joined to the 5′ 'head' of an exon that is normally upstream is referred to as a 'backsplice'. Early studies also detected circRNAs by electron microscopy[3,9], but this approach cannot easily distinguish circRNAs from RNA lariats (which are by-products of RNA splicing)[10].

Subsequent reports identified shuffled transcripts from several other genes, including *ETS-1* (refs. 2,11), *Sry*[12] and cytochrome P450 2C24 (*CYPIIC24*)[13,14] in human, mouse and rat cells. In each case, discovery began with the serendipitous observation of PCR products with backsplice sequences. *Sry* is usually unspliced, but sites with the canonical splice site GT/AG sequence motifs were involved in the backsplice, suggesting the involvement of the canonical spliceosome. The splice junctions used in the exonic circRNA forms of *ETS-1* and *CYPIIC24* used splice donor and acceptor sites also involved in forward splicing[11,13]. A few additional circRNAs were identified in the ensuing two decades[15–18], but they were generally much less abundant than the linear products of their source gene. Therefore, before the era of massively parallel sequencing, circRNAs were considered oddities of uncertain importance.

In this Review, we discuss methods for the identification of endogenous circRNAs, including molecular methods and genome-wide approaches, with a focus on the advantages and disadvantages of various techniques. Next, we consider the findings from these genomic studies, focusing on exonic circRNAs, and describe the biochemical properties of circRNAs *in vivo*, including methods for validation of circularity. Finally, we discuss known and predicted circRNA functions and speculate on possible applications of the newfound understanding of this molecular species.

[1]Department of Genetics, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA. [2]Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA. [3]The Lineberger Comprehensive Cancer Center, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA. Correspondence should be addressed to N.E.S. (nes@med.unc.edu).

Received 22 November 2013; accepted 7 March 2014; published online 8 May 2014; doi:10.1038/nbt.2890

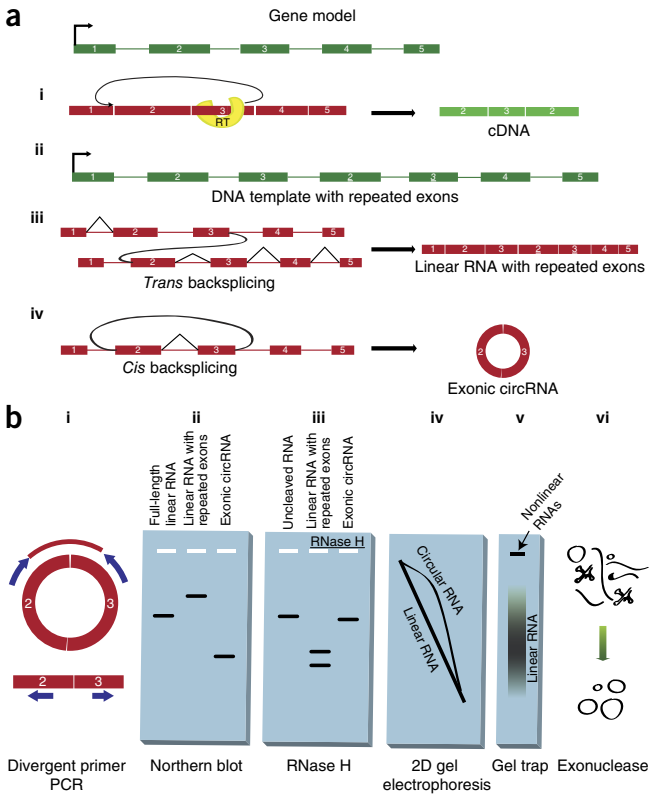

Figure 1 Splicing products and methods for detection. (**a**) Several mechanisms can form an apparent backsplice, illustrated here for a gene model in which exons are shown as rectangles, introns as thin lines and the transcription start site as a right-angled arrow. DNA is in green and RNA is in red. Reverse transcriptase (RT) template switching, in which the RT enzyme transcribes another copy of an upstream exon (**i**). Tandem duplications in the DNA template resulting in repeated exons (**ii**). *Trans*-backsplicing, in which one RNA molecule is spliced to another (shown by curved black line) (**iii**). Regular splicing events are shown as angled black lines. Exonic circRNAs can form by *cis*-backsplicing in which exons from the same RNA molecule are spliced together to form a circle (backsplicing shown by curved black line) (**iv**). (**b**) Molecular assays to distinguish exonic circRNA from other backsplice products and diagrammatic representation of the expected results. Divergent primers that would amplify in outward facing directions with respect to genomic sequence become properly inward facing and produce discreet amplicons when a backsplice connects outside sequences (**i**). The expected migration distance of a canonical linear RNA in a denaturing agarose gel, as well as the relative migration of exonic circRNA and RNAs resulting from *trans*-splicing or tandem duplication (**ii**). Migration of RNA through an agarose gel before and after RNAse H treatment. CircRNA, uniquely, results in a single band after being cut once (**iii**). 2D gel electrophoresis through two differently cross-linked polyacrylamide gels separates circRNAs into an off-diagonal curve (**iv**). Gel-trapping holds circRNAs in the well of an electrophoresis gel as linear RNAs migrate away (**v**). Exonuclease enrichment degrades linear RNAs while leaving a pool enriched for circRNA (**vi**).

Detection methods

Identifying backsplices Although we have become more knowledgeable about circRNAs mainly because of sequencing-based methods, analysis of the molecular characteristics of circRNAs is fundamental to the various strategies that we discuss in this section. Observation of sequences consistent with backsplice formation is crucial evidence of exonic circRNA production. We define an 'apparent backsplice' sequence as any case in which the ordering of the exons in a sequence is reversed relative to the annotated template. Importantly, apparent backsplice sequences may be produced by mechanisms other than formation of exonic circRNA, such as reverse transcriptase template switching, tandem duplication and RNA *trans*-splicing (**Fig. 1a**).

Reverse transcriptase template switching (**Fig. 1a**, **i**) is an artifact of cDNA synthesis, occurring when an extending cDNA molecule dissociates from its template RNA and resumes extension from another RNA template, often in a homology-dependent manner. This effect produces spurious evidence of backsplice-containing products and is known to confound the analysis of rare splicing products[19,20]. However, template switching is largely random and is not expected to produce abundant cDNA molecules of identical sequence. Therefore, high abundance of a particular apparent backsplice sequence in a cDNA library offers evidence that the sequence is also present in the RNA template. This may be assessed either by identification of multiple unique reads in deep sequencing data or by 'divergent' qPCR primers. These divergent primers are oriented to amplify away from each other in a genomic context but become 'convergent' and amplify a discrete amplicon when a backsplicing event brings outside sequences together (**Fig. 1b**, **i**). Alternatively, the presence of the backsplice sequence in the RNA pool may be assessed directly by RNase protection[2] or northern blot probing for the backsplice sequence (**Fig. 1b**).

Apparent backsplice sequences can also arise from tandem DNA duplications that can generate duplicated exons within a gene. When these sequences are transcribed, the mRNA contains an apparent backsplice sequence (**Fig. 1a**, **ii**) owing to the difference between the annotated template sequence and the DNA template present in the cell. In addition, *trans*-splicing—a process in which two distinct molecules participate in splicing—can generate apparent backsplices when it occurs between two RNA molecules originating from the same gene (**Fig. 1a**, **iii**).

There are several approaches to distinguish these species from true exonic circRNA (**Fig. 1a**, **iv**). Linear exonic RNAs will usually have 3′ polyadenylation, whereas circles have no 3′ end. Exonic circRNAs migrate more slowly in a gel than linear RNA of the same length, and this effect is augmented by increased gel cross-linking[21] (**Fig. 1b**). However, exonic circRNAs also contain less total nucleotide sequence than full-length, *trans*-spliced or tandem-duplicated transcripts from the same gene, and therefore will migrate faster in a gel that has low cross-linking (**Fig. 1b**, **ii**). Standard (or virtual[22,23]) northern blot analysis can be used to assess these characteristics. A more conclusive assay uses either weak hydrolysis or targeted RNase H degradation[12]. A circRNA will be linearized into a single product of a predictable size after RNase H degradation or after a single nick by hydrolysis (**Fig. 1b**, **iii**). Additional methods offering strong evidence of circularity are two-dimensional (2D) gel electrophoresis[21] and gel trap electrophoresis[24,25] (**Fig. 1b**, **iv-v**). In 2D gel electrophoresis, circRNAs are revealed by their poor migration through highly cross-linked gels relative to less cross-linked gels. In gel trap electrophoresis, circRNA mixed with melted agarose becomes trapped by cross-links and does not migrate in an applied electric field. Enzymatic methods can also provide evidence of circularity. RNase R exonuclease[26], tobacco acid phosphatase[5] and terminator exonuclease treatment[5] leave circRNA intact but efficiently degrade most linear RNAs. In all three cases, quantification of a specific RNA species before and after treatment should reveal enrichment of circular transcripts. Finally, with sufficiently long sequencing reads or paired-end reads, it should be possible to identify sequences that are inconsistent with circRNA[27]. These could include an apparent backsplice sequence but also include sequences from exons outside of the backsplice coordinates. For example, an apparent backsplice from exon 3 to exon 2 in a longer sequence that

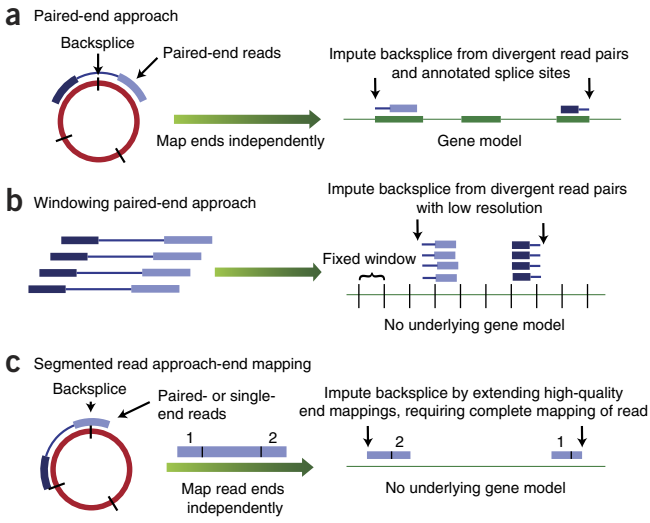

Figure 2 Sequencing-based methods for identification of exonic circRNAs. Several informatics methods have been used to identify the locations of backsplices using deep sequencing data from rRNA-depleted RNA. (**a**) Paired-end reads may be mapped separately to the annotated transcriptome and the location of a backsplice inferred when the pairs have opposite orientations on either side of one or more splice sites. (**b**) Paired-end reads may be mapped directly to the genome, and multiple reads that map out of genomic order suggest there might be a nearby backsplice. Stronger evidence is provided by the presence of multiple such reads accumulating in fixed "windows" of set length tiled across the genome. (**c**) A read can be segmented so that different parts of the read can be mapped to different parts of the genome; this allows backsplices to be mapped at nucleotide resolution without existing annotations.

includes exon 4 or exon 5 (discussed further below) would be best explained by a *trans*-spliced RNA. Each of the above methods have limitations and are best used in combination to validate circRNAs.

It is also important to distinguish exonic circRNA from RNA lariats. Lariat RNA is formed during canonical RNA splicing; it is mostly intronic and is made biochemically distinct from exonic circular species by the presence of a 2′-5′ carbon linkage at the splicing branch point. Recent genomic analysis of lariats by assembly of highly expressed, nonpolyadenylated intronic sequences[7] has made it apparent that many lariat RNAs may be more stable than was previously appreciated. The 3′ tails of these stable lariat RNAs degrade, leaving a remnant molecule[7]. Such lariat products have been called 'circular intronic RNA' (we use the term 'intronic circRNA') and are further distinguished from exonic circRNA by the presence of a 2′-5′ junction[7]. Exonic circRNAs do not feature a 2′-5′ linkage but consist of 3′-5′ links throughout the molecule.

Lariat RNAs behave similarly to exonic circRNA in the assays described above. They are largely exonuclease insensitive[26], migrate more slowly than linear molecules[21] and will form a single band when nicked within the loop component of the lariat. However, exonic circRNA is easily distinguished from intronic circRNA and lariat RNA by the features of the apparent backsplice sequence. Reverse transcription can occur inefficiently across the branch point of a lariat, producing a sequence that is superficially similar to a backsplice in having juxtaposed upstream and downstream sequences. But these branch point–traversing sequences include intronic sequences, specifically the canonical GT of the splice donor site and sequences 5′ of the branch-point nucleotide. In addition, when reverse transcriptase traverses the 2′-5′ junction, one or more untemplated bases are generated, which are readily identifiable in sequencing[23,28]. This signature can be used to rule out exonic circRNA origins for the intronic products. In principle, lariat RNAs could be preferentially depleted from other circular species by treatment with debranching enzyme before exonuclease digestion, as this enzyme selectively hydrolyzes 2′-5′ linkages. Therefore, although lariats are common circRNA molecules, they can be readily distinguished from exonic circRNA.

Genomic methods Recent genome-wide studies of circRNA have been enabled by developments in sequencing technology (deeper sequencing with longer read lengths), better algorithms for mapping RNA to its genomic source and ribosomal RNA depletion strategies that enable sequencing of nonpolyadenylated RNA. In general, two approaches have been taken: first, using a list of candidate junctions generated from existing transcript models[27,29]; or alternatively, identifying junctions by matching reads to the genomic sequence as is done in spliced alignment algorithms (for more on alignment of deep sequencing data, see recent review[30]). Both of these methods have identified circRNAs that have then been validated by sequencing, RNase R exonuclease testing and other methods.

The first identification of exonic circRNAs in a genomic study occurred serendipitously through using independent mapping of paired-end reads sequenced from opposite ends of a single cDNA fragment[27]. This approach identified an unexpected abundance of fragments in which two read pairs mapped to the same gene but were in the opposite order from that expected from the gene annotation (**Fig. 2a**). Realizing that these likely arose from circRNAs, the authors imputed the location of a backsplice using the existing gene annotations. This approach was extended by scanning for out-of-order paired-end reads that were concentrated in fixed windows tiled across a gene (**Fig. 2b**). This strategy has the advantage of being a fast way to analyze rRNA-depleted libraries but it is largely candidate-based, in that it relies on candidate circRNAs constructed from preexisting gene models, and so does not detect circRNAs from unannotated transcripts and does not provide direct evidence of circularity. However, qPCR validation assays on a number of individual species discovered by this method revealed that the transcripts were predominantly resistant to RNase R exonuclease and lacked the expected properties of backsplice-containing linear RNAs (*trans*-spliced or duplication products)[27].

Other methods using rRNA-depleted libraries include identifying reads with apparent backsplice sequence from rRNA-depleted RNA-seq data from mammalian and nematode cells, without using a candidate-based approach[6]. The authors mapped reads to genomic locations *de novo* and identified apparent backsplice sequences in individual reads. They chose reads that could not be mapped directly to the genome and then mapped the two ends of a single read separately. By using single reads, the authors were able to identify the location of the putative backsplice to single-nucleotide resolution (**Fig. 2c**). The authors also selected only apparent backsplice sequences that were flanked by GT/AG splice sites in the genomic context. Therefore, this method can identify unannotated splice sites but it might be less sensitive than a candidate-based approach.

A more nuanced approach to candidate-based analysis was recently developed using rRNA-depleted RNA-seq without exonuclease enrichment[29]. This method uses deep sequencing and 75-bp paired-end reads to identify apparent backsplice sequences that map to a set of candidate junctions generated using existing gene annotations. The authors divided pairs of reads in which one read contained an apparent backsplice into two groups: a group in which the read without a backsplice mapped to an exon between the back-spliced exons, and

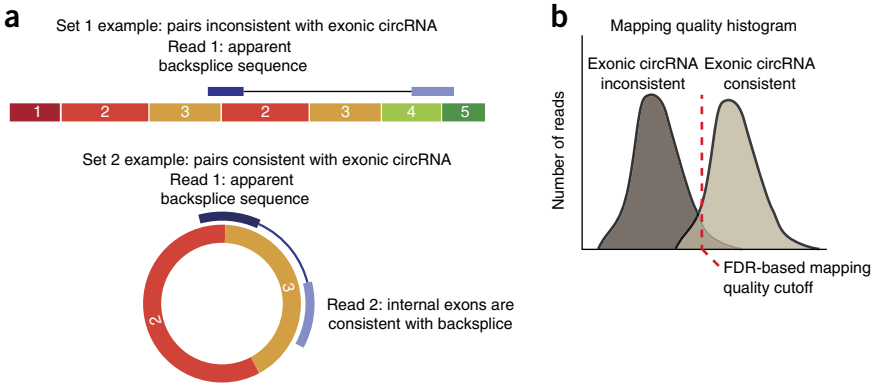

Figure 3 Informatic approach to identifying false-positive backsplices. (**a**) Paired ends that would be inconsistent with underlying exonic circRNA. In the first example (set 1), the first read (dark blue) of the pair maps to a backsplice, but its paired read (light blue) maps to an exon that is inconsistent with a circRNA because it lies in an exon beyond those involved in the backsplice. In the second example (set 2), the first read of the pair maps to the backsplice, but its pair maps to an exon between those two exons involved in the backsplice, consistent with a circRNA. (**b**) The two sets of reads are then used to generate two distributions that may be used to set empirical false-discovery rates based on mapping quality features.

therefore could potentially have arisen from a circRNA (**Fig. 3a**); and a group in which the second read in the pair mapped to an exon outside of the backspliced exons and, therefore, could not be explained by a circRNA. This second group was assumed to represent artifacts of sequencing. The distributions of mapping-quality statistics within the two groups was then used to generate a confidence score for each junction observed (**Fig. 3b**). This approach allowed the use of a false-discovery rate cutoff rather than an arbitrary read depth–based threshold.

In addition to approaches based on sequence only, a biochemical approach for the genome-wide identification of circRNA has been described by others in archaea[31] and by our group in mammals[23]. This technique, "CircleSeq" (**Fig. 4**), uses RNase R digestion before high-throughput sequencing to identify species that are RNase R resistant (**Fig. 4a**). The method uses a mapping algorithm capable of identifying apparent backsplice sequences (MapSplice[32]), rather than an algorithm requiring a choice of exon order. In mammals, but not archaea, rRNA depletion is required. Using this technique, identification of exonic circRNAs is possible on the basis of two features: first, backsplice-containing reads are identified using a segmented mapping approach (**Fig. 4b**); and second, reads derived from circular species should be significantly enriched in the RNase R–treated sample compared to mock treated-control (8- to 16-fold on average, though variable levels of enrichment were observed). The exons of linear RNAs should be depleted by exonuclease digestion, as should splice junctions not present in circRNAs. An example of data for a circRNA (*cANRIL*)[23] exhibiting these features is shown in **Figure 4c**.

CircleSeq also identifies lariat RNAs[23], though these are easily distinguished from circRNAs in the sequencing data. As mentioned, lariats show RNase R enrichment of the loop sequence, RNase R depletion of the lariat tail and the presence of untemplated bases at the branch point. Branch-point sequences resemble apparent backsplice sequences in that parts of the sequence are misordered compared to their genomic annotation. In branch-point sequences, though, the sequences at the backsplice are intronic and contain untemplated bases produced during cDNA synthesis. To date, all abundant species identified by CircleSeq that contain a true backsplice and that have been subjected to specific characterization have been demonstrated to be exonic circRNAs[23,31].

Although CircleSeq generates deep coverage of circular and lariat products, it has some limitations. It requires more input total RNA than sequencing without enrichment and is sensitive to endonuclease contamination. It might also be biased against the detection of longer circRNA products, as a single nicking event would confer exonuclease sensitivity. Finally, exonuclease protection may extend to some linear products with protective 3′ end structures, such as the 3′ triple helix seen in the long noncoding RNAs MALAT1 and Menβ[33,34], thus complicating the interpretation of the results. These caveats must be weighed against CircleSeq's depth of coverage, which lends confidence that backsplice-containing species identified through CircleSeq are circular.

Properties of circRNAs

General features Studies of circRNAs have identified a number of shared features. Exonic circRNA is very stable in cells[2], with most

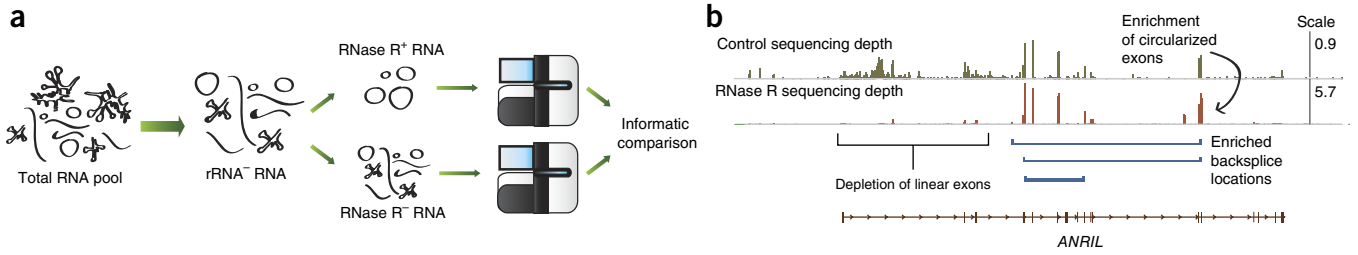

Figure 4 A combined biochemical and informatic approach to identify exonic circRNAs in mammalian cells. This figure illustrates CircleSeq, which was developed in part by our group. (**a**) Total RNA is depleted of rRNA, for example, by using RiboMinus or RiboZero methods. rRNA-depleted samples (rRNA−) are split; one aliquot is treated with the RNase R exonuclease (RNase R+) and the other is subjected to a mock treatment (RNase R−). Sequencing libraries are prepared from each and optimal RNase R digestion is confirmed by checking with quantitative RT-PCR that known exonic circRNAs are enriched in the RNase-treated sample (data not shown). Samples are then sequenced and compared. Informatic approaches, such as that shown in **Figure 2c** can be applied to the data. (**b**) An example of data from CircleSeq in human tissues for the *ANRIL* noncoding RNA locus. Normalized mapped read depth for the mock-treated sample (green, top) and for RNase R-treated samples (brown, bottom) are shown (note different scales). Regions where there is enrichment of a group of exons in the rRNA-treated sample—suggestive of a circRNA—are shown in blue. Exons outside of these regions are depleted in the RNase-treated sample.

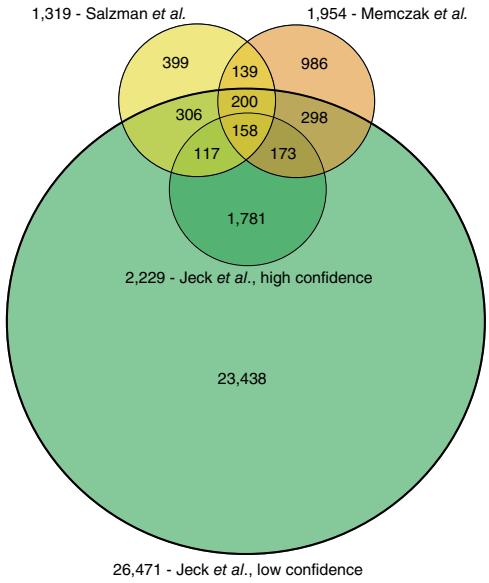

Figure 5 Comparison of circRNAs identified by genomic studies. Numbers of circRNAs identified by three genomic studies[6,23,27], and the numbers of circRNAs overlapping among these studies. Overlapping regions represent the number of species where both the splice donor and splice acceptor of the backsplice were identical in two or more works. Data from Memczak *et al.*[6] were taken from Supplementary Table 2 in that article. Two circles shown for Jeck *et al.* show a high- and low-confidence set of circles reported in that work. High-confidence circles were observed in two sequencing data sets from unenriched sequencing data, whereas low-confidence circles were observed in only one. Sites taken from Salzman *et al.*[27] are from Supplementary Table 2 in that article, which required our inference of the genomic locations using RefSeq gene annotations.

species exhibiting a half-life over 48 h (ref. 23), compared to an average half-life of 10 h for mRNAs[35]. However, exonic circRNA is not stable in serum, with a half-life of <15 s (ref. 36), presumably due to circulating RNA endonucleases[37]. Intracellular stability is likely due to circRNA resistance to RNA exonucleases[2]. Possibly due to this stability, some exonic circRNAs have been shown by sequencing read counting methods and qPCR-based methods to be at higher levels than the linear RNA gene product[23,27,29]. Exonic circRNA species also do not contain the characteristic 2′-5′ linkage of an RNA lariat and therefore are resistant to RNA debranching enzymes. Studies of exonic circRNA localization using different methods have reported cytoplasmic localization[1,2,6,23,25,38,39], though the process of nuclear export remains obscure, and it is possible that circRNA escapes from the nucleus during mitosis. Exonic circRNAs are also susceptible to decay mediated by short interfering RNA[6,23,25], a property that is useful in studying their possible functional roles.

Several shared sequence-based features have been described in exonic circRNAs. First, the exonic circRNAs described to date involve a GT-AG pair of canonical splice sites, although this is biased by discovery methods that favor or require such sites in mapping[6,23]. Furthermore, exonic circRNAs almost always use at least one previously annotated splice site. Introns flanking sites involved in a backsplice tend to be longer than introns generally, but some flanking introns can be smaller than average[29]. As was first shown for the circRNA derived from *Sry*, complementary sequences in the introns flanking backsplice sites seem to promote circularization[23]. In particular, paired Alu repeats in inverted orientations upstream and downstream of backsplice sites are enriched fivefold in sites of human exonic circRNA formation[23]. Likewise, circRNA overexpression constructs, including complementary upstream and downstream sequences, show enhanced circRNA formation relative to constructs without complementary sequences[5,6]. Lastly, the length of a given exon appears to influence circularization. This is most obvious for exonic circRNA comprising a single exon, as the exons forming single-exon circRNA are threefold longer compared to all expressed exons[23,27]. In aggregate, genomic features that appear to promote circularization are longer-than-average exons, flanked by longer-than-average introns containing inverted tandem repeats that likely promote intron pairing (described below).

Prevalence of circRNAs Findings from recent genome-scale studies of circRNA are largely consistent with each other. These analyses have found strong evidence for thousands of circRNAs in diverse human cell types[6,7,23,27]. Statistical analysis of rRNA depleted sequencing data[27], many validation approaches[6] and CircleSeq data[23] show that the majority of apparent backsplice sequences in RNA-seq data arise from circRNAs. Similarly, analyses of mouse RNA with the same techniques have revealed thousands of apparently circular transcripts[6,7,23,29]. Most human and mouse circRNAs arise from coding genes[23,27]. Work in *Caenorhabditis elegans* has found abundant circRNAs and identified life cycle–dependent regulation of some of these[6]. Recent work in human cell lines used in the ENCODE project has shown that exonic circRNA production is regulated independently of the underlying linear RNA gene and that transcription levels vary among cell types[29]. This evidence for regulated production, along with substantial conservation of exonic circRNA expression in mammals[6,23], suggests a functional role for many of these transcripts.

Given the difficulties in separating circular and linear RNAs solely on physical properties, the absolute abundance of circular species in the pool of total RNA is difficult to judge. In control replicates used for CircleSeq[23], some backsplice species were represented by as many as 1 in 300,000 reads and others by as few as 1 in 300,000,000. Thus circRNA analysis requires at least millions of reads, and preferably hundreds of millions, even in libraries with preferential enrichment by RNase R digestion. A recent estimate using paired-end reads coupled to qPCR-based quantification of some exonic circRNA transcripts estimated the relative abundance of exonic circRNA as 1% the amount of poly(A) RNA[29]. An estimate using an alternative method can be obtained using data from our recent study of human fibroblasts by identifying the number of reads with apparent backsplice sequences that were validated by RNase R enrichment. Reads mapping to these junctions comprise 0.1% of all sequencing from rRNA-depleted total RNA. As circRNAs contain sequences in addition to the backsplice, it is necessary to improve this estimate by imputing exonic circRNA length. Each backsplice-traversing read can then be weighted based on predicted circRNA length, (note though that not all intervening exons appear in all exonic circRNAs, for example, *cANRIL*[23,40], **Fig. 4c**); this gives an estimate that exonic circRNA comprises 0.8% of all nonribosomal RNA. Therefore, these different approaches suggest that a substantial fraction of nonribosomal RNA is contained in exonic circRNA.

With regard to the identification of specific circular species, recent genome-wide studies of human exonic circRNAs also have substantially overlapping findings. In total, 358 circRNAs were identified in all three studies and 1,233 were identified by at least two of the three studies (**Fig. 5**). It should be noted that the Rajewsky *et al.*[6] and Salzman *et al.*[27] analyses include some common source data, although the methodological features of these analyses differ. Substantially more circRNAs were found using CircleSeq[23] (shown on **Fig. 5** as

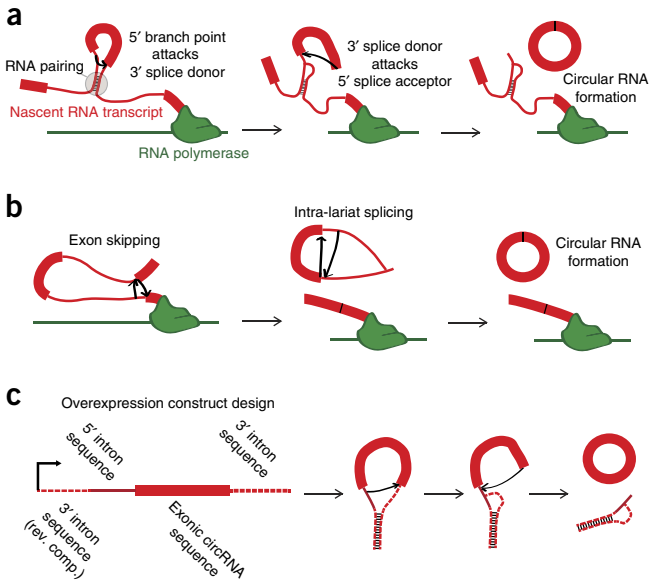

Figure 6 Possible mechanisms for formation of exonic circRNAs. (**a**) Direct backsplicing[2], in which two unspliced introns within a transcript pair and the intervening introns are spliced through the usual mechanism. Here, a branch point in the 5′ intron attacks the splice donor of the 3′ intron. The 3′ splice donor then completes the backsplice by attacking the 5′ splice acceptor. This forms a circRNA. (**b**) Exon skipping resulting in circRNA formation[53]. In this case an exon skipping event creates a lariat containing an exon. This lariat is spliced internally, removing the intronic sequence and producing a circRNA. (**c**) Design schema for an overexpression vector that produces circRNA, which was used for *CDR1as* and *Sry*[5]. Here the exon made to be circularized is included in an overexpression construct with upstream and downstream intron sequence. Additional sequence is placed 5′ in the transcript to produce pairing with the downstream intron, as in the direct backsplicing mechanism (shown). Rev. comp., reverse complement.

'Jeck *et al.*[23], low confidence'), which likely reflects a lower limitation of detection using a biochemical enrichment approach. Assuming complete rRNA depletion, 700 megabases of nonribosomal RNA per cell and equal RNA representation across sequencing libraries, then the least abundant exonic circRNAs in this low confidence data set number one copy per 80 cells. However, many circRNAs at low levels were also identified by approaches used by Rajewsky *et al.*[6] and Salzman *et al.*[27]. Analysis of CircleSeq data with more stringent filtering (requiring backsplices to be observed in multiple replicates and in samples without RNase R treatment) gives a set of tenfold fewer circRNAs (shown on **Fig. 5** as 'Jeck *et al.*[23], high confidence'), and this set has a larger overlap with circRNAs found using bioinformatic approaches (20% versus 4%). In a few instances, an abundant exonic circRNA was found by one approach but not the others, possibly due to the different cell types analyzed.

Exonic circRNA production mechanism Two mechanisms have been proposed for mammalian exonic circRNA formation (**Fig. 6**). Both involve the backsplice being formed by the canonical spliceosome. In the first mechanism—historically termed 'mis-splicing' but referred to here as 'direct backsplicing' to emphasize that these are not always errors of splicing—a downstream splice donor pairs with an unspliced upstream splice acceptor[2] and the intervening RNA is circularized (**Fig. 6a**). The second mechanism—known as the 'lariat intermediate' or 'exon skipping' mechanism—involves splicing within lariats produced from exon skipping[14] (**Fig. 6b**).

Although it is likely that both mechanisms function *in vivo,* some evidence suggests that direct backsplicing may occur more frequently than exon skipping. In some cases a linear mRNA has been found that lacks the exons that are included in a circRNA, providing evidence that exon skipping is at least a plausible mechanism of formation[14,23,27], though such events could also plausibly occur after direct backsplicing. However, genome-wide exonic circRNA discovery studies have failed to find such linear RNAs for the majority of exonic circRNAs[23,27]. It is possible that such an analysis underestimates exon skipping, given that linear products are less stable than most circular species. There are examples of exonic circRNAs with no additional exons upstream or downstream of the circularized product in an annotated linear transcript, such as in *Sry* and *CDR1as*[12,25]. For these transcripts to result from exon skipping, there would need to be unannotated upstream and downstream exons that skipped over the sequence that is eventually circularized. Given the abundance of the exonic circRNAs that result from these loci, the presence of such unannotated exons appears unlikely, further supporting direct backsplicing for the formation of these products. As additional evidence for the direct backsplicing model, constructs for the overexpression of circRNAs usually include the exon that will become circularized and partial sequences of the flanking introns[5,6,41] (**Fig. 6c**), but no additional upstream or downstream exonic sequence. These constructs successfully produce exonic circRNA without the need for additional flanking exons and, therefore, would be inconsistent with an exon skipping event.

We believe the common features of exonic circRNA producing loci suggest that some endogenous circRNAs are produced by a direct backsplicing mechanism. First, long exons are preferentially circularized, perhaps because long exons are sterically more favorable for 3′ to 5′ splicing at canonical splice sites. Also, exonic circRNA–producing genes are enriched for repeat elements in flanking introns within 500 bp of the backsplice sites. These repeat elements are more likely to be in orientations that promote RNA base pairing[23], suggesting that RNA base pairing might bring upstream splice acceptors into physical proximity with downstream splice donors. The human *HIPK3* locus appears typical in this regard (**Fig. 7**). This locus is particularly informative in that mouse *Hipk2* and *Hipk3* and human *HIPK3* genes all harbor similar genomic structures with regard to intron and exon length and also show a high degree of circularization. The human *HIPK3* locus, however, contains tandem Alu repeats, and exhibits a substantially higher degree of circularization. Nevertheless, it is worth noting that some single-exon circRNAs are as small as 204 nucleotides[29], and that many exonic circRNAs are flanked by relatively short introns. Therefore, it seems likely that multiple mechanisms may be involved in exonic circRNA generation.

Putative functions of exonic circRNAs

Genome-wide analyses have identified a large number of abundant exonic circRNAs, and cross-species comparisons have shown that sites of circularization are conserved in orthologous exons at a rate well above that expected by chance[23,27,29]. The abundance and evolutionary conservation suggest that exonic circRNAs might have specific roles in cellular physiology. Several possible functions have been proposed including miRNA binding, protein binding, regulation of translation and translation into proteins.

miRNA sponges Recently, two exonic circRNAs in mammals have been shown to function as miRNA sponges or competing endogenous RNAs[42–44]. Competing endogenous RNAs act as decoys for the binding of miRNA with their coding RNA targets, increasing the expression of those same coding RNAs with the increasing expression of the competing endogenous RNA. The exonic circRNAs of *CDR1as*

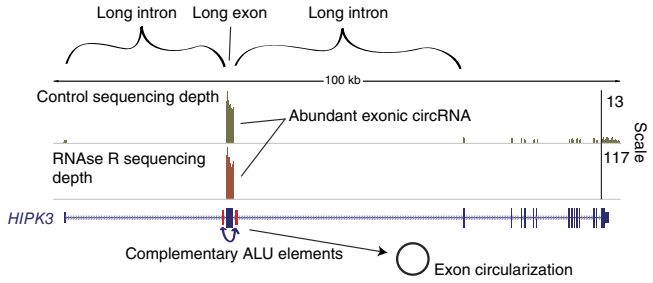

Figure 7 Genomic features of circRNAs. CircRNAs are generated from exons that are longer than average and are flanked by long introns containing inverted Alu repeats, as exemplified by the *HIPK3* gene, shown with CircleSeq data set[23]. In the *HIPK3* gene, ribosomal RNA–depleted sequencing (green) resulted in substantially more coverage of a single exon, which also demonstrated a large number of backsplice sequences. These sequences are also demonstrated to be true circles by their enrichment by RNAse R (brown), whereas all other exons are degraded. This circularized exon is the longest of the gene, is surrounded by the longest introns of the gene, and is flanked by Alu elements in a complementary orientation.

and *Sry* have been shown to bind miRNAs without being degraded, making them excellent candidates for competing endogenous RNA activity. Binding of miRNAs by exonic circRNAs was first shown for *CDR1as*[25], and support for *CDR1as* acting as a miRNA sponge comes from several lines of evidence. *CDR1as* contains 74 miR-7 seed sequence matches and is densely bound by Argonaute proteins (the proteins that bind to miRNAs)[6]. miR-7 and *CDR1as* are co-expressed in the mouse brain and *in vitro* microscopy and co-immunoprecipitation experiments have shown colocalization of *CDR1as* and miR-7 (ref. 5). Knockdown of *CDR1as* or overexpression of a miRNA previously shown to cleave *CDR1as* (miR-671)[25], decreased expression of known miR-7 target genes. By contrast, *CDR1as* overexpression prevented knockdown of miR-7 targets. Furthermore, transgenic expression of mouse *CDR1as* in zebrafish embryos, which lack endogenous *CDR1as*, substantially reduced midbrain size, mimicking the phenotype of morpholino knockdown of miR-7 in zebrafish[6]. Further loss-of-function experiments are needed to test whether endogenous *CDR1as* also regulates brain size. Similarly, the circular *Sry* transcript has 16 binding sites for miR-138 and co-precipitates with Argonaute 2 (AGO2) when miR-138 is overexpressed, and miR-138-mediated knockdown is attenuated when *Sry* is overexpressed in mouse cells[5]. Although these two examples are striking, our analysis of the large set of exonic circRNAs identified by CircleSeq suggests that very few circRNAs in mammalian cells contain more than ten binding sites for an individual miRNA (W.R.J. and N.E.S., unpublished observations). Many exonic circRNAs, however, contain a smaller number of putative miRNA binding sites. Therefore, effective miRNA sponging by exonic circRNA may be relatively unusual, or may not require the large number of miRNA binding sites that characterize *Sry* and *CDR1as*. Efforts are also under way to explore the potential use of circRNA sponges as potential therapeutics, for example, to target oncogenic miRNAs[45,46].

Regulation of transcription Mechanisms have been suggested whereby circularization could regulate transcription. For example, in mice the formin (*Fmn*) gene is essential for limb development. Exonic circRNAs are produced from the *Fmn* transcript through backsplicing involving a splice acceptor upstream of the *Fmn* coding sequence[47]. Knockout mice lacking this splice acceptor have no detectable expression of the exonic circRNAs and have normal limb development, but they have an incompletely penetrant renal agenesis phenotype[47]. The inability to produce an exonic circRNA from the targeted *Fmn* locus therefore appears to lead to aberrant expression of the formin protein, although the 5′ UTR exons deleted in the animal model could have functions unrelated to circularization. The authors propose that the formation of an exonic circRNA acts as an 'mRNA trap' by sequestering the translation start site, leaving a noncoding linear transcript and thereby reducing the expression level of the formin protein.

The mRNA trap mechanism could be widespread. For example, in human and mouse cells, exonic circRNAs are derived from the second exons of the *HIPK2* and *HIPK3* loci and this exon contains the canonical ATG (i.e., a translation start codon). In the case of *HIPK3*, the exonic circRNA is considerably more abundant than the linear, protein-coding transcript[23] (**Fig. 7**). The circularization of the ATG-containing exon would be inconsistent with the production of the canonical protein from the locus, and therefore circularization can be considered a form of 'alternative splicing' that regulates protein translation. We have observed that in human fibroblasts, 34% of single-exon circles contain a translation start compared with only 14% of all exons containing a translation start ($P < 10^{-10}$ by χ^2 test, W.R.J. and N.E.S. unpublished observations). This finding is consistent with a widespread role for circRNAs as mRNA traps that regulate protein expression.

The mRNA trap role of exonic circRNA suggests potential functions for circRNAs through their mechanism of formation, rather than as end products. For example, the dystrophin (*DMD*) gene is known to produce several circular products[48] and a study of *DMD* exonic circRNA in patients with a dystrophinopathy showed that exonic circRNA formation might lead to inactive *DMD* transcripts in individuals with certain deletion mutations[49], further reducing the pool of mRNAs that can be translated. This suggests that the mRNA-trap effect of exonic circRNAs might enhance the disease phenotype in these patients. Splicing regulation is already a target of emerging therapies for these dystrophinopathies. For example, antisense oligonucleotides against certain exons, which induce exon skipping and restore open reading frames, are currently in clinical trials[50]. Circularization might be a future target for therapies, either to decrease the circularization of functional transcripts or to sequester, through an mRNA trap, exons contributing to dysfunctional transcripts.

Interactions with RNA binding proteins It has previously been shown that some linear noncoding RNA transcripts sequester RNA binding proteins[51], and exonic circRNAs might have a similar role. For example, circRNA can stably associate with AGO proteins and RNA polymerase II (refs. 6,23). Exonic circRNAs also have some properties that suggest they might act as 'scaffolding' for RNA binding proteins by binding multiple proteins and facilitating stable interaction through the underlying increased stability of the circRNA transcript. It has also been proposed[6] that they might have roles as sequence targeting elements, binding simultaneously to RNA binding proteins and regions of RNA or DNA that are complementary to the circRNA sequence. CircRNAs could adopt tertiary structures distinct from related linear molecules of the same sequence owing to the limitations of circularization, or new protein binding sites could be generated by the sequences that are brought together in the circRNA; such features might result in circRNAs being able to bind different sets of proteins than related linear RNAs.

Translation of circRNAs It is possible that some circRNAs are translated, as inclusion of an internal ribosome entry site (IRES) allows translation of engineered circRNAs[52]. Presumably an endogenous

circRNA with an IRES and ATG could undergo translation. At least one naturally occurring circRNA is known to encode a protein in mammalian cells: the hepatitis δ agent[3], a circRNA satellite virus of the hepatitis B virus. Hepatitis δ leads to production of a single viral protein associated with pathogenicity, but the mechanism of translation is noncanonical and is probably specific to certain viral agents.

It is interesting to consider the nature of possible proteins encoded by circRNAs. If the number of base pairs is divisible by three, a protein-encoding circRNA would be read recurrently in-frame to produce a repeated polypeptide sequence, as has been experimentally demonstrated[52]. If the circRNA contained a number of base pairs not divisible by three, it would be read in alternative reading frames each time the ribosome passed around the circle, and a stop codon would usually be encountered when read out of frame. In theory, it would be possible to have an exonic circRNA read in all three alternate reading frames without ever encountering a stop (i.e., a 'Möbius protein'), but we have not identified a naturally occurring exonic circRNA with this property. We have considered the protein coding potential of many ATG-containing exonic circRNAs produced in human fibroblasts, but to date we[23] and others[29] have not been able to identify a naturally occurring exonic circRNA that undergoes translation. For example, we have not been able to find exonic circRNA undergoing translation (i.e., bound to polysomes), and we have not identified peptides in mass spectrometry data that could only occur as a result of translation of exonic circRNAs.

Conclusions

Until recently, only a handful of circRNAs had been identified in mammalian cells, and these were largely thought to have arisen from errors in RNA splicing. This view is not compatible with the recent discoveries of thousands of distinct exonic circRNAs in various human cell types and with many of these circRNAs being abundant, stable and evolutionarily conserved. There is evidence of functionality from the finding that some exonic circRNA act as miRNA sponges and apparent regulation of the expression of linear protein-encoding RNA products by 'mRNA trap' mechanisms, and it seems likely that additional functions will be described.

As studies of circRNAs proliferate, it will become increasingly necessary to develop a standard nomenclature, especially if they are to be incorporated into RNA databases, including RefSeq and UCSC genome browser annotations. Although rare exonic circRNAs—such as *CDR1as*—can be defined by their source gene, an alternative naming system is needed for cases where circRNAs arise from the gene bodies of protein-coding linear transcripts, as several genes produce multiple circRNAs. For example, the name 'circular *ANRIL*' would be ambiguous because several circRNAs can arise from *ANRIL*. The recent identification of stable intronic circRNAs further confuses the terminology. We favor a naming convention that identifies the source gene and adds a numeric identifier (e.g., ecirc*HIPK3-1* to designate the first exonic circRNA isoform of *HIPK3*). Name standardization would assist both bioinformatic and experimental research into circRNA origins and function.

Finally, exonic circRNAs have potential therapeutic roles. If properly packaged and delivered, their high cytoplasmic stability could make them long-acting regulators of cellular behavior. CircRNA overexpression constructs could be used to generate high levels of stable RNA circles in cells for a variety of purposes, such as to act as miRNA sponges to reduce the activities of oncogenic miRNAs (e.g., miR-21 and miR-221 (ref. 46)) in the context of cancer. CircRNAs containing IRES sequences could be used to produce unusual peptides, for example, long repeating polypeptides that might be useful in the production of new biologic materials. As more functions of exonic circRNAs are discovered, further uses of this newly appreciated class of RNAs are likely to emerge.

ACKNOWLEDGMENTS

The authors would like to acknowledge W. Marzluff for his assistance in preparing this manuscript, as well as funding from the National Institute on Aging, grants F30 AG041567–3 and RO1 AG024379–11.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Nigro, J.M. *et al.* Scrambled exons. *Cell* **64**, 607–613 (1991).
2. Cocquerelle, C., Mascrez, B., Hétuin, D. & Bailleul, B. Mis-splicing yields circular RNA molecules. *FASEB J* **7**, 155–160 (1993).
3. Kos, A., Dijkema, R., Arnberg, A.C., van der Meide, P.H. & Schellekens, H. The hepatitis delta (delta) virus possesses a circular RNA. *Nature* **323**, 558–560 (1986).
4. Sanger, H.L., Klotz, G., Riesner, D., Gross, H.J. & Kleinschmidt, A.K. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc. Natl. Acad. Sci. USA* **73**, 3852–3856 (1976).
5. Hansen, T.B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384–388 (2013).
6. Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338 (2013).
7. Zhang, Y. *et al.* Circular intronic long noncoding RNAs. *Mol. Cell* **51**, 792–806 (2013).
8. Gilbert, W. Why genes in pieces? *Nature* **271**, 501 (1978).
9. Arnberg, A.C., van Ommen, G.J., Grivell, L.A., Van Bruggen, E.F. & Borst, P. Some yeast mitochondrial RNAs are circular. *Cell* **19**, 313–319 (1980).
10. van der Veen, R. *et al.* Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. *Cell* **44**, 225–234 (1986).
11. Cocquerelle, C., Daubersies, P., Majérus, M.A., Kerckaert, J.P. & Bailleul, B. Splicing with inverted order of exons occurs proximal to large introns. *EMBO J.* **11**, 1095–1098 (1992).
12. Capel, B. *et al.* Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell* **73**, 1019–1030 (1993).
13. Zaphiropoulos, P.G. Differential expression of cytochrome P450 2C24 transcripts in rat kidney and prostate: evidence indicative of alternative and possibly trans splicing events. *Biochem. Biophys. Res. Commun.* **192**, 778–786 (1993).
14. Zaphiropoulos, P.G. Circular RNAs from transcripts of the rat cytochrome P450 2C24 gene: correlation with exon skipping. *Proc. Natl. Acad. Sci. USA* **93**, 6536–6541 (1996).
15. Caudevilla, C. *et al.* Natural trans-splicing in carnitine octanoyltransferase pre-mRNAs in rat liver. *Proc. Natl. Acad. Sci. USA* **95**, 12185–12190 (1998).
16. Frantz, S.A. *et al.* Exon repetition in mRNA. *Proc. Natl. Acad. Sci. USA* **96**, 5400–5405 (1999).
17. Surono, A. Circular dystrophin RNAs consisting of exons that were skipped by alternative splicing. *Hum. Mol. Genet.* **8**, 493–500 (1999).
18. Rigatti, R. Exon repetition: a major pathway for processing mRNA of some genes is allele-specific. *Nucleic Acids Res.* **32**, 441–446 (2004).
19. Cocquet, J., Chong, A., Zhang, G. & Veitia, R.A. Reverse transcriptase template switching and false alternative transcripts. *Genomics* **88**, 127–131 (2006).
20. McManus, C.J., Duff, M.O., Eipper-Mains, J. & Graveley, B.R. Global analysis of trans-splicing in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **107**, 12975–12979 (2010).
21. Tabak, H.F. *et al.* Discrimination between RNA circles, interlocked RNA circles and lariats using two-dimensional polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **16**, 6597–6605 (1988).
22. Hurowitz, E.H. & Brown, P.O. Genome-wide analysis of mRNA lengths in *Saccharomyces cerevisiae*. *Genome Biol.* **5**, R2 (2003).
23. Jeck, W.R. *et al.* Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**, 141–157 (2013).
24. Schindler, C.W., Krolewski, J.J. & Rush, M.G. Selective trapping of circular double-stranded DNA molecules in solidifying agarose. *Plasmid* **7**, 263–270 (1982).
25. Hansen, T.B. *et al.* miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J.* **30**, 4414–4422 (2011).
26. Suzuki, H. *et al.* Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. *Nucleic Acids Res.* **34**, e63 (2006).
27. Salzman, J., Gawad, C., Wang, P.L., Lacayo, N. & Brown, P.O. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS ONE* **7**, e30733 (2012).
28. Gao, K., Masuda, A., Matsuura, T. & Ohno, K. Human branch point consensus sequence is yUnAy. *Nucleic Acids Res.* **36**, 2257–2267 (2008).
29. Salzman, J., Chen, R.E., Olsen, M.N., Wang, P.L. & Brown, P.O. Cell-type specific features of circular RNA expression. *PLoS Genet.* **9**, e1003777 (2013).

30. Fonseca, N.A., Rung, J., Brazma, A. & Marioni, J.C. Tools for mapping high-throughput sequencing data. *Bioinformatics* **28**, 3169–3177 (2012).
31. Danan, M., Schwartz, S., Edelheit, S. & Sorek, R. Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res.* **40**, 3131–3142 (2012).
32. Wang, K. *et al.* MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res.* **38**, e178 (2010).
33. Brown, J.A., Valenstein, M.L., Yario, T.A., Tycowski, K.T. & Steitz, J.A. Formation of triple-helical structures by the 3′-end sequences of MALAT1 and MENβ noncoding RNAs. *Proc. Natl. Acad. Sci. USA* **109**, 19202–19207 (2012).
34. Wilusz, J.E. *et al.* A triple helix stabilizes the 3′ ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev.* **26**, 2392–2407 (2012).
35. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
36. Umekage, S., Uehara, T., Fujita, Y., Suzuki, H. & Kikuchi, Y. in *Innovations in Biotechnology* (ed. Agbo, E.C.) 75–90. http://cdn.intechopen.com/pdfs/28708/InTech-In_vivo_circular_rna_expression_by_the_permuted_intron_exon_method.pdf (InTech, 2012).
37. Haupenthal, J., Baehr, C., Kiermayer, S., Zeuzem, S. & Piiper, A. Inhibition of RNAse A family enzymes prevents degradation and loss of silencing activity of siRNAs in serum. *Biochem. Pharmacol.* **71**, 702–710 (2006).
38. Li, X.F. & Lytton, J. A circularized sodium-calcium exchanger exon 2 transcript. *J. Biol. Chem.* **274**, 8153–8160 (1999).
39. Dubin, R.A., Kazmi, M.A. & Ostrer, H. Inverted repeats are necessary for circularization of the mouse testis Sry transcript. *Gene* **167**, 245–248 (1995).
40. Burd, C.E. *et al.* Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet.* **6**, e1001233 (2010).
41. Pasman, Z., Been, M.D. & Garcia-Blanco, M.A. Exon circularization in mammalian nuclear extracts. *RNA* **2**, 603–610 (1996).
42. Ebert, M.S., Neilson, J.R. & Sharp, P.A. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat. Methods* **4**, 721–726 (2007).
43. Franco-Zorrilla, J.M. *et al.* Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* **39**, 1033–1037 (2007).
44. Poliseno, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033–1038 (2010).
45. Hansen, T.B., Kjems, J. & Damgaard, C.K. Circular RNA and miR-7 in Cancer. *Cancer Res.* **73**, 5609–5612 (2013).
46. Liu, Y. *et al.* Construction of circular miRNA sponges targeting miR-21 or miR-221 and demonstration of their excellent anticancer effects on malignant melanoma cells. *Int. J. Biochem. Cell Biol.* **45**, 2643–2650 (2013).
47. Chao, C.W., Chan, D.C., Kuo, A. & Leder, P. The mouse formin (Fmn) gene: abundant circular RNA transcripts and gene-targeted deletion analysis. *Mol. Med.* **4**, 614–628 (1998).
48. Surono, A. *et al.* Circular dystrophin RNAs consisting of exons that were skipped by alternative splicing. *Hum. Mol. Genet.* **8**, 493–500 (1999).
49. Gualandi, F. Multiple exon skipping and RNA circularisation contribute to the severe phenotypic expression of exon 5 dystrophin deletion. *J. Med. Genet.* **40**, e100 (2003).
50. Spitali, P. & Aartsma-Rus, A. Splice modulating therapies for human disease. *Cell* **148**, 1085–1088 (2012).
51. Romeo, T. Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. *Mol. Microbiol.* **29**, 1321–1330 (1998).
52. Chen, C.Y. & Sarnow, P. Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science* **268**, 415–417 (1995).
53. Zaphiropoulos, P.G. Exon skipping and circular RNA formation in transcripts of the human cytochrome P-450 2C18 gene in epidermis and of the rat androgen binding protein gene in testis. *Mol. Cell. Biol.* **17**, 2985–2993 (1997).