RBM20 Regulates Circular RNA Production From the Titin Gene

Mohsin A.F. Khan,* Yolan J. Reckman,* Simona Aufiero,* Maarten M.G. van den Hoogenhof, Ingeborg van der Made, Abdelaziz Beqqali, Dave R. Koolbergen, Torsten B. Rasmussen, Jolanda van der Velden, Esther E. Creemers, Yigal M. Pinto

Rationale: RNA-binding motif protein 20 (RBM20) is essential for normal splicing of many cardiac genes, and loss of RBM20 causes dilated cardiomyopathy. Given its role in splicing, we hypothesized an important role for RBM20 in forming circular RNAs (circRNAs), a novel class of noncoding RNA molecules.

Objective: To establish the role of RBM20 in the formation of circRNAs in the heart.

Methods and Results: Here, we performed circRNA profiling on ribosomal depleted RNA from human hearts and identified the expression of thousands of circRNAs, with some of them regulated in disease. Interestingly, we identified 80 circRNAs to be expressed from the titin gene, a gene that is known to undergo highly complex alternative splicing. We show that some of these circRNAs are dynamically regulated in dilated cardiomyopathy but not in hypertrophic cardiomyopathy. We generated RBM20-null mice and show that they completely lack these titin circRNAs. In addition, in a cardiac sample from an RBM20 mutation carrier, titin circRNA production was severely altered. Interestingly, the loss of RBM20 caused only a specific subset of titin circRNAs to be lost. These circRNAs originated from the RBM20-regulated I-band region of the titin transcript.

Conclusions: We show that RBM20 is crucial for the formation of a subset of circRNAs that originate from the I-band of the titin gene. We propose that RBM20, by excluding specific exons from the pre-mRNA, provides the substrate to form this class of RBM20-dependent circRNAs. (Circ Res. 2016;119:996-1003. DOI: 10.1161/CIRCRESAHA.116.309568.)

Key Words: alternative splicing ■ cardiovascular system ■ dilated cardiomyopathy ■ hypertrophic cardiomyopathy ■ mutation

Mutations in the RNA-binding motif protein 20 (RBM20) have been shown to cause a clinically aggressive form of dilated cardiomyopathy (DCM).1 A next-generation sequencing study in a large cohort of idiopathic DCM patients revealed that RBM20 belongs to the most frequently affected genes in DCM.2 RBM20 is essential for proper splicing of a large number of genes, and loss of RBM20 induces splicing defects in, for example, titin.3 These splicing defects in titin are thought to be an important reason why these mutations in RBM20 cause DCM.14 However, the pathophysiological role of RBM20 mutations may not be limited to abnormal splicing. Given its essential role in splicing, we hypothesized that RBM20 may also regulate other splicing-dependent processes, like the formation of circular RNAs (circRNAs). If so, this would be of importance because it adds a novel potential disease mechanism.

Editorial, see p 966
In This Issue, see p 965

Despite their discovery over 20 years ago, circRNAs have only recently been recognized as a novel class of noncoding RNA molecules. Because of their unusual properties, they were presumed to be by-products of aberrant RNA splicing.3 Decades later, next-generation sequencing has revealed that thousands of endogenous circRNAs are expressed in mammals, including the cardiovascular system,6 and that some of these circRNAs are even more abundant than their linear counterparts.7
CircRNAs are produced by the canonical spliceosome machinery, by back-splicing of exons of pre-mRNA, which results in covalently closed, single-stranded RNA molecules that lack poly(A) tails.\textsuperscript{7,8} The formation of circRNAs can affect splicing, as it has been shown that the more an exon is circularized, the less it will be represented in the linearly processed mRNA.\textsuperscript{9,10} How RNA circularization is connected to alternative splicing remains largely unknown, but splicing factors such as muscleblind and quaking have been shown to regulate circRNA formation.\textsuperscript{8,11} Recent studies revealed that circRNAs may also regulate gene expression by different mechanisms. Specifically, 2 cytoplasmic circRNAs have been shown to serve as microRNA sponges,\textsuperscript{12,13} and a class of nuclear circRNAs has been shown to promote transcription of their parental genes directly by associating with RNA polymerase II.\textsuperscript{14}

Little is known about circRNA expression and biogenesis in the healthy and diseased human heart. Therefore, we explored circRNAs in cardiac tissue from patients having hypertrophic cardiomyopathy (HCM) or DCM and from non-diseased individuals. We identified thousands of circRNAs expressed in the heart, with some of them regulated in disease. We identified a hotspot of circRNAs produced by the titin gene, precisely within the I-band region, a region known to undergo extensive, RBM20 regulated, alternative splicing.\textsuperscript{15} We show in a patient with an RBM20 mutation and in Rbm20 knockout mice that RBM20 function is required for the production of circRNAs from the I-band region of titin. This suggests that loss of RBM20 may induce myocardial disease not only by abnormal splicing of linear transcripts but possibly also by loss of a specific class of conserved and regulated circRNAs.

**Methods**

Detailed methods are provided in the Online Data Supplement.

**Results**

**Whole Transcriptome Identification of circRNAs in Human Hearts**

To detect circRNAs in diseased and nondiseased human hearts, ribosomal depleted RNA, obtained from left ventricles of 2 control, 2 HCM, and 2 DCM individuals, were used for whole transcriptome sequencing (Figure 1A). We searched for evidence of back-splicing (Figure 1B) by mapping canonical and noncanonical fusion junctions using the software MapSplice\textsuperscript{16} and uncovered a total number of 7130 putative circRNAs, of which 826 back-spliced junctions were commonly identified in all 6 samples (Figure 1C; Online Table I). Regression analysis confirmed that back-spliced junction counts were positively correlated across replicate samples within each group (Online Figure I). The most striking observation was that a total of 80 different circRNAs were identified within the titin gene (80 titin circRNAs expressed in ≥2 individuals and 22 titin circRNAs in all 6 individuals). Similarly, 59 different back-spliced junctions were identified in the RYR2 gene (59 RYR2 circRNAs expressed in ≥2 individuals and 16 RYR2 circRNAs in all 6 individuals; Online Table II). The expression level of circRNAs did not correlate with the level of expression of the host gene (Figure 1D), neither did we observe a relationship between transcript/gene length and number of circRNAs arising from the corresponding host gene (Online Figure II).

There is a general consensus that circRNAs are flanked by significantly longer introns than expected by chance.\textsuperscript{17} Therefore, we compared the median length of introns flanking the back-splicing junctions of the 826 predicted circRNAs with a control set comprising an equal number of introns randomly selected from the human genome. The median length of introns flanking our set of circRNAs was 13,071 nt compared with 1727 nt in the control set, corresponding to a 7- to 8-fold difference in length (Online Figure IIIA).

Introns flanking the back-spliced junctions of circRNAs are known to be enriched for paired Alu repeats in inverted orientation (Online Figure IIIB). We performed de novo motif enrichment analysis in introns flanking the backsplicing junctions of all 826 predicted circRNAs and detected the Alu-Ya5 motif as the candidate with the highest information content (Online Figure IIIC). In total, 16% of the circRNAs were flanked by introns containing at least 1 paired inverted Alu repeat, whereas in a control set of introns only 7% contained inverted Alu repeats (Online Figure IIID and Table III).

Recent studies suggest that circRNAs can act as microRNA sponges.\textsuperscript{17} Using miRbase in conjunction with the RNAhybrid tool, we compared the frequency of putative miRNA-binding sites identified in exons belonging to the set of 826 circRNAs to those identified in 2 control sets of exons, containing either 3' untranslated region or coding sequences. As shown in Online Figure IV, circRNAs expressed in the heart are not globally enriched for miRNA-binding sites.

**Experimental Validation of Next-Generation Sequencing–Derived Human Heart circRNAs**

To test whether the identified transcripts are bona fide circRNAs, we selected 22 of the identified circRNAs for reverse transcription-polymerase chain reaction (RT-PCR) using primers designed to amplify the circRNA-specific back-splice junctions. We selected these candidates on the basis of their absolute expression level, location within the host gene, conservation, or function of their host gene in cardiac biology (Online Table V). To avoid amplification of linear transcripts, primers were designed to diverge on linear cDNA, although being convergent on circRNA-derived cDNA. We tested circRNA expression in poly(A)-negative and poly(A)-positive RNA fractions as well as their expected resistance to exoribonuclease RNase R digestion. Although a linear transcript (αMHC [myosin heavy chain]) seemed sensitive to RNase R digestion and was mainly found in the poly(A)-positive RNA fraction, the
22 circRNAs were all resistant to RNase R digestion and were exclusively found in the poly(A)-negative fraction (Figure 2A; Online Figure VA). Interestingly, several host genes produced alternative circRNAs, as can be appreciated from the additional bands for cTTN and cLAMA2 gels in Figure 2. Sanger sequencing confirmed the presence of back-spliced exons, both in the predicted amplicons and in the additional bands observed on gel. The higher bands mostly contained additional exons, and sometimes introns, upstream of the acceptor exon or downstream of the donor exon, which indicates alternative circularization (Online Figure VI). The precise exon and intron composition of the identified titin circRNAs remains unknown because only back-spliced regions of the circRNAs were mapped. Taken together, experimental validation indicates that our RNA-seq and circRNA prediction with MapSplice is a robust approach to identify bona fide circRNAs.

As shown in Figure 2B and Online Figure VB, RT-PCR of these 22 circRNAs on a human panel of adult and fetal tissues revealed that a subset of circRNAs are widely expressed (eg,
cPDLIM5, cATP2B4), whereas others are expressed in a tissue-specific manner (eg, cLAMA2 and cTTN). Interestingly, the precise alternative circularization of some of the circRNAs seems tissue specific (eg, cSTRN) or developmental specific (eg, cTMEM38b).

Identification of Disease-Regulated circRNAs

Differential expression analysis revealed 43 out of 826 commonly identified circRNAs to be differentially expressed in DCM compared with control samples and 60 circRNAs in HCM compared with control samples (Figure 3A and 3B; Online Table IV A and IVB). Because of the limited number of samples (n=2 per group), a fold change of >2 and a \( P \) value of <0.05 were used as cutoffs, rather than the adjusted \( P \) value (note that only 2 circRNAs survived multiple testing; Online Table IV A and IVB). We selected 10 candidates from the set of 22 experimentally validated circRNAs and examined them by semiquantitative RT-PCR in a larger group of patients (7 control, 7 HCM, and 7 DCM). As shown in Figure 3C and quantified in Online Figure VII, we confirmed the loss of circRNA formation from the host genes CAMK2D (in DCM and HCM) and titin (mainly in DCM) in disease. To investigate whether differential expression of circRNAs was caused by altered expression of host genes, we performed qRT-PCR of the linear transcripts (Online Figure VIII). Interestingly, expression of CAMK2D and titin circRNAs did not correlate with expression changes of their linear host transcripts, indicating that the disease-regulated changes in circRNA production are independent of transcriptional regulation.

RBM20 Is Required for the Production of circRNAs From the Host Gene Titin

We observed a remarkably large number of circRNAs produced from the titin gene, a gene that is known to undergo highly complex alternative splicing. Titin is a protein that spans half of the sarcomere and determines biomechanical properties of the heart. Particularly, the inclusion/exclusion of a large segment within titin, the so-called I-band, which behaves as a molecular spring, importantly determines the passive stiffness of sarcomeres.\(^3,4\) RBM20 is the splicing factor responsible for alternative splicing within the I-band (particularly the elastic PEVK domain and the immunoglobulin (Ig)-rich region), and mutations in RBM20 have been shown to result in the expression of large and highly compliant titin isoforms, suggested to cause DCM.\(^3,4\)

Strikingly, we noted a hotspot of circRNAs exactly within titin’s I-band (Figure 4A). This prompted us to investigate whether there is an enrichment of RBM20-binding sites in the introns flanking the back-spliced junctions of the 80 predicted titin circRNAs. Interestingly, we found a 5-fold increase in RBM20-binding site frequency in the introns flanking (within 100 bp) the 80 titin circRNAs compared with a control set of introns (see also track in Figure 4A). Examination of RBM20-binding sites in flanking introns of all 826 identified circRNAs revealed a \(~2\)-fold enrichment compared with the control intron set (Online Figure IXA and IXB and Table VI).

To investigate whether RBM20 is essential for titin-derived circRNAs, we generated Rbm20-null mice and examined...
their cardiac titin circRNA production. In these mice, loss of Rbm20 resulted in early-onset DCM, which manifested by LV dilatation and impaired cardiac function at an age of 10 weeks and was accompanied by the previously described aberrant titin splicing3 (Online Figure X). From a recent report, which profiled circRNAs in the mouse heart, 19 we selected 4 titin circRNAs that arise from the Ig and PEVK regions, regions that critically depend on Rbm20 for alternative splicing (mus_cTtn1-4), and 2 circRNAs that arise from the Rbm20-independent N2A and Z-disk regions (mus_cTtn5-6, Online Figure XI). Interestingly, we identified both Rbm20-dependent and Rbm20-independent circRNAs arising from the titin gene. CircRNAs produced from the Ig and PEVK regions were absent in the Rbm20 knockout hearts, whereas the circRNAs generated from the Ig and PEVK regions were absent in the Rbm20 knockout hearts, whereas the circRNAs generated from the Rbm20 splicing–independent regions were abundantly expressed in the Rbm20 knockout hearts (Figure 4B). Interestingly, for the Rbm20-dependent circRNAs, the corresponding exons are more included in the linear titin transcript after loss of Rbm20, whereas for Rbm20-independent circRNAs, there is equal exon inclusion between wild-type and Rbm20 knockout mice (Online Figure XII). This implies that when Rbm20-dependent exons are spliced out of the linear titin transcript, they can serve as a substrate for circRNA formation. When, however, Rbm20 is lost and these exons are included in the linear titin transcript, the substrate is lost, and circRNA formation is abolished.

To validate RBM20-dependent circRNA production and its relevance in human disease, we investigated a DCM patient with a heterozygous mutation in RBM20 (E913K). This mutation resulted in a shift in titin from the less compliant N2B toward the highly compliant N2BA isoforms.20 Analysis of circRNA production by RT-PCR in the heart of this patient revealed that titin circRNA production was grossly abnormal, specifically in the Ig and PEVK domain (ie, cTTN1, 2, 4, and 5), when compared with hearts of other idiopathic DCM patients, but not for circRNAs produced partially from the Z-disk region (ie, cTTN3), which is considered RBM20-independent (Figures 3C and 4A).

Our data confirms that alternative splicing and circRNA production are intimately connected. We show that skipping of titin’s I-band region is associated with circRNA formation, likely by providing a substrate for the formation of circRNAs, as illustrated in Figure 4C. This figure depicts the concept based on the established role of RBM20 as a splicing repressor.4 Thus, in the absence of RBM20, these exons are included in the linear titin transcript so that they cannot serve as substrate for circRNA formation.

**Discussion**

We identified thousands of circRNAs in the human heart and show that a subset of these circRNAs is differentially expressed in diseased hearts. The identified cardiac circRNAs

---

**Figure 3. Circular RNAs (circRNAs) are differentially expressed in disease.**

**A.** CircRNA expression in hypertrophic cardiomyopathy (HCM) and (B) dilated cardiomyopathy (DCM) hearts compared with controls. Differentially expressed circRNAs (fold change >2 and P value <0.05) are marked in gray. **C.** RT-PCR on total RNA from healthy, HCM, and DCM human hearts (n=7 per group). One DCM patient carried a mutated RNA-binding motif protein 20 (RBM20) allele (E913K). GAPDH was used as input control. The locations of the titin circRNAs (cTTN) are shown in Figure 4A. See Online Figure VII for gel quantification.
conform to most of the key properties of circRNAs: (1) circRNAs were flanked by 7- to 8-fold longer introns compared with a random set of intron, (2) we found an enrichment of inverted Alu repeats in the introns flanking the predicted circRNAs, (3) cardiac circRNAs were not generally enriched for miRNA-binding sites, (4) circRNAs were resistant to RNase R treatment and lack poly(A) tails.

The main finding of this study is that RBM20 is crucial for the formation of a subset of circRNAs that originate from a specific region within the I-band of the titin gene (ie, PEVK and Ig repeats), a region that is known to undergo extensive alternative splicing to produce titin isoforms with the desired biomechanical properties. Given the known role of RBM20 in exon skipping in titin’s I-band, we propose the concept that RBM20, by excluding specific exons from the pre-mRNA, provides the substrate to form this class of RBM20-dependent circRNAs. It has been described that idiopathic DCM patients have increased N2BA/N2B isoform ratios and thus more inclusion of Ig repeats and a longer PEVK domain in the I-band. In similarity to our observations in the Rbm20 knockout mice, we postulate that increased inclusion of these domains in the linear titin transcript prevents the formation of titin circRNAs. As shown in Figure 3C, this is indeed the case for the I-band circRNAs. Interestingly, we have also identified RBM20-independent titin circRNAs. These were produced from exons closer to the Z-disk and from exons within the N2A domain, regions that are not spliced by RBM20; these
circRNAs were readily expressed in the Rbm20 knockout hearts.

Maatz et al\(^8\) previously identified a correlation between RBM20 mRNA levels and the levels of titin splicing, which suggests that titin circRNA expression is also correlated to RBM20 mRNA levels. In the myocardial tissues used for RT-PCR (Figure 3C), we observed a 29%, nonsignificant downregulation of RBM20 mRNA in the DCM samples (data not shown). However, the small group size of this study was not sufficient to find a correlation between RBM20 and titin circRNA expression. This is not surprising considering the observation by Maatz et al that the expression of RBM20 greatly varies between patients with end-stage heart failure. Therefore, the authors used 10 patients with the highest RBM20 expression and 10 patients with the lowest RBM20 expression, selected from a database of 148 heart failure patients to demonstrate a correlation between titin splicing and RBM20 levels. In line with this observation, the RNA-seq experiment in our study was underpowered to test for a correlation between RBM20 mRNA and cTTN expression because none of the RBM20-dependent titin circRNAs were significantly differentially expressed between DCM and controls after correcting for multiple testing (Online Table VII).

It is known that circRNA processing is intimately connected to alternative splicing\(^5\); however, our findings suggest that circularization of RNA substrates occurs after alternative splicing has taken place. This order of events is supported by a recent study of Zhang et al,\(^2,2\) who used metabolic tagging of nascent RNAs to show that circRNA processing from pre-mRNA occurs post-transcriptionally. Specifically, they show that the majority of circRNAs were produced after transcription and splicing of their parental gene was completed. However, others have shown that the use of 5’ and 3’ splice sites in circRNAs can compete with pre-mRNA splicing, which would imply that circRNA biogenesis occurs simultaneously and thus can affect or even regulate alternative splicing.\(^5\) It therefore is possible that circRNA formation within titin underlies the diversity of titin splicing, and it is tempting to speculate that perturbations in circRNA expression could contribute to the pathophysiology of DCM patients. The general function of circRNAs remains unclear, but an intriguing possibility, which is still under debate, is that circRNAs form a template for protein synthesis. Because most circRNAs are composed of protein-coding exons, it will be interesting to investigate their potential to be translated. If these circRNAs are translatable, it has to be considered that small titin peptides are expressed in the heart.

Taken together, we propose that RBM20 is important in normal cardiac physiology not only by regulating exclusion of specific exons but also by allowing a specific subclass of circRNAs to be generated. This opens the possibility that loss of RBM20 induces disease not only by aberrant splicing of titin and other genes but also by loss of circRNAs.

**Acknowledgments**

We thank Aeilko H. Zwinderman for valuable advice on bioinformatics and Anke J. Tijsen for scientific discussions. This work was supported by grants from the Netherlands Organization for Scientific Research (NWO-836.12.002 and NWO-821.02.021) and the Netherlands Cardiovascular Research Initiative (grants CVON-ARENA-2011-11).

**Disclosures**

None.

**References**

20. Beqalli A, Bollen IA, Rasmussen TB et al. A mutation in the glutamate-rich region of Rbm20 causes dilated cardiomyopathy through
What Is Known?
• Circular RNAs (circRNAs) are a novel class of RNAs that are formed by back-splicing of exons by the spliceosome.
• Thousands of different circRNAs are expressed in mammals, but their function remains largely elusive.
• RNA-binding protein 20 (RBM20) is an important cardiac splicing factor, and mutations therein cause early-onset dilated cardiomyopathy.

What New Information Does This Article Contribute?
• We identified thousands of circRNAs to be expressed in the human heart, with at least 80 originating from the titin gene.
• Using Rbm20 knockout mice and myocardial tissue of a patient with a mutated RBM20 allele, we demonstrate a role for RBM20 in the formation of circRNAs of the titin gene.

Novelty and Significance
• We propose the model that the splicing factor RBM20, by excluding specific exons from titin’s I-band region, provides the substrate to form circRNAs.

Here, we identified thousands of circRNAs from the human heart, with at least 80 circRNAs originating from the titin gene. Our most important finding is that the splicing factor RBM20 is crucial for the formation of a subset of these titin circRNAs. As such, we provide insights into the biogenesis of titin circRNAs, by showing that exon skipping of titin’s I-band region is required for circRNAs to be formed. We think that these findings open novel avenues to understand the role of RBM20 as a cause of hereditary dilated cardiomyopathy, while contributing to our basic understanding of the biogenesis of circRNAs.
RBM20 Regulates Circular RNA Production From the Titin Gene

Mohsin A.F. Khan, Yolan J. Reckman, Simona Aufiero, Maarten M.G. van den Hoogenhof, Ingeborg van der Made, Abdelaziz Beqqali, Dave R. Koolbergen, Torsten B. Rasmussen, Jolanda van der Velden, Esther E. Creemers and Yigal M. Pinto

Circ Res. 2016;119:996-1003; originally published online August 16, 2016;
doi: 10.1161/CIRCRESAHA.116.309568

Circulation Research is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2016 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/119/9/996

Data Supplement (unedited) at:
http://circres.ahajournals.org/content/suppl/2016/08/16/CIRCRESAHA.116.309568.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation Research can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation Research is online at:
http://circres.ahajournals.org//subscriptions/
Online Methods

Human samples
Fresh LV tissue from 2 patients (1 male, 53 y and 1 female, 50 y) with end-stage idiopathic dilated cardiomyopathy (DCM) were obtained during heart transplantation surgery. As controls we used left ventricular (LV) tissue of non-failing donors (not used for transplantation due to logistic reasons) which had no history of cardiac abnormalities (2 males, age: 52 and 37 y). These samples were obtained from the Sydney Heart Bank (Cris dos Remedios) with approval of the local ethical committee (St. Vincent's Hospital Human Research Ethics committee, Sydney, Australia: File number: H03/118: Title: Molecular analysis of human heart failure) and by The University of Sydney HREC number 12146. LV tissue of HCM patients (2 males, age 47 and 64 y) were obtained during septal reduction myectomy at the AMC, Amsterdam, the Netherlands. The LV tissue sample of the heterozygous RBM20-E913K patient (male, age 19 y) was obtained during heart transplantation surgery at Aarhus University Hospital. The investigation conforms with the principles outlined in the Declaration of Helsinki (1997) and in accordance with institutional guidelines.

RNA isolation
Total RNA was extracted from the above described human tissue samples and from LV tissue samples of 6 months old Rbm20 KO mice (3 wildtype, 3 heterozygous Rbm20 KO and 3 Rbm20 KO; littermates) with TRI Reagent (Sigma-Aldrich) according to manufacturer’s protocol. To analyze circRNA expression across different fetal and adult tissues, RNA from commercially available Human Total RNA Master Panel II (636643, Clontech) was used.

Library preparation and whole transcriptome RNA sequencing
RNA quality was assessed with the Agilent 2100 Bioanalyser. All samples had a RIN score of > 8.0. Total RNA samples (500 ng) were treated with biotin-streptavidin based bead systems (Exiqon) to minimize ribosomal contamination. Ribosomal-depleted RNA libraries were sequenced on a illumina NextSeq 500 platform in paired-end mode and with a read length of 101bp. Sequencing depth was approximately 90 million raw reads per sample. Base-calling was performed using the bcl2fastq 2.0 Conversion Software from Illumina.

Quality Control and processing of RNA-Seq samples
Quality control of fastq files was performed using FASTQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). Trimmomatic version 0.35\textsuperscript{1} was used to remove Illumina adapters, using a phred score cut-off of 30 whilst discarding reads with a length below 25 bases. Reads passing quality control were then utilized for circRNA identification as described below.

Identification of putative circular RNAs
Exonic circRNA back-spliced junctions were identified using MapSplice version 2.2.0\textsuperscript{2}. MapSplice aligns reads to the genome using Bowtie\textsuperscript{3} and further attempts to map unaligned reads by introducing gap alignments. Both canonical and non-canonical fusion junctions were detected as evidence for circRNAs using the following options --min-fusion-distance 200 (as suggested by the authors), --filtering 1 and --min-map-len 25. The junction coverage and total number of supporting reads reported by MapSplice for each sample were then formatted into matrices where the value in the i-th row and the j-th column of each matrix represented the total number of reads successfully mapped to the back-splice junctions of the host gene i in sample j.
Identification of inverted Alu repeats flanking introns of circRNAs
To quantify inverted Alu repeats in candidate circRNAs, genomic positions of all known Alu repeats were downloaded from the UCSC genome table browser. Bedtools together with custom PERL scripts were then used to determine for each candidate circRNA, the genomic positions and strand (sense or anti-sense) of Alu repeats occurring within introns flanking 100bp, 200bp and 500bp upstream and downstream of the back-splicing acceptor and donor sites respectively. For each flanking distance, the number of paired inverted (reverse complementary) Alu repeats occurring in flanking introns of all circRNAs were then tallied. As a control set, an equal number of pairs of exons known to be alternatively spliced were randomly selected from the genome as follows. First, BioMart from Ensembl was used to extract genomic features (i.e. transcripts and exons) and positions of all genes consisting of 3 or more transcripts. For each gene, exons exclusively occurring in only 1 transcript out of at least 3 were then identified (as evidence of alternatively spliced exons). Finally, n pairs (determined by the total number of exon pairs represented in all predicted circRNAs) of alternatively spliced exons were randomly selected whilst ensuring that each pair corresponded to the same transcript. The Alu enrichment procedure was repeated as described above and the number of inverted Alu repeats occurring within introns flanking 100bp, 200bp and 500bp upstream and downstream of these 'control' exon pairs were tallied.

miRNA binding site analysis of circRNAs
To determine whether candidate circRNAs identified in this study have the potential to act as sponges for microRNAs, the RNAhybrid tool was used to predict miRNA target sites in each of the predicted circRNAs. The parameters were set as follows; -p 0.05, -s 3UTR_human, allowing a mismatch in the seed region. A total of 2,588 human mature miRNA sequences were downloaded from miRBase (release 21). Exon sequences located within the predicted circRNAs were downloaded from the UCSC Table Browser, using the GRCh37/hg19 assembly of the human genome. As a control set, an equal number of exons were randomly selected from the genome together with a random set of 3' UTR sequences. The miRNA enrichment procedure was repeated as described above.

Differential circular RNA expression analysis
Differential expression analysis of circRNAs was performed using the R Bioconductor package, DESeq. To investigate the role of circRNAs in normal, hypertrophic and dilated hearts, three sets of differential expression analyses were conducted. First, hearts from control individuals were compared with those from DCM patients; second, control hearts were compared with HCM hearts. Owing to the availability of replicates, the dispersion method “pooled” from DESeq was used to accurately estimate dispersion between each comparison and the negative binomial model was used to estimate differentially expressed circRNAs for each analysis. At the end, only those circRNAs passing a fold-change (log2) cut-off of 1 together with a p-value cut-off of 0.05 were deemed significantly differentially expressed.

Differential gene expression analysis
To calculate differential expression of linear junction reads across DCM, HCM and controls, RNA-Seq reads from all samples were first aligned to the human genome (hg19 build) using TopHat. The dispersion method “pooled” from DESeq was used to accurately estimate dispersion between each comparison and the negative binomial model was used to estimate differentially expressed genes for each analysis. Transcript assembly and quantification of linear junctions were performed using Cufflinks. Only transcripts with an FPKM of > 0 were retained.
Rbm20 binding site analysis
Genomic sequences of introns flanking circRNAs were obtained from the UCSC genome table browser. Custom PERL scripts were then used to determine for each candidate circRNA, the genomic positions of the Rbm20 motif (UCUU) occurring within introns flanking 100bp, 200bp and 500bp upstream and downstream of the back-splicing acceptor and donor sites respectively. Rbm20 motifs occurring in flanking introns of all circRNAs were then tallied. The same control set used for Alu enrichment (described above) was utilized as a background set. The number of Rbm20 motifs occurring within introns flanking 100bp, 200bp and 500bp upstream and downstream of these 'control' exon pairs were tallied. To calculate the density of Rbm20 binding sites within TTN gene, the total number of Rbm20 motifs occurring in a given intron was divided by the length of the corresponding intron. Exons were excluded from the analysis.

Preparation of poly(A)-positive and poly(A)-negative RNA fractions
Total RNA isolated from LV tissue from three HCM patients was used to generate poly(A) enriched and depleted fractions using 200 µl oligo d(T)25 Magnetic Beads (NEB). After washing with 500 µl washing buffer (300 mM NaCl, 20 mM Tris-HCL, 0.01% NP-40), beads were incubated with 6 µg total RNA from left ventricles of HCM patients in 2x binding buffer (300 mM NaCl, 40 mM Tris-HCL, 0.02% NP-40) in a 100 µl reaction for 10 minutes with gentle agitation. Beads were pulled to the side of the tube with a magnetic rack and supernatant containing poly(A) negative RNA was collected. After washing three times with 500 µl washing buffer, beads were incubated in 100 µl elution buffer (150 mM NaCl, 20 mM Tris-HCl, 0.01% NP-40) at 50°C for 2 minutes. Beads were pulled to the side of the tube with a magnetic rack and supernatant containing poly(A) positive RNA was collected. Poly(A) negative RNA was also treated at 50°C for 2 minutes.

RNase R digestion
Total RNA isolated from LV tissue from the same three HCM patients used for poly(A) enrichment/depletion was used to generate RNase R digested RNA and control samples by incubating 1 µg total RNA in 1x RNase R buffer in a 20 µl reaction with or without 5 units of RNase R (Epicentre) at 37°C for 10 minutes followed by heat inactivation at 95°C for 3 minutes.

RT-PCR and qRT-PCR
After poly(A) enrichment/depletion or RNase R treatment, 1 µg RNA was DNA depleted using DNase I (Invitrogen) in a 20 µl reaction. Afterwards, cDNA was generated using SuperScript II Reverse Transcriptase (Invitrogen) in a 40 µl reaction. For circRNA detection reverse transcription was primed with random hexamers (Invitrogen) and for mRNA detection oligo(dT) primers were used. For RT-PCR, 1-4.5 µl of 5-times diluted cDNA was amplified in a 25 or 37.5 µl reaction with HOT FIREpol DNA polymerase (Solis Biodyne) using the following program: 15 minutes pre-incubation at 95°C, 30-35 cycli of 30 seconds 95°C, 30 seconds 58°C, and 10 seconds-3 minutes 72°C and 10 minutes final elongation at 72°C. PCR products were size separated on 1-2% agarose gels. Primers are designed to amplify the smallest band depicted on the gels shown in the (Online) Figures. Sanger sequencing confirmed the presence of the expected backsplice junction in these bands. A majority of the additional bands was also sequenced. Band density was quantified using ImageJ software. Two different DNA ladders were used: 1) Lambda DNA digested with BstEII, which produces bands with the following sizes (bp): 8453, 7242, 6369, 4822, 4324, 3675, 2323, 1929, 1371, 1264 and 702. 2) pBS digested with MspI, which produces bands with the following sizes (bp): 710, 469, 404, 328, 242, 190, 157, 147, 110, 67, 57 and 34.

For qRT-PCR, 2 µl of 5-times diluted cDNA was amplified with SYBR Green I Master (Roche) on a LightCycler480 system II (Roche) using the following program: 5 minutes pre-
incubation at 95°C and 45 cycli of 10 seconds 95°C, 20 seconds 55°C, and 20 seconds 72°C. Primer sequence are included in Online Table V.

**Rbm20 knockout mice**

The International Knockout Mouse Consortium (Knockout-first allele, University of California, Davis) has generated a Rbm20 conditional mouse allele, in which a neomycin and LacZ cassette together with loxP sites flanking exon 4 and 5 are inserted by homologous recombination in C57BL/6N embryonic stem (ES) cells. Correct ES cell targeting was confirmed by Southern blot analysis (not shown). Pronuclear injections of these targeted ES cells in (FVB) blastocysts, Flp-mediated excision of the neomycin cassette, and crossing with a CMV-Cre mouse line was performed at the AMC medical center. Rbm20 heterozygous mice were crossed at least 6 generations with FVB wildtype mice to obtain a pure background. All animal studies were approved by the Institutional Animal Care and Use Committee of the University of Amsterdam, and in accordance with the guidelines of this institution and the Directive 2010/63/EU of the European Parliament.

Homozygous Rbm20 knockout mice did not express Rbm20 mRNA, as shown by qPCR, with a primer set designed in exon 2 (Fwd) and 3 (Rev), and revealed abnormal titin splicing in their hearts (see Online Figure IX), thereby phenocopying the previously described Rbm20 knockout rat model. Inclusion of addition 1-band exons, resulted in a shift from the less compliant N2B towards the highly compliant N2BA titin isoforms.

**Data visualization and statistical analysis**

All results generated from differential circRNA analysis were visualized using the R bioconductor package, ggplot2. Venn diagrams were constructed using the R package, VennDiagram. For miRNA enrichment analysis, a cumulative distribution function (CDF) was computed using the ecd function of the R Bioconductor package, Stats. The graphical representation of the TTN gene and circRNAs were constructed using the R package, GenomicRanges and custom R scripts respectively. The Rbm20 density track was constructed using the R package, Sushi.

For differential circRNA expression across conditions, circRNA back-spliced junction data were statistically analyzed using the negative binomial test in the R Bioconductor package, DESeq. All correlation tests were performed in R, using the Pearson's correlation as the method of choice. Intron lengths flanking ecircRNAs and the control set were compared using the unpaired t test in R. The frequency of inverted Alu repeats flanking ecircRNAs and the control set were compared as follows. For each flanking distance (100, 200 and 500bp up and downstream of the back-spliced junctions), a 2 x 2 contingency table was constructed, comprising of the number of inverted and non-inverted Alu repeats in the ecircRNAs and control set. The Fisher's exact test in R was then used to calculate a p-value for each flanking distance. The frequency of Rbm20 binding sites flanking ecircRNAs were statistically analyzed as follows. For each flanking distance (100, 200 and 500bp up and downstream of the back-spliced junctions), the proportion of observed Rbm20 binding sites flanking the ecircRNAs (X) and control set (Y) were calculated using the equations, \( X = a/a+b \) and \( Y = b/a+b \) respectively, where \( a = \) total number of Rbm20 binding site hits flanking the set of ecircRNAs and \( b = \) total number of Rbm20 binding site hits flanking the control set. Differences in enrichment of Rbm20 sites between ecircRNAs and control sequences (across all flanking distances) were compared using the t test.

qPCR data were analyzed using LinRegPCR quantitative PCR data analysis software and normalized for expression of GAPDH or HPRT. Values are expressed as mean ± standard error. Differences between groups were compared using Student's t-test and p-values <0.05 were considered statistically significant.
Online Figure I. Regression analysis of circRNA expression in replicate samples of control, DCM and HCM hearts. (A) The X-axis represents back-spliced junction counts from sample 1 and the Y-axis from sample 2 of control hearts. CircRNA detection positively correlated with respect to each replicate sample (R = 0.702, p < 2.2e – 16). (B) CircRNA detection positively correlated across both replicate DCM samples (R = 0.721, p < 2.2e – 16), and (C) circRNA detection positively correlated across both replicate HCM samples (R = 0.744, p < 2.2e – 16).
Online Figure II. Relationship between transcript length and number of circRNAs. A) A scatter plot illustrating the correlation between number of circRNAs detected across all 6 samples (X-axis) and transcript length (Y-axis). Overall there was no global correlation (Pearson's $R = 0.15$, $p=0.00026$). B) A scatter plot illustrating the correlation between number of circRNAs detected across all 6 samples (X-axis) and gene length (Y-axis). Overall there was no global correlation (Pearson's $R = 0.16$, $p=0.00011$).
Online Figure III. Flanking intron length and Alu enrichment analysis. (A) A boxplot comparing the length of introns flanking the back-spliced junctions of the subset of 826 circRNAs, with the length of an equal number of randomly selected introns from the human genome (control set). Introns flanking back-spliced junctions of circRNAs were significantly longer than expected by chance (p < 2.2e-16). (B) A conceptual illustration of how inverted Alu repeats occurring in flanking introns of circRNAs might be important in facilitating circularization. (C) De novo motif enrichment analysis of introns flanking the set of 826 circRNAs, revealing the Alu motif as the top candidate. (D) A bar graph comparing the fraction of inverted Alu repeats flanking the set of 826 circRNAs with those flanking a set of randomly selected transcripts known to be alternatively spliced. Introns flanking back-spliced junctions of circRNAs were significantly enriched for Alu repeats occurring in inverted orientation compared to introns flanking a set of randomly selected linear transcripts (p = 0.0009, p < 0.0001 and p < 0.0001; flanking distance = 100bp, 200bp and 500bp respectively).
Online Figure IV. miR binding site enrichment analysis. Cumulative distribution function (CDF) showing the abundance of miRNA binding in the exons of 826 circRNAs (red), a set of randomly selected coding sequences (CDS) from the human genome (green) and a set of randomly selected 3' UTR sequences from the human genome (blue). As a whole, circRNAs are not densely populated with miR binding sites, compared to 3' UTR and CDS sequences of mRNA.
Online Figure V. Validation of selected circRNAs and tissue expression analysis (A) Total RNA from three human hearts was either fractionated in a poly(A) negative(-) and positive(+) fraction using oligo-dT beads or treated with(+) or without(-) RNase R and amplified using divergent primers. αMHC was used as linear control. UT: untreated sample. (B) CircRNA expression in a human tissue panel. GAPDH was used as input control. Primers are designed to amplify the smallest band on gel, see Online Figure VI for the composition of additional bands.
Online Figure VI. Alternative circularization detected by RT-PCR. (A) By alternative circularization extra exons, either upstream of the acceptor exon or downstream of the donor exon, are included in the circRNA, which leads to additional larger amplicons in an RT-PCR reaction. RT-PCR primers are divergent on linear mRNA and convergent on circular RNA (red arrows). A: acceptor exon. D: donor exon. X: additional exon included in the circRNA. (B) The amplicon composition is depicted for a majority of the bands detected with RT-PCR on the poly(A) negative (-) fraction of three human heart samples (complete gels in Figure 2 and Online Figure V) or three mouse heart samples (complete gels in Online Figure XI). *Please note that for two alternatively circularized circRNAs (cLAMA2 and cCCSER2) an intron is (partially) included instead of an up- or downstream exon. Exon numbers are counted using transcript ENST00000421865 for cLAMA2, ENST00000589042 for cTTN, ENST00000264808 cPRDM5, ENST00000374692 for cTMEM38b, ENST00000263918 for cSTRN, ENST00000224756 for cCCSER2 and ENSMUST00000099981.8 for mouse cTtn. Sequence reactions on additional bands for some circRNAs were not successful and therefore not shown here.
Online Figure VII. Band density quantification of circRNAs examined in a larger group of patients. Gel images are depicted in Figure 3C. Band densities were quantified using ImageJ software and normalized to GAPDH. Only the lowest band on each gel was quantified as this band corresponds to the circRNA identified in the RNAseq. Mean ± SEM. Differences between Control (n=7) and HCM (n=7) or DCM (n=6) groups were compared using Student’s t-test. Band densities for RBM20 E913K were not tested for significance compared to other groups, because this group constitutes only one sample. *p<0.05 compared to Control group.
Online Figure VIII. Expression of linear transcripts of CAMK2D and total titin. (A) Linear CAMK2D (B) Linear total titin. qRT-PCR was performed on total RNA from healthy and diseased (HCM or DCM) human hearts (n=7 per group). Normalized to HPRT expression. Mean ± SEM. Differences between groups were compared using Student’s t-test. No statistically significant differences were observed.
Online Figure IX. RBM20 binding site analyses. (A) RBM20 sites flanking introns of 826 circRNAs. A bar graph comparing the number of predicted RBM20 binding sites (UCUU) in the introns flanking the set of 826 circRNAs with a control set. Introns flanking back-spliced junctions of circRNAs were ~2 fold enriched with RBM20 binding sites compared to the control set (p<0.0001).

(B) RBM20 sites flanking introns of 80 titin derived circRNAs. A bar graph comparing the number of predicted RBM20 binding sites (UCUU) in the introns flanking the set of 80 titin circRNAs with a control set. Introns flanking back-spliced junctions of circRNAs were 5 fold enriched with RBM20 binding sites compared to the control set (p<0.00007). Details of the frequency of RBM20 binding sites flanking (100bp up and downstream) the back-splice junctions of each of the 826 circRNAs are provided in Online Table VII.
Online Figure X. Loss of Rbm20 in mice results in a switch from N2B to N2BA titin isoforms. qRT-PCR was performed on wildtype (WT), heterozygous (HET) and knockout (KO) mouse hearts (n=3 per group). Rbm20 expression is corrected for Gapdh. N2B and N2BA expression are corrected for total titin expression. Note that N2BA primers are designed in the N2A element and therefore reflect both N2BA and N2BA-giant transcript levels. Mean ± SEM. Differences between groups were compared using Student’s t-test. *p<0.05 compared to WT.
Online Figure XI. Experimental validation of selected mouse titin circRNAs (cTtn). Total RNA from three mouse hearts was either fractionated in a poly(A) negative (-) and positive (+) fraction using oligo-dT beads or treated with (+) or without (-) RNase R and amplified using divergent primers. αMHC was used as linear control. UT: untreated sample. The location of the titn circRNAs is shown in Figure 4A. Primers are designed to amplify the smallest band on gel, see Online Figure VI for the composition of additional bands.
Online Figure XII. Rbm20 produces a substrate for titin circRNA (cTtn) formation. (A+B) RT-PCR of mouse circRNAs (as depicted in Figure 4A) and linear titin (Ttn) transcript in Rbm20 knockout hearts. (A) In the wildtype hearts Rbm20 splices the Ig-repeats and PEVK domain (linear Ttn [exon 192-202] and linear Ttn [exon 120-133]) and thus provides a substrate for circRNA formation (mus_cTtn1, 2, 3 and 4). However, when Rbm20 is absent, exons form the Ig-repeats and PEVK domain are no longer spliced out (as evidenced by larger amplicons for linear Ttn [exon 192-202] and linear Ttn [exon 120-133]) and thereby prevent circRNA formation (mus_cTtn1, 2, 3 and 4). (B) The circRNAs derived form the N2A and Z-disk regions (muscTtn5 and 6) are not affected by loss of Rbm20. As expected, the linear counterpart of these circRNAs (linear Ttn [exon 2-9] and linear Ttn [exon 106-112]) are also not affected by loss of Rbm20. Please note that, apart from mus_cTtn6, the distance between acceptor and donor exon is too larger to PCR the complete linear counterpart. Therefore, a primer in the donor or acceptor exon and an additional primer 5-10 exons up or downstream was used in the PCR reaction. For circRNAs, primers are designed to amplify the smallest band on gel. See Online Table V for predicted amplicon sizes. bspl: backsplice junction.
Online Table I:

EXCEL FILE named “Online Table I”

Online Table I: A total of 7130 putative circRNAs were identified with a back-spliced junction count of at least 1 in any of the 6 samples. Chromosome numbers, positions and sequences of donor and acceptor sites, host gene symbols together with back-spliced junction counts of predicted circRNAs in each sample are tabulated. Entries highlighted in blue represent the set of 826 circRNAs with identical back-spliced junctions commonly identified in all 6 samples.

Online Table II:

EXCEL FILE named “Online Table II”

Online Table II: A table illustrating details of 80 circRNAs identified in the TTN gene and 59 circRNAs identified in the RYR2 gene. Chromosome numbers, positions and sequences of donor and acceptor sites, host gene symbols together with back-spliced junction counts of predicted circRNAs in each sample are tabulated.

Online Table III:

EXCEL FILE named “Online Table III”

Online Table III: Genomic positions of paired Alu repeats detected within a flanking distance of 500bp up and downstream of back-spliced junctions of the 826 predicted circRNAs. Entries highlighted in yellow represent paired inverted Alu repeats.

Online Table IVA:

EXCEL FILE named “Online Table IVA”

Online Table IVA: 43 circRNAs differentially expressed in DCM Vs. Controls. Entries highlighted in green represent circRNAs upregulated in DCM, whereas those in red represent circRNAs downregulated in DCM. A p-value cut-off of < 0.05 was used as the significance threshold. NOTE: Entries in bold represent circRNAs that pass a FDR cut-off of 0.05.

Online Table IVB:

EXCEL FILE named “Online Table IVB”

Online Table IVB: 60 circRNAs differentially expressed in HCM Vs. Controls. Entries highlighted in green represent circRNAs upregulated in HCM, whereas those in red represent circRNAs downregulated in HCM. A p-value cut-off of < 0.05 was used as the significance threshold.
Online Table V:

EXCEL FILE named “Online Table V”

**Online Table V**: Primers used for RT-PCR and qRT-PCR. TTN exons counted in ENST00000589042 for human and in ENSMUST00000099981.8 for mouse.

Online Table VI:

EXCEL FILE named “Online Table VI”

**Online Table VI**: The frequency of RBM20 binding sites flanking the back-spliced junctions of each of the 826 circRNAs. Entries highlighted in yellow represent circRNAs flanked by at least 1 RBM20 binding site, either up or downstream of the back-spliced junction.

Online Table VII:

EXCEL FILE named “Online Table VII”

**Online Table VII**: Differential expression analysis between DCM and Controls for RBM20-dependent TTN-derived circRNAs. cTTN1 (highlighted in blue) expression is significantly different with a p-value cut-off of <0.05, which was used due to the limited number of samples (n=2 per group). However, after correction for multiple testing the significance is lost.
Online References