ARTICLE

doi:10.1038/nature11928

Circular RNAs are a large class of animal RNAs with regulatory potency

Sebastian Memczak[1]*, Marvin Jens[1]*, Antigoni Elefsinioti[1]*, Francesca Torti[1]*, Janna Krueger[2], Agnieszka Rybak[1], Luisa Maier[1], Sebastian D. Mackowiak[1], Lea H. Gregersen[3], Mathias Munschauer[3], Alexander Loewer[4], Ulrike Ziebold[1], Markus Landthaler[3], Christine Kocks[1], Ferdinand le Noble[2] & Nikolaus Rajewsky[1]

Circular RNAs (circRNAs) in animals are an enigmatic class of RNA with unknown function. To explore circRNAs systematically, we sequenced and computationally analysed human, mouse and nematode RNA. We detected thousands of well-expressed, stable circRNAs, often showing tissue/developmental-stage-specific expression. Sequence analysis indicated important regulatory functions for circRNAs. We found that a human circRNA, antisense to the cerebellar degeneration-related protein 1 transcript (CDR1as), is densely bound by microRNA (miRNA) effector complexes and harbours 63 conserved binding sites for the ancient miRNA miR-7. Further analyses indicated that CDR1as functions to bind miR-7 in neuronal tissues. Human CDR1as expression in zebrafish impaired midbrain development, similar to knocking down miR-7, suggesting that CDR1as is a miRNA antagonist with a miRNA-binding capacity ten times higher than any other known transcript. Together, our data provide evidence that circRNAs form a large class of post-transcriptional regulators. Numerous circRNAs form by head-to-tail splicing of exons, suggesting previously unrecognized regulatory potential of coding sequences.

Mature messenger RNAs are linear molecules with 5′ and 3′ termini that reflect start and stop of the RNA polymerase on the DNA template. In cells, different RNA molecules are sometimes joined together by splicing reactions (trans-splicing), but covalent linkage of the ends of a single RNA molecule to form a circular RNA (circRNA) is usually considered a rare event. circRNAs were discovered in plants and shown to encode subviral agents[1]. In unicellular organisms, circRNAs mostly stem from self-splicing introns of pre-ribosomal RNA[2], but can also arise from protein-coding genes in archaea[3]. In the few unambiguously validated circRNAs in animals, the spliceosome seems to link the 5′ and downstream 3′ ends of exons within the same transcript[4–10]. Perhaps the best known circRNA is antisense to the mRNA transcribed from the *SRY* (sex-determining region Y) locus and is highly expressed in testes[6]. Evidence from computational analyses of expression data in Archaea and Mammalia suggests that circRNAs are more prevalent than previously thought[3,10]; however, it is unknown whether animal circRNAs have any biological function.

In comparison to circRNAs, miRNAs are extremely well studied. miRNAs are ~21-nucleotide-long non-coding RNAs that guide the effector protein Argonaute (AGO) to mRNAs of coding genes to repress their protein production[11–14]. In humans, miRNAs directly regulate expression of most mRNAs[15–18] in a diverse range of biological functions. However, surprisingly little is known about how and if mRNAs can escape regulation by a miRNA. A recently discovered mechanism for miRNA removal in a sequence-specific manner is based on target sites acting as decoys or miRNA sponges[19,20]. RNA with miRNA binding sites should, if expressed highly enough, sequester away the miRNA from its target sites. However, all reported mammalian miRNA sponges have only one or two binding sites for the same miRNA and are not highly expressed, limiting their potency[21–24].

To identify circRNAs across animal cells systematically, we screened RNA-seq data for circRNAs. Compared to previous approaches[10] our computational pipeline can find circRNAs in any genomic region, takes advantage of long (~100 nucleotides) reads, and predicts the acceptor and donor splice sites used to link the ends of the RNAs. We do not rely on paired-end sequencing data or known splice sites. Using published[10,25,26] and our own sequencing data, our method reported thousands of circRNAs in human and mouse tissues as well as in different developmental stages of *Caenorhabditis elegans*. Numerous circRNAs appear to be specifically expressed across tissues or developmental stages. We validated these data and showed that most tested circRNAs are well expressed, stable and circularized using the predicted splice sites. circRNA sequences were significantly enriched in conserved nucleotides, indicating that circRNAs compete with other RNAs for binding by RNA binding proteins (RBPs) or miRNAs. We combined biochemical, functional and computational analyses to show that indeed a known human circRNA, *CDR1* antisense (CDR1as)[9], can function as a negative regulator of miR-7, a miRNA with perfect sequence conservation from annelids to human. Together, our data provide evidence that circRNAs form an important class of post-transcriptional regulators.

circRNAs have complex expression patterns

To comprehensively identify stably expressed circRNAs in animals we screened RNA sequencing reads for splice junctions formed by an acceptor splice site at the 5′ end of an exon and a donor site at a downstream 3′ end (head-to-tail) (Fig. 1a). As standard RNA expression profiling enriches for polyadenylated RNAs, we used data generated after ribosomal RNA depletion (ribominus) and random priming. Such data were used before to detect scrambled exons in mammals[10] (see Methods for comparison). However, this approach was not specifically designed to detect circRNAs and (1) only used existing exon–intron annotations, thus missing RNAs transcribed from introns or unannotated transcripts; (2) did not explicitly identify

[1]Systems Biology of Gene Regulatory Elements, Max-Delbrück-Center for Molecular Medicine, Robert-Rössle-Strasse 10, 13125 Berlin, Germany. [2]Angiogenesis and Cardiovascular Pathology, Max-Delbrück-Center for Molecular Medicine, Robert-Rössle-Strasse 10, 13125 Berlin, Germany. [3]RNA Biology and Post-Transcriptional Regulation, Max-Delbrück-Center for Molecular Medicine, Robert-Rössle-Strasse 10, 13125 Berlin, Germany. [4]Signaling Dynamics in Single Cells, Max-Delbrück-Center for Molecular Medicine, Robert-Rössle-Strasse 10, 13125 Berlin, Germany.
*These authors contributed equally to this work.

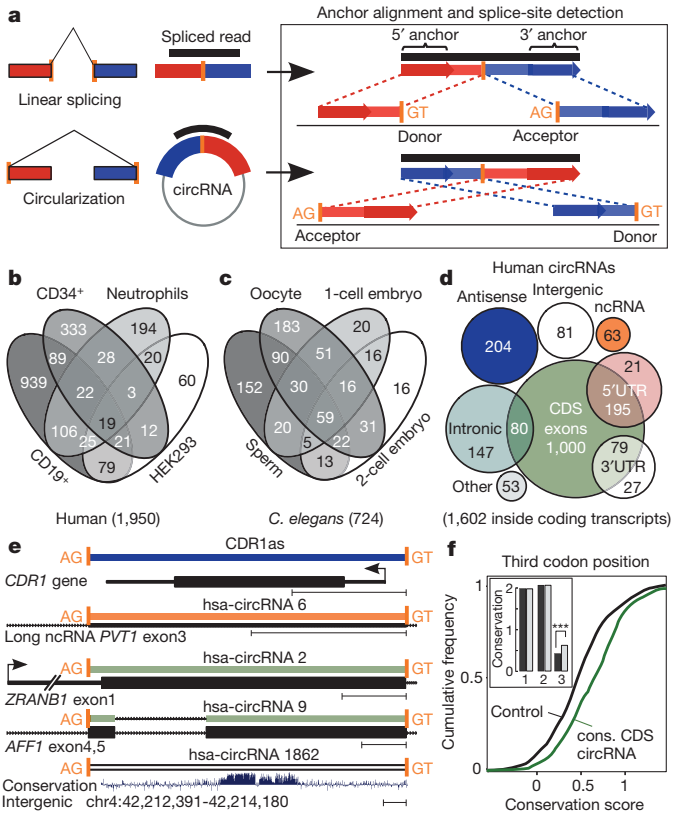

Figure 1 | Detection, classification and evolutionary conservation of circRNAs. **a**, The termini of junction-spanning reads (anchors) align sequentially to the genome for linear (top) but in reversed orientation for head-to-tail spliced reads (bottom). Spliced reads must distribute completely to anchors, flanked by AG/GU (Methods). **b**, **c**, circRNAs in human cell types (**b**) and nematode stages (**c**). **d**, Genomic origin of human circRNAs. A total of 96% of circRNAs overlap known transcripts. **e**, Examples of human circRNAs. The *AFF1* intron is spliced out (Supplementary Fig. 2e). Sequence conservation: placental mammals phyloP score (Methods), scale bar, 200 nucleotides. **f**, A total of 223 human coding sequence circRNAs with mouse orthologues (green) and controls (black) with matched conservation level (inset: mean conservation for each codon position (grey), controls (black); *x* axis, codon positions; *y* axis, placental mammals phyloP score; see also Methods and Supplementary Fig. 1j, k). Third codon positions are significantly more conserved ($P < 4 \times 10^{-10}$, Mann–Whitney *U*-test, $n = 223$).

the splice sites used for circularization; and (3) assumed that each pair of mates in paired-end sequencing derives from the same RNA molecule. To search in a more unbiased way for circRNAs, we designed an algorithm (Methods) that identifies linear and circular splicing events in ribominus data. First, we filtered out reads that aligned contiguously to the genome, retaining the spliced reads. Next, we mapped the terminal parts of each candidate read independently to the genome to find unique anchor positions. Finally, we demanded that (1) anchor alignments can be extended such that the original read sequence aligns completely, and (2) the inferred breakpoint is flanked by GU/AG splice signals. Non-unique mappings and ambiguous breakpoints were discarded. We detected circularization splicing from the reversed (head-to-tail) orientation of the anchor alignments (Fig. 1a). Our method also recovered tens of thousands of known linear splicing events (Methods and Supplementary Fig. 1a, b). We estimated sensitivity (>75%) and false-discovery rate (FDR <0.2%) using simulated reads and various permutations of real sequencing data (Methods and Supplementary Fig. 1c). However, the efficiency of ribominus protocols to extract and sequence circRNAs is limited, reducing overall sensitivity.

We generated ribominus data for HEK293 cells and, combined with human leukocyte data[10], detected 1,950 circRNAs with support from at least two independent junction-spanning reads (Fig. 1b). The expression of genes predicted to give rise to circRNAs was only slightly shifted towards higher expression values (Supplementary Fig. 1d), indicating that circRNAs are not just rare mistakes of the spliceosome. We also identified 1,903 circRNAs in mouse (brains, fetal head, differentiation-induced embryonic stem cells; Supplementary Fig. 1e)[25,26]; 81 of these mapped to human circRNAs (Supplementary Fig. 1f). To explore whether circRNAs exist in other animal clades, we used sequencing data that we produced from various *C. elegans* developmental stages (Stoeckius, M. *et al.*, manuscript in preparation) (Methods) and detected 724 circRNAs, with at least two independent reads (Fig. 1c).

Numerous circRNAs seem to be specifically expressed in a cell type or developmental stage (Fig. 1b, c and Supplementary Fig. 1e). For example, hsa-circRNA 2149 is supported by 13 unique, head-to-tail spanning reads in $CD19^+$ leukocytes but is not detected in $CD34^+$ leukocytes (which were sequenced at comparable depth; Supplementary Table 1), neutrophils or HEK293 cells. Analogously, a number of nematode circRNAs seem to be expressed in oocytes but absent in 1- or 2-cell embryos.

We annotated human circRNAs using the RefSeq database and a catalogue of non-coding RNAs[27–29]. 85% of human circRNAs align sense to known genes. Their splice sites typically span one to five exons (Supplementary Fig. 1g) and overlap coding exons (84%), but only in 65% of these cases are both splice sites that participate in the circularization known splice sites (Supplementary Table 2), demonstrating the advantage of our strategy. 10% of all circRNAs align antisense to known transcripts, smaller fractions align to UTRs, introns, unannotated regions of the genome (Fig. 1d). Examples of human circRNAs are shown in Fig. 1e.

We analysed sequence conservation within circRNAs. As genomic sequence is subject to different degrees of evolutionary selection, depending on function, we studied three subtypes of circRNAs. Intergenic and a few intronic circRNAs display a mild but significant enrichment of conserved nucleotides (Supplementary Fig. 1h, i). To analyse circRNAs composed of coding sequence and thus high overall conservation, we selected 223 human circRNAs with circular orthologues in mouse (Methods) and entirely composed of coding sequence. Control (linear) exons were randomly selected to match the level of conservation observed in first and second codon positions (Methods, Fig. 1f inset and Supplementary Fig. 1k for conservation of the remaining coding sequence (CDS)). circRNAs with conserved circularization were significantly more conserved in the third codon position than controls, indicating evolutionary constraints at the nucleotide level, in addition to selection at the protein level (Fig. 1f and Supplementary Fig. 1j, k). In summary, we have confidently identified a large number of circRNAs with complex expression patterns, which derive often but not always from coding exons. Sequence conservation suggests that at least a subset contains functional sequence elements.

Characterization of 50 predicted circRNAs

We experimentally tested our circRNA predictions in HEK293 cells. Head-to-tail splicing was assayed by quantitative polymerase chain reaction (qPCR) after reverse transcription, with divergent primers and Sanger sequencing (Fig. 2a, b). Predicted head-to-tail junctions of 19 out of 23 randomly chosen circRNAs (83%) could be validated, demonstrating high accuracy of our predictions (Table 1). In contrast, 5 out of 7 (71%) candidates exclusively predicted in leukocytes could not be detected in HEK293 cells, validating cell-type-specific expression.

Head-to-tail splicing could be produced by trans-splicing or genomic rearrangements. To rule out these possibilities as well as potential PCR artefacts, we successfully validated the insensitivity of human circRNA candidates to digestion with RNase R—an exonuclease that degrades linear RNA molecules[30]—by northern blotting with probes which span the head-to-tail junctions (Fig. 2c). We quantified RNase R resistance for 21 candidates with confirmed head-to-tail splicing by

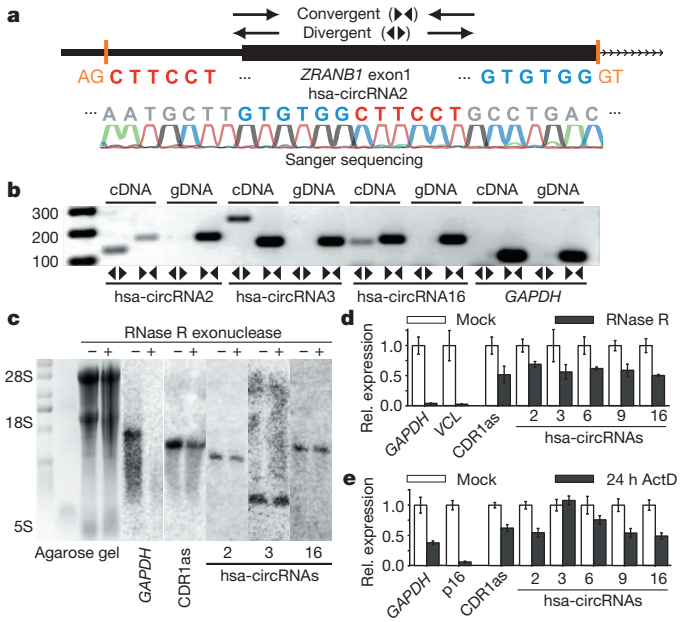

Figure 2 | CircRNAs are stable transcripts with robust expression. **a**, Human (hsa) *ZRANB1* circRNA exemplifies the validation strategy. Convergent (divergent) primers detect total (circular) RNAs. Sanger sequencing confirms head-to-tail splicing. **b**, Divergent primers amplify circRNAs in cDNA but not genomic DNA (gDNA). *GAPDH*, linear control, size marker in base pairs. **c**, Northern blots of mock (−) and RNase R (+) treated HEK293 total RNA with head-to-tail specific probes for circRNAs. *GAPDH*, linear control. **d**, **e**, circRNAs are at least 10-fold more RNase R resistant than *GAPDH* mRNA (**d**) and stable after 24 h transcription block (**e**) (qPCR; error bars indicate standard deviation).

qPCR. All of these were at least 10-fold more resistant than GAPDH (Fig. 2d and Supplementary Fig. 2a). We reasoned that circRNAs should generally turn over more slowly than mRNAs. Indeed, we found that 24 h after blocking transcription circRNAs were highly stable, exceeding the stability of the housekeeping gene *GAPDH*[31] (Fig. 2e and Supplementary Fig. 2b). We also validated 3 out of 3 tested mouse circRNAs with human orthologues in mouse brains (Supplementary Fig. 2c). In *C. elegans* 15 out of 20 (75%) of the predictions from gametes and early embryos were validated in a mixed stage sample (Supplementary Fig. 2d and Supplementary Table 3).

circRNA CDR1as is densely bound by AGO

Stable transcripts with many miRNA-binding sites could function as miRNA sponges. We intersected our catalogue of circRNAs with transcript annotations, assuming that introns would not occur in mature circRNAs (as observed for 3 out of 3 tested circRNAs, Supplementary Fig. 2e). We screened for occurrences of conserved miRNA family seed matches (Methods). When counting repetitions of conserved matches to the same miRNA family, circRNAs were significantly enriched compared to coding sequences ($P < 2.96 \times 10^{-22}$, Mann–Whitney *U*-test, $n = 3{,}873$) or 3′ UTR sequences ($P < 2.76 \times 10^{-21}$, Mann–Whitney *U*-test, $n = 3{,}182$) (Supplementary Fig. 3a, b).

As an extreme case, we discovered that the known human circRNA CDR1as (ref. 9) harboured dozens of conserved miR-7 seed matches. To test whether CDR1as is bound by miRNAs, we analysed biochemical, transcriptome-wide binding-site data for the miRNA effector AGO proteins. We performed four independent PAR-CLIP (photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation) experiments for human AGO (Methods) and analysed them together with published, lower-depth data[32]. PAR-CLIP[32–34] is based on ultraviolet crosslinking of RNA to protein and subsequent sequencing of RNA bound to a RBP of interest. The ~1.5-kilobase (kb) *CDR1as* locus stood out in density and number of AGO PAR-CLIP reads (Fig. 3a), whereas nine combined PAR-CLIP libraries for other RBPs gave virtually no signal. Of note, there is no PAR-CLIP read mapping to the sense coding transcript of the *CDR1* gene, which was originally identified as a target of autoantibodies from patients with paraneoplastic cerebellar degeneration[35].

Sequence analysis across 32 vertebrate species revealed that miR-7 is the only animal miRNA with conserved seed matches that can explain the AGO binding along the CDR1as transcript (Methods). Human CDR1as harbours 74 miR-7 seed matches of which 63 are

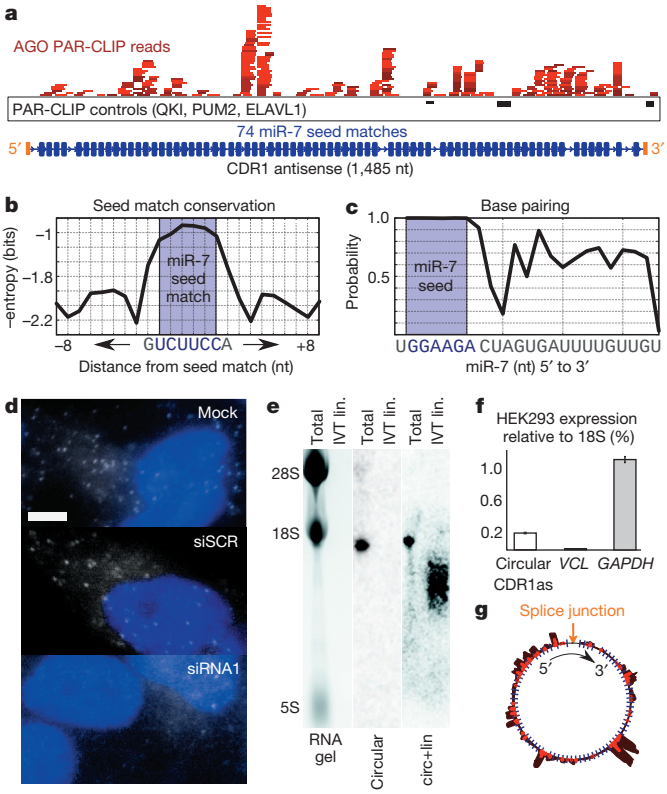

Figure 3 | The circRNA CDR1as is bound by the miRNA effector protein AGO, and is cytoplasmic. **a**, CDR1as is densely bound by AGO (red) but not by unrelated proteins (black). Blue boxes indicate miR-7 seed matches. nt, nucleotides. **b**, **c**, miR-7 sites display reduced nucleotide variability across 32 vertebrate genomes (**b**) and high base-pairing probability within seed matches (**c**). **d**, CDR1as RNA is cytoplasmic and disperse (white spots; single-molecule RNA FISH; maximum intensity merges of *Z*-stacks). siSCR, positive; siRNA1, negative control. Blue, nuclei (DAPI); scale bar, 5 μm (see also Supplementary Fig. 10 for uncropped images). **e**, Northern blotting detects circular but not linear CDR1as in HEK293 RNA. Total, HEK293 RNA; circular, head-to-tail probe; circ+lin, probe within splice sites; IVT lin., *in vitro* transcribed, linear CDR1as RNA. **f**, Circular CDR1as is highly expressed (qPCR, error bars indicate standard deviation). **g**, CDR1as. Blue, seed matches; dark red, AGO PAR-CLIP reads; bright red, crosslinked nucleotide conversions.

Table 1 | Summary of the validation experiments

Sample	Validation experiment	Validation success
Human (HEK293)	Head-to-tail splicing	19 of 23
	Circularity	21 of 21
	Expression >3% vinculin	12 of 21
	Expression specificity (leukocyte specific)	5 of 7
Mouse (adult brain)	Head-to-tail splicing	3 of 3
	Circularity	3 of 3
	Expression >1% β-actin	2 of 3
C. elegans	Head-to-tail splicing	15 of 20
	Circularity	13 of 13
	Expression >1% *eif-3.d*	12 of 15

Most experimentally tested circRNAs are validated.

conserved in at least one other species (Supplementary Fig. 4). Interspaced sequences were less conserved, indicating that miR-7 binding sites are probably functional (Fig. 3b). Secondary structure analysis of predicted circRNA–miRNA duplexes (Methods) showed reduced base-pairing of miR-7 beyond the seed (Fig. 3c). None of the ~1,500 miR-7 complementary sites across 32 vertebrate sequences was complementary beyond position 12 of miR-7 (only three could form an 11-nucleotide duplex) (Supplementary Table 4). Slicing by mammalian Argonaute requires complementarity of positions 10 and 11 and depends on extended complementarity beyond position 12 (ref. 36). Thus, CDR1as seems optimized to be densely bound but not sliced by miR-7.

Single-molecule imaging (Methods) revealed disperse and mostly cytoplasmic CDR1as expression (HEK293 cells), consistent with miRNA sponge function (Fig. 3d and Supplementary Table 5). CDR1as circularization was assayed by northern blotting (Fig. 3e). Nicking experiments confirmed that CDR1as circRNA can be linearized and degraded (Supplementary Fig. 5a). In RNA from HEK293 cells, circularized but no additional linear CDR1as was detected (Supplementary Fig. 5b). Circular expression levels were quantified by qPCR with divergent primers calibrated by standard curves (Supplementary Table 6). CDR1as was highly expressed (~15% to ~20% of *GAPDH* expression, Fig. 3f). Estimating *GAPDH* mRNA copy number from HEK293 RNA-seq data (~1,400 molecules per cell, data not shown) suggests that CDR1as may bind up to ~20,000 miR-7 molecules per cell (Fig. 3g).

If CDR1as functions as a miR-7 sponge, its destruction could trigger downregulation of miR-7 targets. We knocked down CDR1as in HEK293 cells and monitored expression of published miR-7 targets by qPCR with externally spiked-in standards (Methods and Supplementary Fig. 5c, d). All eight miR-7 targets assayed, but also housekeeping genes, were downregulated. Nanostring technology[37] additionally indicated downregulation of many genes (data not shown). Furthermore, stable loss of CDR1as expression by virally delivered small hairpin RNAs led to significantly reduced migration in an *in vitro* wound closure assay (Methods, Supplementary Fig. 5e, f and Supplementary Table 7). Thus, knockdown of CDR1as affects HEK293 cells, but we could not delineate miR-7-specific effects, potentially because of indirect or miR-7-independent CDR1as function (see below).

Co-expression of miR-7 and CDR1as in brain

If CDR1as indeed interacts with miR-7, both must be co-expressed. miR-7 is highly expressed in neuronal tissues, pancreas and pituitary gland[38]. Apart from HEK293 cells, a cell line probably derived from neuronal precursors in embryonic kidney[39], we quantified miR-7 and CDR1as expression across mouse tissues and pancreatic-island-derived MIN6 cells (Methods and Fig. 4a). CDR1as and miR-7 were both highly expressed in brain tissues, but CDR1as was expressed at low levels or absent in non-neuronal tissues, including tissues with very high miR-7 expression. qPCR suggested that CDR1as is exclusively circular in adult and embryonic mouse brain (Supplementary Fig. 5g, h). Thus, CDR1as and miR-7 seem to interact specifically in neuronal tissues. Indeed, when assaying CDR1as and miR-7 in mouse brains by *in situ* hybridizations (Methods), we observed specific, similar, but not identical, expression patterns in the brain of mid-gestation (embryonic day 13.5 (E13.5)) embryos (Fig. 4b). Specifically, CDR1as and miR-7 were highly co-expressed in areas of the developing midbrain (mesencephalon)[40,41]. Thus, CDR1as is highly expressed, stable, cytoplasmic, not detectable as a linear RNA and shares expression domains with miR-7. Together with extensive miR-7 binding within CDR1as, CDR1as has hallmarks of a potent circular miR-7 sponge in neuronal tissues.

Effects of miR-7 and CDR1as in zebrafish

It would be informative to knock out CDR1as in an animal model system. However, a knockout would also affect CDR1 protein, with

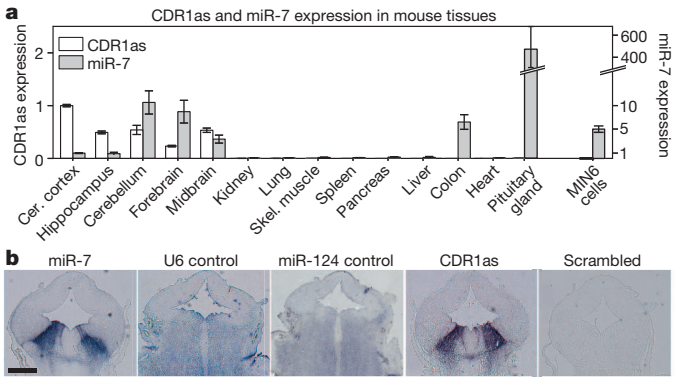

Figure 4 | CDR1as and miR-7 have overlapping and specific expression in neuronal tissues. a, Among mouse tissues and MIN6 cells (qPCR, relative to cerebral cortex expression; error bars indicate standard deviations; see Supplementary Fig. 9a for miR-122 control) neuronal tissues co-express miR-7 and CDR1as. **b**, *In situ* staining of CDR1as and miR-7 in mouse embryo brain E13.5 (U6 and miR-124, positive control; scrambled probe, negative control). Scale bar, 1 mm.

unknown consequences. This problem is circumvented when using zebrafish (*Danio rerio*) as an animal model. According to our bioinformatic analyses (not shown) zebrafish has lost the *cdr1* locus, whereas miR-7 is conserved and highly expressed in the embryonic brain[42]. Thus, we can test whether miR-7 has a loss-of-function phenotype and if this phenotype can be induced by introduction of mammalian CDR1as RNA. We injected morpholinos to knock down mature miR-7 expression in zebrafish embryos (Methods). At a dose of 9 ng of miR-7 morpholino, the embryos did not show overall morphological defects but reproducibly, and in two independent genetic backgrounds (Supplementary Fig. 6a–c), developed brain defects (Fig. 5a, b). In particular, ~70% showed a consistent and clear reduction in midbrain size, and an additional ~5% of animals had almost completely lost their midbrains. Of note, the telencephalon at the anterior tip of the brain was not affected in size. Brain volumes were also measured based on confocal three-dimensional stacks (Fig. 5c and Supplementary Fig. 7). Reduction of the midbrain size correlated with miR-7 inhibition in the respective animals (Supplementary Fig. 6d). These data provide evidence that miR-7 loss-of-function causes a specific reduction of midbrain size.

To test whether CDR1as can function as a miR-7 sponge *in vivo*, we injected embryos with plasmid DNA that expressed a linear version of the full-length human *CDR1as* sequence (Supplementary Fig. 6e, f) or a plasmid provided by the Kjems laboratory that can produce circular CDR1as in human cells (Fig. 5d, e). qPCR analysis detected circular RNA in zebrafish embryos injected with the latter plasmid (Supplementary Fig. 8), which reproducibly and in independent genetic backgrounds lead to reduced midbrain sizes (Fig. 5g, h). Similarly, animals injected with *in vitro*-transcribed partial mouse CDR1as RNA, but not with RNA from the other strand, showed significant midbrain reduction (Supplementary Fig. 6g–i). Thus, the phenotype is probably caused by CDR1as RNA and not by an unspecific effect of RNA or DNA injection. These results provide evidence that human/mouse CDR1as transcripts are biologically active *in vivo* and impair brain development similarly to miR-7 inhibition. The midbrain reduction could be partially rescued by injecting miR-7 precursor (Fig. 5f, g), arguing that the biological effect of CDR1as expression is caused at least in part by interaction of CDR1as with miR-7.

Discussion

We have shown that animal genomes express thousands of circRNAs from diverse genomic locations (for example, from coding and non-coding exons, intergenic regions or transcripts antisense to 5′ and 3′ UTRs) in a complex tissue-, cell-type- or developmental-stage-specific manner. We provided evidence that CDR1as can act as a

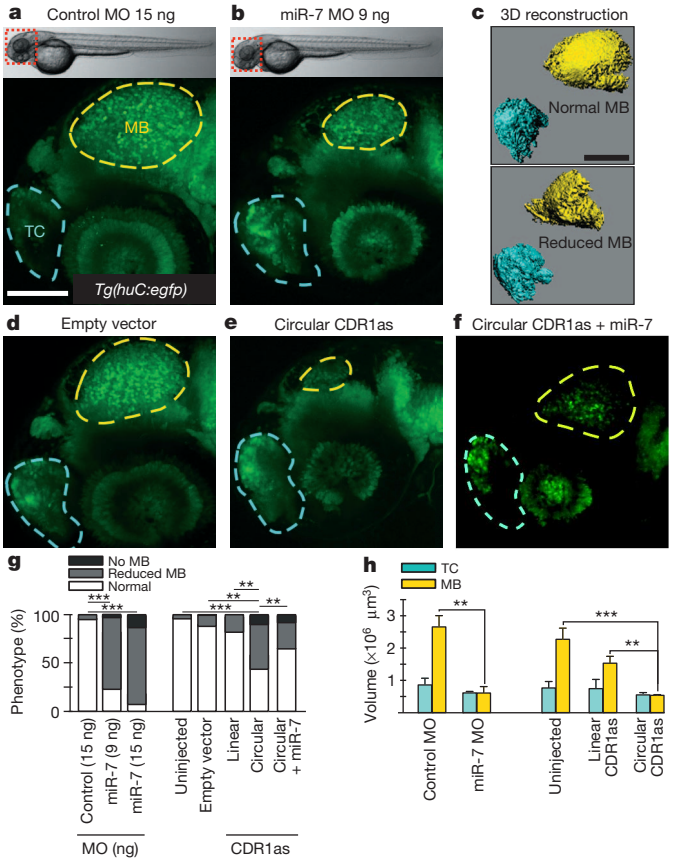

Figure 5 | In zebrafish, knockdown of miR-7 or expression of CDR1as causes midbrain defects. **a**, **b**, Neuronal reporter (*Tg(huC:egfp)*) embryos (top, light microscopy) 48 h post fertilization (bottom, representative confocal *z*-stack projections; blue dashed line, telencephalon (TC) (control); yellow dashed line, midbrain (MB)). Embryos after injection of 9 ng miR-7 morpholino (MO) (**b**) display a reduction in midbrain size. Panel **a** shows a representative embryo injected with 15 ng control morpholino. **c**, Three-dimensional volumetric reconstructions. **d**, Empty vector control. **e**, Expression vector encoding human circular CDR1as. **f**, Rescue experiment with miR-7 precursor. **g**, Phenotype penetrance (% of embryos, miR-7 MO, $n = 135$; uninjected, $n = 83$; empty vector, $n = 91$; linear CDR1as, $n = 258$; circular CDR1as, $n = 153$; circular CDR1as plus miR-7 precursor, $n = 217$). Phenotype distribution derived from at least three independent experiments. Scale bar, 0.1 mm. $^{**}P < 0.01$; $^{***}P < 0.001$ in Students *t*-test for normal midbrain, reduced midbrain (see also Supplementary Fig. 6). **h**, Phenotype quantification (Methods). Error bars indicate standard deviation $n = 3$ per group.

post-transcriptional regulator by binding miR-7 in brain tissues: (1) CDR1as is densely bound by miRNA effector molecules; (2) CDR1as harbours 74 miR-7 seed matches, often deeply conserved; (3) CDR1as is expressed highly, stably and mostly cytoplasmic; (4) CDR1as and miR-7 share specific expression domains in mouse embryonic brain; (5) human/mouse CDR1as is circularized *in vivo* and is not detectable as a linear molecule; (6) human/mouse CDR1as sequences, when injected into zebrafish, and miR-7 knock down have similar phenotypes in brain. While zebrafish circularization of human CDR1as may be incomplete, the midbrain phenotype was stronger compared to expressing linear CDR1as RNA that lacks circularization splice sites. Although the two DNA plasmids used carry identical promoters and were injected in equal concentrations, we cannot rule out the possibility that the difference in midbrain phenotype strength may be explained by other factors. However, because of the observed extreme stability of CDR1as and circRNAs in general, our data argue that circRNAs can be used as potent inhibitors of miRNAs or RBPs. Future studies should elucidate how CDR1as can be converted into a linear molecule and targeted for degradation. miR-671 can trigger destruction of CDR1as[9]. Thus, CDR1as may function to transport miR-7 to subcellular locations, where miR-671 could trigger release of its cargo. Known functions of miR-7 targets such as PAK1 and FAK1 support these speculations[43,44].

The phenotype induced by CDR1as expression in zebrafish was only partially rescued by expressing miR-7, indicating that CDR1as could have functions beyond sequestering miR-7. This idea is supported by *in situ* hybridization in mouse adult hippocampus (Supplementary Fig. 9b) where areas staining for CDR1as but not miR-7 were observed. What could be additional functions of circRNAs beyond acting as sponges? As a single-stranded RNA, CDR1as could, for example, bind *in trans* 3′ UTRs of target mRNAs to regulate their expression. It is even possible that miR-7 binds CDR1as to silence these trans-acting activities. Alternatively, CDR1as could be involved in the assembly of larger complexes of RNA or protein, perhaps similar to other low-complexity molecules[45].

How many other circRNAs exist? In this study, we identified approximately 2,000 human, 1,900 mouse and 700 nematode circRNAs from sequencing data, and our validation experiments confirmed most of the 50 tested circRNAs. However, we analysed only a few tissues/developmental stages with stringent cutoffs. Thus, the true number of circRNAs is almost certainly much larger. Although CDR1as is an extreme case, many circRNAs have conserved seed matches. For example, circRNA from the *SRY* locus[6] has seed sites for murine miRNAs. Therefore, circRNAs probably compete with other RNAs for miRNA binding. Sequence analyses indicated that coding exons serve additional, presumably regulatory functions when expressed within circRNAs, whereas intergenic or intronic circRNAs generally showed only weak conservation. Because we detected thousands of circRNAs, it is appealing to speculate that occasional circularization of exons is easy to evolve and may provide a mechanism for rapid evolution of stably and well expressed regulatory RNAs. Of note, we detected multiple seed matches for viral miRNAs within human circRNAs (not shown). However, there is no reason to think that circRNAs function predominantly to bind miRNAs. As known in bacteria, the decoy mechanism underlying miRNA sponges could be important also for RBPs[46,47]. Similarly, circRNAs could function to store, sort, or localize RBPs. In summary, our data suggest that circRNAs form a class of post-transcriptional regulators which compete with other RNAs for binding by miRNAs and RBPs and may generally function in modulating the local free concentration of RBPs, RNAs, or their binding sites.

Note added in proof: While this paper was under review, circular RNAs in fibroblasts were described[51].

METHODS SUMMARY

Computational pipeline for predicting circRNAs from ribominus sequencing data. A detailed description of the computational methods is given in the Methods.

Cell culture and treatments. HEK293, HEK293TN and HEK293 Flp-In 293 T-REx (Life Technologies) were cultured following standard protocols. Transcription was blocked by adding 2 µg ml^{-1} actinomycin D (Sigma). RNase R (Epicentre Biotechnologies) treatment (3 U µg^{-1}) was performed on total RNA (5 µg) at 37 °C for 15 min. qPCR primers are listed in Supplementary Table 8.

Single-molecule RNA fluorescence *in situ* hybridization (smRNA FISH). Stellaris Oligonucleotide probes complementary to CDR1as were designed using the Stellaris Probe Designer (Biosearch Technologies). Probe pools were obtained from BioCat GmbH as conjugates coupled to Quasar 670. Probes were hybridized at 125 nM at 37 °C. Images were acquired on an inverted Nikon Ti microscope.

Mouse strains and *in situ* hybridization. *In situ* hybridization (ISH) was performed on paraffin tissue sections from B6129SF1/J wild-type mice as described[48] using locked nucleic acid (LNA) probes or RNAs obtained by *in vitro* transcription on PCR products.

Zebrafish methods. *Tg(huC:egfp)* and *Tg(Xia.Tubb:dsRED)* transgenic zebrafish lines were used[49,50]. Morpholino antisense oligomers were injected into the yolk of single-cell-stage embryos. Furthermore, two pCS2+ plasmids coding for full-length linear CDR1as or CDR1as plus upstream and downstream sequence that can express circular CDR1as in human cells (courtesy of the Kjems laboratory)

were injected. Confocal imaging was performed using Carl Zeiss MicroImaging. Reduced midbrain development was defined as >50% smaller than the mean size of controls. Each experimental group was evaluated in at least three independent experiments; a minimum of 80 individual embryos per group was examined.

Full Methods and any associated references are available in the online version of the paper.

Received 11 September 2012; accepted 24 January 2013.
Published online 27 February 2013.

1. Sanger, H. L., Klotz, G., Riesner, D., Gross, H. J. & Kleinschmidt, A. K. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc. Natl Acad. Sci. USA* **73,** 3852–3856 (1976).
2. Grabowski, P. J., Zaug, A. J. & Cech, T. R. The intervening sequence of the ribosomal RNA precursor is converted to a circular RNA in isolated nuclei of Tetrahymena. *Cell* **23,** 467–476 (1981).
3. Danan, M., Schwartz, S., Edelheit, S. & Sorek, R. Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res.* **40,** 3131–3142 (2012).
4. Nigro, J. M. *et al.* Scrambled exons. *Cell* **64,** 607–613 (1991).
5. Cocquerelle, C., Mascrez, B., Hetuin, D. & Bailleul, B. Mis-splicing yields circular RNA molecules. *FASEB J.* **7,** 155–160 (1993).
6. Capel, B. *et al.* Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell* **73,** 1019–1030 (1993).
7. Chao, C. W., Chan, D. C., Kuo, A. & Leder, P. The mouse formin (Fmn) gene: abundant circular RNA transcripts and gene-targeted deletion analysis. *Mol. Med.* **4,** 614–628 (1998).
8. Burd, C. E. *et al.* Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet.* **6,** e1001233 (2010).
9. Hansen, T. B. *et al.* miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J.* **30,** 4414–4422 (2011).
10. Salzman, J., Gawad, C., Wang, P. L., Lacayo, N. & Brown, P. O. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS ONE* **7,** e30733 (2012).
11. Ambros, V. The functions of animal microRNAs. *Nature* **431,** 350–355 (2004).
12. Baek, D. *et al.* The impact of microRNAs on protein output. *Nature* **455,** 64–71 (2008).
13. Selbach, M. *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* **455,** 58–63 (2008).
14. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136,** 215–233 (2009).
15. Krek, A. *et al.* Combinatorial microRNA target predictions. *Nature Genet.* **37,** 495–500 (2005).
16. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120,** 15–20 (2005).
17. Xie, X. *et al.* Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* **434,** 338–345 (2005).
18. Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19,** 92–105 (2009).
19. Ebert, M. S., Neilson, J. R. & Sharp, P. A. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nature Methods* **4,** 721–726 (2007).
20. Franco-Zorrilla, J. M. *et al.* Target mimicry provides a new mechanism for regulation of microRNA activity. *Nature Genet.* **39,** 1033–1037 (2007).
21. Poliseno, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465,** 1033–1038 (2010).
22. Tay, Y. *et al.* Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell* **147,** 344–357 (2011).
23. Cesana, M. *et al.* A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147,** 358–369 (2011).
24. Ebert, M. S. & Sharp, P. A. Emerging roles for natural microRNA sponges. *Curr. Biol.* **20,** R858–R861 (2010).
25. Vivancos, A. P., Guell, M., Dohm, J. C., Serrano, L. & Himmelbauer, H. Strand-specific deep sequencing of the transcriptome. *Genome Res.* **20,** 989–999 (2010).
26. Huang, R. *et al.* An RNA-Seq strategy to detect the complete coding and non-coding transcriptome including full-length imprinted macro ncRNAs. *PLoS ONE* **6,** e27288 (2011).
27. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12,** 996–1006 (2002).
28. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33,** D501–D504 (2005).
29. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25,** 1915–1927 (2011).
30. Suzuki, H. *et al.* Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. *Nucleic Acids Res.* **34,** e63 (2006).
31. Iwai, Y., Akahane, K., Pluznik, D. H. & Cohen, R. B. Ca^{2+} ionophore A23187-dependent stabilization of granulocyte-macrophage colony-stimulating factor messenger RNA in murine thymoma EL-4 cells is mediated through two distinct regions in the 3′-untranslated region. *J. Immunol.* **150,** 4386–4394 (1993).
32. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141,** 129–141 (2010).
33. Lebedeva, S. *et al.* Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Mol. Cell* **43,** 340–352 (2011).
34. Baltz, A. G. *et al.* The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **46,** 674–690 (2012).
35. Dropcho, E. J., Chen, Y. T., Posner, J. B. & Old, L. J. Cloning of a brain protein identified by autoantibodies from a patient with paraneoplastic cerebellar degeneration. *Proc. Natl Acad. Sci. USA* **84,** 4552–4556 (1987).
36. Wee, L. M., Flores-Jasso, C. F., Salomon, W. E. & Zamore, P. D. Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell* **151,** 1055–1067 (2012).
37. Geiss, G. K. *et al.* Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature Biotechnol.* **26,** 317–325 (2008).
38. Landgraf, P. *et al.* A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129,** 1401–1414 (2007).
39. Shaw, G., Morse, S., Ararat, M. & Graham, F. L. Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *FASEB J.* **16,** 869–871 (2002).
40. Kaufman, M. H. & Bard, J. B. L. *The Anatomical Basis of Mouse Development* (Academic, 1999).
41. Schambra, U. *Prenatal Mouse Brain Atlas* (Springer, 2008).
42. Kapsimali, M. *et al.* MicroRNAs show a wide diversity of expression profiles in the developing and mature central nervous system. *Genome Biol.* **8,** R173 (2007).
43. Jacobs, T. *et al.* Localized activation of p21-activated kinase controls neuronal polarity and morphology. *J. Neurosci.* **27,** 8604–8615 (2007).
44. Chacon, M. R. *et al.* Focal adhesion kinase regulates actin nucleation and neuronal filopodia formation during axonal growth. *Development* **139,** 3200–3210 (2012).
45. Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149,** 753–767 (2012).
46. Romeo, T. Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. *Mol. Microbiol.* **29,** 1321–1330 (1998).
47. Gottesman, S. The small RNA regulators of *Escherichia coli*: roles and mechanisms. *Annu. Rev. Microbiol.* **58,** 303–328 (2004).
48. Huelsken, J., Vogel, R., Erdmann, B., Cotsarelis, G. & Birchmeier, W. β-Catenin controls hair follicle morphogenesis and stem cell differentiation in the skin. *Cell* **105,** 533–545 (2001).
49. Park, H. C. *et al.* Analysis of upstream elements in the HuC promoter leads to the establishment of transgenic zebrafish with fluorescent neurons. *Dev. Biol.* **227,** 279–293 (2000).
50. Peri, F. & Nusslein-Volhard, C. Live imaging of neuronal degradation by microglia reveals a role for v0-ATPase a1 in phagosomal fusion *in vivo*. *Cell* **133,** 916–927 (2008).
51. Jeck, W. R. *et al.* Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19,** 1–17 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Feldkamp and C. Langnick (laboratory of W. Chen) for Illumina sequencing runs. We thank J. Kjems for sending us a plasmid encoding circular human CDR1as for our zebrafish experiments. We thank K. Meier for technical assistance with zebrafish experiments and A. Sporbert from the confocal imaging facility. We thank A. Ivanov for assisting in bioinformatic analysis. N.R. thanks E. Westhof for useful discussions. We acknowledge the following funding sources: PhD program of the Max-Delbrück-Center (MDC) (S.M., F.T., L.H.G.); the MDC-NYU exchange program (M.M.); BMBF project 1210182, 'MiRNAs as therapeutic targets' (A.E.); DFG for KFO218 (U.Z.); Helmholtz Association for the 'MDC Systems Biology Network', MSBN (S.D.M.); BMBF support for the DZHK (F.I.N. and N.R.); Center for Stroke Research Berlin (J.K., F.I.N.). Funding for the group of M.L. is supported by BMBF-funding for the Berlin Institute for Medical Systems Biology (0315362C).

Author Contributions S.M., M.J., A.E. and F.T. contributed equally. S.M. performed many experiments, assisted by L.M. M.J. and A.E. carried out most of the computation, with contributions from N.R. and S.D.M. F.T. performed the circRNA validation experiments. A.R. performed all northern experiments. L.H.G. and M.M. contributed AGO PAR-CLIP experiments and HEK293 ribominus data, supervised by M.L. C.K. designed and carried out the single molecule experiments, in part together with A.L. U.Z. performed the mouse experiments. J.K. contributed the zebrafish experiments, supervised by F.I.N. N.R. designed and supervised the project. N.R. and M.J. wrote the paper.

Author Information Sequencing data have been deposited at GEO under accession number GSE43574. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.R. (rajewsky@mdc-berlin.de).

METHODS

Computational pipeline for predicting circRNAs from ribominus sequencing data. Reference genomes (human hg19 (February 2009, GRCh37), mouse mm9 (July 2007, NB137/mm9), *C. elegans* ce6 (May 2008, WormBase v. WS190)) were downloaded from the UCSC genome browser (http://genome.ucsc.edu/)[27]. In a first step, reads that aligned contiguously and full length to the genomes were discarded. From the remaining reads we extracted 20mers from both ends and aligned them independently to find unique anchor positions within spliced exons. Anchors that aligned in the reversed orientation (head-to-tail) indicated circRNA splicing (compare main Fig. 1a). We extended the anchor alignments such that the complete read aligns and the breakpoints were flanked by GU/AG splice sites. Ambiguous breakpoints were discarded. We used the short-read mapper Bowtie 2 (ref. 52). Initially, ribominus reads were aligned in end-to-end mode to the genome:

$ bowtie2 -p16 --very-sensitive --phred64 --mm -M20 --score-min=C,-15,0 -q -x <index> -U reads.qfa 2> bowtie2.log | samtools view -hbuS - | samtools sort - sample_vs_genome

The unmapped reads were separated and run through a custom script to split the reads as indicated in Fig. 1a to obtain 20-nucleotide anchors from both ends of the read:

$ samtools view -hf 4 sample_vs_genome.bam | samtools view -Sb - > unmapped_sample.bam

$ ./unmapped2anchors.py unmapped_sample.bam | gzip > sample_anchors.qfa.gz

Here is an example of two anchor pairs in the FASTQ format; the original read was kept as part of the first anchors identifier to simplify downstream analysis:

```
@s_8_1_0001_qseq_14_A__NCCCGCCTCACCGGGTCAGTGAAAAAACGATCAGAGTAGTGGTCTTCTTCCGGCGGCCCCGCGCGCGCCGCGCTGC
NCCCGCCTCACCGGGTCAGT
+
#BB@?@AB@;=@B;B@@58(
@s_8_1_0001_qseq_14_B
CCCCGCGCGCGCCGCGCTGC
+
;.;((>)>0;.8########
```

Next the anchors were aligned individually to the reference, keeping their paired ordering. The resulting alignments were read by another custom script that jointly evaluates consecutive anchor alignments belonging to the same original read, performs extensions of the anchor alignments, and collects statistics on splice sites. After the run completes, the script outputs all detected splice junctions (linear and circular) in a UCSC BED-like format with extra columns holding quality statistics, read counts etc. The original full-length reads that support each junction are written to stderr:

$ bowtie2 -p16 –reorder --mm -M20 --score-min=C,-15,0 -q -x genome -U sample_anchors.qfa.gz | ./find_circ.py -S hg19 -p sample_ -s sample/sites.log > sample/sites.bed 2> sample/sites.reads

The resulting BED-like file is readily filtered for minimal quality cutoffs to produce the reported circRNA candidates. In particular, we demanded the following: (1) GU/AG flanking the splice sites (built in); (2) unambiguous breakpoint detection; (3) a maximum of two mismatches in the extension procedure; (4) the breakpoint cannot reside more than 2 nucleotides inside an anchor; (5) at least two independent reads (each distinct sequence only counted once per sample) support the junction; (6) unique anchor alignments with a safety margin to the next-best alignment of at least one anchor above 35 points (~more than two extra mismatches in high-quality bases); and (7) a genomic distance between the two splice sites of no more than 100 kb (only a small percentage of the data). As the ribosomal DNA cluster is part of the *C. elegans* genome assembly (ce6) and ribosomal pre-RNAs could give rise to circular RNAs by mechanisms independent of the spliceosome, we discarded 130 candidates that mapped to the rDNA cluster on chrI:15,060,286-15,071,020.

Permutation testing. To test the robustness of the circRNA detection pipeline we altered the sequence of real sequencing reads in different ways at the step of anchor generation. We (1) reversed either anchor; (2) reversed the complete read; (3) randomly reassigned anchors between reads; or (4) reverse complemented the read (as a positive control). Although the reverse complement recovered the same output as expected, the various permutations led to only very few candidate predictions, well below 0.2% of the output with unpermuted reads and in excellent agreement with the results from simulated reads (Supplementary Fig. 1c).

HEK293 RNA-seq after rRNA depletion (RibominusSeq). Total HEK293 RNA was isolated using Trizol as recommended by the manufacturer. Ribosomal RNA was depleted from total RNA using the Ribominus kit (Invitrogen). A cDNA library was generated from rRNA-depleted RNA according to the Illumina RNA-seq protocol. The cDNA library was sequenced on an Illumina GAIIx by a 2×76 bp run.

***C. elegans* oocyte isolation.** Oocytes were isolated from worms carrying a temperature-sensitive (TS) allele for *fem-1* (unovulated oocytes BA17[*fem-1*(*hc17ts*)] strain) and *spe-9* (partially ovulated oocytes BA671[*spe-9*(*hc88ts*)]) as described previously[53]. Oocytes were washed at least four times in PBS containing protease inhibitors (Sigma-Aldrich) to separate from worm debris. Oocyte purity was observed under the dissection scope (Zeiss). Oocytes were extracted from young adults to enrich for non-endomitotic oocytes, which was also checked by fluorescence microscopy (Zeiss) with a nuclear dye. Oocytes isolated from *fem-1* or *spe-9* mutant background worms are hereafter referred to as *fem-1* oocytes and *spe-9* oocytes, respectively.

***C. elegans* sperm isolation.** Sperm was isolated in principle as described previously[54] from male worms obtained from a *fog-2*(q71) mutant background. Males were cut in cold PBS containing protease inhibitors (Sigma-Aldrich). Sperm was subsequently purified by filtration (3×40 µm nylon mesh, 2×10 µm nylon mesh) and a series of differential centrifugations (30 min 300*g*, 10 min 450*g*) and washed twice in cold PBS. Sperm was subsequently activated by incubation in PBS containing 200 µg ml^{-1} Pronase (Sigma-Aldrich) for 30 min at 25 °C. Sperm purity is around 70% spermatids and spermatozoa contaminated with around 30% primary and secondary spermatocytes, as observed under oil immersion microscope.

***C. elegans* isolation of 1-cell- and 2-cell-stage embryos.** 1-cell and 2-cell-stage embryos were obtained by fluorescence-activated cell sorting as described previously[55]. Microscopic examination of the sorted embryos indicated that the 1-cell-stage sample was virtually pure (>98% one-cell stage embryos), whereas the 2-cell-stage embryo sample was a mixture of 1-cell-stage (40%), 2-cell-stage (55%) and older (<5%) embryos. Moreover, purity of the stages was further validated by checking for marker gene expression.

Ribominus RNA preparation from *C. elegans* samples. We used a kit that was developed for human and mouse samples, but still performs sufficiently to enrich mRNAs up to 30% in *C. elegans*. Most of the remaining reads mapped to ribosomal RNAs. 1 µg of total RNA per sample was depleted from rRNAs with the Ribominus Transcriptome kit (Invitrogen) according to the manufacturer's instructions with the modification that annealing of LNA probes to total RNA was performed in a thermocycler (Eppendorf) with a temperature decrease from 70 to 37 °C at a rate of 1 °C per min. Depletion of rRNAs was validated by capillary gel electrophoresis on a Bioanalyzer (Agilent). The ribominus RNA was then processed for sequencing library preparation according to the Illumina protocol.

Cluster generation and sequencing of *C. elegans* libraries. Cluster generation as well as sequencing of the prepared libraries was performed on the Illumina cluster station (Illumina) and sequenced on the HiSeq2000 according to the manufacturer's protocols (Illumina).

Human gene models. We obtained gene models for RefSeq transcripts (12 December 2011), non-coding RNAs[29,56] and the rnaGene and tRNA tracks from the UCSC table browser (23 April 2012)[27].

Intersection of circRNAs with known transcripts. Our computational screen identifies only the splice sites that lead to circularization but not the internal exon/intron structure of circular RNAs. To perform analyses of the sequence content of circRNAs we therefore inferred as much as possible from annotated transcripts. The conservative assumption was that as little as possible should be spliced out. On the other hand, coincidence of circRNA splice sites with exonic boundaries inside a transcript were considered as an indicator for relevant agreement and internal introns appear to be spliced out (Supplementary Fig. 2e). We therefore sorted all overlapping transcripts hierarchically by (1) splice-site coincidence (2, 1, or 0); (2) total amount of exonic sequence between the splice sites; (3) total amount of coding sequence. The latter was used to break ties only and helped the annotation process. If one or both splice sites fell into an exon of the best matching transcript, the corresponding exon boundary was trimmed. Likewise, if it fell

into an intron or beyond transcript bounds, the closest exon was extended to match the circRNA boundaries. circRNA start/end coordinates were never altered. If no annotated exons overlapped the circRNA we assumed a single-exon circRNA. The resulting annotation of circRNAs is based on the best matching transcript and may in some cases not represent the ideal choice. Changing the annotation rules, however, did not substantially change the numbers in Fig. 1d.

Finding circRNAs conserved between human and mouse. We reasoned that when comparing two species, the cutoff of two independent reads in each of them could be dropped, as orthologous circRNAs would automatically be supported by two independently produced reads via the intersection. We therefore mapped all mouse circRNA candidates with less stringent filtering to human genome coordinates using the UCSC liftOver tool[57]. The mapped mouse circRNAs were compared with independently identified human circRNAs, yielding 229 circRNAs with precisely orthologous splice sites between human and mouse. Of these, 223 were composed exclusively of coding exons and were subsequently used for our conservation analysis (Fig. 1f). When intersecting the reported sets of circRNAs supported by two independent reads in each species, we found 81 conserved circRNAs (supported by at least 4 reads in total).

Conserved element counting. We downloaded genome-wide human (hg19) phyloP conservation score[58] tracks derived from genome alignments of placental mammals from UCSC[27]. We interrogated the genome-wide profile inside circRNAs in two different ways. (1) Intergenic and intronic circRNAs. We read out the conservation scores along the complete circRNA and searched for blocks of at least 6-nucleotide length that exceeded a conservation score of 0.3 for intergenic and 0.5 for intronic circRNAs. The different cutoffs empirically adjust for the different background levels of conservation and were also used on the respective controls. For each circRNA, we computed the cumulative length of all such blocks and normalized it by the genomic length of the circRNA. Artefacts of constant positive conservation scores in the phyloP profile, apparently caused by missing alignment data, were removed with an entropy filter (this did not qualitatively affect the results). circRNAs annotated as intronic by the best-match procedure explained above that had any overlap with exons in alternative transcripts on either strand (five cases) were removed from the analysis. The resulting distributions are shown in Supplementary Fig. 1h, i. (2) Coding exon circRNAs. We used the best-match strategy outlined above to construct an estimated 'exon-chain' for the circRNAs that overlapped exclusively coding sequence. Using this chain we *in silico* 'spliced' out the corresponding blocks of the conservation score profile. We kept track of the frame and sorted the conservation scores into separate bins for each codon position. In addition to this, we also recorded conservation scores in the remaining pieces of coding sequence ('outside' the circRNA) as a control. However, we observed that the level of conservation is systematically different between internal parts of the coding sequence and the amino- or carboxy-terminal parts (not shown). We therefore randomly generated chains of internal exons, mimicking the exon-number distribution of real circRNAs, as a control. When analysing the circRNAs conserved between human and mouse, it became furthermore apparent that we also needed to adjust for the higher level of overall conservation. High expression generally correlates with conservation and thus, an expression cutoff was enforced on the transcripts used to generate random controls. This resulted in a good to conservative match with the actual circRNAs (Supplementary Fig. 1j, k).

Overlap of identified circRNAs with published circular RNAs. A number of studies in human have reported evidence for circRNAs which derive from exons of DCC[4], ETS1[5] and a non-coding RNA from the human *INK4/ARF* locus[8] and the CDR1as locus[9]. Additionally, circRNAs from exons of the genes *CAMSAP1*, *FBXW4*, *MAN1A2*, *REXO4*, *RNF220* and *ZKSCAN1* have been recently experimentally validated[10]. For the four genes from the latter study, where we had ribominus data from the tissues in which these circRNAs were predicted (leukocytes), we recovered validated circRNAs from all of them (*ZKSCAN1*, *CAMSAP1*, *FBXW4*, *MAN1A2*).

Cell culture and treatments. HEK293 (Fig. 3f), HEK293TN (for virus production) and HEK293 Flp-In T-REx 293 (Life Technologies, all other experiments) were cultured in Dulbecco's modified Eagle medium GlutaMax (Gibco) 4.5 g l^{-1} glucose, supplemented with 10% FCS, 20 U ml^{-1} penicillin and streptomycin (Gibco) at 37 °C, 5% CO_2. Whereas CDR1as/*GAPDH* ratios were within the given range, we observed two- to fivefold variation of CDR1as/vinculin ratios between different HEK lines. Transcription was blocked by adding 2 μg ml^{-1} actinomycin D or DMSO as a control (Sigma-Aldrich) to the cell culture medium. For *in vitro* wound healing assays, cells were grown to confluency, the cell layer was disrupted using a 300 μl pipette tip and cells were washed once with medium. Bright-field images of cells were taken using a Axio Observer.Z1 (Zeiss) right after setting the scratch and 24 h later. The relative scratch areas were measured using ImageJ software.

Quantitative PCR. Total RNA from cell lines was isolated using Trizol (Invitrogen) extraction following the manufacturer's protocol. Adult B6129SF1/J mice were dissected and tissue samples were collected directly into ice-cold Trizol for RNA preparation. *Caenorhabditis elegans* RNA was isolated from about 7,000 mixed stage worms by two rounds of freeze–thaw lysis in Trizol LS reagent (Invitrogen) according to the manufacturer's protocol. RNA was extracted from aqueous phase with phenol:chloroform (Ambion). RNA was precipitated with isopropanol and Glycoblue (Ambion) overnight at −20 °C or for 30 min at −80 °C, respectively. Reverse transcription was performed using M-MLV (Promega) or Superscript III with oligo(dT) primer (all Invitrogen) or random primer (Metabion). For assaying mRNA expression level, qRT–PCR was performed using SYBR-Green Fluorescein (Thermo Scientific, Fermentas) and a StepOnePlus PCR System (Applied Biosystems). Expression data in CDR1as knockdown experiments, transcriptional block and RNase R assays were normalized to *C. elegans* spike-in RNA. To this end 5–10% *C. elegans* total RNA was added to the respective Trizol sample and qPCR primer for *ama-1* or *eif-3.d* were used. Mouse expression data were normalized to *Actb*. miRNA expression levels were assayed using TaqMan microRNA assays (Applied Biosystems) and normalized to sno-234. Expression levels of circRNAs described in this study were measured by qPCR using divergent primers. A list of primer sequences is available in Supplementary Table 8.

PCR amplification and Sanger sequencing. DNA templates were PCR amplified using BioRad Mastercyclers and ThermoScientific DreamTaq Green PCR Master Mix according to the manufacturer's protocol. We performed 35 cycles of PCR. PCR products were visualized after electrophoresis in 2% ethidium bromide-stained agarose gel. To confirm the PCR results, the PCR products were purified through Agencourt AMPure XP PCR purification kit. Direct PCR product Sanger sequencing was performed by LGC Genomics Ready2 Run services. Primer P1 was provided for sequencing the product for each candidate.

Primer design. Divergent primers were designed for each candidate (P1, P2) to anneal at the distal ends of its sequence. As negative controls we used divergent primers for *GAPDH* and *ACTB* linear transcript in HEK293 cells, and eIF-3.D in *C. elegans*. As a further negative control for divergent primers, we used genomic DNA extracted through Qiagen DNeasy Blood & Tissue kit. As positive controls, we used convergent primers for the corresponding linear transcripts or for housekeeping genes (eIF-3.D for *C. elegans*).

RNase R treatment. HEK293 DNase-treated total RNA (5 μg) was incubated 15 min at 37 °C with or without 3 U μg^{-1} of RNase R (Epicentre Biotechnologies). RNA was subsequently purified by phenol-chloroform extraction, retro-transcribed through Superscript SSIII (Invitrogen) according to the manufacturer's protocol, and used in qPCR.

RNA nicking assay. For partial alkaline hydrolysis (nicking) 1 μg μl^{-1} of HEK293 total RNA was incubated in 50 mM $NaHCO_3$ for 2.5 or 5 min at 90 °C or 5 min on ice for controls. After incubation the samples were immediately re-suspended in denaturing RNA sample buffer and analysed on northern blots.

Northern blotting. Total RNA (10–20 μg) was loaded on a 1.2% agarose gel containing 1% formaldehyde and run for 2–2.5 h in MOPS buffer.

The gel was soaked in 1×TBE for 20 min and transferred to a Hybond-N^+ membrane (GE Healthcare) for 1 h (15 V) using a semi-dry blotting system (Bio-Rad). Membranes were dried and ultraviolet-crosslinked (at 265 nm) 1× at 200,000 μJ cm^{-2}. Pre-hybridization was done at 42 °C for 1 h and ^{32}P-labelled oligonucleotide DNA probes were hybridized overnight. The membranes were washed briefly in 2× SSC, 0.1% SDS at room temperature and two additional times at 55 °C for 30 min, followed by two 30-min washes in 0.2× SSC, 0.1% SDS at 50–55 °C. For data collection, the membrane was exposed to a phosphoimager screen.

Genome alignments for detecting miRNA seed complementary sites. Multiple species alignments for the genomic intervals, corresponding to circRNAs predicted in *C. elegans* (ce6), human (hg19) and mouse (mm9), were generated via the Galaxy server at UCSC[59–61]. In case that a circle was overlapping with an annotated transcript, the inferred spliced sequence was used for retrieving the alignments.

The alignments included *C. elegans*, *C. briggsae* and *C. remanei* in the first case and *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Bos taurus* and *Canis familiaris* in the latter two.

***C. elegans* human and mouse miRNAs.** Fasta files with *C. elegans*, human and mouse miRNAs were obtained from miRBase release 16 (ref. 62). Only mature miRNAs were considered for the seed analysis. According to miRBase 16 a mature miRNA is the predominant miRNA between the two species arising from the two arms of the precursor hairpin (information that is not included in more recent versions).The miRNAs were grouped into families that share a common seed (nucleotides 2–7). There are 117, 751 and 723 miRNA families for *C. elegans*, human and mouse, respectively.

Detecting putative miRNA seed matches. The *C. elegans*, human and mouse multiple species alignments were scanned for putative conserved miRNA target sites for each of the mature miRNA families. A putative target site of a miRNA is a

6-nucleotide-long sequence in the genome that is the reverse complement of nucleotides 2–7 of the mature miRNA sequence. A putative target site is called conserved if it is found in *C. elegans*, *C. briggsae* and *C. remanei* in the first case or in human, mouse, rat, cow and dog in the latter.

AGO PAR-CLIP. Generation and growth conditions of human embryonic kidney (HEK) 293 cells and HEK293 stably expressing Flag/HA–AGO1 and Flag/HA–AGO2 were reported previously[63]. Stably transfected and parental HEK293 cells were labelled with 100 μM 4-thiouridine for 16 h. After labelling, procedure followed the PAR-CLIP protocol as described[32]. Briefly, ultraviolet-irradiated cells were lysed in NP-40 lysis buffer. Immunoprecipitation was carried out with protein G magnetic beads (Invitrogen) coupled to anti-Flag antibody (Sigma) and to anti-AGO2 antibody[64] from extracts of stably transfected and parental HEK293 cells, respectively, for 1 h at 4 °C. Beads were treated with calf intestinal phosphatase (NEB) and radioactively end-labelled by T4 polynucleotide kinase (Fermentas). The crosslinked protein–RNA complexes were resolved on 4–12% NuPAGE gel (Invitrogen), and a labelled protein–RNA complex of close to 100 kDa was excised. The protein–RNA was isolated by electroelution. RNA was isolated by proteinase K treatment and phenol-chloroform extraction, reverse transcribed and PCR-amplified. The amplified cDNA was sequenced on a GAIIx (Illumina) with 36 cycles.

Human Argonaute PAR-CLIP analysis. We obtained Argonaute PAR-CLIP reads from ref. 32. We additionally produced 4 PAR-CLIP libraries. In total, we analysed the following PAR-CLIP data sets: AGO1_4su_1 (SRR048973), AGO3_4su_1 (SRR048976) from ref. 32; AGO1_4su_ML_MM_6, AGO2_4su_ML_MM_7, AGO2_4su_ML_MM_8, and AGO2_4su_3_ML_LG (our own data, published under GEO accession GSE43574).

Redundant reads were collapsed (such that each distinct read sequence appears only once), aligned to the human genome (assembly hg19) using bwa 0.6.1-r104 (ref. 65), and analysed by our in-house PAR-CLIP analysis pipeline (Jens, M. *et al.*, unpublished), essentially as described in ref. 33. Briefly, reads uniquely aligning to the genome are grouped into clusters contiguously covering the reference, assigning each cluster a number of quality scores (T conversions, number of independent reads, etc). Clusters with less than 3 reads from 3 of 6 independent AGO PAR-CLIP libraries or lacking T conversions were discarded. Remaining clusters are annotated against a comprehensive list of transcript models (see below) and collected into 'only sense', 'only antisense' and 'intergenic/overlapping transcription' categories based on their annotation. As PAR-CLIP sequencing preserves the directionality of RNA fragments we assume 'only antisense' clusters to predominantly represent false positives due to mapping artefacts (PAR-CLIP RNA is mutated and fragments are often short), and choose quality cutoffs for all clusters such that the fraction of kept 'only antisense' clusters is reduced to below 5%. Remaining 'only antisense' clusters were discarded. For Fig. 3a, uniquely aligning, collapsed reads are shown.

AGO binding sites in *C. elegans*. Sequencing reads from the Zisoulis Alg-1 HITS-Clip data were obtained from http://yeolab.ucsd.edu/yeolab/Papers_files/ALG1_MT-WT_raw.tar.gz (ref. 66). The raw sequencing data of the wild-type Alg-1 HITS-CLIP was pre-processed and mapped with the mapper module from miRDeep2 (ref. 74). The pre-processed reads were mapped with bowtie version 0.12.7 (ref. 67) to the *C. elegans* genome (ce6). All reads that overlapped when mapped to the genome were merged into bigger regions (islands). Read counts were averaged. This resulted in 24,910 islands in the *C. elegans* genome.

Analysis of sequence conservation in CDR1as. Genome alignments of 32 vertebrates were downloaded from the UCSC database (hg19)[27] and analysed for the *CDR1as* locus. Primate species other than human were discarded to not bias the analyses. The one species (cow) with more than 50% gaps in the *CDR1as* locus was also discarded. The alignments for the seed regions were then corrected. Specifically, bases that would clearly align with the seed but had been separated in the alignment by runs of gaps were re-aligned. These corrections were necessary in less than 1% of all seed sites.

For an in-depth analysis we BLATed[68] the human *CDR1as* sequence with 20-nucleotide flanking region against all vertebrate genomes in the UCSC genome browser and kept only hits that in turn aligned best to the human locus. The resulting sequences were used to build a multiple species alignment with MUSCLE[69]. The same corrections were applied as described above. This alignment was also used for Supplementary Fig. 4. Entropy was calculated in $\log_2$ units and averaged across all alignment columns bracketing each human seed site by maximally 8 nucleotides.

Analysis of miR-7 base-pairing within CDR1as. RNAcofold[70] was used to co-fold miR-7 with each of the 74 binding regions within CDR1as defined as the miR-7 seed match TCTTCC and the next 16 bases upstream.

Single-molecule RNA fluorescence *in situ* hybridization (smRNA FISH). 48 oligonucleotide probes (20 nucleotides length; spacing 2 nucleotides) complementary to the CDR1as transcript were designed using the Stellaris Probe Designer version 2.0 (Biosearch Technologies) with a masking level of 4 on the human genome to achieve high probe specificity (Supplementary Table 8). Stellaris probe pools were obtained from BioCat GmbH as conjugates coupled to Quasar 670 (a Cy 5 replacement). Flp-In T-REx 293 cells (Life Technologies) were grown exponentially and seeded into LabTek 4-well chambered coverslips (1 to 2×10^5 cells per well). Hybridizations were performed according to the manufacturer's instructions with 50 ng ml^{-1} DAPI as nuclear counterstain; Stellaris probes were hybridized at 125 nM concentration with a stringency of 10% formamide in overnight hybridizations at 37 °C. Images were acquired on an inverted Nikon Ti microscope with a Hamatsu ORCA R2 CCD camera, a 60× NA 1.4 oil objective and Nikon NIS-Elements Ar software (version 4), using an exposure time of 50 ms for DAPI and 1–1.5 s for Quasar 670. Groups of cells for imaging were chosen in the DAPI channel; *Z*-stacks were acquired in the Quasar 670 channel using 0.3 μm spacing and comprised a total depth of 6.5 μm (5 μm below and 1.5 μm above the middle of the nucleus) and merged using maximum intensity.

Mouse strains and *in situ* hybridizations. All mice were bred and maintained in the animal facility of the Max Delbrück Centrum under specific pathogen-free conditions, in plastic cages with regular chow and water ad libitum. All aspects of animal care and experimental protocols were approved by the Berlin Animal Review Board (REG 0441/09). B6129SF1/J wild-type adult, newborns (postnatal day 1) or pregnant females (plug detection at day 0.5; embryo collection at day 13.5) were used, as indicated for each experiment, to obtain the tissues needed for RNA analysis and *in situ* hybridizations (ISH). After death, embryos or tissues were immediately frozen in liquid nitrogen and stored at −70 °C, or fixed for ISH.

Mouse brain structures were collected and named according to the anatomical guidelines of the Gene Expression Nervous System Atlas of the Rockefeller University (http://www.gensat.org) and the Mouse Brain Atlas (http://www.mbl.org/mbl_main/atlas.html).

For the RNA analysis and to clone CDR1as-specific RNA probes, two adult 1-year-old mice of both sexes were dissected, total RNA prepared and analysed. If embryos or newborns were sectioned, a minimum of two specimens were evaluated; in some instances up to 5 specimens were used.

For ISH, samples were fixed in formalin (1×PBS; 4% formaldehyde) for 12 h and post-fixed (70% ethanol, 18 h) before dehydrating and paraffin-embedding. Next, the organs were perfused with a standard protocol using a Shandon XP Hypercentre. For ISH mouse embryos or organs were cut in RNase-free conditions at 6 μm and ISH was performed as described[48] with digoxigenin (DIG)-labelled RNA probes. All DIG–RNA probes were hybridized at 58 °C overnight. A total of 600 ng of the labelled probes was used per slide.

To amplify *Cdr1* sense and antisense sequences for ISH probe preparation a standard PCR-amplification was performed using mouse cerebellum cDNA. Three *Cdr1as* amplicons were generated, two of which probes are meant for the detection of both linear and circular forms using mmuCdr1_1f 5′-TGCCAGTACCAAGGTCTTCC-3′ and mmuCdr1_1r 5′-TTTTCTGCTGGAAGATGTCAA-3′, as well as mmuCdr1_2f 5′-CCAGACAATCGTGATCTTCC-3′ and mmuCdr1_2r 5′-ATCTTGGCTGGAAGACTTGG-3′. In addition a probe was generated, specific to the circular probe, using the divergent primers mmuCdr1_as_7f 5′-CCACATCTTCCAGCATCTTT-3′ and mmuCdr1_as_7r 5′-TGGATCCCTTGGAAGACAAA-3′ (CDR1_as head to tail probe). All ensuing fragments were subcloned into pCR-BluntII-TOPO (Invitrogen) and verified by sequencing. Linearized plasmids were amenable for *in vitro* transcription using the T7 (antisense) or SP6 (sense) polymerase and a DIG-label nucleotide mixture according to manufacturer's instruction (Roche Applied Science).

LNA ISHs were performed according to a protocol suggested by the manufacturer (Exiqon) with minor modifications. For individual LNAs, specific protocols were run at 51 °C (miR-7; 38485-15) or 58 °C (miR-124; 88066-15) on an InsituPro VS robot (Intavis). A pre-hybridization step was added, which consisted of an incubation of the slides at 15 °C lower than the hybridization temperature for 30 min using hybridization buffer. The antibody-blocking step was performed in the presence of 1% mouse blocking reagent (Roche 11096176001) and 10% sheep serum. The LNA probes were used at the following concentrations: miR-7 40 nM; miR-124 20 nM; U6 snRNA 1 nM; scrambled 40 nM, as suggested by miRCURY LNA microRNA ISH Optimization kit (Exiqon; 90004). Before detection all slides were washed 4× in NTMT including 1 mM Levamisole. The doubly DIG-labelled LNAs were detected by the alkaline phosphatase using the substrate BM-purple (Roche; 11442074001) at 37 °C.

siRNA- and shRNA-mediated knock down. CDR1as was knocked down using custom designed siRNA oligonucleotides (Sigma) and Lipofectamine RNAiMax (Invitrogen). 2×10^6 HEK293 cells were transfected with 10 nM siRNA duplex following the manufacturer's protocol. After 12–16 h cells were harvested and subjected to RNA analysis. For stable knock down of CDR1as, 293TN cells were co-transfected with the packaging plasmids pLP1, pLP2 and the VSV-G plasmid

(Invitrogen) and pSicoR constructs[71] (sequences available in the Supplementary Table 8) by calcium phosphate transfections. Viral supernatants were harvested after 24 h and 48 h post transfection and filtered through a 0.44 μm filter. For infection the viral supernatants supplemented with fresh medium and 6 μg ml^{-1} polybrene was added to target cells. After overnight infection cells were allowed to recover for 12 h and subjected to a second round of infection. Cells were collected 48–72 h after the first infection. The list of siRNA oligonucleotides is provided in Supplementary Table 8.

Zebrafish methods. Zebrafish and their embryos were handled according to standard protocols[72] and in accordance with Max Delbrück Centrum institutional ethical guidelines. The *Tg*(*huC:egfp*) and the *Tg*(*Xia.Tubb:dsRED*) transgenic zebrafish lines have been described elsewhere[49,50]. Morpholino antisense oligomers (Gene Tools) were prepared at a stock concentration of 1 mM according to the manufacturer's protocol. Sequences: control morpholino, 5′-CTC TTACCTCAGTTACAATTTATA-3′ (control morpholino) and morpholino targeting miR-7, 5′-ACAACAAAATCACAAGTCTTCCACA-3′ (miR-7 morpholino). For titration experiments we used 15 ng of control morpholino and 9 and 15 ng of miR-7 morpholino; for all other experiments we used 9 ng miR-7 morpholino. 3 nl of morpholinos were injected into the yolk of single-cell-stage embryos.

A 673-nucleotide mouse *Cdr1as* fragment was amplified from mouse cerebellar cDNA and the amplicon was subcloned into a pCR-Blunt II Topo vector (Invitrogen). The vector was linearized with KpnI or ApaI (Fermentas) *in vitro* transcribed (IVT) using T7 and SP6 RNA polymerases (Promega) and the resulting *Cdr1as* and reverse complement *Cdr1as*_control products were used for injections (1.5 nl of 100 ng nl^{-1}) into the cell of single-cell-stage embryos. In a repetition of these experiments the *Cdr1as* fragment amplicon was directly used as a template for IVT by exploiting T7-promoter extended forward and reverse primer.

Approximately 1.5 nl of a 50 ng μl^{-1} construct (backbone pCS2+) expressing the human linear or the human circular CDR1as was injected into the cell of single-cell-stage embryos (provided by the Kjems laboratory). For rescue experiments the construct containing the circular CDR1as was injected together with 1.5 nl pre-miR-7 precursor (7 μM, pre-miR miRNA precursor ID PM10047 from Applied Biosystems). The negative control was the vector pCS2+ without insert (empty vector, 50 ng μl^{-1}).

Confocal imaging was performed using a Zeiss LSM 510 microscope (Carl Zeiss MicroImaging) equipped with a 25× objective (NA = 0.8). Embryos were anaesthetized using 0.1% tricaine and mounted in 1% agarose as described[73]. Confocal stacks were acquired of the brain region and processed using Zeiss ZEN software. Midbrain and telencephalon volumes were calculated using Imaris 64×7.6.1 software based on high-resolution three-dimensional stacks obtained from *Tg*(*Xia.Tubb:dsRED*) embryos. Reduced midbrain development was defined as >50% smaller than the mean size of controls.

Each experimental group was evaluated in at least three independent experiments; a minimum of 80 individual embryos per group were examined. Data are expressed as mean ± standard deviation. Statistical analysis was performed using Students'*t*-test, and a $P < 0.05$ was considered statistically significant.

Expression of miR-7 in zebrafish embryos at 48 hours post fertilization was normalized to expression of β-actin. In the miR-7 morpholino group, only embryos with a midbrain phenotype were used for the RNA expression analysis. dre β-actin forward primer, 5′-TGCTGTTTTCCCCTCCATTG-3′; reverse primer, 5′-TTCTGTCCCATGCCAACCA-3′; probe sequence FAM-5′-TGGAC GACCCAGACATCAGGGAGTG-3′-TAMRA.

For measuring the expression of dre-miR-7a/b we used Applied Biosystems TaqMan miR assays (ID000268, ID001088).

52. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9,** 357–359 (2012).
53. Aroian, R. V., Field, C., Pruliere, G., Kenyon, C. & Alberts, B. M. Isolation of actin-associated proteins from *Caenorhabditis elegans* oocytes and their localization in the early embryo. *EMBO J.* **16,** 1541–1549 (1997).
54. L'Hernault, S. W. & Roberts, T. M. Cell biology of nematode sperm. *Methods Cell Biol.* **48,** 273–301 (1995).
55. Stoeckius, M. *et al.* Large-scale sorting of *C. elegans* embryos reveals the dynamics of small RNA expression. *Nature Methods* **6,** 745–751 (2009).
56. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28,** 511–515 (2010).
57. Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34,** D590–D598 (2006).
58. Pollard, K. S. *et al.* Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20,** 110–121 (2010).
59. Blankenberg, D. *et al.* Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.* Ch. 19, Unit 19 10 11-21 (2010).
60. Giardine, B. *et al.* Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* **15,** 1451–1455 (2005).
61. Goecks, J., Nekrutenko, A., Taylor, J. & Galaxy, T. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **11,** R86 (2010).
62. Griffiths-Jones, S. The microRNA Registry. *Nucleic Acids Res.* **32,** D109–D111 (2004).
63. Landthaler, M. *et al.* Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. *RNA* **14,** 2580–2596 (2008).
64. Rudel, S., Flatley, A., Weinmann, L., Kremmer, E. & Meister, G. A multifunctional human Argonaute2-specific monoclonal antibody. *RNA* **14,** 1244–1253 (2008).
65. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
66. Zisoulis, D. G. *et al.* Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*. *Nature Struct. Mol. Biol.* **17,** 173–179 (2010).
67. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
68. Kent, W. J. BLAT–the BLAST-like alignment tool. *Genome Res.* **12,** 656–664 (2002).
69. Goujon, M. *et al.* A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **38,** W695–W699 (2010).
70. Bernhart, S. H. *et al.* Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol. Biol.* **1,** 3 (2006).
71. Ventura, A. *et al.* Cre-lox-regulated conditional RNA interference from transgenes. *Proc. Natl Acad. Sci. USA* **101,** 10380–10385 (2004).
72. Westerfield, M. *The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish (Brachydanio rerio)* 2nd edn (Univ. Oregon Press, 1993).
73. Krueger, J. *et al.* Flt1 acts as a negative regulator of tip cell formation and branching morphogenesis in the zebrafish embryo. *Development* **138,** 2111–2120 (2011).
74. Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W. & Rajewsky, N. miRDeep2 accurately indentifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* **40,** 37–52 (2012).