LETTERS

A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA

Monika Martick[1]*, Lucas H. Horan[1]*, Harry F. Noller[1] & William G. Scott[1]

Structured RNAs embedded in the untranslated regions (UTRs) of messenger RNAs can regulate gene expression. In bacteria, control of a metabolite gene is mediated by the self-cleaving activity of a ribozyme embedded in its 5′ UTR[1]. This discovery has raised the question of whether gene-regulating ribozymes also exist in eukaryotic mRNAs. Here we show that highly active hammerhead ribozymes[2,3] are present in the 3′ UTRs of rodent C-type lectin type II (*Clec2*) genes[4–7]. Using a hammerhead RNA motif search with relaxed delimitation of the non-conserved regions, we detected ribozyme sequences in which the invariant regions, in contrast to the previously identified continuous hammerheads[8–10], occur as two fragments separated by hundreds of nucleotides. Notably, a fragment pair can assemble to form an active hammerhead ribozyme structure between the translation termination and the polyadenylation signals within the 3′ UTR. We demonstrate that this hammerhead structure can self-cleave both *in vitro* and *in vivo*, and is able to reduce protein expression in mouse cells. These results indicate that an unrecognized mechanism of post-transcriptional gene regulation involving association of discontinuous ribozyme sequences within an mRNA may be modulating the expression of several CLEC2 proteins that function in bone remodelling and the immune response of several mammals.

The hammerhead ribozyme is a small, self-cleaving motif composed of a three-helical junction with a core of invariant nucleotides required for activity. To identify hammerhead ribozymes in mammalian mRNAs, we searched mRNA sequence databases using a pattern descriptor that allowed for insertions of up to 5,000 nucleotides at the ends of stem 1 or stem 3 (Supplementary Fig. 1)[11,12]. Three hammerhead ribozymes were identified in the 3′ UTRs of known rodent mRNAs. Two are found embedded in the transcripts of mouse *Clec2d* (osteoclast inhibitory lectin, also known as *Ocil*, *Clr-b* and *Clec2d8*)[13] and its paralogue *Clec2e* (also known as *Clra* and *Clec2d7*)[14], genes that belong to a group of phylogenetically related sequences within the natural killer receptor gene complex of chromosome 6. The third ribozyme is found in rat *CLEC2D11* (ref. 7)—a homologue of mouse *Clec2d*—which resides in the syntenic natural killer receptor gene complex region on chromosome 4. We extended our search to the genomic sequences of other organisms using the UCSC genome browser's comparative genomics tool[15]. Alignments using the natural killer receptor gene complex regions of mouse and rat led to the identification of nine candidate hammerhead ribozymes: four in the 3′ regions of predicted rat, horse and platypus *Clec2*-like genes, and five in the unannotated regions of five other mammalian genomes (Supplementary Fig. 2).

Unlike most known self-cleaving RNA motifs that are contiguous[8–10,16–19], the hammerhead ribozymes identified here (referred to as CLEC2 ribozymes) are split into two fragments separated by a long ribozyme-unrelated insertion in the stem-1-capping loop. The insertion (which is 250 and 696 nucleotides in mouse *Clec2d* and *Clec2e*, respectively, and 145–1,739 nucleotides in other candidate ribozyme sequences) segregates the upstream substrate region, residing within 7–44 nucleotides of the stop codon, from the downstream enzyme fragment (Fig. 1 and Supplementary Fig. 2). When assembled from substrate and enzyme fragments, the secondary structure reflects a characteristic hammerhead ribozyme core of fifteen conserved nucleotides flanked by three helices. Comparison of all twelve CLEC2 ribozymes showed conservation of stem 2 as well as the presence of compensatory mutations maintaining stem 1 and the atypically long stem 3. Stem 1 and stem 2 contain secondary structure elements required to form a tertiary contact known to enhance catalysis greatly (Fig. 1 and Supplementary Fig. 2)[3,20,21]. Six loop and bulge residues (Fig. 1)—necessary for the active structure

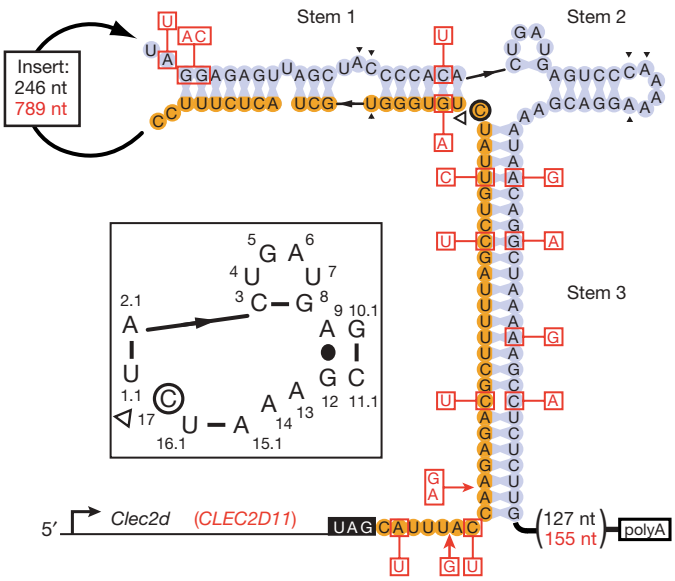


Figure 1 | Sequence arrangement and secondary structure model of the rodent *Clec2d* hammerhead ribozymes. The mouse ribozyme sequence is shown in black and the rat ribozyme sequence length, single nucleotide, and base pair differences are denoted in red. The stop codon is shown in white. The substrate and enzyme sequences are shown on orange and blue backgrounds, respectively. The insertion sequence separating two ribozyme parts is abridged with a thick arrow. The predicted cleavage site (white arrowhead) is 3′ of the active site cytosine (circled). The three-helical junction, composed of conserved (with the exception of 2.1 and 1.1) nucleotides (nt) that make catalytically important interactions, is shown in greater detail (together with canonical numbering scheme) in the inset. The small black arrowheads indicate conserved nucleotides of the catalytically important loop/bulge interactions.

[1]Center for Molecular Biology of RNA, University of California, Santa Cruz, California 95064, USA.
*These authors contributed equally to this work.

stabilization and optimal catalysis in the *Schistosoma mansoni* hammerhead ribozyme[22]—are also found in most CLEC2 ribozyme sequences. Taken together, the secondary structure features suggest that the 3′ UTR-embedded hammerhead ribozyme motifs form catalytically active structures despite their discontinuity.

To test whether the hammerhead ribozymes embedded within the 3′ UTRs possess catalytic activity, we examined the *in vitro* transcription products of mouse 3′ UTRs from *Clec2d* (D-hcu, hammerhead containing UTR) and *Clec2e* (E-hcu). Cleavage of both 3′ UTR RNAs occurs during the course of the transcription reaction (data not shown) and at the predicted hammerhead ribozyme cleavage sites (Fig. 2). The DM1-hcu construct—which is expected to create an inactive hammerhead ribozyme due to a triple mutation that prevents the 'substrate' from being cleaved—yields an intact 3′ UTR transcript, thus confirming that the cleavage is due to the embedded hammerhead ribozyme activity. The mouse *Clec2e*-derived 3′ UTR cleaves to a lesser extent than D-hcu, perhaps because of misfolding effects due to a ~450 nucleotide longer intervening sequence and/or weakened secondary structure in the stem 1 and stem 3 regions (Supplementary Fig. 3). These results demonstrate that the discontinuous 'substrate' and 'enzyme' fragments form a fully active hammerhead ribozyme.

To determine whether the D-hcu sequence can reduce gene expression by forming an active ribozyme *in vivo*, it was incorporated into a dual-luciferase gene expression vector and assayed for activity in NIH 3T3 cells (Fig. 3). D-hcu reduces luciferase expression by 80% when placed downstream from the luciferase reporter (Fig. 3b). Nonspecific downregulation effects were ruled out because inversion of the same sequence (R-D-hcu) restores luciferase expression levels to that of the positive control. To verify that the reduction of gene expression was caused by the embedded ribozyme, mutations specifically targeting hammerhead ribozyme activity were introduced into D-hcu. Both a single transversion mutation at position G8 in the enzyme fragment (DM2-hcu), previously shown to compromise hammerhead ribozyme activity[3], and a triple mutation in the substrate fragment (DM1-hcu) restore gene expression to control levels, indicating that the hammerhead ribozyme embedded in the mouse *Clec2d* 3′ UTR is responsible for reducing reporter protein expression in mouse cells.

RNA from cells transfected with either wild-type (D-hcu) or mutant (DM1-hcu) dual luciferase reporter vectors was compared using reverse transcription–polymerase chain reaction (RT–PCR) to assess the effect of the discontinuous ribozyme-containing 3′ UTR on mRNA integrity. A mixture of two primers, each recognizing one of two reporter genes transcribed from the same vector, was used for first-strand complementary DNA synthesis (Fig. 4a). Amplification of the region spanning the ribozyme cleavage site indicates that

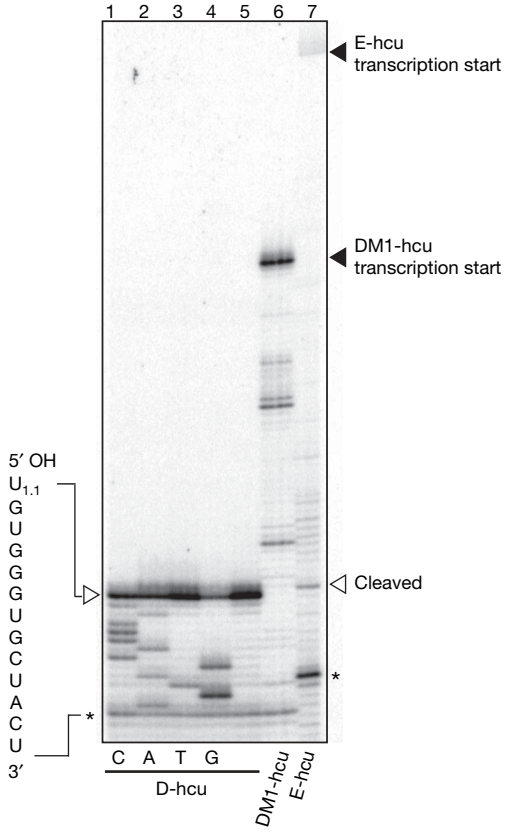


Figure 2 | The discontinuous hammerhead ribozyme-containing 3′ UTRs (hcu) self-cleave *in vitro*. Reverse transcriptase primer extensions of *in vitro* transcribed mouse *Clec2d* (D-hcu, lane 5) and *Clec2e* (E-hcu, lane 7) 3′ UTRs show that both RNAs are cleaved at a single location (open arrowheads). D-hcu cleavage site 5′ of $U_{1.1}$ was identified by dideoxy sequencing of transcription products. Uncleaved E-hcu and triple mutant DM1-hcu are indicated with a black arrowhead. Asterisks indicate non-extended labelled primers.

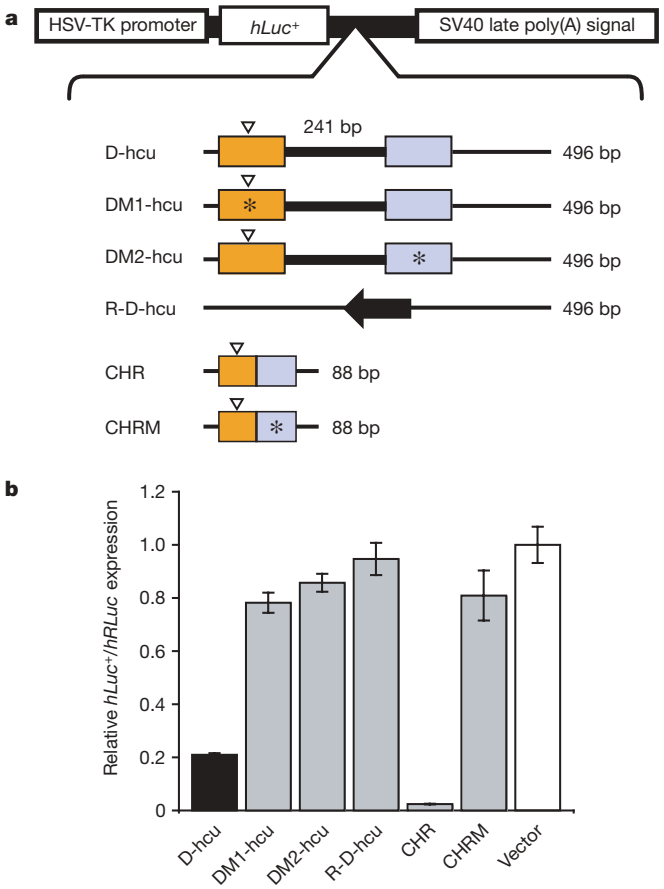
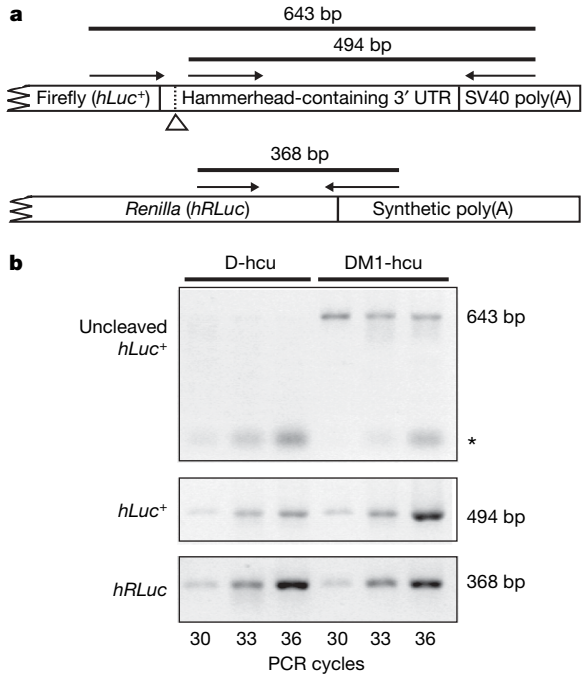


Figure 3 | The discontinuous mouse *Clec2* hammerhead ribozymes embedded in the 3′ UTRs downregulate protein expression *in vivo*. **a**, Layout of the constructs (top) and the schematic of hammerhead ribozyme-containing 3′ UTR (hcu) sequences used for expression. The colours of substrate (orange) and enzyme (blue) regions correspond to those of the structure in Fig. 1 and the insertion is highlighted by a thick line. The cleavage sites are indicated by white arrowheads and asterisks denote the positions of mutations. CHR (control hammerhead ribozyme) and CHRM (G8 to C8 mutant of CHR) are the controls; bp, base pairs. **b**, *In vivo* analysis of firefly luciferase protein expression from different 'hcu' constructs. Black denotes expression from the vector containing wild type 3′ UTR; grey represents mutants and controls; and white indicates a vector without an insert as a negative control. Relative firefly luciferase ($hLuc^+$) expression was determined using a dual luciferase assay using the *Renilla* luciferase (*hRLuc*) expression for normalization. The results shown are means ± s.e.m. of triplicate experiments; the error bars for the vector value are averages of s.e.m. from three assays.

mouse *Clec2d* 3′ UTR self-cleaves via the hammerhead ribozyme activity (Fig. 4b). The uncleaved product of D-hcu is at least tenfold reduced from that of DM1-hcu, in accordance with results of the protein expression assay (Fig. 4c). Previous studies have shown that ribozymes artificially engineered into 3′ UTRs can downregulate gene expression[23] by promoting rapid destruction of the transcript by cytoplasmic RNA degradation machinery[24]. Consistent with these observations, the catalytic activity of mouse *Clec2d* 3′ UTR correlates with the overall reduction of firefly luciferase transcript levels, as compared to those of the mutant UTR-containing mRNA, indicating that the cleaved *Clec2* transcript is degraded *in vivo*.

Structural conservation observed throughout the identified sequences, in conjunction with the tight association of the ribozyme with *Clec2*-like genes, suggests that the discontinuous ribozymes are orthologous to the mouse hammerhead motif characterized here. Within the rodent lineage, the rat *CLEC2D11* ribozyme sequence shares 86% homology with the mouse *Clec2d* ribozyme. Similarly, two predicted rat *Clec2d* paralogues are closely (61% and 67%) related to the mouse *Clec2e* ribozyme (Supplementary Fig. 2). These sequence relationships positively correlate with phylogenetic evidence for multiple duplication events of rodent *Clec2* genes[7]. Further genome searches using separate parts of the hammerhead ribozyme detected four additional mouse *Clec2* paralogues that contain substrate-like, but not enzyme, sequences in their 3′ UTRs (data not shown). These partial motifs are also positioned proximally to the stop codon and share extensive primary sequence identity with the substrate regions of functional ribozymes of either mouse *Clec2d* or *Clec2e*. However, it is not clear if the incomplete ribozyme sequences are functional substrates for the *Clec2d*- and *Clec2e*-derived trans-acting ribozyme elements, or are non-functional evolutionary relics of the rodent *Clec2* gene expansion.



c

RT–PCR product	643 bp	494 bp	368 bp
Intensity ratio D-hcu/DM1-hcu	0.05 ± 0.04	0.57 ± 0.16	1.14 ± 0.09

Figure 4 | RT–PCR analysis of the *in vivo* expression products. a, Schematic of the relative positions of the RT–PCR primers. The hammerhead-specific cleavage site is indicated by an open arrowhead. The 368 base pair (bp) product acts as the internal control for transfection efficiency and the extent of transcription. **b**, RT–PCR of RNA isolated from NIH 3T3 cells that have been transfected with pD-hcu or pDM1-hcu. The asterisk indicates nonspecific PCR products. **c**, Quantification of the difference in PCR product intensity between D-hcu and DM1-hcu. The product intensity was used to determine the ratio of D-hcu/DM1-hcu at each listed cycle. Each value represents the mean ± s.d. of intensity ratios from three listed cycles.

Identification of CLEC2 ribozymes in the genomes of mammalian species other than rodents (notably platypus) indicates that these ribozymes share an early mammalian ancestor. Active hammerhead motifs have been found in several divergent species, such as newts[10], plants[8] and viroids[10]. However, only the *Schistosoma* satellite DNA is known to encode hammerhead ribozyme sequences that have significant homology to the CLEC2 ribozymes[3,18]. Although schistosomes parasitize rodents, their ribozymes are found in lineages more modern than the time of divergence of mouse and rat species[18,25], suggesting a horizontal gene transfer of the hammerhead sequence from host to the parasite genome as the probable scenario.

In eukaryotes, post-transcriptional regulation of gene expression through the 3′ UTR involves an interaction between mRNA-stabilizing and -destabilizing protein-based processes[26]. Together with evidence of cross-species secondary structure conservation, our results indicate that the CLEC2 hammerhead ribozymes have roles analogous to those of destabilizing protein factors. Paired with a yet undiscovered antagonistic process, the hammerhead motif may constitute a system of rapid regulation for CLEC2d, a transmembrane C-type lectin that has a dual role: the inhibition of natural killer (NK)-cell-mediated lysis through the interaction with the NK cell receptor NKR-P1, and the inhibition of osteoclast formation[4,6].

The core sequences of CLEC2 ribozymes resemble those of the smaller members of its ribozyme class, but the discontinuous arrangement of the substrate and enzyme regions exhibit properties previously associated only with catalytic introns. Although functioning through a fundamentally different mechanism and residing in transcripts distinct from processed mRNA, catalytic sequences of some self-splicing group I and II introns span long stretches of RNA[19] on a scale analogous to that of the ribozymes described here. A question remains as to whether CLEC2 ribozymes can also process substrates encoded on detached transcripts similarly to RNase P, which possesses trans-cleaving properties unique among naturally occurring ribozymes[19]. Most self-cleaving RNAs, however, are considered to be compact sequences suitable for single-step functions such as multimer-to-monomer conversion of genomes in the course of viroid and satellite RNA replication, or for response to small molecule binding such as in metabolite-dependent gene regulation in bacteria. However, in the context of complex gene regulatory systems such as those of mammalian cells, the discontinuous configuration of the CLEC2 ribozyme sequences may provide an opportunity for allosteric regulation.

METHODS SUMMARY

mRNA sequence data from the Ensembl database (release 48, http://www.ensembl.org) were searched using the RNABOB program[11,12] with a descriptor (Supplementary Fig. 1) for the hammerhead ribozyme motif. Orthologues of the rodent CLEC2 ribozymes were identified using BLAT and the UCSC Genome Browser (http://genome.ucsc.edu/)[15,27]. For *in vitro* RNA synthesis and analysis, the *Clec2* 3′ UTR-containing plasmids were linearized and transcribed using T7 RNA polymerase at a pH of 8.0 and in the presence of 28 mM $MgCl_2$. Reverse transcriptase primer extensions and RNA sequencing were adapted for a [^{32}P]-end-labelled primer from previously published methods[28]. For the expression assays, the NIH 3T3 mouse fibroblast cells were transiently transfected with *Clec2* 3′ UTR-containing psiCHECK-2 reporter plasmids (Promega) and analysed using the Dual-Glo Luciferase Assay System (Promega). Firefly reporter expression was normalized using expression of the *Renilla* luciferase encoded on the same plasmid. RNA from the transfected cells was purified using Trizol (Sigma). For RT–PCR, cDNA was synthesized using primers specific to plasmid-derived transcripts using previous methods[28]. PCR was performed with Accuprime Taq (Invitrogen). PCR products were resolved on an agarose gel and the product quantification was performed using ImageQuant 5.2 (Molecular Dynamics).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 January; accepted 30 May 2008.
Published online 9 July 2008.

1. Winkler, W. C. *et al.* Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428,** 281–286 (2004).
2. Khvorova, A., Lescoute, A., Westhof, E. & Jayasena, S. D. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nature Struct. Biol.* **10,** 708–712 (2003).
3. Martick, M. & Scott, W. G. Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell* **126,** 309–320 (2006).
4. Zhou, H. *et al.* Osteoclast inhibitory lectin, a family of new osteoclast inhibitors. *J. Biol. Chem.* **277,** 48808–48815 (2002).
5. Plougastel, B., Dubbelde, C. & Yokoyama, W. M. Cloning of *Clr*, a new family of lectin-like genes localized between mouse *Nkrp1a* and *Cd69*. *Immunogenetics* **53,** 209–214 (2001).
6. Carlyle, J. R. *et al.* Missing self-recognition of Ocil/Clr-b by inhibitory NKR-P1 natural killer cell receptors. *Proc. Natl Acad. Sci. USA* **101,** 3527–3532 (2004).
7. Hao, L., Klein, J. & Nei, M. Heterogeneous but conserved natural killer receptor gene complexes in four major orders of mammals. *Proc. Natl Acad. Sci. USA* **103,** 3192–3197 (2006).
8. Przybilski, R. *et al.* Functional hammerhead ribozymes naturally encoded in the genome of *Arabidopsis thaliana*. *Plant Cell* **17,** 1877–1885 (2005).
9. Rojas, A. A. *et al.* Hammerhead-mediated processing of satellite pDo500 family transcripts from Dolichopoda cave crickets. *Nucleic Acids Res.* **28,** 4037–4043 (2000).
10. Flores, R. *et al.* Hammerhead ribozyme structure and function in plant RNA replication. *Methods Enzymol.* **341,** 540–552 (2001).
11. Gautheret, D., Major, F. & Cedergren, R. Pattern searching/alignment with RNA primary and secondary structures: an effective descriptor for tRNA. *Comput. Appl. Biosci.* **6,** 325–331 (1990).
12. Eddy, S. R. RNABOB: a program to search for RNA secondary structure motifs in sequence databases. ⟨http://selab.janelia.org/software.html⟩ (2008).
13. Zhou, H. *et al.* A novel osteoblast-derived C-type lectin that inhibits osteoclast formation. *J. Biol. Chem.* **276,** 14916–14923 (2001).
14. Carlyle, J. R. *et al.* Molecular and genetic basis for strain-dependent NK1.1 alloreactivity of mouse NK cells. *J. Immunol.* **176,** 7511–7524 (2006).
15. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12,** 996–1006 (2002).
16. Salehi-Ashtiani, K., Luptak, A., Litovchick, A. & Szostak, J. W. A genomewide search for ribozymes reveals an HDV-like sequence in the human *CPEB3* gene. *Science* **313,** 1788–1792 (2006).
17. Teixeira, A. *et al.* Autocatalytic RNA cleavage in the human β-globin pre-mRNA promotes transcription termination. *Nature* **432,** 526–530 (2004).
18. Ferbeyre, G., Smith, J. M. & Cedergren, R. Schistosome satellite DNA encodes active hammerhead ribozymes. *Mol. Cell. Biol.* **18,** 3880–3888 (1998).
19. Serganov, A. & Patel, D. J. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nature Rev. Genet.* **8,** 776–790 (2007).
20. Canny, M. D. *et al.* Fast cleavage kinetics of a natural hammerhead ribozyme. *J. Am. Chem. Soc.* **126,** 10848–10849 (2004).
21. Osborne, E. M., Schaak, J. E. & Derose, V. J. Characterization of a native hammerhead ribozyme derived from schistosomes. *RNA* **11,** 187–196 (2005).
22. Martick, M. *Structural Study of how the Hammerhead Ribozyme Solves the Problem of Catalysis*. PhD thesis, Univ. California, ⟨http://www.proquest.com/⟩ (2007).
23. Yen, L. *et al.* Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. *Nature* **431,** 471–476 (2004).
24. Meaux, S. & Van Hoof, A. Yeast transcripts cleaved by an internal ribozyme provide new insight into the role of the cap and poly(A) tail in translation and mRNA decay. *RNA* **12,** 1323–1337 (2006).
25. Lockyer, A. E. *et al.* The phylogeny of the Schistosomatidae based on three genes with emphasis on the interrelationships of Schistosoma Weinland, 1858. *Parasitology* **126,** 203–224 (2003).
26. Garneau, N. L., Wilusz, J. & Wilusz, C. J. The highways and byways of mRNA decay. *Nature Rev. Mol. Cell Biol.* **8,** 113–126 (2007).
27. Kent, W. J. BLAT–the BLAST-like alignment tool. *Genome Res.* **12,** 656–664 (2002).
28. Stern, S., Moazed, D. & Noller, H. F. Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension. *Methods Enzymol.* **164,** 481–489 (1988).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Hall and K. Chakrabarti for assistance with cell culture, members of Haussler and Ares laboratories for sharing tissue culture space, M. Robertson for the transcriptase and A. Zahler for discussion. This work was supported by the National Institutes of Health grant R01043393.

Author Contributions L.H.H. and M.M. did the sequence searches, designed the study, performed the experiments, analysed the data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.M. (mmartick@yahoo.com).

METHODS

Motif search and secondary structure. The RNABOB program[11,12] was used to search mRNA sequence data from the Ensembl database (release 48, http://www.ensembl.org) using a descriptor (Supplementary Fig. 1) for the hammerhead ribozyme motif. The secondary structure was built using information from the *Schistosoma* hammerhead ribozyme crystal structure[3]. Mouse and rat *Clec2* paralogues encoding partial and full ribozyme sequences were identified using BLAT and the UCSC Genome browser[15,27]. Alignment was generated using ClustalW[29]. *Clec2d* orthologues were identified using the 30-way Multiz alignment and conservation tool in the UCSC genome browser[15]. Sequences with similarity to stem 2 regions of the rodent CLEC2 ribozymes were selected for further analysis. These sequences were first screened for completeness of stem 2 and conserved hammerhead ribozyme nucleotides in the enzyme strand. The candidate ribozymes were then selected on the basis of the expected sequence (the ability to form stems 1 and 3 together with the enzyme strand) and orientation of the substrate within the 2,000 bp regions flanking stem 2.

Constructs. The hammerhead ribozyme-containing 3′ UTR sequences (GenBank AF321553 nucleotides 684–1,179 for D-hcu; NM_153506 nucleotides 695–1,710 for E-hcu) were cloned from total DNA (BioChain Institute, Inc.) of BALB/c mice. The DM1-hcu construct is the same as D-hcu with the following exception: the hammerhead ribozyme has been inactivated by a triple mutation spanning the cleavage site (5′-A(16.1)G(17)A(1.1)-3′). The DM2-hcu construct contains the hammerhead ribozyme inactivated by a single base (G8 to C8) mutation (refer to Fig. 1 for residue identification), with the rest of the sequence identical to D-hcu. The R-D-hcu construct contains D-hcu sequence in the antisense direction. The positive controls CHR (control hammerhead ribozyme) and CHRM (G8 to C8 mutant of CHR) were derived from *Schistosoma* ribozyme[3] and were oriented analogously to the CLEC2 ribozymes—with stem 3 connecting the structure to the rest of the transcript. Each of the two controls was capped with an artificial tetraloop closing stem 1 (full CHR sequence with tetraloop underlined and the CHRM mutation position preceded by an asterisk: 5′-CCGGCGT CCTGGTATCCAATCCTTCGGGATGTACTACCAGCTGAT*GAGTCCCAAA TAGGACGAAACGCCGG-3′). For *in vitro* transcription assays, 3′ UTR sequences were cloned into the XbaI site of the pET-20(b) vector (Novagen). For expression assays, 3′ UTR sequences and control ribozymes were placed 7 nucleotides downstream of the firefly luciferase open reading frame (XbaI site) in psiCHECK-2 dual luciferase vector (Promega). Cognate poly A sites have been omitted from all 3′ UTR sequences and replaced with vector-encoded SV40 Poly(A) signal. The psiCHECK-2 vector also contained a *Renilla* luciferase gene, the expression product of which was used as an internal control for transfection efficiency and the extent of expression.

***In vitro* transcription and reverse transcriptase primer extension.** pET-20(b) constructs were linearized and transcribed for 1 h at 37 °C using T7 RNAP in the following conditions: 50–60 μM double-stranded plasmid, 5 mM ribonucleotide triphosphates, 28 mM $MgCl_2$, 2 mM spermidine-HCl, 100 mM dithiothreitol, 0.01% Triton X-100 and 40 mM Tris-HCl buffer (pH 8.0). Transcripts were purified using phenol and chloroform after the transcription reaction and then again after the treatment with RQ-DNase (Promega). Reverse transcriptase primer extensions and RNA sequencing were adapted for a [^{32}P]-end-labelled primer from previously published methods[28].

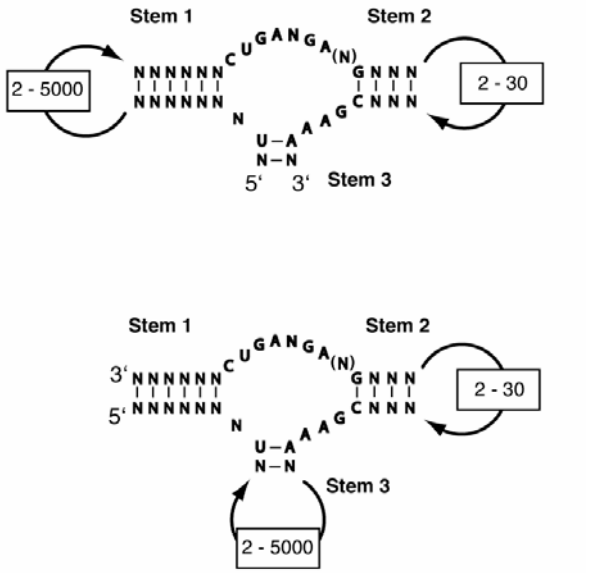
Reporter expression assays. NIH 3T3 mouse fibroblast cells were transiently transfected using Lipofectamine LTX (Invitrogen) according to the manufacturer's instructions and assayed using the Dual-Glo Luciferase Assay System (Promega). The firefly/*Renilla* luciferase expression ratios were normalized to those of the psiCHECK-2 vector.

RT–PCR. NIH 3T3 cells were transfected with mouse *Clec2d* 3′ UTR-containing psiCHECK-2 plasmids in 10-cm plates. Whole-cell RNA was purified using Trizol (Sigma), treated with RQ-DNase (Promega), and purified with phenol and chloroform. cDNA was synthesized using two primers specific to two different (experimental and control) plasmid-derived transcripts using previous methods[28]. Primer sequences for cDNA synthesis were 5′-CATGTCTGCTCGAAGCGGC-3' for the ribozyme-containing firefly luciferase (*hLuc*$^+$) transcript, and 5′-TCCTCAGGCTCCAGTTTCC-3' for the control *Renilla* luciferase (*hRLuc*) transcript. Primer extension products were treated with RNase A and RNase H, purified with phenol and chloroform, and resuspended in water. For PCR reactions, Accuprime Taq along with Buffer II (Invitrogen) were used according to manufacturer's instructions. For PCR of the ribozyme-containing *hLuc*$^+$, the *hLuc*$^+$ cDNA primer was used together with 5′-GTGGACGAGGTGCCCAAGG-3′ (forward) for the 643 nucleotide product, and 5′-GCTACTCTTTCCCCTATGGTC-3′ (forward) for the 494 nucleotide product. For PCR of the control *hRLuc*, the *hRLuc* cDNA primer was used together with 5′-CTGATCTGATCGGAATGGG-3′ (forward). PCR reactions were placed on ice after the designated number of cycles, resolved on a 2% agarose gel and stained with ethidium bromide. PCR product intensities were quantified using ImageQuant 5.2 (Molecular Dynamics).

29. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).

Supplementary Figure 1

a



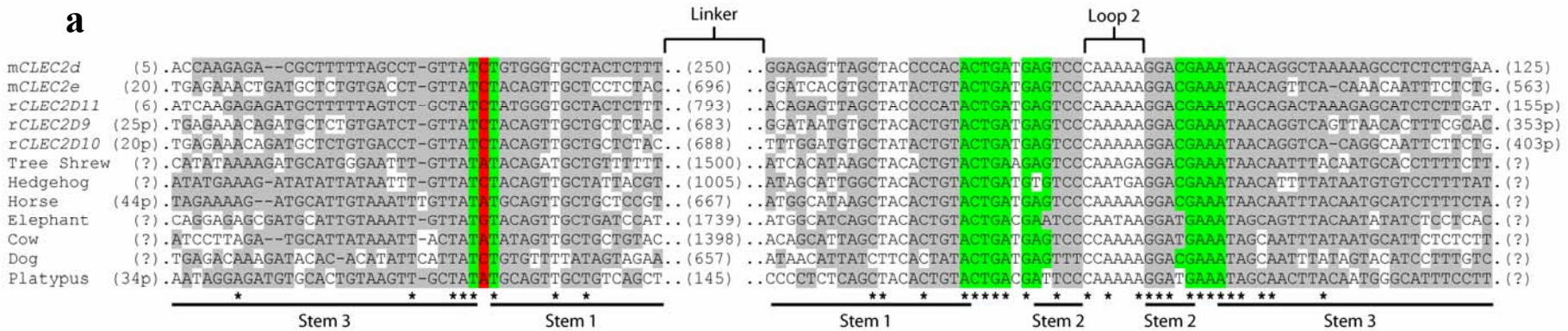
b

```
h1 s1 h2 s2 h2' s3 s4 h3 h4 s5 h4' h3' s6 h1'
h1 0:0 NT:AN
s1 0 N
h2 0:0 NNNNNN:NNNNNN
s2 0 N[4998]N
s3 0 CTGANGA
s4 0 *
h3 0:0 G:C
h4 0:0 NNN:NNN
s5 0 N[28]N
s6 0 GAA
```

Supplementary Figure 1. Descriptor for the hammerhead ribozyme motif used in RNABOB program. (A) Graphic representation of the descriptors for the hammerhead ribozymes embedded through Stem 3 (top) and Stem 1 (bottom). (B) The RNABOB descriptor used to search for hammerhead ribozymes embedded through Stem 3. Similar descriptor was used for the ribozyme embedded through Stem 1.

Supplementary Figure 2

a



b

m*CLEC2d* Mouse (*Mus musculus*)
>chromosome:NCBIM37:6:129136035:129136400:1
Genbank: AF321553

m*CLEC2e* Mouse (*Mus musculus*)
>chromosome:NCBIM37:6:129042600:129043403:-1
RefSeq: NM_153506

r*CLEC2D11* Rat (*Rattus norvegicus*)
>chromosome:RGSC3.4:4:166425123:166426033:1
RefSeq: NM_001048075

r*CLEC2D9* Rat (*Rattus norvegicus*)
>chromosome:RGSC3.4:4:166290554:166291351:-1
(*Predicted*) **Ensembl Gene**: ENSRNOG00000026682

r*CLEC2D10* Rat (*Rattus norvegicus*)
>chromosome:RGSC3.4:4:166331348:166332163:-1
(*Predicted*) **Ensembl Gene**: ENSRNOG00000037073

Tree Shrew *(Tupaia belangeri)*
>scaffold:TREESHREW:scaffold_9056:3264:4881:-1

Hedgehog *(Erinaceus europaeus)*
>scaffold:HEDGEHOG:scaffold_303363:11370:12491:1

Horse *(Equus caballus)*
>chromosome:EquCab2:6:36560178:36560961:1
(*Predicted*) **Ensembl Gene:** ENSECAG00000019259

Elephant *(Loxodonta africana)*
>scaffold:BROADE1:scaffold_60948:6692:8548:-1

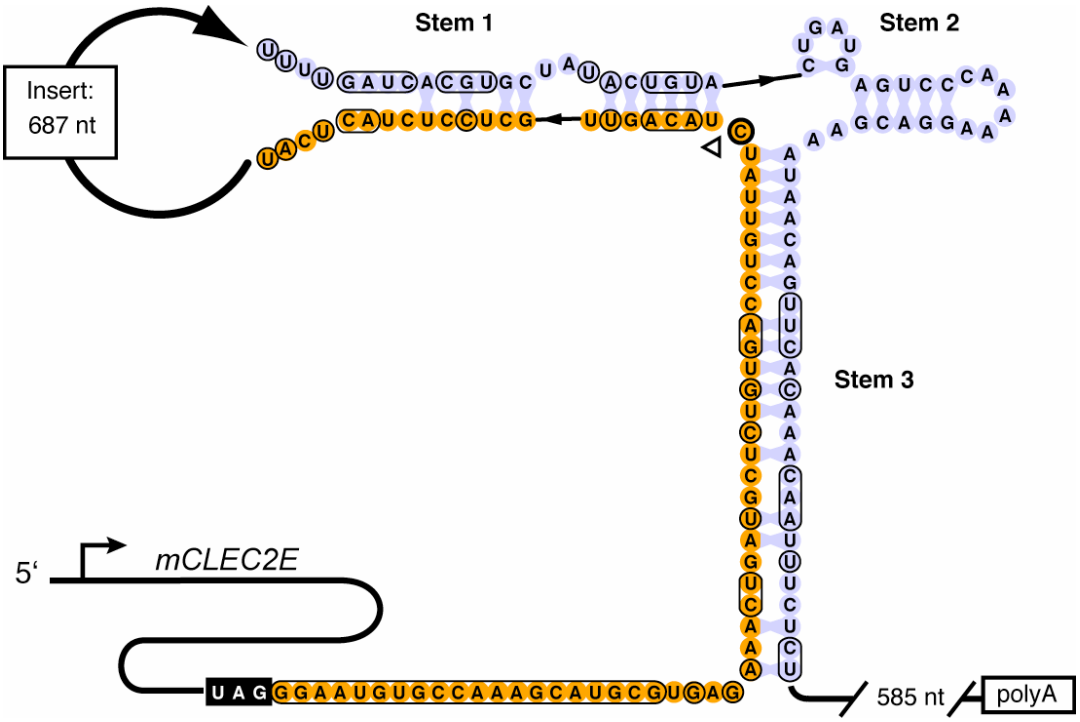
Cow *(Bos taurus)*
>chromosome:Btau_3.1:5:99619543:99621056:1

Dog *(Canis familiaris)*
>chromosome:BROADD2:27:39360384:39361158:-1

Platypus *(Ornithorhynchus anatinus)*
>supercontig:OANA5:Contig10595:5483:5745:-1
(*Predicted*) **Ensembl Gene:** ENSOANG00000006510

Supplementary Figure 2. Comparison of the CLEC2 hammerhead ribozyme sequences. (a) Alignments of verified and predicted CLEC2 hammerhead ribozyme sequences. The sequences of the substrate and enzyme regions were aligned using ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/index.html)[29]. The remainders of 3' UTRs are denoted as length of sequence in parentheses, with lower case *p* indicating the distance to predicted stop codon and Poly(A) signal. For reference, the active site cytosine is highlighted in red, and other conserved catalytic core nucleotides are highlighted in green. Residues predicted to form base pairs in the double helices (including G-U pairs) are highlighted in gray and correspond to stems indicated in the labels below the alignment. Asterisks mark nucleotides that are identical in all sequences. (b) References for hammerhead ribozyme sequences aligned in (a). Ensembl genomic locations (release 49, http://www.ensembl.org) of the aligned hammerhead ribozymes and IDs of the associated genes are noted below the respective species' name.

Supplementary Figure 3



Supplementary Figure 3. Sequence arrangement and secondary structure model of mouse *CLEC2e* hammerhead ribozyme. The ribozyme (black letters) is embedded in the 3' UTR of mouse *CLEC2e* gene (stop codon denoted in white). The secondary structure was built using information from the *Schistosoma* hammerhead ribozyme crystal structure[3]. The "substrate" and "enzyme" sequences are shown on orange and blue backgrounds, respectively. The insertion sequence separating two ribozyme parts is abridged with a thick arrow and the actual insertion length is indicated in the box. The predicted cleavage site (indicated by a white arrow) is the phosphate linkage just 3' of the active site cytosine (circled by thick line). Differences in residues from mouse *CLEC2d* ribozyme are outlined by thin black line.