RNA splicing in the heart
van den Hoogenhof, M.M.G.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
CAR DiAC CIRCRNAs ARISE MAINLY FROM CONSTITUTIVE RATHER THAN ALTERNATIVELY SPLICED EXONS

Simona Aufiero
Maarten M.G. van den Hoogenhof
Yolan J. Reckman
Abdelaziz Beqqali
Ingeborg van der Made
Jolanda Kluijn
Mohsin A.F. Khan
Yigal M. Pinto
Esther E. Creemers

In Revision
Abstract

We recently provided evidence that alternative splicing underlies the formation of circular RNAs (circRNA) arising from the Titin (TTN) gene. Since it is still unclear how circRNAs are generally formed, we hypothesized that alternative splicing, and in particular exon skipping, is a major driver of circRNA production. We performed RNA sequencing on human and mouse hearts, mapped alternative splicing events and overlaid these with expressed circRNAs at exon level resolution. In addition, we performed RNA sequencing on hearts of Rbm20 KO mice to address how important Rbm20-mediated alternative splicing is in the production of cardiac circRNAs. In human and mouse hearts, we show that cardiac circRNAs are mostly (~90%) produced from constitutive exons and less (~10%) from alternatively spliced exons. In Rbm20 KO hearts, we identified 38 differentially expressed circRNAs of which 12 were produced from the Ttn gene. Even though Ttn appeared the most prominent target of Rbm20 for circularization, we also detected Rbm20-dependent circRNAs arising from other genes including Fan1, Stk39 and Bcl2l13. Interestingly, only Ttn circRNAs seemed to arise from Rbm20-mediated skipped exons. In conclusion, cardiac circRNAs are mostly derived from constitutive exons, suggesting that these circRNAs are generated at the expense of their linear counterpart and that circRNA production impacts the accumulation of the linear mRNA.
Circular RNAs arise from constitutive exons

Introduction

Circular RNAs (circRNAs) have only recently been added to the family of non-coding RNAs, despite their discovery over 20 years ago (Nigro et al. 1991). This species of RNA molecules was largely ignored due to their unusual splicing behaviour in which exons are joined at consensus splice sites, but in a scrambled order relative to the primary transcript. Decades later, with the advent of next generation sequencing, thousands of endogenous circRNAs were found to be expressed in various tissues and cells types (Salzman et al. 2012; Memczak et al. 2013; Jeck and Sharpless 2014), including the heart (Khan et al. 2016; Tan et al. 2017; Jakobi et al. 2016; Werfel et al. 2016). Recent studies have revealed that circRNAs can regulate gene expression by different mechanisms; specifically, by serving as miRNA sponges (Memczak et al. 2013; Zheng et al. 2016; Hansen et al. 2013), facilitating transcription of their host gene by directly associating with RNA polymerase II (Zhang et al. 2013) or forming platforms for protein interactions (Du et al. 2016). Emerging evidence also suggests that some circRNAs can act as templates for protein synthesis (Pamudurti et al. 2017; Legnini et al. 2017; Yang et al. 2017).

The biogenesis of circRNAs is linked to pre-mRNA splicing, a process carried out by the spliceosome (Starke et al. 2015; Ashwal-Fluss et al. 2014). Pre-mRNA splicing represents the process of removal of introns and joining of exons in a linear fashion to produce a mature mRNA. CircRNAs, however, are formed by a back-splicing event, in which a donor site of an exon is not connected with an acceptor site of a downstream exon as observed in linear splicing, but rather with an acceptor site of an upstream exon. This gives rise to a single-stranded RNA loop with a unique exon-exon junction not present in the linear transcript. Mechanistically, back-splicing requires that the donor and acceptor site of the back-spliced exons are brought in close proximity to each other, which can be accomplished by direct RNA base-pairing of reverse complementary sequences in the introns flanking the back-spliced exons, or by the interaction of RNA binding proteins that dock on these flanking introns (Ashwal-Fluss et al. 2014; Conn et al. 2015; Zhang et al. 2014).

Ashwal-Fluss et al. (2014) recently demonstrated that there is strong competition between linear splicing and circRNA biogenesis. By introducing strong 5’ and 3’ splice sites in the exons that flank the circRNAs they observed a dramatic reduction in circularization efficiency in minigene experiments in HEK293 cells. Moreover, flies carrying a mutation in RNA polymerase II, which has been shown to increase co-transcriptional splicing efficiency, produced significantly lower numbers of circRNAs. Both interventions decreased circRNA abundance in a context where linear splicing was more efficient, indicating that inclusion of exons in the linear transcript competes with the circRNAs production of the same exons. Thus circRNA production seems to have a negative effect on linear splicing and therefore on gene expression (Ashwal-Fluss et al. 2014).

How RNA circularization is mechanistically connected to alternative splicing remains largely unknown, but splicing factors such as Muscleblind and Quaking have been shown to promote circularization in Drosophila S2 cells and human mesenchymal cells, respectively (Ashwal-Fluss et al. 2014; Conn et al. 2015). RNA Binding Motif protein 20 (RBM20) is yet another splicing factor that has recently been
implicated in circRNA production, specifically from the TTN gene (Khan et al. 2016). RBM20 is highly expressed in the heart and regulates alternative splicing of a large number of genes, including titin (TTN); a structural giant in the cardiomyocyte (Guo et al. 2012; Labeit and Kolmerer 1995). Mutations in the gene encoding RBM20 have been implicated in a clinically aggressive form of dilated cardiomyopathy (DCM), characterised by progressive dilation and dysfunction of the left ventricle (Brauch et al. 2009; Beqqali et al. 2016). The missplicing of TTN is considered an important reason why RBM20 mutation carriers develop DCM. We recently discovered that a remarkably large number of circRNAs are produced from the TTN gene and that RBM20 is required for the production of a subset of these TTN circRNAs (Khan et al. 2016).

In the current study, we hypothesised that alternative splicing, and in particular exon skipping, drives the formation of cardiac circRNAs, and that Rbm20-mediated alternative splicing drives the formation of circRNAs that arise from Rbm20 target genes. To investigate to what extent alternative splicing underlies circRNA formation in the heart, we performed RNA sequencing on human and mouse hearts, and on the hearts of Rbm20 knockout (KO) mice. We integrated in silico analyses of alternative splicing and circRNA production in human heart and revealed that ~90% of cardiac circRNAs are produced from constitutive exons and ~10% from alternatively spliced exons. Strikingly, of the circRNAs that were formed from alternatively spliced exons, a large fraction (~26%) stemmed from the TTN gene. In addition, we show that the strong correlation observed between exon skipping and circular RNA formation in the Ttn gene does not appear as a common mechanism for other Rbm20 target genes, suggesting that circRNA formation is not a general function of Rbm20. In conclusion, we show that alternative splicing drives the formation of only a small subset of cardiac circRNAs, as most cardiac circRNAs arise from exons that are constitutively spliced.

Materials and Methods

Rbm20 knockout mice
Generation of the Rbm20 KO mice has been described previously (Khan et al. 2016). In line with other Rbm20 KO models (Guo et al. 2012; Methawasin et al. 2014), homozygous Rbm20 knockout mice revealed abnormal Ttn splicing and a DCM phenotype, which is manifested by LV dilatation and impaired cardiac function. Total RNA was extracted from LV tissue samples of 6 months old Rbm20 KO mice (3 wild-type and 3 Rbm20-KO) with TRI Reagent (Sigma-Aldrich) according to manufacturer’s protocol. 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.

Library preparation and whole transcriptome RNA sequencing
RNA quality was assessed with the Agilent 2100 Bioanalyzer. All samples had a RIN score of ~8. Total RNA samples (500 ng) were treated with biotin-streptavidin-based bead systems (Exiqon) to minimise ribosomal contamination. Ribosomal-depleted RNA libraries were sequenced on an Illumina NextSeq
Circular RNAs arise from constitutive exons

500 platform in paired-end mode and with a read length of 101bp at Exiqon A/S (Vedbaek, Denmark). Sequencing depth was approximate 70 million raw reads per sample (Supplemental Table S12). Base-calling was performed using the bcl2fastq 2.0 Conversion Software from Illumina.

**Quality Control and processing of RNA-Seq data**

Quality control of fastq files was performed using FASTQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). Trimmomatic version 0.351 (Bolger et al. 2014) 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 utilised for the downstream analyses.

**Conservation analysis between human and mouse circRNAs**

To determine homologous exons groups, we used UCSC liftOver (Karolchik et al. 2004) to convert mouse splice coordinates (mm9) to human coordinates (hg19). For the analysis, we used the 848 unique circRNAs detected in wild-type mice (1 read counts in at least 1 mouse) and the 3487 unique circRNAs detected in human controls (1 read counts in at least 1 control) from our previous work (Khan et al. 2016). For a more stringent analysis, we filtered the human and mouse circRNAs for the expression level using the cut-off of at least 2, 10, 40 and 70 normalised read counts in at least 2 samples.

**Reverse complementary sequences (RCS) analysis in the flanking introns of the predicted circRNAs**

For each circRNA we aligned the downstream and reversed complement of the upstream intron, using the pairwiseAlignment function in the R package Biostrings and retrieved the score of the best local alignment. To calculate a p-value associated with each match, we compared the actual score to the distribution of scores obtained by aligning the sequence of the best local alignment (without gap) with 1000 random introns. The p-value was then adjusted for multiple testing using the FDR methods. The following alignment parameters were used: match = 2; mismatch = -3, gap opening = -5, gap extension = -2. To calculate the percent sequence identity for a pairwise sequence alignment we used the following formula: 100 * (identical positions) / (aligned positions + internal gap positions). We repeated the procedure for each pair intron flanking circRNA. As the last step, we blasted the identified RCS with Repeat Masker Database using ABBlast algorithm (Smit, AFA, Hubley, R & Green, P.).

**Identification of back-spliced and linear junction reads – MapSplice**

Back-spliced and linear junction reads were identified using MapSplice version 2.2.02 (Wang et al. 2010). For circRNAs detection we set the following options --min-fusion-distance 200 (as suggested by the authors), --filtering 1 and --min-map-len 25. The reported linear and back-spliced junction coverage 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 or linear junctions of the host gene i in sample j. The raw counts were then normalised using DESeq2 (Love et al. 2014). GenomicRanges function and custom R scripts were used to overlay linear splicing events with back-splicing events.
Differential circular RNA expression analysis – DESeq2

Differential expression analysis of circRNAs was performed using the R Bioconductor package, DESeq2 (Love et al. 2014). To investigate circRNAs expression in Rbm20 KO and wild-type mice, 2 sets of differential expression analyses were conducted. The first analysis was performed using the total number of circRNAs detected in the hearts of wild-type mice and Rbm20 KO mice. To increase sensitivity, the second analysis was performed considering only circRNAs detected in at least two wild-type or two Rbm20 KO mice with normalised reads counts ≥2. The negative binomial model was used to estimate differentially expressed circRNAs for each analysis. In the end, only those circRNAs passing a log2 fold change cut-off of 1 together with a p-value cut-off of 0.05 were deemed significantly differentially expressed. Owing to the limited number of samples (n=3 per group), the adjusted p-value was not used as a criterion to select candidates and the final set comprised of 38 circRNAs, which were selected for downstream analyses.

Rbm20 binding site enrichment analysis

Genomic sequences of introns flanking the differentially expressed circRNAs were obtained from the UCSC genome table browser. Gregexpr2 function in the R package Biostrings and custom R scripts were then used to determine for each candidate circRNA, the genomic positions of the Rbm20 motif (UCUU) occurring within introns flanking 100bp upstream and downstream of the back-splicing acceptor and donor sites respectively. Rbm20 motifs occurring in flanking introns of the circRNAs were then tallied. An equal number of randomly selected introns was utilised as a control set. The number of Rbm20 motifs occurring within introns flanking 100bp upstream and downstream of these ‘control’ introns pairs were tallied.

Differential exon usage analysis- DEXSeq

The paired-end RNA-seq reads passing the quality controls from the 3 Rbm20 KO mice and wild type mice were first aligned against the mouse genome, Gencode annotation release M1 (NCBI37), using TopHat2 (Kim et al. 2013) version 2.0.14 with default values. To test for exon usage differences between Rbm20 KO and wild-type mice we used the R Bioconductor package DEXSeq version 1.20.0 (Anders et al. 2012), using GenomicRanges infrastructure to count the number of aligned reads overlapping with each exonic region.

Percentage Spliced In of the back-spliced exons – PSI

The ‘percentage spliced in’ index (PSI) was computed by executing the script PSI.sh of Schafer et al. (2017) (Schafer et al. 2015) on exonic parts derived from the Gencode annotation release M1 (NCBI37) for mouse and Gencode annotation release 19 (GRCh37.p13) for human. GenomicRanges function and custom R scripts were used to overlay the coordinates of the back-spliced exons with the coordinates of the exon bins identified during the PSI calculation. Back-spliced exons of circRNAs detected in at least two samples, wild-type or Rbm20 KO mice, and two human controls with normalised reads counts ≥ 2 were used in the analysis. If an exon’s boundary is not the same in all transcripts, the exon is cut into two
Circular RNAs arise from constitutive exons or more parts, referred to as exon bins. We considered exons to be constitutive if they were included in the linear mRNA in at least 90% of the transcripts (PSI 0.9) (Schafer et al. 2017).

**Data visualisation and statistical analysis**
All results generated from differential circRNA analysis were visualised using the R Bioconductor package, ggplot2. The graphical representation of the TTN gene and circRNAs were constructed using the R package, GenomicRanges and custom R scripts, respectively. For differential circRNA expression across conditions, circRNA back-spliced junction data were statistically analysed using the negative binomial test in the R Bioconductor package, DESeq2. All correlation tests were performed in R, using the Spearman’s correlation as the method of choice. Differences in the enrichment of Rbm20 sites between circRNAs and control sequences were compared using the Fisher-exact test. To be able to analyse circRNAs that were not expressed in the RBM20 KO mice or wild-type mice, a count of 1 was added to avoid problems with infinite values when log transforming. The p-value for delta PSI difference was calculated using a one-tail t-test unequal variance. Scatter plots were used to plot the mean values of the samples.

**Preparation of poly(A)-positive and poly(A)-negative RNA fractions**
Poly(A)-positive and poly(A)-negative RNA fractions were generated as described previously (Khan et al. 2016). Briefly, 6 μg total RNA from three wild-type mouse hearts and 200 μl oligo d(T)25 Magnetic Beads (NEB) were used to generate poly(A) enriched and depleted fractions. The beads were washed first and subsequently incubated with the RNA in a 100 μl reaction for 10 minutes with gentle agitation. Afterwards, the poly(A) negative fraction was collected as supernatant by pulling the beads to the side of the tube. After three times washing, beads were incubated in 100 μl elution buffer at 50°C for 2 minutes to recover the poly(A) positive RNA fraction. The poly(A) negative RNA fraction was also treated at 50°C for 2 minutes.

**RNase R treatment**
RNase R treatment was performed as described previously (Khan et al. 2016). Briefly, 1 μg total RNA from three wild-type mouse hearts was treated with or without 5 units of RNase R (Epicentre) at 37°C for 10 minutes in a 20 μl reaction followed by heat inactivation at 95°C for 3 minutes.

**RT-PCR**
RT-PCR was performed as described previously (Khan et al. 2016). Briefly, 1 μg of poly(A) enriched/depleted or RNase R treated/non-treated RNA was DNA depleted with DNase I (Invitrogen) and subsequently used to generate cDNA using SuperScript II Reverse Transcriptase (Invitrogen) and random hexamers (Invitrogen). 5-times diluted cDNA was amplified with HOT FIREpol DNA polymerase together with divergent primers for circRNA detection or normal convergent primers for mRNA detection. Primers are designed to amplify the smallest band depicted on the gels. Sanger sequencing confirmed the presence of the expected back-splice junctions in these bands.
Results

Identification of circRNAs in mouse hearts
To detect circRNAs in the mouse heart, ribosomal-depleted RNA obtained from left ventricles of three wild-type and three Rbm20 KO mice was used for whole transcriptome sequencing (Fig. 1A). The MapSplice tool was used to identify back-splice junctions in the RNA-seq data (Wang et al. 2010; Khan et al. 2016). We detected a total of 1283 unique circRNAs in these 6 mouse samples (Table 1; Supplemental Table S1). However, the expression of the circRNAs was rather variable between hearts, as there were only 156 circRNAs commonly detected in all three wild-type hearts (Table 1). Most circRNAs were predicted to comprise between 2 and 6 exons within the back-spliced exons, but we also detected circRNAs potentially harbouring as many as 30 to 100 exons within the back-spliced exons, albeit less frequently (Supplemental Fig. S1A). We confirmed the previous observation (Salzman et al. 2012) that particularly exon 2 of the host gene is overrepresented in circRNAs (Supplemental Fig. S1B). Furthermore, half of the host genes were predicted to produce a single circRNA, while the other half generated between 2 to 9 different circRNAs, with the exception of Ttn, which generated a total of 38 putative circRNAs (Supplemental Fig. S1C). These Ttn specific circRNAs were mostly formed from the I-band region of the gene; a region known to undergo complex alternative splicing. This is in line with our previous report showing robust circRNA production from titin’s I-band region in human hearts (Khan et al. 2016) (Fig. 1B).

Only a small subset of circRNAs expressed in the heart are evolutionarily conserved

We have previously identified a total of 7130 unique circRNAs in human hearts, of which 3478 were found in hearts of healthy individuals. In the current study, using the exact same methodology, we detected a total of 1283 circRNAs in the mouse hearts. Overall, we found that ~9% and ~3% of the genes expressed in the human and mouse hearts produced circRNAs, respectively. To investigate to what extent circRNAs in human and mouse hearts are evolutionarily conserved, we converted mouse circRNA back-splice genomic coordinates to human coordinates and compared the genomic positions of circRNAs expressed in human and mouse hearts. We found that only 5% of human circRNAs were conserved in the mouse (Fig. 1C; Supplemental Table S2). Interestingly, when filtering human circRNAs for high expression levels (≥70 reads), the overlap with mouse circRNAs increases from 5% to 19% (Fig. 1C). This indicates that the highly expressed human circRNAs show a higher degree of conservation. Vice versa, up to 50% of the mouse circRNAs were also identified in human hearts (Supplemental Table S3).
Circular RNAs arise from constitutive exons

Figure 1: Characterization of circRNAs in mouse hearts. (A) Strategy used to elucidate the relation between the circRNAs formation and alternative splicing. (B) Schematic representation of two splice isoforms of the Ttn gene (N2BA-G and N2B) and the location of the 38 identified circRNAs in the mouse hearts and the previously identified TTN circRNAs in human hearts5 (C) Table showing conservation of cardiac circRNAs derived from mouse and human after filtering the human circRNAs for the expression level. Between brackets the number of circRNAs passing the filter are indicated. (D) Scatter plot showing the relation between the percentage of complementarity and the length of the identified reverse complementary sequences (RCS) passing the p-value cut-off (FDR ≤ 0.05). RCS sharing high sequence similarity with the mouse B1 elements are depicted in black.

Table 1. Summary of the circRNAs detected in mouse hearts.

<table>
<thead>
<tr>
<th></th>
<th>wt1</th>
<th>wt2</th>
<th>wt3</th>
<th>ko1</th>
<th>ko2</th>
<th>ko3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique circRNAs identified in each sample</td>
<td>451</td>
<td>461</td>
<td>421</td>
<td>550</td>
<td>180</td>
<td>525</td>
</tr>
<tr>
<td>Commonly identified circRNAs per genotype</td>
<td>156</td>
<td></td>
<td></td>
<td>105</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unique circRNAs identified in all samples</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1283</td>
</tr>
</tbody>
</table>

Number of circRNAs identified in each sample, per genotype (at least 1 read in all 3 samples) and in all samples (at least 1 read in at least 1 sample).
Chapter 4

Mouse cardiac circRNAs are flanked by Reverse Complementary Sequences (RCS)

There is a consensus that the presence of RCS in introns flanking back-spliced exons is associated with the formation of circRNAs (Zhang et al. 2014; Ivanov et al. 2015). To investigate the presence of these RCS, we retrieved 1000 bp of intronic sequence flanking each of the 1283 mouse circRNAs and investigated complementarity (Supplemental Fig. S2A). All the introns flanking the detected circRNAs had RCS, and 802 were significantly enriched over randomly selected introns (FDR ≤ 0.05) (Supplemental Table S4). It is noteworthy that some of the matches (complementarity level of >65%) have a length of up to 300 nucleotides (Fig. 1D). Analysis of the identified RCS showed that ~13% of them share sequence similarity with inverted B1 elements (mouse equivalent of Alu-repeats in human) (Fig. 1D), ~10% with other elements like short interspersed sequence elements (SINEs) B2-B4, long interspersed sequence elements (LINEs) and simple repeats, and ~77% with none of the known elements (Supplemental Fig. S2B). The sequences of all RCS are shown in Supplemental Table S4, however, experimental studies are needed to provide evidence whether base-pairing events of the identified sequences are necessary and/or sufficient for circularization of these mouse circRNAs.

38 circRNAs are differentially expressed in hearts of RBM20 knockout mice

To identify Rbm20-dependent circRNAs expressed in the heart, we performed differential expression analysis comparing circRNAs in Rbm20 KO with wild-type mice. We identified 38 out of 1283 circRNAs to be differentially expressed, of which 26 were down-regulated and 12 were up-regulated in Rbm20 KO mice compared to wild-type mice (Fig. 2A; Supplemental Table S1). Notably, a subset of 19 down-regulated circRNAs was completely absent in Rbm20 KO mouse hearts, while they were readily expressed in wild-type hearts. Eleven of those circRNAs were generated from the Ttn gene and the remaining from the host genes Unc13b, Stk39, Arhgap10, Tfdp2, Sorbs1, Fan1, Insr and Xdh (Table 2). A subset of 7 up-regulated circRNAs was uniquely expressed in the hearts of Rbm20 KO mice, and they were generated from the host genes Sorbs1, D19Ertd386e, Ehmt1, Gcom1, Bcl2l13, Smad1 and Ttn (Table 2).

To validate these differentially expressed circRNAs experimentally, we selected 6 circRNA candidates, based on p-value and expression level, that were down-regulated (cTtn_2, cTtn_3, cTtn_4, cFan1, cStk39 and cXdh) and 4 circRNA candidates that were up-regulated (cTtn_7, cSorbs1, cBcl2l13, cEhmt1) in Rbm20 KO hearts (Fig. 2B; Table 2). We designed divergent primers flanking the circRNA-specific back-splice junctions and performed RT-PCR on RNA isolated from hearts of 6 wild-type, 6 heterozygous and 6 Rbm20 KO mice. Sanger sequencing confirmed the presence of back-splice junctions in the amplicons, and RT-PCR on poly(A)-negative and -positive RNA, and on RNase R-treated RNA and mock samples revealed that we detected bona fide circRNAs (Supplemental Fig. S3; Supplemental Table S5). In line with our previous report (Khan et al. 2016), we confirmed the loss of the three Ttn circRNAs in the Rbm20 KO hearts, but also the circRNAs produced from Fan1 and Stk39. Interestingly, the circRNA derived from the Fan1 gene seemed to depend critically on Rbm20 expression, as we found that cFan1 was substantially reduced in the heterozygotes and completely absent in the Rbm20 KO hearts (Fig. 2B). Of the circRNAs that were up-regulated in the Rbm20 KO hearts, we could experimentally validate cTtn_7, cSorbs1 and cBcl2l13. Expression of cXdh and cEhmt1 did not seem to be affected in the
Circular RNAs arise from constitutive exons

absence of Rbm20 in this set of 18 mouse hearts. Overall, 8 out of 10 differentially expressed circRNAs could be experimentally validated.

Rbm20 binding site enrichment analysis showed that introns flanking the 38 differentially expressed circRNAs are 2-fold enriched for Rbm20 binding sites compared to a random set of introns; with the most significant enrichment occurring within a flanking distance of 100bp (p<0.01) (Supplemental Fig. S4). Interestingly, even after removing Ttn circRNAs from the analysis, a significant enrichment remained. For instance, we identified a total of 4 potential Rbm20 binding sites in the introns flanking cFan1. In conclusion, the RNA-seq in wild-type and Rbm20 KO hearts in combination with the Rbm20 binding site prediction revealed that the Ttn gene seems to be the most prominent target of Rbm20-regulated circRNA production. Nevertheless, of the 38 differentially expressed circRNAs, 26 originated from other genes, suggesting that circRNAs arising from Ttn gene are not the only ones that are regulated by Rbm20.

Figure 2: A subset of circRNAs are differentially expressed in hearts of Rbm20 knock-out mice. (A) Volcano plot showing differentially expressed circRNAs in Rbm20 KO hearts compared to wild-type hearts. Differentially expressed circRNAs (Log2 fold change ≥ 1 and p-value ≤ 0.05) are marked in grey. (B) RT-PCR of circRNAs on hearts of 6 wild-type, 6 heterozygous and 6 Rbm20 KO mice. GAPDH was used as input control.
Table 2. Normalized read counts of differentially expressed circRNAs in hearts of Rbm20 knock-out mice.

<table>
<thead>
<tr>
<th>circRNA</th>
<th>gene</th>
<th>wt1</th>
<th>wt2</th>
<th>wt3</th>
<th>ko1</th>
<th>ko2</th>
<th>ko3</th>
<th>Log2 FC *</th>
<th>p-value</th>
<th>circRNA id (mm9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cTtn_4</td>
<td>Ttn</td>
<td>906</td>
<td>1379</td>
<td>1649</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-11.3</td>
<td>1.3E-27</td>
<td>chr2:76695961-76719896</td>
</tr>
<tr>
<td>cTtn_3</td>
<td>Ttn</td>
<td>137</td>
<td>122</td>
<td>105</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-8.3</td>
<td>1.1E-14</td>
<td>chr2:76695961-76722254</td>
</tr>
<tr>
<td></td>
<td>Ttn</td>
<td>25</td>
<td>38</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-6.1</td>
<td>4.0E-07</td>
<td>chr2:76685445-76719896</td>
</tr>
<tr>
<td></td>
<td>Ttn</td>
<td>18</td>
<td>15</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-5.7</td>
<td>4.8E-06</td>
<td>chr2:76686827-76719896</td>
</tr>
<tr>
<td>cTtn_2</td>
<td>Ttn</td>
<td>15</td>
<td>23</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-5.3</td>
<td>5.6E-05</td>
<td>chr2:76640288-76656873</td>
</tr>
<tr>
<td></td>
<td>Ttn</td>
<td>11</td>
<td>13</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-5.0</td>
<td>2.6E-04</td>
<td>chr2:76700407-76719896</td>
</tr>
<tr>
<td></td>
<td>Ttn</td>
<td>16</td>
<td>15</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-4.9</td>
<td>3.4E-04</td>
<td>chr2:76686827-76722254</td>
</tr>
<tr>
<td></td>
<td>Tfdp2</td>
<td>15</td>
<td>12</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-4.8</td>
<td>7.6E-04</td>
<td>chr9:96147507-96102550</td>
</tr>
<tr>
<td>cStk39</td>
<td>Stk39</td>
<td>5</td>
<td>19</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-4.5</td>
<td>2.5E-03</td>
<td>chr2:68196871-68248198</td>
</tr>
<tr>
<td></td>
<td>Ttn</td>
<td>9</td>
<td>9</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-4.5</td>
<td>2.4E-03</td>
<td>chr2:76695961-76731234</td>
</tr>
<tr>
<td>cFan1</td>
<td>Fan1</td>
<td>10</td>
<td>11</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-3.8</td>
<td>1.6E-02</td>
<td>chr7:71506826-71516545</td>
</tr>
<tr>
<td>cXdh</td>
<td>Xdh</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-3.8</td>
<td>1.8E-02</td>
<td>chr17:74255350-74257085</td>
</tr>
<tr>
<td></td>
<td>Ttn</td>
<td>0</td>
<td>7</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-3.7</td>
<td>2.7E-02</td>
<td>chr2:76700857-76719896</td>
</tr>
<tr>
<td></td>
<td>Ttn</td>
<td>0</td>
<td>14</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-3.6</td>
<td>2.8E-02</td>
<td>chr2:76699154-76719896</td>
</tr>
<tr>
<td>Unc13b</td>
<td>0</td>
<td>8</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-3.6</td>
<td>2.9E-02</td>
<td>chr4:43128032-43100775</td>
</tr>
<tr>
<td>Arhgap10</td>
<td>10</td>
<td>3</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-3.6</td>
<td>2.7E-02</td>
<td>chr8:79868539-79875571</td>
</tr>
<tr>
<td></td>
<td>Ttn</td>
<td>0</td>
<td>18</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-3.4</td>
<td>4.1E-02</td>
<td>chr2:76640288-76655920</td>
</tr>
<tr>
<td>Insr</td>
<td>16</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-3.4</td>
<td>4.7E-02</td>
<td>chr8:3158696-3174888</td>
</tr>
<tr>
<td>Sorbs1</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-3.3</td>
<td>4.3E-02</td>
<td>chr19:40411476-40439656</td>
</tr>
</tbody>
</table>
### Upregulated

<table>
<thead>
<tr>
<th>gene</th>
<th>wt1</th>
<th>wt2</th>
<th>wt3</th>
<th>ko1</th>
<th>ko2</th>
<th>ko3</th>
<th>Log2 FC</th>
<th>pvalue</th>
<th>circRNA id</th>
</tr>
</thead>
<tbody>
<tr>
<td>cEhmt1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>11</td>
<td>3</td>
<td>4,2</td>
<td>1,00E-02</td>
<td>chr2:24669047-24680587</td>
</tr>
<tr>
<td>Smad1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>13</td>
<td>0</td>
<td>3,8</td>
<td>2,00E-02</td>
<td>chr8:81878732-81879240</td>
</tr>
<tr>
<td>cSorbs1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>9</td>
<td>0</td>
<td>3,8</td>
<td>2,00E-02</td>
<td>chr19:40437692-4039656</td>
</tr>
<tr>
<td>cTtn_7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>11</td>
<td>3,8</td>
<td>2,00E-02</td>
<td>chr2:76700169-76700484</td>
</tr>
<tr>
<td>D19Erd3 86e</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>4</td>
<td>3,4</td>
<td>5,00E-02</td>
<td>chr9:71362876-71403544</td>
</tr>
<tr>
<td>cBcl2l13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>7</td>
<td>3</td>
<td>3,4</td>
<td>5,00E-02</td>
<td>chr6:120826364-120820784</td>
</tr>
</tbody>
</table>

*A finite estimate was added to the zero count to allow fold change calculation.

### Rbm20-dependent alternative splicing and circRNA production

We have previously shown that Rbm20-dependent exon skipping within Ttn’s I-band region is associated with circRNA formation (Khan et al. 2016). To assess whether Rbm20-controlled alternative splicing and circRNA formation are connected beyond the Ttn gene, we calculated the Percentage Spliced In (PSI) (Schafer et al. 2015; Kakaradov et al. 2012) of the back-spliced exons of all 38 differentially expressed circRNAs. The PSI indicates the efficiency of splicing of a specific exon into the transcript population of a gene. It ranges from 0 to 1, indicating an exon is never (0) to always (1) spliced into the expressed transcripts of a gene. In Figure 3A and B, we plotted the PSI of back-spliced exons of the 38 differentially expressed circRNAs against the expression of the corresponding circRNAs. Figure 3A shows that particularly the circRNAs from the Ttn gene arise from exons that are spliced out of the linear transcripts. The back-spliced exons of almost all other differentially expressed circRNAs, except those arising from Arhgap10 and 2310046A06Rik host genes, have a PSI close to 1, indicating that these exons are always included in the linear transcripts. In Figure 3B, where the PSI is plotted against the expression of the circRNAs in the Rbm20 KO hearts, it can be appreciated that PSI levels of the Ttn back-spliced exons increase to 1, indicating that these exons are now included in the linear Ttn transcripts. To quantify the difference in exon skipping between Rbm20 KO and wild-type mice, we calculated the delta-PSI of the back-spliced exons and plotted these against the expression changes of the circRNAs (Fig. 3C, Supplemental Table S6). Again, this shows that the exons used for circulization are mostly constitutive exons since the exons that are differentially back-spliced upon Rbm20 loss are not differentially spliced in the linear transcript. Also in this approach, the most notable exceptions are Ttn circRNAs.

In addition to these PSI calculations, we used an alternative strategy to assess whether circRNA formation
and alternative splicing are concurrently regulated. We used MapSplice to retrieve the linear junction reads (lj reads) stemming from the back-spliced exons in the Rbm20 KO and wild-type mice. We then plotted the fold changes of the linear junction reads against the fold changes of back-spliced junction reads (bsj reads) in Rbm20 KO compared to wild-type mice (Supplemental Fig. S5). This also shows that reduced numbers of Ttn circRNAs in Rbm20 KO hearts associates with increased linear junction read counts, indicative of a negative correlation between circRNA production and linear exon usage.

Besides regulating Ttn splicing, Rbm20 is known to regulate splicing of many other cardiac genes, including Camk2d, Ldb3 and Myh7 (Maatz et al. 2014). We next performed a dedicated search for circRNAs that may have arisen by exon skipping events in these other Rbm20 target genes. Genome-wide differential exon usage analysis (using the R Bioconductor package DEXSeq) in our Rbm20 mouse model confirmed that loss of Rbm20 leads to missplicing of several genes, including Ttn, Myh7, Obscn, Slc8a1, Sorbs2, Camk2d, Ldb3 and Rtn4 (Supplemental Table S7). We retrieved the exons that are alternatively spliced by Rbm20 and overlapped these with the back-spliced exons to assess whether these exons generated circRNAs. Interestingly, the exons that were found to be differentially spliced in the Rbm20 KO mice do not produce detectable circRNAs, except the ones arising from the Ttn gene (Supplemental Table S8). As described in the previous section we identified cFan1 as a circRNA that also critically depends on Rbm20 expression. To investigate whether Rbm20-dependent alternative splicing may underlie cFan1 production we generated a splicing plot for the Fan1 gene showing exon usage in Rbm20 KO and wild-type mice. As shown in Supplemental Figure S6, several exons in the Fan1 mRNA are alternatively spliced but not the two exons that circularise (Supplemental Table S9). Therefore, it remains unclear at this point how Rbm20 regulates the production of cFan1. In conclusion, Rbm20-mediated exon skipping results in circRNA production from the Ttn gene, but not from other bona fide splicing targets of Rbm20.

**Genome-wide analysis of alternative splicing and circRNA production in the healthy heart**

With the availability of RNA datasets of the mouse and human heart, we have the opportunity to investigate the connection between alternative splicing and circRNAs formation, in an unbiased and genome-wide manner. Therefore, we calculated the PSI of the back-spliced exons of all circRNAs passing the filtering step (read count ≥2 in both samples) in 2 control human hearts from our previous study (Khan et al. 2016). Figure 4A shows the PSI of back-spliced exons of the 1289 circRNAs detected in human hearts, plotted against the expression level of the corresponding circRNAs. We considered exons to be constitutive if they were included in the linear mRNA in at least 90% of the transcripts (PSI>0.90) (Schafer et al. 2017). The results show that only ~10% of the circRNAs had back-spliced exons with a PSI ≤ 0.90 and thus may have been derived from exon skipping events (Supplemental Table S10). This indicates that most circRNAs are not derived from alternatively spliced exons but from constitutive exons. Of the 10% of the circRNAs that we found to associate with alternative splicing, a large proportion (26%) derived from the TTN gene (Fig. 4A, blue dots). Interestingly, we also found several other genes where exon skipping was associated with circRNA production. To visualise the relation between exon skipping and circRNA production in greater detail, we plotted the PSI of four of these genes, PALM2,
Circular RNAs arise from constitutive exons

YAF2, CUX1, and ZEB1 and marked the exons used for circularization in orange (Fig. 4B-E). In the mouse hearts around 13% of the circRNAs were associated with exon skipping events, when considering a PSI of 0.90 as a cutoff (Supplemental Fig. S7; Supplemental Table S11). In conclusion, many of the cardiac circRNAs that coincide with alternative splicing are derived from the Ttn gene. However, particularly in the human heart, we found numerous examples of non-TTN circRNAs that were produced from regions that undergo extensive alternative splicing.

Based on these results we propose a model in which there are two mechanisms for circRNA production: competition-based and exon skipping-based circRNA production (Fig. 4F). In the competition-based mechanism, which is the most frequently observed mechanism (~90% of the circRNAs), a pre-mRNA is used to produce a circRNA, but not a mature linear mRNA. In this case, circRNA biogenesis may be an important regulator in the control of gene expression and depending on the gene in question; there might be a favoured production of linear mRNAs or circular RNAs. In the exon skipping-based mechanism, which is less frequently observed (~10%), pre-mRNA undergoes alternative splicing and results in the formation of both a linear exon-skipped mRNA and a circular RNA.

Figure 3: Rbm20-dependent alternative splicing and circRNA production. Scatterplots showing the relation between the PSI of the back-spliced exons and the expression of the circRNAs (Log2 bsj reads) in (A) wild-type mice and (B) Rbm20 KO mice. Only the 38 differentially expressed circRNAs are depicted. The names of the circRNA host genes are plotted. (C) Scatterplot showing the relation between the dPSI of the back-spliced exons of the 38 differentially expressed circRNAs (dPSI back-spliced exons) and the expression fold changes (Log2 bsj reads FC) of the corresponding circRNAs between Rbm20 KO hearts and WT hearts.
Figure 4: CircRNAs in the human heart are derived from exons that are typically not alternatively spliced. (A) Scatterplot showing the relation between the PSI of the back-spliced exons with the expression of the corresponding circRNAs (Log2 bsj reads) in the human hearts. There was no significant correlation between extent of exon skipping (i.e. PSI) and the expression level of the circRNAs. Grey dots represent back-spliced exons with PSI > 0.90. Blue dots represent back-spliced exons from TTN gene. Black dots represent back-spliced exons with PSI ≤ 0.90. (B-E) The PSI of the 4 host genes (PALM2, YAF2, CUX1 and ZEB1, indicated in orange in Figure A) are plotted, in which exon skipping is intimately connected with circRNA production. Black dots represent all exon bins of the corresponding genes and orange dots represent the exon bins that give rise to the circRNAs (e.g. back-spliced exons). (F) Proposed model showing two mechanisms for circRNA production: competition-based and exon skipping-based circRNA production.
Discussion

It is known that the splicing machinery can circularise exons and that linear splicing is directly involved in circRNA production (Zaphiropoulos 1996). However, to what extent alternative splicing, and in particular exon skipping, drives circRNA formation is currently unknown. Here, we analysed RNA-seq data from mouse and human hearts on alternative splicing and circRNA production, and show in humans that ~90% of the circRNAs are derived from exons that are typically not alternatively spliced (i.e. constitutive exons). This suggests that these circRNAs are mostly generated at the expense of linear mRNA isoforms, and that circRNA production competes with the accumulation of linear mRNA. Conversely, ~10% of the human circRNAs may have been produced from exon skipping events, as these circRNAs are derived from exons that are subjected to alternative splicing (Figure 4F). Interestingly, a large proportion (~26%) of the circRNAs that are associated with exon skipping events in the human heart is derived from the TTN gene.

Our finding that circRNAs can be derived from exon-skipping events is in line with other studies (Salzman et al. 2012; Jeck et al. 2013), although the described frequency of these events differs. Jeck et al. (2013) determined exon-skipped linear transcripts from human fibroblasts and found that 45% of the expressed circRNAs are derived from linear splicing reads with characteristics of an exon-skipping event. Salzman et al. (2012) provided circumstantial evidence that in a variety of normal and malignant human cells, 2% of the circRNAs are derived from exons predicted to undergo alternative splicing. In the current study, we used an integrated approach to evaluate alternative splicing and circRNA production by calculating PSI, back-spliced junction reads and linear junction reads and provide evidence that at most 10% of the circRNAs are derived from exon skipping events. This difference in frequencies between the studies may be explained by differences in examined tissue, or the difference in methods / cut-offs to detect exon skipping. Unfortunately, we were unable to examine the underlying differences since it is unclear what methods and cut-offs were exactly used for the detection of alternative splicing in the other two studies. We were rather strict in setting the cut-off for the detection of alternative splicing (PSI≤0.90); with a more lenient cut-off for exon skipping (PSI≤0.98), we find that up to 25% of the circRNAs can be derived from exon skipping.

We recently discovered a causal link between exon skipping regulated by the splice factor RBM20 and circular RNA formation in the human TTN gene (Khan et al. 2016). These results led us to hypothesise that this phenomenon may not be restricted to a single gene but that RBM20 may also regulate the formation of circRNAs arising from other RBM20 target genes. Therefore, we designed this study to investigate whether Rbm20 is required for the production of circRNAs beyond those of the Ttn gene. We identified 38 differentially expressed circRNAs in Rbm20 KO hearts, and the Ttn gene appeared as the most prominent target of Rbm20 for circRNA production, as we found that 11 Ttn circRNAs were completely absent in Rbm20 KO hearts. Interestingly, RT-PCR of 10 of the most significantly regulated circRNAs, revealed, that besides several Ttn circRNAs, cFan1 was completely lost in the absence of Rbm20.
We next investigated whether the set of 38 circRNAs that were regulated in the Rbm20 KO hearts could have been formed by exon skipping events. We used two different strategies (PSI and linear junction read count of back-spliced exons) to interrogate alternative splicing of linear transcripts, and both methods similarly showed that predominantly the circRNAs from the Ttn gene are associated with exon skipping events. This is a remarkable finding since Rbm20 is known to act as a splicing repressor of many cardiac genes, including Camk2d, Ldb3, Sorbs1 and Myh7 (Maatz et al. 2014). Dedicated searches for circRNAs in the mapped reads of these well-described Rbm20 target genes did not yield any evidence of back-spliced junctions in the exons that are targeted by Rbm20 for splicing. Also, the Sorbs1 and Ryr2 circRNAs that we found to be differentially expressed in the Rbm20 KO hearts were derived from different exons than the ones that are normally repressed for splicing by Rbm20. A differential exon usage plot of the Fan1 gene in wild-type and Rbm20 KO hearts revealed that Rbm20 regulates alternative splicing (i.e. exon inclusion) in the region of the back-spliced exons, but not in the back-spliced exons specifically. The presence of 4 Rbm20 binding sites in the introns flanking the back-spliced exons of cFan1 further suggest that Rbm20 underlies circRNA formation from this gene directly, however, the precise mechanism remains elusive. Overall, the tight correlation between exon skipping and circular RNA formation observed in Ttn gene does not appear a common mechanism for Rbm20 target genes indicating that circRNA formation is not a general function of Rbm20. This also indicates that additional regulators or functional interactions are required for efficient circularization following exon skipping.

We performed a comparative analysis and showed that the normal human heart expresses 4 times more circRNAs than the mouse heart (3478 circRNAs in humans and 848 in mice). This is in line with a study by Tan et al. (2017), where 5 times more circRNAs were detected in the human heart compared to the mouse heart. Werfel et al. (2016) only found 1.6 times more circRNAs in human hearts compared to the mouse hearts; which is probably caused by the relatively high number of circRNAs they detected in the mouse hearts. We further interrogated evolutionary conservation of the circRNAs by converting the mouse genomic coordinates of the back-spliced junctions to the human genome. We show that only 184 (i.e. 5%) of the human circRNAs are also expressed in mouse hearts. When discarding the lowly expressed human circRNAs, the overlap with mouse circRNAs increased from 5% to 19%. This level of evolutionary conservation is in agreement with the study of Werfel et al. (2016), who showed that approximately 13% of the human cardiac circRNAs are conserved to mouse and rats. Vice versa, due to the lower numbers of circRNAs expressed in the mouse heart, a much higher percentage (between 20-50%) of the mouse circRNAs are conserved to human. Clearly, during evolution, an elusive mechanism has led to an astounding increase in the number of circRNAs in the human transcriptome. The observation that the expression of circRNAs is highly variable between hearts from the same species suggests that occasional circularization of exons is easy to evolve and may provide a mechanism for rapid evolution of stably expressed circRNAs. It has been shown that inverted Alu repeats in the introns flanking back-spliced exons are associated with the formation of circRNAs (Jeck et al. 2013). Alu repeats are primate-specific SINE retrotransposons and thus not present in the mouse genome (Kriegs et al. 2007). In the mouse genome, the B1 element is the most abundant class of repetitive elements. They belong to the SINE family and, like Alu repeats in human, originated from an initial duplication of the 7SL RNA and
amplified and duplicated independently in the two genomes while accumulating specific mutations (Quentin 1994). The discrepancy in the numbers of identified circRNAs in human and mouse hearts may relate to the fact that regulatory sequences for circularization (for instance Alu and B1 elements) are not well conserved (Tsirigos and Rigoutsos 2009). Analysis of the introns flanking the back-spliced exons of the mouse circRNAs revealed besides the B1 elements, also reverse complementary sequences that do not share any sequence similarity with transposable elements. Further investigation is required to evaluate whether the identified sequences are necessary for circularization.

In conclusion, this study revealed that the splice factor Rbm20 is not a global regulator of circRNA formation in the heart. The tight correlation between exon skipping and circular RNA formation that we observed for the Ttn gene does not appear a common mechanism for other Rbm20 target genes. In addition, we have shown that cardiac circRNAs are mostly (~90%) produced from constitutive exons, indicating that these circRNAs are generated at the expense of their linear counterpart. Future studies investigating how circRNA production from these constitutive exons affect gene expression and protein production of the host gene will be awaited with great interest.
Chapter 4

References

Circular RNAs arise from constitutive exons

Figure S1: CircRNA characteristics in the mouse heart. (A) Frequency of the number of exons found between the back-spliced coordinates of the predicted circRNAs is plotted. The majority of the back-spliced exons are spaced from multiple exons, most frequently 2–6 exons. (B) Frequency of the exons that are back-spliced in the predicted circRNAs. The most frequently back-spliced exon is the second exon of the circRNA host genes. (C) Frequency of the number of circRNAs produced from the host genes are plotted. The black bar refers to the number of circRNAs produced from the Ttn gene.
Circular RNAs arise from constitutive exons

Figure S2: Reverse complementary sequences (RCS) identified in the introns flanking the mouse circRNAs. (A) Schematic representation showing the methodology used to identify RCS. (B) Bar graph showing the BLAST results of the identified RCS in the introns of the predicted circRNAs against the Repeat masker DB. Approximately 23% of the RCS comprised of transposable elements (SINEs, LINEs, LTR and DNA elements), simple repeats, satellite and low complexity region and the rest contained other unknown motifs (RCS).
**Figure S3: Validation of the selected mouse circRNAs.** 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.
Circular RNAs arise from constitutive exons

Figure S4: Rbm20 enrichment analysis. Bar graph comparing the predicted number of Rbm20 binding sites found in introns (≤100bp distance of the back-splice exons) of the differentially expressed circRNAs, with and without circRNAs from the Ttn gene. A 2- and 1.8-fold enrichment of Rbm20 binding sites was identified in the differentially expressed circRNAs, with and without circRNAs from the Ttn gene, respectively (p<0.01).

Figure S5: Changes in exon usage in linear RNA and circRNAs, in Rbm20 KO versus WT hearts. Scatterplot showing the relation between the expression changes of the exons in the linear transcripts (Log2 lj reads FC) with the expression changes of the exons in the circRNAs (Log2 bsj reads FC). A negative correlation was found between the expression changes of the back-spliced exons in the linear transcripts with the expression changes of the circRNAs mainly due to the presence of Ttn circRNAs (R = -0.37, p-value=0.02). The regression line is depicted in blue.
Figure S6: Differential exon usage of the circRNA host gene Fan1. Splicing plot for Fan1 mRNA showing the differential exon usage in Rbm20 KO mice compared to wild-type mice. Exon bins and cFan1 are depicted in grey and orange respectively. *Exon bins passing the p-value cut-off (≤0.05). † The back-spliced exons overlap with exon bin 8 (E008) and 14 (E014).
Figure S7: CircRNAs in the mouse heart are derived from exons that are typically not alternatively spliced. Scatterplot showing the relation between the PSI of the back-spliced exons (PSI back-spliced exons) with the expression of the corresponding circRNAs (Log2 bsj reads) in wild type mice. Grey dots represent back-spliced exons with PSI > 0.90; blue dots represent back-spliced exons from Ttn gene; black dots represent back-spliced exons with PSI ≤0.90. There was no significant correlation between extent of exon skipping (i.e. PSI) and the expression level of the circRNAs.