Non-coding AUG circRNAs constitute an abundant and conserved subclass of circles.

Lotte VW Stagsted¹, Katrine M Nielsen¹, Iben Daugaard¹, and Thomas B Hansen^{1*}

1. Department of Molecular Biology and Genetics (MBG), and Interdisciplinary Nanoscience Center (iNANO), Aarhus University, Aarhus, Denmark

* To whom correspondence should be addressed:

T.B.H.: tbh@mbg.au.dk;

Running title: AUG circRNAs

Keywords: circRNA, circRNA translation, circRNA biogenesis, non-coding RNA, dhx9, ribosome profiling, riboseq

Abstract

Circular RNAs (circRNAs) are a subset of non-coding RNAs (ncRNAs) previously considered as products of missplicing. Now, circRNAs are considered functional molecules, although to date, only few functions have been experimentally validated, and therefore the vast majority of circRNAs are without known relevance. Here, based on RNA sequencing from the ENCODE consortium, we identify and characterize a subset of circRNAs, coined AUG circRNAs, encompassing the annotated translational start codon from the protein-coding host genes. AUG circRNAs are more abundantly expressed and conserved than other groups of circRNAs, and they display an *Alu*-independent mechanism of biogenesis. The AUG circRNAs contain part of bona fide ORF, and in the recent years, several studies have reported cases of circRNA translation. However, using thorough cross-species analysis, extensive ribosome profiling, proteomics analyses, and experimental data on a selected panel of AUG circRNAs, we observe no indications of translation of AUG circRNAs or any other circRNAs. Our data provide a comprehensive classification of circRNAs and, collectively, the data suggest that the AUG circRNAs constitute an abundant subclass of circRNAs produced independently of primate-specific *Alu* elements. Moreover, AUG circRNAs exhibit high cross-species conservation and are therefore likely to be functionally relevant.

Introduction

Non-coding RNAs (ncRNAs) constitute the vast majority of the human transcriptome as only a few percent of the produced transcripts are translated into proteins (ENCODE Project Consortium 2012). NcRNAs represent a highly heterogeneous group of molecules that besides including essential elements of protein synthesis, ribosomal RNA (rRNA) and transfer RNA (tRNA), also comprise small RNAs, such as microRNAs (miRNAs), which are involved in regulation of mRNA stability and protein synthesis (Bartel 2009), as well as long non-coding RNAs (lncRNAs) with many established functionalities (Böhmdorfer and Wierzbicki 2015; Noh et al. 2018). Recently, by means of high throughput tools, circular RNAs (circRNAs) were added to the rapidly expanding list of non-coding RNAs (Ebbesen et al. 2016). CircRNAs are typically derived from annotated protein-coding genes, but due to their relatively low abundance compared to their linear messenger RNA (mRNA) counterparts, circRNA molecules were first presumed to be missplicing events of the spliceosome with little to no relevance (Cocquerelle et al. 1993; Zaphiropoulos 1997). While this may be the case

for a substantial subset of circRNAs, the identification and functional characterization of the highly conserved circRNA and miR-7-sponge, CDR1as/ciRS-7 (Hansen et al. 2013b; Memczak et al. 2013), as well as extensive profiling of differentially expressed circRNAs from RNA sequencing analyses (Rybak-Wolf et al. 2015; Memczak et al. 2013; Salzman et al. 2012; Veno et al. 2015) strongly supports circRNAs as biologically relevant RNA species in eukaryotic cells. CircRNAs are generated by non-linear splicing (coined backsplicing) where an upstream splice acceptor (SA) is covalently joined to a downstream splice donor (SD) resulting in a circular structure (Hansen et al. 2011; Jeck et al. 2013). This results in a very high intracellular stability due to the lack of free ends, which protects them from normal exonucleolytic decay. CircRNAs are mostly comprised of exonic regions (most commonly 2-3 exons) derived from annotated protein-coding transcripts (Zhang et al. 2014). The current model of biogenesis suggests that backsplicing is stimulated by bringing the involved splice sites into close proximity (Ebbesen et al. 2016). This is conventionally facilitated by inverted *Alu* elements (IAE) (Jeck et al. 2013; Zhang et al. 2014), however trans-acting RNA-binding factors have also been implicated in circRNA formation (Ashwal-Fluss et al. 2014; Conn et al. 2015; Li et al. 2017).

With the exception of the exon-intron circRNAs (ElciRNA) (Li et al. 2015), circRNAs are exported to the cytoplasm (Jeck et al. 2013) with a recent study reporting the *Drosophila* Hel25E and its human homologs to regulate nuclear export of circRNAs in a length-dependent manner (Huang et al. 2018). In the cytoplasm, circRNAs have been shown to tether and ‘sponge’ miRNAs, initially exemplified by CDR1as/ciRS-7 harbouring >70 miR-7 binding sites (Hansen et al. 2013a; Memczak et al. 2013). Since then, several other examples have been published showing anti-miR effects of circRNA expression (Peng et al. 2016; Chaiteerakij et al. 2017; Zheng et al. 2016), although, bioinformatics analysis indicates that - apart from ciRS-7 - miRNA binding sites are generally not enriched in circRNA more than expected by chance (Guo et al. 2014). CircRNAs can also sequester RNA binding proteins and hereby modulate protein activity (Ashwal-Fluss et al. 2014). Additionally, synthetic circRNAs have been engineered to express protein by the use of internal ribosome entry sites (IRESs) allowing cap-independent translation (Wang and Wang 2015). Recently, it was shown that open reading frames (ORFs) within endogenously expressed circRNAs give rise to circRNA-specific peptides (Legnini et al. 2017; Pamudurti et al. 2017; Yang et al. 2017; Zhang et al. 2018a; Yang et al. 2018; Zhang et al. 2018b) suggesting that circRNAs are not necessarily exclusively non-coding.

In this study, publicly available RNA sequencing datasets from the ENCODE consortium are used to characterize the circRNA transcriptomes in 378 human and 75 murine samples, and the most abundant circRNAs in each dataset are identified, analyzed and stratified based on their genomic features. These analyses reveal that a substantial fraction of highly abundant circRNAs derives from exons encoding the translational start codon, here coined AUG circRNAs. In addition, the AUG circRNAs are more conserved than other groups of circRNAs and generally rely on an IAE-independent mode of biogenesis. Lastly, to determine the protein-codon ability of AUG circRNAs, we conduct extensive analyses of cross-species conservation and ribosome profiling (RiboSeq). This shows that ORF-associating features are not preserved in evolution and that backsplice-spanning reads found in RiboSeq datasets are not derived from translating ribosomes. Consistently, we fail to detect any peptides derived from AUG circRNAs using ectopic overexpression in cell-lines. Collectively, these results suggest that circRNAs are generally not subjected to translation and thus the functional relevance of the most conserved and abundant AUG circRNAs remains elusive.

Results

The ENCODE circRNA landscape

To obtain a comprehensive overview of circRNA expression across multiple tissues and cell lines, we took advantage of the total RNA sequencing datasets on human and mouse samples made available from the ENCODE consortium (see Supplementary Table 1). We conducted circRNA prediction and quantification using two established pipelines; find_circ (Memczak et al. 2013) and circexplorer2 (Zhang et al. 2016a). In total, find_circ and circexplorer2 identify 140,304 and 235,179 unique circRNAs using slightly modified settings (see methods), respectively, of which 81,589 are shared by both algorithms (**Fig. 1A**). The notable fraction of circRNAs only predicted by one algorithm – the so-called exotic circRNAs – is in general lowly expressed (**Fig. 1B**), which is also reflected by a small subset of exotic circRNAs in the top 1000 expressed circRNA candidates predicted by each algorithm (1-8%, data not shown). Consistently, we observe a high positive correlation between the algorithms for the abundant circRNA species (**Fig. 1C**, similar analyses for mouse samples are shown in **Supplementary Figs. 1A-C**). We have previously shown that exotic circRNAs are more likely to be

104 false positives (Hansen et al. 2015), and therefore we decided to focus only on the circRNAs jointly
105 predicted by both algorithms.

106 The used ENCODE data comprises 378 samples derived from 218 different human tissues and cell
107 lines (or 75 samples from 26 tissues in mouse, **Supplementary Table 1**). Plotting the expression of
108 circRNAs in each sample reveal a marked difference in circRNA expression between the samples
109 with particular abundant circRNA expression levels in tissues (**Supplementary Figs. 2A and C**),
110 whereas the similar analysis for mRNAs show similar expression in all biosample types
111 (**Supplementary Figs. 2B and D**). Moreover, even though the detected diversity of circRNAs is much
112 lower in mouse, the overall expression levels are comparable (**Supplementary Fig. 3A**). CircRNA
113 levels have previously been correlated with proliferation, i.e. circRNAs tend to accumulate in slow
114 or non-proliferative tissue (Bachmayr-Heyda et al. 2015). Thus, circRNA profiling from non-dividing
115 cells may dominate the average expression levels of circRNA across samples. Instead of comparing
116 expression across samples, we instead focused on the highest expressed circRNA in each sample
117 (the alpha circRNA). Here, the alpha circRNA in many samples exhibits disproportionally high
118 expression compared to the bulk of circRNAs. In fact, assuming a log-normal distribution of circRNA
119 expression, the alpha circRNAs are significant outliers in more than half of the samples (246 out of
120 378, $\text{fdr} < 0.05$, one-tailed Grubbs test), whereas only 4 of 378 samples show similar significant outlier
121 mRNAs. This tendency is also observed in mouse (**Supplementary Fig. 3B**).

122 Based on the ENCODE data, circHIPK3 is the most predominant alpha circRNA followed by the miR-
123 7 sponge, ciRS-7 (**Fig. 1D**). Even though most of the top 10 alpha circRNAs are found in the mouse
124 dataset, only circSLC8A1 and circCDYL are shared in the top 10 between mouse and human (**Fig. 1D**
125 and **Supplementary Fig. 4A**).

126 We then zoomed in on the top 10 alpha circRNAs, i.e. the ten circRNAs most often seen as the
127 highest expressed in a given sample, to determine the genomic features associating with these
128 highly abundant circRNA species (see **Fig. 1E**). Here, the human alpha circRNAs are flanked with very
129 distal inverted *A/u* elements (IAE), which is in stark contrast to the bulk of circRNAs (**Fig. 1F**) and the
130 prevalent model of biogenesis (Jeck et al. 2013; Zhang et al. 2014). Moreover, no significant
131 association between circRNA producing loci and other inverted repeat elements are observed for
132 the alpha circRNAs specifically or circRNAs in general compared to host gene exons (**Fig. 1F**). In

mouse, no repetitive elements are selectively demarcating circRNAs from host exons (Supplementary Fig. 4B), although a slight tendency towards proximal B1/alu SINE elements was detected. Instead, for both species, we observe a clear tendency for alpha circRNAs to have very long flanking introns (Fig. 1F and Supplementary Fig. 4B, ciRS-7, circRMST and circFAT3 are without annotated flanking intron and thus excluded in this analysis). Moreover, a positive correlation between intron length and IAE distance is detected (Fig. 1G and Supplementary Fig. 4C), indicating that circRNAs either utilize an *Alu*-dependent mechanism of biogenesis or require long flanking introns to favour back-splicing.

AUG circRNAs are highly expressed and conserved

The vast majority of circRNAs derive from annotated splice-sites (Zhang et al. 2016a), and we decided to stratify circRNAs by host gene annotation (see Fig. 2A). Here, circRNAs derived from exons containing annotated start codons, coined AUG circRNA, comprise 5 out of the top 10 alpha circRNAs in both human and mouse samples, whereas the percentage of AUG circRNA in general is 7-11% (Fig. 2B and Supplementary Fig. 5A). In fact, the AUG circRNAs in both human and mouse also show a significant over-representation in subsets of highly expressed circRNAs, which is not seen for other circRNA sub-classes (Fig. 2B and Supplementary Fig. 5A). Consistently, in both human and mouse AUG circRNAs are generally and significantly more abundant than other circRNAs (Supplementary Figs. 6A-B and D-E) both in terms of absolute expression but also regarding the circular-to-linear ratios, whereas the host genes, from which the AUG circRNAs are derived, are not significantly more abundant than for other circRNAs (Supplementary Figs. 6C and F).

Using liftover (UCSC) we evaluated the number of human circRNAs re-identified in the mouse dataset of circRNAs as a measure of conserved biogenesis. Similar to AUG circRNAs, the fraction of conserved circRNAs increase with expression (Fig. 2C). Focusing specifically on the top 1000 most abundant human circRNAs based on total backsplice junction (BSJ)-spanning reads, 39% of all AUG circRNAs are conserved comprising almost twice as many conserved species compared to the other circRNA sub-groups (Fig. 2D, $p=9.4e-4$, Fisher's exact test), despite the fact that the 5'UTR-embedded splice acceptors in human AUG circRNAs are often (21%) not annotated as splice sites in mouse. As with the top 10 alpha circRNAs, AUG circRNAs generally exhibit distal IAE and longer

flanking introns (**Fig. 2E-F**). In fact, AUG circRNAs, conserved circRNAs and host gene exons exhibit an overall similar distribution of IAEs (**Fig. 2E**). In contrast, the flanking intron lengths effectively demarcate AUG circRNA and conserved circRNAs from host gene exons (**Fig. 2F**), which is also supported by the analysis of AUG circRNAs in mouse (**Supplementary Fig. 5B**). In fact, using an atlas of circRNAs from *Drosophila* (Westholm et al. 2014), AUG circRNAs also increase in frequency with circRNAs expression and associate with significantly longer introns compared to other circRNAs (**Supplementary Figs. 5C-D**), whereas fruit fly circles in general show no evidence of flanking inverted elements (Westholm et al. 2014). Thus, based on the observations that AUG circRNAs are more conserved and overall devoid of flanking inverted *Alu* elements, and the fact that AUG circRNA features may extent all the way to invertebrates, we propose that AUG circRNAs are more likely to be biologically relevant and to utilize an *Alu*-independent biogenesis pathway.

To demarcate *Alu*-dependent from *Alu*-independent circRNAs, we empirically determined the distance to nearest IAE by which the cumulative fraction of circularizing exons differed the most from non-circularizing host exons. Here, approximately 45% of all human circRNAs has an inverted *Alu* element within 2300 nucleotides total distance, whereas this cut-off only applies to 22% of host exons (**Supplementary Fig. 7A**). We thus defined this 45% subset as the *Alu*-dependent circRNAs. Based on this demarcation, only 13% of *Alu*-dependent circRNAs are observed in mouse, which is consistent with the fact that *Alu* elements are primate specific (**Supplementary Fig. 7C**). In contrast, 41% of *Alu*-independent and 31% of circRNAs with long flanking introns (defined empirically as flanking intron > 6500 nts, **Supplementary Fig. 7B**) are conserved (**Supplementary Fig. 7B**), which suggests that at least for the evolutionary relevant circRNAs, biogenesis relies more on having long flanking introns instead of proximal inverted *Alu* repeats. However, the requirement for long flanking introns in circRNA biogenesis is currently unclear, and therefore the mechanism governing production of the majority of abundant and conserved circRNAs remains undisclosed.

Recently, the RNA resolvase, Dhx9, was shown to inhibit circRNA production by unwinding and destabilizing RNA structures formed by inverted *Alu* elements in flanking regions of circRNAs (Aktaş et al. 2017). Dhx9 is proposed to protect cells from adverse secondary structures in the nucleus. As a consequence, circRNAs sensitive to Dhx9 depletion are considered products of aberrant backsplicing mediated by random insertion of inverted repeat elements, and therefore these

circRNAs are more likely to be functionally irrelevant. Based on RNAseq from Dhx9-depleted HEK293 cells (**Supplementary Table 3**), it is possible to determine the subset of circRNAs sensitive to Dhx9 expression. We selected the top 1000 expressed circRNAs from this experiment (**Supplementary Table 4**). Here, roughly 25% (275 circRNAs out of 1000) responds significantly ($\text{fdr} < 0.05$) to the Dhx9 depletion (**Fig. 3A**), with a clear tendency towards proximal IAE and short flanking introns (**Figs. 3B-C**). Consistently, 39% of circRNAs designated as *Alu* dependent (IAE distance $< 2300\text{nt}$) are Dhx9-sensitive compared to only 12% of the non-*Alu* circRNAs, whereas long flanking introns are generally insensitive to Dhx9 compared to short introns (43% vs 23%, **Supplementary Fig. 8D**). Interestingly, in alignment with the analyses described above, AUG circRNAs are significantly reduced in the Dhx9-sensitive fraction (5% vs 15%, $p=4.8\text{e}7$, **Fig. 3D**). In fact, only 10% of the AUG circRNAs compared to 26% of non-AUG circRNAs is affected significantly in expression upon Dhx9 depletion (**Fig. 3E**).

CircHIPK3 has previously been characterized as an *Alu*-dependent circRNA (Zheng et al., 2016), however, it is also an AUG circRNA, as well as the overall highest expressed circRNA in the ENCODE data. As such, the most proximal IAEs are within 2300 nt (see **Fig. 3B**), and while these *Alu* elements could stimulate biogenesis, circHIPK3 carries more than 60 *Alu* elements in the immediate flanking introns and is insensitive to Dhx9 depletion (**Fig. 3A**). Dhx9-insensitive biogenesis is also observed for the two additional top 10 expressed AUG circRNAs, circSETD3 and circVRK1. Instead, the three AUG circRNAs in the top 10 fraction all associate with very long flanking introns (**Fig. 3B**), and notably, circZBTB44, which in this analysis is termed ‘ambiguous’ because it overlaps both an AUG and a non-coding transcript, share the same features. Consistently, HITS-CLIP analysis of Dhx9 occupancy shows a clear selection for binding in the immediate flanking regions of *Alu*-dependent compared to *Alu*-independent circRNAs (**Fig. 3F** and **Supplementary Fig. 8E**), but also a clear preference for non-AUG over AUG circRNAs (**Fig. 3F** and **Supplementary Fig. 8F**). Collectively, this strongly indicates that the AUG circRNAs are generally not affected by Dhx9 helicase activity and thus not depending on IAE for biogenesis.

No detectable protein production from ectopically expressed AUG circRNAs

Recent studies have shown that circRNAs despite lacking a 5'cap and 3'poly(A)-tail are still capable of recruiting ribosomes and act as templates for protein synthesis (Legnini et al. 2017; Yang et al.

2017; Pamudurti et al. 2017). This is most likely facilitated by IRES-like elements in the circRNA required for cap-independent translation. The AUG circRNAs all contain a 5' part of a bona fide ORF, and translation of this putative ORF will in most cases produce a truncated protein mimicking the N-terminal part of the host gene encoded protein. To test the hypothesis that AUG circRNAs are in fact protein-coding circRNAs, we initially focused on the two-exon AUG circRNAs derived from the LPAR1 gene (**Fig. 4A**). In the ENCODE data, circLPAR1 is the highest expressed circRNA in 21 samples, and it was the most abundant circRNA in one of the first global analyses of circRNA expression in a human fibroblast cell line, hs68 (Jeck et al. 2013). Here, circLPAR1 was shown to be 3-fold higher expressed than the second highest circRNA and resistant towards RNase R treatment. We constructed a minigene expression vector including the two exons of LPAR1 and a portion of the flanking introns (**Fig. 4B**). As with most other AUG circRNA, LPAR1 has no IAE in close proximity, and the mode of biogenesis is therefore currently unclear. To overcome this, we artificially inverted and inserted part of the upstream intron downstream of the splice donor (**Fig. 4B**), which results in clean and efficient expression of circLPAR1 exhibiting RNase R resistance (**Figs. 4C- D**). The protein-coding ability of circLPAR1 was determined by the insertion of an eGFP tag just upstream of the stop codon in the putative ORF (**Fig. 4B**). Here, detection of the tags would only be possible if translation proceeds across the BSJ. First, we tested whether the insertion of eGFP would impede or alter the circularization of circLPAR1. As expected, the circLPAR1-eGFP shows changed migration but remains RNase R resistant (**Fig. 4D**). Then, we over-expressed the untagged and GFP-tagged variants of circLPAR1 in HEK293 cells and performed fluorescent microscopy and western blot analyses, however, in both cases we were unable to obtain any signal (**Figs. 4E- F**). Based on AUG circRNA, circSLC8A1, and the ambiguous AUG circRNA, circCDYL, similar vector-designs were constructed (**Supplementary Figs. 9A-B and 10A-B**) and effective circRNA production was observed (**Supplementary Figs. 9C-E and 10C-E**), however, once again no GFP-positive signal by western blotting or fluorescent microscopy was obtained (**Supplementary Figs. 9F-G and 10F-G**). Collectively, this suggests that these specific circRNAs are not subjected to translation under normal conditions in HEK293 cells.

In a parallel experiment using the largeT antigen transformed HEK293T cells, which normally show higher expression of ectopic transgenes, we surprisingly observed a faint GFP-positive band on the western and few GFP positive cells when overexpressing circLPAR1-eGFP. However this was seen

both for LPAR1 vectors with or without artificial inverted elements and thus irrespective of circularization (**Supplementary Figs. 11A-E**). Moreover, vector-linearization prior to transfection reduced the GFP output, although this was also observed for the canonical GFP expression vector. Collectively, this indicates that in extreme conditions, exon-repeats are likely produced from ‘rolling circle’ read-through transcription on plasmid templates (see schematics in **Supplementary Fig. 11F**), and, presumably, this is particularly prevalent in circRNA expression vectors, as the vector-encoded poly(A) signal is situated downstream the SD and therefore subjected to U1-mediated repression (Kaida et al. 2010). Thus, vector-based overexpression may generate false positive protein products from capped mRNA indistinguishable from the predicted circRNA-derived peptides, and conclusions based on ectopic expression setups should be drawn with utmost caution.

Putative ORFs are not conserved features in AUG circRNAs

As shown above, AUG circRNAs are generally more conserved across species than other circRNAs (i.e. more often found in mouse as a circRNA). If the functional relevance of these AUG circRNAs is to encode protein, features specific for translation should also exert increased conservational restraint. We focused this analysis on the AUG circRNAs within the top 1000 expressed circRNAs from the ENCODE analysis, and used the AUG-containing exon from ‘other circRNA’-associated host genes (termed the ‘AUG exon’) to include exons with comparable expression level but without any evidence of circularization as a control.

In theory, the circular topology allows for infinite ORFs without stop-codons. However, this is only predicted for a very small subset of AUG circRNAs (8% in both human and mouse, **Supplementary Figs. 12A-B**). For the remaining circRNAs, the predicted ORF terminates shortly after the BSJ (median length of 10 amino acids (aa) after BSJ, **Supplementary Figs. 12C-D**), which is very close to the expected geometric distribution of stop-codon frequency considering the overall 5’UTR nucleotide composition (**Supplementary Figs. 12E-F**). This suggests that the predicted lengths of the circRNA derived peptides are very close to what would be expected by chance.

The mRNA ORFs are typically highly conserved between species. In contrast, the 5’UTRs generally exhibit much lower evolutionary constraints. We compared the overall conservation of 5’UTRs, but

only considering the AUG-containing exon (see schematics in **Fig. 5A**, upper panel). Here, AUG circRNAs show a significantly higher cross-species conservation compared to the control AUG exons (**Fig. 5A**). Next, to elucidate whether the increased conservation coincides with a putative ORF from the annotated AUG across the BSJ and into the 5'UTR, we determined the relative conservation of predicted stop-codons. As a positive control, we included the annotated stop-codon from the circRNA-derived host genes (termed 'Host mRNA') in the analysis. Based on phastCons scores obtained from the UCSC genome browser, the relative conservation of stop-codon vs downstream triplet was plotted (**Fig. 5B**). This shows, as expected, a notable conservational enrichment of host mRNA stop-codons, however no significant difference between AUG circRNAs and AUG exons is observed, and not even when focusing the analysis on the conserved subset of AUG circRNAs. In agreement with our analysis of ORF lengths, this suggests that the putative stop-codon sequence is not under evolutionary constraints. Finally, we determined the third-nucleotide (the wobble) conservation relative to the two other nucleotides in every codon within the putative ORF after the BSJ (**Fig. 5C**, schematic). Based on phyloP basewise conservation, the annotated mRNA ORFs show a clear and significant decrease in wobble nucleotide conservation, in accordance with previous analyses (Chamary et al. 2006). However, for the AUG circRNA, again, no differential conservation between the wobble position and the two other bases is observed, not even for the conserved subset of AUG circRNA (**Fig. 5C**), supporting the preliminary conclusion that the coding properties of AUG circRNAs are not conserved. Consistently, similar analyses on the murine repertoire of AUG circRNAs produce almost identical results (**Supplementary Figs. 13A-C**), suggesting little or no preservation of the circRNA-specific ORF, however, the exact peptide sequence encoded after the BSJ could be of less significance and therefore not under evolutionary pressure, but still the protein output could be functionally important.

AUG circRNAs are generally not templates for translation

To assess the translational potential of circRNA globally, we took advantage of the wide range of ribosome profiling (RiboSeq) data currently available online: ~500 and ~1300 samples from human and mouse origin (**Supplementary Table 5**). After adapter-trimming, we obtained a total of approximately 22 and 24 billion reads between 25 and 35 nucleotides in length from human and

307 mouse, respectively, consistent with an approximate ribosome footprint of ~30nts (Ingolia et al.
308 2009).

309 In general, the distance from the 5' end of footprinting reads to the ribosome P-site (P-site offset) is
310 12 nts (Ingolia et al. 2009; Bazzini et al. 2014), however, for shorter and longer reads this offset
311 varies (Dunn and Weissman 2016). Moreover, we noted that there was a large degree of p-site offset
312 variation between samples, and consequently, we initially analyzed each sample individually by
313 mapping all the 25-35 nt reads onto the *GAPDH* and *ACTB* (beta-actin) mRNAs. For each read-length,
314 we determined the amount of on-frame P-sites for 12, 13, and 14 nts offsets and the associated p-
315 value by binomial test (see example in **Supplementary Fig. 14**), which was used as a measure of
316 dataset fidelity. Thus, in all samples, the efficiency by which each read length (25-35 nts) is able to
317 demarcate translation from noise was determined (**Supplementary Fig. 15**). Here, for both species,
318 the 28-31 nt reads show the highest abundance and fidelity (**Supplementary Fig. 16**).

319 We then applied the reads to circRNAs using the 5' offset with the lowest p-value in the above
320 analysis. To evaluate translation of circRNA, the BSJ is the only circRNA-specific sequence. Therefore,
321 we concatenated the circRNA exons on all the top 1000 expressed circRNAs from the ENCODE data
322 to display the BSJ in a linear manner compatible with short read mapping. As above, we also included
323 the AUG-containing exon from 'Other circRNA'-derived host genes ('AUG exon'), as well as
324 'ambiguous AUG circRNAs', i.e. circRNAs derived from ambiguous host gene isoforms of which at
325 least one is annotated to contain the AUG start site (see schematics on **Fig. 6A**). By plotting the
326 distribution of reads across the BSJ, only a small fraction of reads spans the BSJ compared to the
327 immediate upstream regions (**Fig. 6B**). However, in contrast to previous reports (Guo et al., 2014;
328 You et al., 2015), there is a notable fraction of BSJ-spanning reads defined here as P-site position
329 from -8 to +6 relative to the BSJ (**Fig. 6B**) comprising a 15 nt stretch (5 codons). To ensure that BSJ-
330 spanning reads are in fact likely derivatives of circRNAs, the reads were re-aligned to an assembled
331 transcriptome allowing one mismatch. Particularly for the human RiboSeq data, this discards the
332 majority of reads spanning the BSJ on the 'AUG Exon' subset. In contrast, almost all reads mapping
333 perfectly to the BSJ of bona fide circRNA have no detectable mRNA alignment (**Supplementary Figs.**
334 **17A and B**). Now, when considering the likelihood of each read to actually derive from translating
335 ribosomes, only ~20% of BSJ-spanning reads are from high quality samples (with $\text{fdr} < 0.01$), whereas

~70% of the upstream-derived reads are of high quality (**Fig. 6C**). This suggests that across the BSJ, the quality of ribosome profiling data is of particular high relevance and that noise consumes most of the BSJ-spanning reads. This difference in quality is corroborated in mouse although here the difference in quality is less pronounced (**Supplementary Figs. 18A-C**). Nonetheless, to address translation across the BSJ, we filtered out the low quality reads ($\text{fdr} > 0.01$) and used the remaining reads to determine whether phasing in accordance with translation of the putative ORFs is evident. Here, we simply counted the number of reads in-frame and out-of-frame on the 5 codon BSJ-spanning stretch and compared this to a 5 codon stretch immediately upstream the BSJ. For all subtypes of circRNA ('AUG circRNA', 'Ambiguous AUG circRNA', and 'Other circRNA') a roughly equal distribution of reads between all three frames is observed across the BSJ, whereas for the upstream region, approximately 50-60% of reads are in-frame, both for humans and mouse (**Fig. 6D** and **Supplementary Fig. 18D**). This strongly suggests that the AUG circRNAs as a whole, or any of the other circRNAs for that matter, are not subjected to translation as evidenced by RiboSeq analysis. However, it is likely that a small and restricted subset of circRNAs is acting as templates for translation, and therefore the signal from these drown in the noise from others. To evaluate this, we analyzed all top 1000 circRNAs with at least 10 BSJ-spanning RiboSeq reads individually. Here, in humans only circUBXN7 – an ambiguous AUG circRNA for which the annotated host gene has multiple start codons – shows an enrichment of in-frame reads although not significant when evaluating unique reads only ($\text{fdr}(\text{unique reads}) = 0.057$, fisher's exact test, **Fig. 6E**, unique reads shown in parentheses). Similarly, in mouse, no significant phasing of unique reads is observed across the circRNA BSJ (**Supplementary Fig. 18E**).

No circRNA-specific peptides are found in proteomics data

To strengthen the conclusion drawn from the previous analyses, we conducted an extensive mining of proteomics data. Based on the top 1000 expressed circRNAs derived from protein-coding host genes, we predicted the resulting peptide from each circRNA, and compiled an associated database of decoys (**Fig. 7A**). Using the established mass-spec search tool, comet (Eng et al. 2012), we searched through 35 mio. spectral data (see **Supplementary Table 8**) derived from the nci60 cell lines, the human proteome draft map, as well as the human brain proteome. All hits and the

corresponding e-values were then assigned to either host gene, circRNA, or decoy. Here, in contrast to host genes, the e-values of circRNA-derived peptides are not significantly different from the decoy peptides (**Fig. 7B** and **Supplementary Fig. 19A**). In fact, using a relaxed *fdr* cutoff of 0.2, not a single circRNA-derived peptide is identified (**Fig. 7B**), and the AUC in a ROC analysis was for all circRNA subgroups close to 0.5 (**Supplementary Fig. 19B**). This adheres with the above conclusions that circRNAs are unlikely to be substrates for translation. Moreover, we focused specifically on circRNAs shown to produce protein in the literature and subjected these candidates to both RiboSeq and Mass-Spec analysis (summarized in **Supplementary Table 9**). Here, in all cases, no supportive Riboseq reads are found, and not a single circRNA specific peptide with e-values below 0.05 is observed. In fact, for many of the putative peptides derived from circRNAs in Yang et al, 2017, the associated ORFs have no canonical start codon, or the circRNAs are without any reads in the ENCODE RNAseq. This indicates to us that many of these identified peptides are likely false positives.

CircRNAs are not enriched with 8mer miRNA target sites

Finally, to determine whether AUG circRNAs show evidence of miRNA *sponging*, we characterized the frequency of miRNA target sites on circRNAs. Consistent with other analyses, we focused on the top 1000 expressed circRNA with splice sites derived from protein-coding genes (i.e. ciRS-7 is not included in this analysis), the AUG exons from 'Other circRNA'-producing host genes, and the unique 3'UTRs from the top 1000 expressed circRNA host genes. We conducted an 8mer target site analysis based on all the conserved and confident-assigned miRNAs from miRbase (n=338). Here, we observe for all target categories that the overall numbers of 8mers are enriched compared to expected based on permutations (**Supplementary Fig. 20A**). However, this is also observed for the reverse complement sequence, suggesting that mere nucleotide frequency is not a suitable null hypothesis for in silico target site prediction. Instead, considering the phastCons scores of all predicted target sites, we observe, as expected, that the 3'UTR targets are significantly more conserved compared to the reverse complement and the permuted sequences (**Supplementary Fig. 20B**). This, in fact, is also observed for AUG exon, i.e. non-circularized exons with annotated start codons, and to a lesser degree the AUG circRNAs. This could indicate that preserved miRNA targeting in 5'UTRs is occurring, but not in a circRNA specific manner. To elucidate any putative circRNA sponges, we

counted the frequency of all circRNA:miRNA pairs. Here, no circRNAs were observed with more than 3 target sites for any particular miRNA, in contrast to ciRS-7 with 59 8mer sites for miR-7. In one outlier instance, however, hsa-miR-342-5p has 23 distinct target sites in the 3'UTR of ANKRD11 (uc002fnf.1) (**Supplementary Fig. 20C**). As a last resort, based on the Smith-Waterman alignment score for each putative target site, we sought to disclose whether circRNAs were more prone to miRNA-mediated cleavage than expected by chance, however here no significant differences were observed between the actual target sequences and the controls (**Supplementary Fig. 20D**). Conclusively, this aligns with previous studies showing no enrichment of miRNA target sites on circRNAs – except for ciRS-7 (Guo et al. 2014), and suggests that circRNAs are unlikely to serve as miRNA sponges.

Discussion and conclusion

Here, by thorough disclosure of the circRNA landscape across human and murine tissues and cell lines, we have shown that a certain subclass of circRNAs, namely the AUG circRNAs, are abundantly expressed and conserved across species. The high abundance of AUG circRNAs is partly reflected in increased circular-to-linear ratio suggesting increased backsplicing efficiency. In fact, for the top 1000 expressed circRNAs, there is no significant difference in linear spliced reads between AUG circRNA and Other circRNAs. This indicates that the cross-species conservation is not merely a consequence of abundant host gene expression but may instead reflect sequence preservation of the AUG circRNAs specifically. Moreover, the AUG circRNAs associate with very long flanking introns and are devoid of flanking inverted *Alu* elements in contrast to most other circRNAs. The *Alu*-independent biogenesis is consolidated by analysis of RNAseq on Dhx9 depleted cells and Dhx9 HITS-CLIP, where AUG circRNAs generally are insensitive to Dhx9 perturbation and exhibit reduced Dhx9 binding in the flanking regions compared to other circRNAs. This strongly suggests that AUG circRNAs are not aberrant RNA species occasionally escaping the Dhx9-mediated surveillance. CircHIPK3, the most abundant circRNA in humans, has proximal IAE and is considered an *Alu*-dependent circRNA. Our analysis shows that circHIPK3 escape regulation by Dhx9 suggesting *Alu*-independent biogenesis. Moreover, circHipk3 is also highly expressed in mouse, but here devoid of any inverted *Alu*-elements in the flanking regions. Instead, the mouse and human loci share the long

flanking introns, suggesting that the biogenesis of circHIPK3 may not rely solely on inverted *Alu* elements in the flanking regions. Here, the combination of long flanking introns and proximal IAE could in part explain the high overall circHIPK3 expression levels in cells and tissue. Thus, we conclude that most AUG circRNAs use *Alu*-independent routes of biogenesis that most likely depend on the length and features of the flanking introns. In *Drosophila*, a well-established circRNA from the laccase locus is flanked by inverted repeats required for biogenesis (Kramer et al. 2015), however consistent with our results, circRNAs from *Drosophila* associate in general with very long introns and without any obvious flanking sequences engaging in intron-intron base-pairing (Westholm et al. 2014). Instead, splicing kinetics and backsplicing seem to be closely related features; It has been shown that increased polymerase speed results in decelerated linear splicing (Carrillo Oesterreich et al. 2016) but stimulates backsplicing (Zhang et al. 2016b). Similarly, depletion of splicing factors increase circRNA production by affecting splicing kinetics negatively (Liang et al. 2017). Therefore, we propose that some embodiment of slow splicing kinetics – which may be commonly associated with long introns - is a prerequisite for the production of many circRNAs. While in higher eukaryotes, the duration of splicing does not generally correlate with intron length (Hollander et al. 2016), the duration of intron transcription and the concomitant delay of splicing could in itself be of importance. Whether the intrinsic features of long intron splicing are sufficient for efficient backsplicing or whether certain trans-acting factors are required as well is currently unknown, and as such, the exact production of the most abundant and conserved circRNAs remains elusive.

Based on cross-species sequence analysis, the 5'UTR sequences contained within AUG circRNAs are overall more conserved than other non-circular 5'UTR elements. However, based on single nucleotide constraints, this conservation is not due to circRNA-specific ORFs. Consistently, upon overexpression of three different AUG circRNAs, no circRNA-specific peptides were detected. Additionally, thorough and extensive RiboSeq analysis suggests that BSJ-spanning reads are not derived from translating ribosomes assuming that the p-site offsets deduced from house-keeping genes (*GAPDH* and *ACTB*) also apply to circRNA translation. Finally, analysis of 35 mio. peptide spectra from mass spectrometry fails to identify a single circRNA-derived peptide with a *fdr* below 0.2. Collectively, this argues that circRNAs in general and AUG circRNAs specifically are not subjected to translation. It should be emphasized that our analyses solely focus on human and mouse samples

and therefore the functional features of circRNA in other organisms, such as *Drosophila*, has not been addressed here. Moreover, our results do not necessarily exclude inefficient translation of circRNAs or restricted translation solely under specific conditions. Consequently, it is possible that published circRNAs, such as circMbl in *Drosophila* (Pamudurti et al. 2017), and circZNF609, circSHPRH, circFBXW7, and circLINC-PINT in human cells (Legnini et al. 2017; Zhang et al. 2018a; Yang et al. 2018; Zhang et al. 2018b), indeed engage in protein production, however, our data strongly point towards translation of circRNAs as being a rare and uncommon process.

Since the functional characterization of ciRS-7 as a dedicated miR-7 sponge or regulator, many examples of circRNAs acting as miRNA sponges have been proposed. Apart from the 70+ selectively conserved miR-7 sites on ciRS-7, the miRNA sponge potential seems not to be a conserved feature of circRNAs (Guo et al. 2014). Moreover, stoichiometric analysis of circRNA:miRNA:mRNA ratios suggests little or no overall effect upon circRNA-mediated miRNA inactivation (Denzler et al. 2014; Jens and Rajewsky 2015), highlighting that the notion of circRNAs as miRNA sponges is very controversial. This is consistent with our analysis of putative 8mer target sites. Here, the circRNAs showed no overall enrichment of target site frequency, target site conservation, target site alignment scores, nor any emerging circRNA:miRNA pairs, suggesting that from a bioinformatics point of view, circRNAs do not classify as miRNA regulators.

Instead, specifically for the AUG circRNAs, expressing translationally inert canonical start codons in its natural sequence context could be useful in certain scenarios, e.g. as binding platforms for or regulators of translation factors in the cytoplasm, and using the circular topology for this purpose seems plausible, although the same stoichiometric issues as for the miRNA sponge hypothesis may apply here. It is possible that the functional relevance of one particular AUG circRNA is very subtle, but that the accumulated contribution of all the circRNAs are of physiological importance. As such, the highly stable and durable circRNAs could constitute a background of non-responsive RNA entities in the cell to ensure robustness by transiently associating with RNA-binding proteins in the cytoplasm thereby reducing non-specific and potentially detrimental RNA-protein interactions. In any case, future research will undoubtedly shed light on the elusive mechanism by which these highly abundant circRNA species are produced and more interestingly elucidate the functional capabilities of AUG circRNAs.

482

483 **Materials and Methods**

484 *Plasmids*

485 All plasmids were generated by PCR with subsequent restriction digest and ligation into pcDNA3
486 (Invitrogen). Primers are listed in **Supplementary Table 6**.

487 *Cell lines and transfection*

488 HEK293 *Flp-In™ T-Rex™* cells (Invitrogen) or HEK293T cells (ATCC) were used for all experiments.
489 The cells were cultured in Dulbecco's modified Eagle's media (DMEM) with GlutaMAX (Thermo
490 Fischer Scientific) supplemented with 10 % foetal bovine serum (FBS) and 1 %
491 penicillin/streptomycin sulphate. The cells were kept at 37 °C and 5 % CO₂. Transient transfections
492 were carried out using calcium phosphate as transfection reagent using standard procedures or
493 Lipofectamine Reagent 2000 (Invitrogen) accordingly to manufacturers protocol. After 24 hours, the
494 media was changed and 48 hours post transfection cells were harvested either by resuspension in
495 a) 2xSDS loading buffer (for western blotting) or b) TRIzol Reagent (Thermo Fisher Scientific) (for
496 RNA purification) adhering to manufacturer's protocol (see below). For linearization of plasmids
497 prior to transfections, the respective plasmids were digested with FastDigest PvuI (Thermo Fisher
498 Scientific).

499 *Northern blotting*

500 Northern blotting was performed as described in (Hansen 2018a). Briefly, 10 µl RNA (1g/l) and 20
501 µL northern loading buffer [58,8% formamide, 6,5% formaldehyde, ethidium bromide, 1,18% MOPS,
502 bromophenol blue] were mixed and denatured at 65°C for 5 minutes. The RNA was separated by
503 electrophoresis on a 1,2% agarose gel containing 3% formaldehyde and 1x MOPS at 75 V. After
504 electrophoresis, the gel was briefly washed in water and exposed to UV to visualize the EtBr stained
505 rRNA bands. The gel was transferred to a Hybond N+ membrane (GE Healthcare) over-night (O/N)
506 in 10xSSC. Then, the membrane was UV cross-linked and pre-hybridized in Church buffer [0,158 M
507 NaH₂PO₄, 0,342 M Na₂HPO₄, 7 % SDS, 1 mM EDTA, 0.5 % BSA, pH 7.5] for one hour at 55°C and
508 subsequently probed with a 5' radioactively labelled DNA oligonucleotide (see **Supplementary**

Table 6) at 55 °C O/N. The next day, the membrane was washed twice in 2 x SSC, 0.1 % SDS for 5 minutes and twice in 0,2xSSC, 0,1% SDS for 15 minutes. All washes were carried out at 50°C. Finally, the membrane was exposed on a phosphoimager screen and analyzed using Quantity One[®] or Image Lab[™] software (Bio Rad). Membranes were stripped in boiling stripping buffer [0,1% SDS, 1 mM EDTA].

RNase R

For RNase R experiments, 4 µg RNA was digested with 4 U RNase R (Epicentre) in a total reaction volume of 10 µL for 10 minutes at 37 °C. Then, 20 µL northern loading buffer was added, heated at 65°C for 5 minutes before loading on an agarose northern gel, see above.

Western blotting

Cells were harvested in 1xPBS and centrifuged at 1200 rpm at 4°C for 5 min. The supernatant was removed and cell pellet was lysed directly and resuspended in 2xSDS loading buffer [125 mM Tris-HCl, pH 6.8, 20 % glycerol, 5 % SDS, 0,2 M DTT]. After resuspension, the samples were boiled at 95°C for 5 minutes before loading 1 % on a 10 % Tris-Glycine SDS-PAGE gel with 10 µL PageRuler Plus Prestained Protein Ladder (Thermo Scientific). After approximately 1 ½ hours electrophoresis proteins were immobilized on an Immobilon-P Transfer Membrane (EMD Milipore) O/N in a wet-blotting chamber. The next day, the membrane was pre-incubated at RT with 20 % skim milk to block unspecific binding for 1 hr. Then, primary antibody (**Supplementary Table 6**) in blocking solution was added and incubated 1 hour at RT, followed by 1 hour incubation with secondary antibody in blocking solution. After each antibody incubation, the membrane was washed 3x5 min in 1xPBS with 0,05 % Tween 20 and subsequently with 1x5 min wash with 1xPBS. Exposure was performed using SuperSignal West Femto Maximum Sensitivity Substrate kit (Thermo Scientific) and Amersham Hyperfilm ECL (GE Healthcare).

Microscopy

48 hours after transfection, cells were imaged live using phase contrast and fluorescence microscopy with normal FITC filter set (using 1-second exposure, ISO200) on an Olympus IX73 microscope. Images were merged using ImageJ.

536 *RNAseq datasets and CircRNA detection*

537 Raw RNA sequencing data was downloaded from the ENCODE Consortium
 538 (www.encodeproject.org) or the NCBI Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/) (See
 539 **Supplementary Tables 1 and 3**). CircRNA prediction was performed by find_circ, version 1.0
 540 (Memczak et al. 2013) and circexplorer2.0.1 (Zhang et al. 2016a) adhering to the recommendation
 541 by the authors. For find_circ, an increased stringency threshold was used requiring that both
 542 adaptor sequences map with highest possible mapping quality (mapq=40) (Hansen 2018b).
 543 Moreover, for find_circ and CIRCexplorer2, only circRNAs supported by at least two reads in a given
 544 sample was kept. CircRNAs found by both algorithms with the above-mentioned stringency were
 545 used in subsequent analyses. CircRNA expression was based on BSJ-spanning reads according to
 546 find_circ quantification. Likewise, the circular-to-linear ratio was determined by the total number
 547 of BSJ-spanning reads multiplied by two and divided by the total number of linear spliced reads
 548 spanning the upstream and downstream splice sites, respectively, as determined by find_circ. RPM
 549 values were calculated for each sample as the number of BSJ-spanning reads divided by the total
 550 number of reads. mRNA expression (FPKM values) was quantified using cufflinks (Trapnell et al.
 551 2010).

552 *circRNA annotation*

553 Annotation of circRNAs was based on UCSC Genes tracks (hg19 and mm10). First, the annotation
 554 database was queried for host genes sharing both the circRNA-specific splice sites. If none were
 555 found, genes sharing at least one splice-site were queried, and as a last resort, genes fully covering
 556 the circRNA locus were retrieved. Similarly, host gene exons were retrieved from annotated
 557 isoforms sharing both circRNA producing splice-sites but omitting first and last exons as well as the
 558 exons with splice-sites coinciding with the circRNA, and duplicate exons were discarded. The
 559 circRNA subclass, i.e. 'AUG circRNA', 'CDS circRNA' etc, was only determined based on host gene
 560 ORF annotation, and if multiple host gene entries were recovered with divergent annotation, the
 561 circRNA was categorized as 'Ambiguous'. To guide detection of circRNAs by circExplorer2 and to
 562 facilitate proper annotation of known circRNAs, two additional entries were manually added to the
 563 gene annotation database (shown in **Supplementary Table 7**). Flanking intron lengths were based
 564 on the host gene exon-intron structure immediately upstream and downstream the back-splicing

splice sites. In case of multiple isoforms with varying intron length, the mean of all flanking introns found was calculated, whereas circRNAs with no flanking introns annotated were discarded. To extract the distance to nearest flanking inverted *Alu* element, the UCSC Repeatmasker tracks (hg19 and mm10) were used. Here, the 20 most proximal but flanking *Alu* elements were retrieved irrespective of intron-exon structure on either side of the circRNA, and based on these, the closest possible inverted pair was determined.

Conservation of circRNA

Human circRNAs were converted to mouse (mm10) coordinates using the UCSC liftover tool. Based on the predicted circRNAs from ENCODE mouse data, and on the UCSC gene annotation database, the coordinates coinciding perfectly with mouse circRNA or annotated splice sites were assessed. Converted coordinates not found as splice sites as well as unmapped coordinated were grouped together as ‘non-conserved’.

HITS CLIP analysis

For HITS-CLIP analyses, reads were adaptor-trimmed using trim_galore and barcodes were subsequent removed. The reads were pair-wised mapped on to the human genome (hg19) using bowtie2 using default settings, and mapped reads were extracted using samtools. Reads mapping in the flanking introns within 1000bp of the back-splicing splice-sites of the top 1000 expressed human circRNAs were counted and compared.

Conservation of Open reading frame (ORF)

Conservation analysis was performed for all top 1000 circRNAs with both splice sites annotated in a protein-coding host gene. For ‘AUG circRNAs’ and ‘AUG exons’, the ORF was predicted based on the annotated start codon whereas for ‘Other circRNAs’, the longest finite ORF traversing the BSJ was used. Infinite ORFs were not considered in this analysis. In case of multiple isoforms, i.e. alternative exons within the circRNAs, all isoforms were considered equally in the analysis. The 5’UTR conservation and stop-codon conservation were based on the phastCons scores from 100 species (UCSC) using the hg19 or mm10 reference for human and mouse, respectively. For 5’UTRs, only the AUG-containing exon was analyzed, and for stop-codons, the stop-codon triplet and the immediate downstream triplet irrespective of position was evaluated. For wobble position analysis, only the

circRNA-specific ORF within the 5'UTR was analyzed. Here, for each position (1st, 2nd and wobble) in each codon, the PhyloP scores from 100 species (UCSC) was retrieved and analyzed. The ORF lengths were based on number of codons from BSJ to stop, and the geometric distribution of stop codon probability considering the frequency of the individual nucleotides in the 5'UTR region was used to determine the expected ORF lengths.

RiboSeq analysis

Ribosome profiling (RiboSeq) datasets (see **Supplementary Table 5**) were trimmed using trim_galore and only reads between 25 and 35 nucleotides in length were kept and mapped onto *GAPDH* and *ACTB* mRNA (UCSC accessions; uc001qop.2 and uc003sot.4, or uc009dts.2 and uc009ajk.2, for human and mouse, respectively). For each read-length in each RiboSeq sample, an offset of 12, 13 or 14 nucleotides was tested to determine the best possible offsetting based on one-tailed binomial tests, e.g. how many codons exhibit more on-frame than off-frame reads (see **Supplementary Figs. 14-16**). The offset with the lowest mean p-value (obtained from *GAPDH* and *ACTB*) was used for each given read-length in a given sample. The mature sequence of all top 1000 circRNAs with both splice sites annotated in a protein-coding host gene were then concatenated to allow mapping across the BSJ. All RiboSeq reads were then bowtie-mapped to the concatenated circRNA sequences allowing no mismatches using the following arguments: *bowtie -S -a -v 0*. The offset from the quality assessment on *GAPDH* and *ACTB* was used to obtain p-site position. P-sites within [-8;6] relative to the BSJ was defined as BSJ-spanning whereas P-sites immediately upstream the SD [-31;-17] was defined as linear upstream reads. Putative BSJ-spanning reads were mapped against an mRNA reference (build on UCSC annotations) with one mismatch tolerance: *bowtie -f -v 1*, and omitted from downstream analysis if mapped. The annotated AUG was used to predict the circRNA-specific ORF, however in 'Other circRNAs', the longest possible ORF traversing the BSJ was used. Again, infinite ORFs were not considered. P-site positions relative to the predicted frames were counted and analyzed. The statistical significance of the proportion of in-frame reads was determined by Fisher's exact test.

Mass-spec analysis

First, a reference proteome database was compiled. This contained all well-known contaminants, the uniprot protein collection (June 27, 2018, n=73099), and all the putative proteins produced from

the circRNAs subjected to RiboSeq analysis. Then, decoys were generated from all the circRNA-derived proteins using the decoyPyRat python script (<https://www.sanger.ac.uk/science/tools/decoypyrat>) and concatenated to the proteome reference database. Raw MS files (see **Supplementary table 8**) were converted to mgf format using MSconvert (Chambers et al. 2012). Then, the spectra were searched against the reference database using the comet search engine (Eng et al. 2012) adhering to the suggested fixed and variable modifications for each sample. Only the peptides derived from circRNA-based ORFs and the decoys were analysed, and the peptides were stratified by overlap with the Uniprot assembly indicating host gene or circRNA origin. Scores and e-values were extracted and analysed. Fdr values were calculated as the cumulative fraction of decoys ordered by e-value for circRNAs and host genes, respectively.

circRNA sponge analysis

To determine miRNA target sites on circRNAs, all ‘confident’ human mature miRNA were retrieved from miRBase version 21. Only miRNAs with phastCons scores > 0.95 were kept (n=338). Then 8mer seed-matches (using miRNA positions 2-8 and a downstream A) were searched in the sequence from the top 1000 expressed human circRNAs with both splice-sites annotated in protein-coding host genes. In addition, all unique 3’UTRs from circRNA producing host genes were included. For each analyzed target sequence, the reverse complement sequence as well as 20 random permutations were included in the analysis. The number of putative targets identified were normalized to the expected number of targets based on the nucleotide compositions of target sequence and miRNA seedmatch. Conservation of target sites were based on the phastCons score of the seed-match. Smith-Waterman alignments (using 2, -1, -1 scores for match, mismatch and gap, respectively) were generated between the full miRNA sequence excluding the first nucleotide, and the 8mer target site including an additional 18 nucleotides upstream. Finally, for each target species, all circRNA:miRNA pairs were grouped by counts and summarized.

Statistical analyses

650 All statistical analyses are based on Wilcoxon Rank-sum tests except if explicitly noted otherwise.
651 Fdr-values reflect Benjamini-Hochberg-adjusted p-values, except for the mass-spec analysis (see
652 above).

653

654 Figure Legends

655 **Figure 1: Abundant circRNAs.** **A)** Venndiagram showing the number of exotic (orange) and shared
656 (blue) circRNAs found by find_circ and circexplorer2 algorithms in the ENCODE datasets. **B)**
657 Smoothed fraction of shared circRNAs found by find_circ and circexplorer2 as a function of ranked
658 expression. **C)** Scatterplot depicting the number of backsplice junction (BSJ)-spanning reads
659 obtained from find_circ and circexplorer2 across all the samples analyzed. The points are color-
660 coded as shared (blue) or exotic (orange). **D)** The alpha frequency of the top 10 most commonly
661 found alpha circRNAs as well as the frequency of being an abundant circRNA (i.e. one of the top 10
662 expressed circRNAs in a sample) are plotted as a stacked barplot. **E)** Schematic illustration of the
663 flanking intron length and inverted *Alu* element (IAE). **F)** Boxplot comparing the distance to inverted
664 repeat element and flanking intron length for circRNAs in general (n=81589), host gene exons
665 (n=131002) and the top 10 alpha circRNAs. *, $p < 0.05$; **, $p < 0.01$, Wilcoxon rank-sum test. **G)**
666 Density-colored scatterplot showing relationship between IAE and flanking intron length for all
667 circRNAs (left) and host gene exons (right). The top 10 alpha circRNAs are highlighted to the left.

668 **Figure 2: The AUG circRNAs.** **A)** Schematics showing circRNA annotation. The genic features are not
669 drawn to scale. **B)** Frequency of circRNA- annotations for either all circRNA, the top 1000, top 100,
670 top 10 circRNAs based on overall expression (RPM), and the top 10 alpha circRNAs, color-coded as
671 denoted. P-values are calculated using Fisher's exact test. **C)** Smoothed relationship between
672 circRNA expression and frequency of AUG containing circRNA and conservation to mouse, i.e. found
673 as circRNA in mouse ENCODE RNAseq. **D)** Based on the top 1000 expressed circRNAs, the fraction
674 of circRNA coordinates found as circRNA in the mouse ENCODE data (Conserved circRNA) stratified
675 by annotation is shown. Moreover, the fraction of conserved splice sites, i.e corresponding to
676 annotated splice sites in mouse, conserved splice donor (SD), conserved splice acceptor (SA), and
677 sites without annotation are depicted. P-value is calculated using Fisher's exact test using conserved

678 circRNAs from AUG and non-AUG stratification. **E-F**) Inverted *Alu* element (IAE) distance (E) and
679 flanking intron length (F) for circRNAs stratified by annotation or by conservation, as well as host
680 gene exons. For flanking intron length (F), only circRNAs and exons with annotated up- and
681 downstream introns were included in the analysis. P-values are based on Wilcoxon rank-sum tests,
682 and where 'AUG circRNAs' are compared to 'Other circRNAs' and 'Ambiguous circRNAs', only the
683 highest obtained p-value is denoted.

684 **Figure 3: AUG circRNAs are DHX9-resistant.** **A)** Volcanoplot on top 1000 expressed circRNAs from
685 Aktas *et al*, 2017, showing circRNA deregulation upon DHX9 knockdown color-coded by fold-change
686 significance ($\text{fdr} < 0.05$). The top 10 expressed circRNAs are highlighted and color-coded by
687 annotation. **B)** Scatterplot showing flanking intron length by flanking *Alu* distance. Here, as in (A)
688 the top 10 expressed circRNAs are highlighted. **C)** Boxplot showing the distribution of IAE distance
689 and flanking intron length for circRNAs stratified by Dhx9 sensitivity. **D)** Frequency of circRNA-
690 annotations for circRNAs resistant or sensitive towards Dhx9 knockdown. P-value is calculated by
691 Fisher's exact test using AUG and non-AUG stratification. **E)** Within the top 1000 expressed circRNA,
692 the fraction of Dhx9-sensitive species are grouped by annotation and plotted. **F)** Cumulative plot
693 showing the number of Dhx9 HITS-CLIP reads in the flanking vicinity (within 1kb upstream and
694 downstream of the SA and SD, respectively) of circRNAs. Here, the circRNAs were stratified by either
695 genic annotation or Dhx9 sensitivity. P-values are calculated by Wilcoxon rank-sum test, and refer
696 to the 'AUG circRNA' vs 'Other circRNA' and 'DHX9 sensitive' vs 'DHX9 resistant' subgroups.

697 **Figure 4: No evidence of circLPAR1 translation.** **A)** Genomic representation of the *LPAR1* host gene
698 locus. The exons are not drawn to scale. **B)** Schematic representation of expression vectors
699 comprising the CMV promoter, exons 2 and 3 known to circularize, the putative circRNA-specific
700 stop-codon (cSTOP), the insertion of eGFP ORF, the flanking regions (divergent arrows indicate
701 artificially introduced inverted element) and the BGH pA signal. **C-D)** Northern blot analysis of total
702 RNA from ectopic overexpression of circLPAR1 vectors as denoted in HEK293 cells (C) or RNA with
703 or without RNase R treatment (D). The membranes were probed for circLPAR1 as denoted to the
704 right (top panels), and 18S (obtained by ethidium bromide stain) serves as loading and RNase R
705 control (bottom panels). **E)** Western blot showing GFP expression in HEK293 cells transfected with

706 positive control (pcDNA-eGFP) or circLPAR1-eGFP fusion. **F)** Merged phase contrast and GFP
707 fluorescence images (PC/GFP) obtained from HEK293 cells transfected with vectors as denoted.

708 **Figure 5: No evolutionary preservation of Open-Reading-Frames in AUG circRNAs.** **A)** PhastCons
709 analysis of 5'UTRs within the AUG-containing exon performed on AUG-containing exons in
710 conserved and non-conserved AUG circRNAs, as well as AUG-containing exons from non-circular
711 AUG exons. The 5'UTRs (transparent red and transparent grey) and representative phastCons tracks
712 are depicted for 'AUG circRNAs' and 'AUG exons', respectively, in the above schematics. **B)**
713 PhastCons scores of putative circRNA-derived stop-codon positions were extracted and compared
714 to the PhastCons scores of the immediate downstream triplet (+3) to obtain a measure of selective
715 conservation (Relative Stop-codon conservation). The analysis was stratified by annotation as in (A)
716 and visualized as a cumulative fraction plot. The stop-codon and +3 triplet is exemplified by
717 circSLC8A1 to the right. Similar analysis on bona-fide stop-codons within host gene ORFs is included
718 (black line). **C)** PhyloP analysis of single-position conservation for 1st, 2nd and wobble-position for
719 bona-fide ORFs within host genes and putative ORFs after BSJ as exemplified to the right by
720 circSLC8A1. N denotes number of codons analyzed.

721 **Figure 6: Ribosome profiling reads across BSJ.** **A)** Schematics showing circRNA annotation of 'AUG
722 circRNA', 'Ambiguous circRNA', and 'Other circRNA'. **B)** Based on Ribosome profiling datasets, the
723 number of ribosome P-sites around the backsplice junction (BSJ) were counted for each subclass of
724 circRNA ('AUG circRNA', 'Ambiguous AUG circRNAs', and 'Other circRNA', see text for more detail).
725 The AUG-containing exon from non-AUG ('AUG exon') circRNA host genes. The plot is color-scaled
726 according to the associated read-class p-value (See **Supplementary Figs. 14 and 15**). The grey box
727 denotes the defined P-site position of BSJ-spanning reads, from pos -8 to +6 relative to the BSJ. **C)**
728 Based on all BSJ-spanning reads (P-sites from -8 to +6 relative to BSJ) and upstream reads (-31 to -
729 17 relative to BSJ), the p-value distribution based on RiboSeq quality assessment is shown. **D)**
730 Phasing of reads across BSJ. Here, based solely on high quality reads (p<0.01), the fraction of P-sites
731 in-frame and out-of-frame across the BSJ (-8 to +6) are shown for each subclass of circRNA. The
732 number below the plot represent total number of reads analyzed, whereas the numbers inside the
733 plot reflect total or unique (in parenthesis) counts within each frame. **E)** As in (D), but for each

individual circRNA with 10+ reads across the BSJ. P-values in (D) and (E) represent one-tailed exact binomial tests.

Figure 7: circRNA-derived peptides. **A)** Schematics outlining the proteomics analysis pipeline. For the top 1000 expressed circRNAs, the putative ORFs were deduced. From this, a reference proteome was established comprising the entire circRNA-derived peptides as well as a matched list of decoys. Then, proteomics data was queried against the reference database using comet and hits were stratified by ‘circRNA’, ‘Host gene’ and ‘Decoy’. **B)** Boxplot showing the distribution of obtained e-values from ‘circRNA’, ‘Host gene’ or ‘Decoy’ subclassified into circRNA annotation as shown. **C)** Cumulative sum of peptide hits with incrementing false discovery rate (FDR) cutoffs. FDRs are calculated as the cumulative fraction of decoys on e-value-sorted data with circRNAs and Host genes analyzed separately.

Acknowledgements

We would like to thank Karoline Krogh Ebbesen for the insightful discussions and critical reading of the manuscript. This work was supported by the Novo Nordisk Foundation (NNF16OC0019874 to T.B.H)

References

- Aktaş T, Avşar İlik İ, Maticzka D, Bhardwaj V, Pessoa Rodrigues C, Mittler G, Manke T, Backofen R, Akhtar A. 2017. DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. *Nature* **544**: 115–119.
<http://www.nature.com/doi/10.1038/nature21715> (Accessed September 5, 2017).
- Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, Evtal N, Memczak S, Rajewsky N, Kadener S. 2014. circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell* **56**: 55–66.

760 Bachmayr-Heyda A, Reiner AT, Auer K, Sukhbaatar N, Aust S, Bachleitner-Hofmann T, Mesteri I,
761 Grunt TW, Zeillinger R, Pils D. 2015. Correlation of circular RNA abundance with proliferation-
762 -exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human
763 tissues. *Sci Rep* **5**: 8057. <http://www.ncbi.nlm.nih.gov/pubmed/25624062>.

764 Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* **136**: 215–233.

765 Bazzini AA, Johnstone TG, Christiano R, MacKowiak SD, Obermayer B, Fleming ES, Vejnar CE, Lee
766 MT, Rajewsky N, Walther TC, et al. 2014. Identification of small ORFs in vertebrates using
767 ribosome footprinting and evolutionary conservation. *EMBO J* **33**: 981–993.

768 Böhmdorfer G, Wierzbicki AT. 2015. Control of Chromatin Structure by Long Noncoding RNA.
769 *Trends Cell Biol* **25**: 623–32. <http://www.ncbi.nlm.nih.gov/pubmed/26410408> (Accessed
770 September 1, 2017).

771 Carrillo Oesterreich F, Herzel L, Straube K, Hujer K, Howard J, Neugebauer KM. 2016. Splicing of
772 Nascent RNA Coincides with Intron Exit from RNA Polymerase II. *Cell*.

773 Chaiteerakij R, Zhang X, Addissie BD, Essa A, Harmsen WS, Theobald PJ, Peters BE, Joseph G, Ward
774 MM, Giama NH, et al. 2017. Circular RNA MTO1 acts as the sponge of miR-9 to suppress
775 hepatocellular carcinoma progression. *Hepatology* **2**–43.

776 Chamary J V., Parmley JL, Hurst LD. 2006. Hearing silence: Non-neutral evolution at synonymous
777 sites in mammals. *Nat Rev Genet* **7**: 98–108.

778 Chambers MC, MacLean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt
779 B, Egertson J, et al. 2012. A cross-platform toolkit for mass spectrometry and proteomics. *Nat*
780 *Biotechnol*.

781 Cocquerelle C, Mascrez B, Hetuin D, Bailleul B, Héтуin D, Bailleul B. 1993. Mis-splicing yields
782 circular RNA molecules. *FASEB J* **7**: 155–160.

783 Conn SJ, Pillman KA, Toubia J, Conn VM, Salmanidis M, Phillips CA, Roslan S, Schreiber AW,
784 Gregory PA, Goodall GJ. 2015. The RNA binding protein quaking regulates formation of
785 circRNAs. *Cell* **160**: 1125–1134.

786 Denzler R, Agarwal V, Stefano J, Bartel DP, Stoffel M. 2014. Assessing the ceRNA Hypothesis with
787 Quantitative Measurements of miRNA and Target Abundance. *Mol Cell* **54**: 766–776.

788 Dunn JG, Weissman JS. 2016. Plastid: Nucleotide-resolution analysis of next-generation
789 sequencing and genomics data. *BMC Genomics* **17**.

790 Ebbesen KK, Kjems J, Hansen TB. 2016. Circular RNAs: Identification, biogenesis and function.
791 *Biochim Biophys Acta* **1859**: 163–168.

792 ENCODE Project Consortium TEP. 2012. An integrated encyclopedia of DNA elements in the human
793 genome. *Nature* **489**: 57–74. <http://www.ncbi.nlm.nih.gov/pubmed/22955616> (Accessed
794 September 1, 2017).

795 Eng JK, Jahan TA, Hoopmann MR. 2012. Comet: an open source tandem mass spectrometry
796 sequence database search tool. *Proteomics*.

797 Guo JU, Agarwal V, Guo H, Bartel DP. 2014. Expanded identification and characterization of
798 mammalian circular RNAs. *Genome Biol* **15**: 409–014–0409–z.

799 Hansen TB. 2018a. Characterization of circular RNA concatemers. In *Methods in Molecular Biology*,
800 Vol. 1724 of, pp. 143–157 http://link.springer.com/10.1007/978-1-4939-7562-4_12 (Accessed
801 January 26, 2018).

802 Hansen TB. 2018b. Improved circRNA Identification by Combining Prediction Algorithms. *Front Cell*
803 *Dev Biol* **6**. <http://journal.frontiersin.org/article/10.3389/fcell.2018.00020/full>.

804 Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. 2013a. Natural
805 RNA circles function as efficient microRNA sponges. *Nature* **495**: 384–388.
806 <http://www.ncbi.nlm.nih.gov/pubmed/23446346>.

807 Hansen TB, Kjems J, Damgaard CK. 2013b. Circular RNA and miR-7 in Cancer. *Cancer Res* **73**: 5609–
808 5612.

809 Hansen TB, Venø MT, Damgaard CK, Kjems J, Venø MT, Damgaard CK, Kjems J. 2015. Comparison
810 of circular RNA prediction tools. *Nucleic Acids Res* **gkv1458**.

- 811 Hansen TB, Wiklund ED, Bramsen JB, Villadsen SB, Statham AL, Clark SJ, Kjems J. 2011. miRNA-
812 dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA.
813 *EMBO J* **30**: 4414–4422. <http://www.ncbi.nlm.nih.gov/pubmed/21964070>.
- 814 Hollander D, Naftelberg S, Lev-Maor G, Kornblihtt AR, Ast G. 2016. How Are Short Exons Flanked
815 by Long Introns Defined and Committed to Splicing? *Trends Genet* **32**: 596–606.
816 <http://linkinghub.elsevier.com/retrieve/pii/S0168952516300798> (Accessed September 5,
817 2017).
- 818 Huang C, Liang D, Tatomer DC, Wilusz JE. 2018. A length-dependent evolutionarily conserved
819 pathway controls nuclear export of circular RNAs. *Genes Dev*.
820 <http://www.ncbi.nlm.nih.gov/pubmed/29773557> (Accessed May 23, 2018).
- 821 Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. 2009. Genome-wide analysis in vivo of
822 translation with nucleotide resolution using ribosome profiling. *Science (80-)* **324**: 218–223.
- 823 Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, Sharpless NE. 2013.
824 Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**: 141–57.
825 [http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3543092&tool=pmcentrez&rend](http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3543092&tool=pmcentrez&rendertype=abstract)
826 [ertype=abstract](http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3543092&tool=pmcentrez&rendertype=abstract).
- 827 Jens M, Rajewsky N. 2015. Competition between target sites of regulators shapes post-
828 transcriptional gene regulation. *Nat Rev Genet* **16**: 113–126.
- 829 Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, Dreyfuss G. 2010. U1 snRNP protects pre-
830 mRNAs from premature cleavage and polyadenylation. *Nature* **468**: 664–668.
831 <http://www.ncbi.nlm.nih.gov/pubmed/20881964> (Accessed May 23, 2018).
- 832 Kramer MC, Liang D, Tatomer DC, Gold B, March ZM, Cherry S, Wilusz JE. 2015. Combinatorial
833 control of Drosophila circular RNA expression by intronic repeats, hnRNPs, and SR proteins.
834 *Genes Dev* **29**: 2168–2182.
- 835 Legnini I, Di Timoteo G, Rossi F, Morlando M, Briganti F, Sthandier O, Fatica A, Santini T,
836 Andronache A, Wade M, et al. 2017. Circ-ZNF609 Is a Circular RNA that Can Be Translated and
837 Functions in Myogenesis. *Mol Cell* **66**: 22–37.e9.

838 Li X, Liu C-X, Xue W, Zhang Y, Jiang S, Yin Q-F, Wei J, Yao R-W, Yang L, Chen L-L. 2017. Coordinated
839 circRNA Biogenesis and Function with NF90/NF110 in Viral Infection. *Mol Cell* **67**: 214–227.e7.
840 <http://www.ncbi.nlm.nih.gov/pubmed/28625552> (Accessed September 1, 2017).

841 Li Z, Huang C, Bao C, Chen L, Lin M, Wang X, Zhong G, Yu B, Hu W, Dai L, et al. 2015. Exon-intron
842 circular RNAs regulate transcription in the nucleus. *Nat Struct Mol Biol* **22**: 256–264.
843 <http://www.nature.com/nsmb/journal/v22/n3/full/nsmb.2959.html%5Cnhttp://www.nature.com/nsmb/journal/v22/n3/pdf/nsmb.2959.pdf>.

845 Liang D, Tatomer DC, Luo Z, Wu H, Yang L, Chen L-L, Cherry S, Wilusz JE. 2017. The Output of
846 Protein-Coding Genes Shifts to Circular RNAs When the Pre-mRNA Processing Machinery Is
847 Limiting. *Mol Cell* **68**: 940–954.e3. <http://www.ncbi.nlm.nih.gov/pubmed/29174924>
848 (Accessed May 23, 2018).

849 Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen
850 LH, Munschauer M, et al. 2013. Circular RNAs are a large class of animal RNAs with regulatory
851 potency. *Nature* **495**: 333–338. <http://www.ncbi.nlm.nih.gov/pubmed/23446348>.

852 Noh JH, Kim KM, McClusky WG, Abdelmohsen K, Gorospe M. 2018. Cytoplasmic functions of long
853 noncoding RNAs. *Wiley Interdiscip Rev RNA*.

854 Pamudurti NR, Bartok O, Jens M, Ashwal-Fluss R, Stottmeister C, Ruhe L, Hanan M, Wyler E, Perez-
855 Hernandez D, Ramberger E, et al. 2017. Translation of CircRNAs. *Mol Cell* **66**: 9–21.e7.
856 <http://www.ncbi.nlm.nih.gov/pubmed/28344080> (Accessed August 31, 2017).

857 Peng L, Chen G, Zhu Z, Shen Z, Du C, Zang R, Su Y, Xie H, Li H, Xu X, et al. 2016. Circular RNA ZNF609
858 functions as a competitive endogenous RNA to regulate AKT3 expression by sponging miR-
859 150-5p in Hirschsprung’s disease. *Oncotarget* **5**: 808–818.
860 <http://www.oncotarget.com/abstract/13656>.

861 Rybak-Wolf A, Stottmeister C, Glažar P, Jens M, Pino N, Hanan M, Behm M, Bartok O, Ashwal-Fluss
862 R, Herzog M, et al. 2015. Circular RNAs in the Mammalian Brain Are Highly Abundant,
863 Conserved, and Dynamically Expressed. *Mol Cell* **58**: 870–885.

864 Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. 2012. Circular RNAs are the predominant

transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* **7**: e30733.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**: 511–515.

Veno MT, Hansen TB, Veno ST, Clausen BH, Grebing M, Finsen B, Holm IE, Kjems J, Venø MT, Hansen TB, et al. 2015. Spatio-temporal regulation of circular RNA expression during porcine embryonic brain development. *Genome Biol* **16**: 243–245.
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4635978&tool=pmcentrez&rendertype=abstract>.

Wang Y, Wang Z. 2015. Efficient backsplicing produces translatable circular mRNAs. *RNA* **21**: 172–179.

Westholm JO, Miura P, Olson S, Shenker S, Joseph B, Sanfilippo P, Celniker SE, Graveley BR, Lai EC. 2014. Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell Rep* **9**: 1966–1980.

Yang Y, Fan X, Mao M, Song X, Wu P, Zhang Y, Jin Y, Yang Y, Chen L-L, Wang Y, et al. 2017. Extensive translation of circular RNAs driven by N6-methyladenosine. *Cell Res* **27**: 626–641.
<http://www.ncbi.nlm.nih.gov/pubmed/28281539> (Accessed August 31, 2017).

Yang Y, Gao X, Zhang M, Yan S, Sun C, Xiao F, Huang N, Yang X, Zhao K, Zhou H, et al. 2018. Novel Role of FBXW7 Circular RNA in Repressing Glioma Tumorigenesis. *J Natl Cancer Inst*.

Zaphiropoulos PG. 1997. Exon skipping and circular RNA formation in transcripts of the human cytochrome P-450 2C18 gene in epidermis and of the rat androgen binding protein gene in testis. *Mol Cell Biol* **17**: 2985–2993.

Zhang M, Huang N, Yang X, Luo J, Yan S, Xiao F, Chen W, Gao X, Zhao K, Zhou H, et al. 2018a. A novel protein encoded by the circular form of the SHPRH gene suppresses glioma tumorigenesis. *Oncogene*.

Zhang M, Zhao K, Xu X, Yang Y, Yan S, Wei P, Liu H, Xu J, Xiao F, Zhou H, et al. 2018b. A peptide

encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. *Nat Commun* **9**: 4475. <http://www.nature.com/articles/s41467-018-06862-2> (Accessed November 23, 2018).

Zhang X-O, Dong R, Zhang Y, Zhang J-L, Luo Z, Zhang J, Chen L-L, Yang L. 2016a. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res* **26**: 1277–87. <http://genome.cshlp.org/lookup/doi/10.1101/gr.202895.115> (Accessed September 1, 2017).

Zhang XO, Wang H Bin, Zhang Y, Lu X, Chen LL, Yang L. 2014. Complementary sequence-mediated exon circularization. *Cell* **159**: 134–147.

Zhang Y, Xue W, Li X, Zhang J, Chen S, Zhang JL, Yang L, Chen LL. 2016b. The Biogenesis of Nascent Circular RNAs. *Cell Rep*.

Zheng Q, Bao C, Guo W, Li S, Chen J, Chen B, Luo Y, Lyu D, Li Y, Shi G, et al. 2016. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nat Commun* **7**: 11215. <http://www.nature.com/doi/10.1038/ncomms11215> <http://www.nature.com/ncomms/2016/160406/ncomms11215/full/ncomms11215.html>.

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7