Genome-Wide Dual-Selection Unveils Novel Self-Cleaving Ribozymes in the Human Genome

Zhe Zhang[1], Jian Zhan[1,2,3*] and Yaoqi Zhou[1*]

[1]Institute for Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen 518107, China
[2]Ribopeutic (Shenzhen) Co., Ltd., Futian, Shenzhen, Guangdong Province, 518000, China
[3]Ribopeutic Inc., Qiantang, Hangzhou, Zhejiang Province, 310018, China
[*] Co-corresponding authors (zhanjian@szbl.ac.cn and zhouyq@szbl.ac.cn)

Abstract

The landscape of catalytic RNAs in complex eukaryotes remains poorly charted. Although self-cleaving ribozymes are widespread in microbial and viral genomes, their existence and functional roles in humans remain largely unknown. Here, we introduce a generalizable, genome-wide discovery platform that integrates high-throughput signals from two complementary adapter ligation assays—3P-seq and 5OH-seq—to specifically capture RNA fragments bearing cleavage signatures (2′,3′-cyclic phosphate/3′-phosphate and 5′-hydroxyl termini). By applying a dedicated computational scoring algorithm to human genomic data, we systematically identified four previously unrecognized self-cleaving ribozymes. These ribozymes localize to diverse genomic features: an exon of WDFY1, an intron of PLD5 embedded within a repetitive element, an LTR retrotransposon, and the antisense strand of an SYNJ2BP intron. Truncation analyses and mutational profiling defined minimal functional cores of 21–55 nucleotides, and experiment-assisted secondary structure prediction suggests that they adopt simple structural architectures. Despite their short sequences and simple architectures, we demonstrated that these ribozymes possess robust cleavage activity under multiple biochemical conditions. Most importantly, our experimental results indicate that three of them exhibit significant self-cleavage activity within the cellular environment. Our work establishes a powerful strategy for genome-wide mining of self-cleaving ribozymes, expands the catalogue of human ribozymes from 6 to 10, and reveals new layers of functional complexity in the human RNA world.

Main

Self-cleaving ribozymes are an important class of non-coding RNAs with catalytic functions[1] that catalyze the site-specific self-cleavage of phosphodiester backbone[2]. Recent advancements in computational and high-throughput techniques have significantly accelerated the pace of ribozyme discovery[3,4]. However, the vast majority of self-cleaving ribozymes identified to date originate from lower organisms. Only 6 self-cleavage ribozymes are found in human so far (hammerhead ribozymes (HH9, HH10) located with introns[5], the HDV-like cytoplasmic polyadenylation element-binding protein 3 (CPEB3) ribozyme[6], 2 twister sister-like (TS-like) ribozymes[7] (LINE-1 and OR4K15), and the hovlinc ribozyme found in very long intergenic non-coding RNAs (vlincRNAs)[8]. Although these self-cleaving ribozymes lack structural similarity, they all accomplish site-specific cleavage through a nucleophilic attack by a 2′-oxygen on the adjacent phosphodiester bond. This reaction produces cleavage termini characterized by a 2′,3′-cyclic phosphate group and a 5′-hydroxyl group, respectively.

Previously, the Kapranov group utilized the RppH 5′ pyrophosphohydrolase and the XRN-1 5′-3′ exoribonuclease to degrade the other transcriptional products, except for the RNA fragments containing the 5′-hydroxyl[8]. This led to the discovery of the hovlinc ribozyme among 28 candidates. Although the low discovery rate of self-cleaving ribozymes suggests the rarity in complex organisms, the question of whether other self-cleaving ribozymes exist in the human genome merits further study.

To probe the human genome for potential self-cleaving ribozymes, we developed a strategy that leverages the unique specificity of RtcB ligase, an enzyme that joins a 5′-hydroxyl terminus to a 3′-phosphate or 2′,3′-cyclic phosphate terminus[9]. As this type of terminus are uniquely generated by self-cleavage, we can capture RNA molecules resulted from self-cleavage by utilizing RtcB ligase in conjunction with specifically designed adapter RNAs. Among 96 candidate sequences with possible cleavage products captured, six were confirmed to exhibit self-cleaving activity. Two of these active sequences are known self-cleaving ribozymes, whereas the other four represent novel ones. In particular, one self-cleaving ribozyme was identified within an exonic region of the WDFY1 (WD repeat and FYVE domain containing 1) protein, suggesting that it may play a direct role in regulating protein expression. To the best of our knowledge, this is the first report of a self-cleaving ribozyme residing in a protein-coding exon, and thus, expands our understanding of the functional repertoire and mechanistic diversity of self-cleaving ribozymes.

Results

Genome-wide discovery of self-cleaving ribozymes. Genomic libraries were prepared using the Illumina tagmentation kit, which employs a Tn5 transposase to simultaneously fragment and tagmate DNA, streamlining the library construction process. Briefly, the tagmentation reaction fragmented genomic DNA and ligated Nextera s7 and Nextera s5 adapter sequences to both ends of the fragments in a single step. Then, the termini of the genomic library fragments were appended with T7 promoter through PCR amplification, facilitating downstream in vitro transcription (Fig. 1a). These RNA transcripts along with self-cleavage products obtained from the genomic library offers the distinct advantage of being unaffected by intracellular proteases and RNases. Unlike prior methods[8] that solely enriched for 5′-hydroxyl ends from self-cleavage, our approach specifically captures cleavage products containing either 2′,3′-cyclic phosphate or 5′-hydroxyl termini from self-cleavage. The 3P-seq is designed to isolate RNA fragments possessing 2′,3′-cyclic phosphate or 3′ -phosphate ends. These 3′-ends were joined with a synthetic RNA linker that features a 5′-hydroxyl group by taking advantage of the unique ligation activity of RtcB ligase. Meanwhile, the 5OH-seq captures RNAs with 5′-hydroxyl termini by employing RtcB ligates these 5′-hydroxyl ends to a separate RNA linker that itself carries a 3′-phosphate group. A desthiobiotin modification was strategically engineered at the distal ends of both RNA linkers with a dual purpose: first, to minimize the formation of off-target ligation products, and second, to enable highly specific capture and enrichment by using streptavidin-based purification methods. The enriched products were subsequently processed into sequencing libraries for high-throughput analysis.

To minimize false positives, we employed paired-end sequencing and implemented a stringent quality filter based on the Mapping Quality (MAPQ) score from the SAM files, retaining only alignments with a MAPQ $\geqslant$ 30, which is widely regarded as a threshold for high-confidence mappings. The precise cleavage sites were inferred from the alignment boundaries, corresponding to the junctions between the inserted fragments and the adapter sequences. The distribution of these putative cleavage sites was then quantified by counting the supporting reads. During data processing, we observed that a substantial number of cleavage sites remained clustered within specific genomic regions. Given that self-cleaving ribozymes typically catalyze hydrolysis at unique sites, we rationalized that sites supported by an exceptionally low number of reads within these clusters were likely artifacts, potentially resulting from random RNA degradation. Consequently, a sophisticated denoising step was employed, which eliminated sites based on an intra-window signal intensity ratio. Furthermore, to enhance the reliability of our findings, we integrated the results from two independent experiments—3P-seq and 5OH-seq. A composite scoring system was designed to prioritize cleavage sites detected in both assays, under the premise that such "consensus" sites exhibit a lower probability of being false positives.

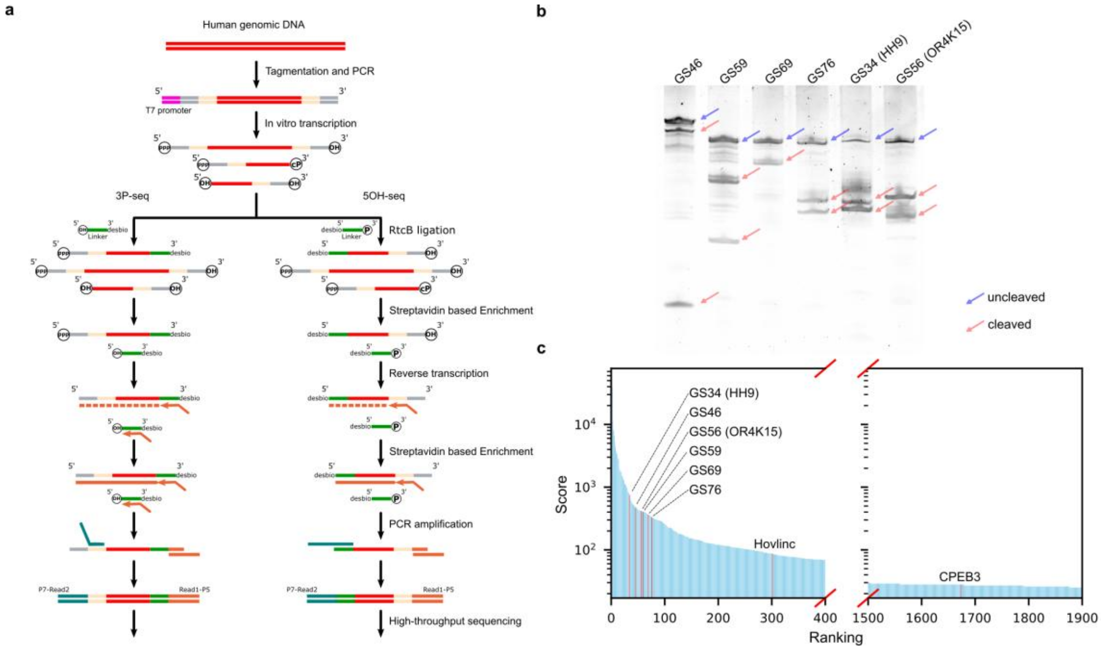


Fig. 1: Genome-scale discovery of self-cleaving ribozymes in human genome.

a, The discovery flowchart: To profile RNA cleavage products, we generated a protease/RNase-free genomic RNA library through in vitro transcription. RtcB ligase and terminally modified RNA linkers were utilized to selectively capture and distinguish fragments bearing 2′,3′-cyclic phosphate or 5′-OH ends, followed by streptavidin bead enrichment and sequencing library construction. **b**, Urea PAGE analysis of the six experimentally validated ribozyme candidates from the top 96 hits. Hit nomenclature reflects its ranking (e.g., GS46 corresponds to the hit ranked 46th). **c**, Scoring distribution of the top 1,900 hits. Hits corresponding to known ribozymes and novel ribozymes that have been validated through experiments are explicitly marked in red.

Self-cleaving ribozymes identified in HeLa genome. Applying this integrated approach, we successfully identified four out of the five known self-cleaving ribozymes, with HH9 and OR4K15 ranking within the top 96 hits (Fig. 1b). However, the hovlinc ribozyme, discovered more recently, was ranked 301st (Fig. 1c), primarily due to its low detection count (2 reads) in the 5OH-seq assay. The CPEB3 ribozyme was ranked 1,673rd (Fig. 1c) and was not detected in the 5′OH-seq assay.

To gain some understanding of low priority in hovlinc and CPEB3 ribozymes, we generated normalized coverage files from the BAM files and visualized the signal intensity across the targeted gene regions by using the deepTools suite[10]. Analysis of the signal profiles across identified ribozymes shows a spectrum: from those with a pronounced signal asymmetry (Extended Data Figs. 1a, 1c, 1f, 1g and 1h) to those with relatively balanced signals at both ends (Extended Data Figs. 1b, 1d and 1e). For hovlinc and CPEB3 ribozymes, their characteristically stable 3′ terminal structures may have impeded efficient adapter ligation, resulting in negligible 3′ end capture signal and thus a substantially lower ranking in our analysis. These inherent structural differences in self-cleavage ribozymes are a major source of the significant variation in observed sequencing depth, highlighting a specific limitation of the current ranking method in detecting self-cleaving ribozymes with certain structural features.

New ribozymes with potential novel functions. We characterized four novel self-cleaving ribozymes (GS46, GS59, GS69, GS76), each occupying a distinct genomic locus that suggests unique regulatory potentials. Particularly notable is GS46, which was located within the eighth exon of the protein-coding gene WDFY1 (WD repeat and FYVE domain containing 1) (Extended Data Fig. 2a); its self-cleavage activity is thus supposed to truncate the host mRNA, potentially triggering nonsense-mediated decay and leading to reduced expression of the WDFY1 protein. Furthermore, GS59 and GS69 were identified within repetitive genomic architectures: GS59 was situated in the first intron of PLD5 (Phospholipase D Family Member 5) (Extended Data Fig. 2b) and embedded within a DNA repeat element, while GS69 was found within an intron of a novel transcript (Extended Data Fig. 2c) that itself resides in a Long Terminal Repeat (LTR) retrotransposon. Additionally, GS76 was mapped to the reverse (antisense) strand of the first intron of the SYNJ2BP (Synaptojanin 2 Binding Protein) gene (Extended Data Fig. 2d), an intriguing location that raises the possibility of its involvement in regulating the sense transcript.

To elucidate the structural basis of the self-cleaving activity, we conducted a preliminary investigation of the four newly identified ribozymes. It is important to note that the RNA fragments captured in our sequencing libraries do not necessarily represent the minimal structural module required for catalytic activity as we demonstrated previously[7]. Therefore, to delineate the core functional domain of each ribozyme, we systematically generated a series of truncated variants. Given that the 3P-seq assay yielded stronger signals for most of these ribozymes (Extended Data Fig. 1), indicating a more defined or stable 3′ end post-cleavage, the truncation strategy primarily involved progressive shortening from the 3′ terminus. The catalytic competence of these truncated constructs was then assessed by analyzing their self-cleavage efficiency via polyacrylamide gel electrophoresis (PAGE). This empirical approach allowed us to identify the minimal functional core for each ribozyme, with optimized lengths of approximately 70 nt, 100 nt, 80 nt, and 80 nt, respectively (Figs. 2a-d).

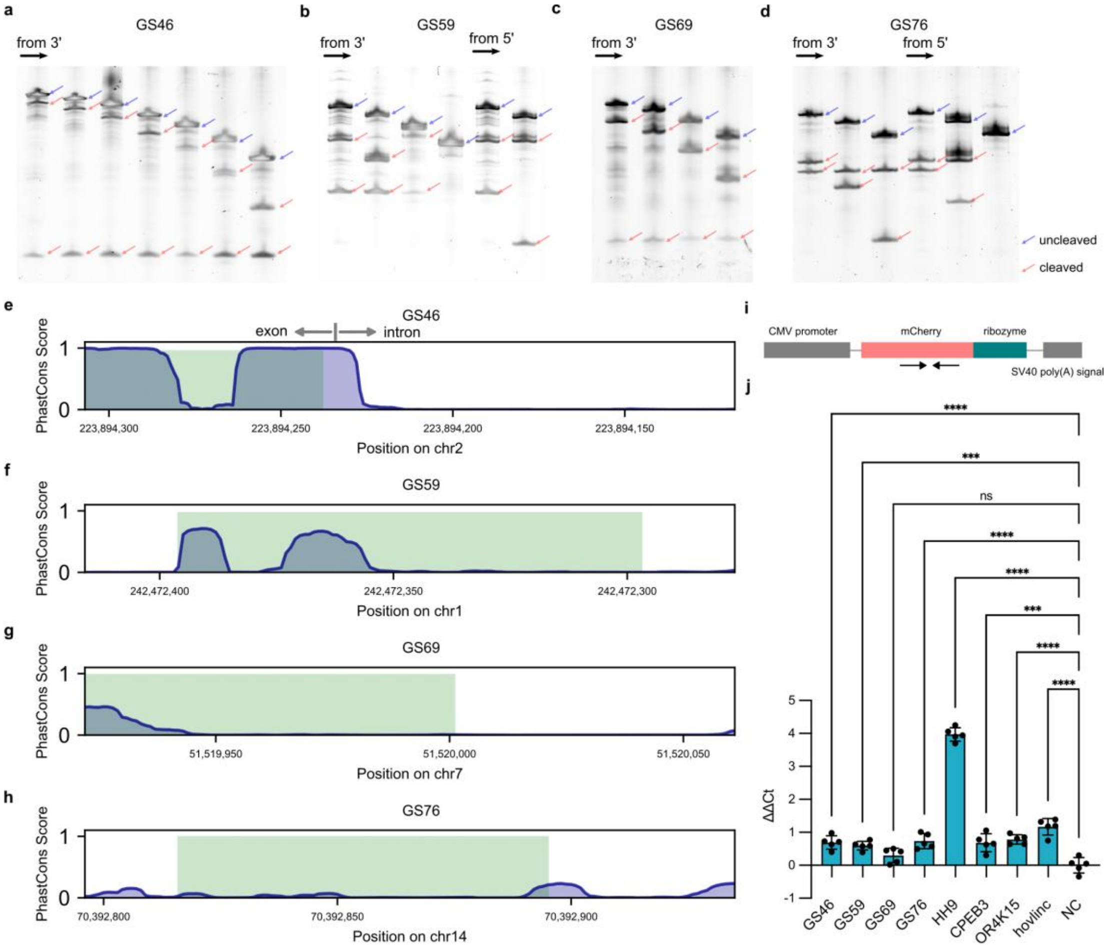


Fig. 2: Truncations and in vivo activities of the 4 new ribozymes.

a-d, PAGE analysis of the truncated variants of the 4 new ribozymes. A series of truncated variants was generated from the original sequence for the activity assay. The 3′ end (or 5′ end) was anchored, and sequential 20-nt deletions were made from the opposite terminus. This process continued until the whole fragment reached a minimum length of 100 nt, which was determined to be optimal for efficient molecular cloning. **e-h**, PhastCons score distribution of 4 ribozymes from the multiple alignment of 30 species. Higher scores (closer to 1) indicate stronger evolutionary constraint, suggesting the region is likely under negative selection and may be functionally important. Lower scores (closer to 0) suggest neutral evolution or lack of constraint. The green shading denotes the functionally essential region, as determined by sequential truncation analysis. **i**, Schematic diagram of the plasmid used for in vivo activity assay. Self-cleavage of the downstream ribozyme is expected to reduce the stability of mCherry mRNA. Thus, the in vivo activity of the ribozyme could be monitored by quantifying the mCherry mRNA level. **j**, Expression levels of mCherry containing different inserted ribozyme sequences. The y-axis shows the ΔΔCt values relative to the mean value of the designated control group (NC, empty vector). Statistical analysis was performed using ordinary one-way ANOVA followed by Dunnett's multiple comparisons test (comparing all groups to the control group). Exact P values from Dunnett's test are as follows: GS59 vs. NC, $P = 0.0007$; GS69 vs. NC, $P = 0.174$ (not significant); CPEB3 vs. NC, $P = 0.0001$. When the calculated P value was below 0.0001, it was indicated as $P < 0.0001$. Graph asterisks correspond to the following significance levels: ***$P < 0.001$,****$P < 0.0001$.

Interestingly, the evolutionarily conserved regions predicted by PhastCons scores[11] from UCSC aligned closely with the functionally essential cores identified experimentally for GS46, GS59, and GS69. This correlation validates the functional importance of these conserved sequences. The exception was GS76; its lack of sequence conservation (Fig. 2h) may be attributed to its location on the reverse genomic strand, potentially subject to different selective pressures. Furthermore, the PhastCons score correlated with the genomic context and inferred functional constraints of each ribozyme. GS46, located within an exonic region, exhibited the highest level of sequence conservation (Fig. 2e), consistent with strong purifying selection in protein-coding sequences. GS59, which resides in an intron, showed moderate conservation (Fig. 2f). GS69 displayed the lowest conservation score (Fig. 2g); it is situated within an intron of an uncharacterized novel transcript, which may reflect more relaxed selective pressure or a dispensable regulatory role.

To assess whether the truncated ribozymes retain self-cleavage activity in a cellular environment, we employed a previously reported, straightforward reporter system[8]. Briefly, each ribozyme sequence was cloned into a plasmid downstream of an mCherry reporter gene (Fig. 2i), under the control of a single promoter, such that both are transcribed as a single mRNA. Ribozyme-mediated self-cleavage within this transcript is expected to destabilize the mCherry segment, leading to a reduction in mCherry RNA levels. The plasmid also carries a kanamycin resistance gene transcribed from an independent promoter, whose expression should be unaffected by ribozyme activity. We transfected these constructs into HEK-293 cells and quantified the RNA levels of mCherry and kanamycin by RT-qPCR. Using kanamycin RNA as an internal reference, we compared the relative abundance of mCherry RNA across constructs to evaluate the intracellular activity of each ribozyme variant. We evaluated the intracellular self-cleavage activity of eight ribozyme sequences, including the four novel ribozymes identified in this study and four previously reported ribozymes that were also captured in our screen (Fig. 2j). With the exception of GS69, all seven other ribozymes showed significant cleavage activity compared to the control. The most active ribozyme was a hammerhead ribozyme (HH9). Notably, the three new ribozymes (GS46, GS59, and GS76) displayed intracellular activity comparable to that of the CPEB3 ribozyme, which has been previously shown to regulate CPEB3 protein expression and influence human episodic memory[6,12]. Although GS69 exhibited the weakest activity, it still showed a trend toward cleavage ($P = 0.174$) relative to the control. These results indicate that all four novel ribozymes can fold into catalytically active structures in a cellular environment.

However, due to the 20-nt step size used in our truncation analysis, the precise boundaries of the minimal functional region could not be definitively mapped. Furthermore, as indicated by our prior work[7], internal regions within a ribozyme sequence may also be dispensable for catalytic activity. This notion is supported by the PhastCons conservation profile of GS46 (Fig 2e), which suggests the presence of non-conserved and likely functionally non-essential segments within the truncated region. Therefore, to precisely delineate functionally critical residues, we plan to perform systematic random mutagenesis across the truncated sequence followed by mutational scanning.

Mutational profiling reveals critical residues in four ribozymes.

As described in our previous study[7], the mean relative activity of all single-nucleotide substitutions at each position provides a quantitative measure of its importance for self-cleavage. Figs. 3a-d show the per-position mean fitness (natural logarithm of its relative activity value) profiles for ribozymes

GS46, GS59, GS69, and GS76, respectively. Positions with low mean fitness (e.g., below zero values) indicate high mutational sensitivity and are thus essential for catalysis, whereas positions tolerant to mutation (higher mean fitness, fitness ≥ 0) are likely structurally or functionally dispensable. Analysis of the per-position fitness profiles revealed distinct but functionally interpretable patterns for each of the four novel ribozymes (Figs. 3a-d). GS46 is characterized by a central, essential catalytic core (positions 21-54) flanked by terminal regions of high mutational tolerance; notably, a short segment within this core exhibits relative insensitivity, suggesting a structurally permissive spacer. Similarly, GS59 displays a bimodal profile with two separate sensitivity troughs, suggesting a bipartite functional architecture. In contrast, GS69 exhibits the most compact profile, with a sharp, single sensitivity peak defining a short, uninterrupted catalytic domain. Finally, GS76 presents a more complex landscape featuring several discrete sensitivity peaks distributed across its length. This pattern likely reflects a structurally intricate ribozyme architecture; alternatively, some of the observed variations could arise from experimental noise in the fitness measurements. The low signal-to-noise ratio in these measurements can likely be accounted for by the inherently low self-cleavage activity of the GS59 and GS76 wild-type sequences, combined with potential inefficiencies during template switching in RNA-seq library construction (Extended Data Fig. 3).

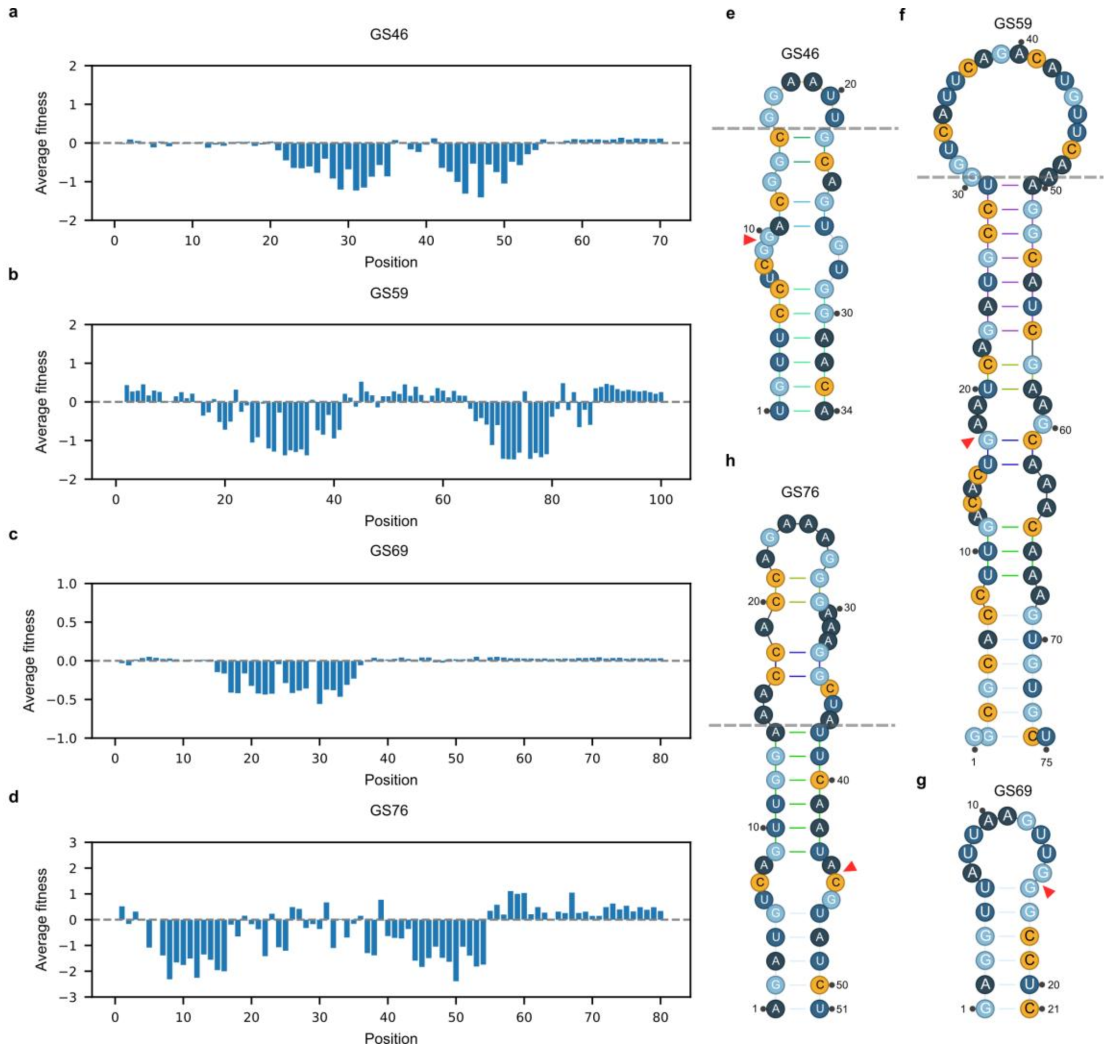


Fig. 3: Mutational analysis and predicted secondary structures of the 4 new ribozymes.

a-d, Average fitness of single mutations at each position of the ribozymes. A fitness of 0 indicates the equal self-cleavage activity to the wild-type sequence. Positions with values significantly less than 0 indicate that these nucleotides are highly sensitive to mutations and are critical for maintaining specific structures and catalytic functions. **e-h**, Predicted secondary structures of the minimal regions for four newly identified ribozymes. Predicted secondary structures were generated using RNAstructure[13], with single-mutation fitness values applied as constraints. The red arrowhead marks the cleavage site. Regions above the grey dashed line are dispensable for catalytic activity, whereas the region below the line constitutes the catalytically essential core, which was used for subsequent biochemical assays.

Beyond identifying functionally essential regions, mutational fitness data can also inform secondary structural information. In prior work, base-pair interactions were inferred from the deviation between the fitness of a double mutant and the expected additive effect of its two single mutants[14]. Generally, double-mutant libraries suffer from low sequencing depth relative to single mutants. Critically, the in vitro self-cleavage activity of all four ribozymes is inherently low (as will be shown in subsequent biochemical assays), which precludes obtaining high-quality fitness measurements for double mutants. We therefore turned to using single-mutant fitness values as constraints for structure prediction with RNAstructure[13]. RNAstructure typically employs direct structural probes—such as enzymatic cleavage, traditional chemical modification, or SHAPE reactivity—as folding constraints. Fitness, in contrast, reports on function: low fitness indicates a position critical for activity, but does not reveal whether that nucleotide is paired or unpaired. Conversely, a high fitness value can occur for either paired or unpaired nucleotides; however, even if a nucleotide is base-paired, a high fitness value suggests that the specific pairing is not essential for the core functional fold. We therefore treated high-fitness positions as structurally unconstrained (single-stranded). Specifically, we masked positions with mean fitness < -0.5 and retained those with mean fitness ≥ -0.5. The fitness values of the retained positions were rescaled to resemble SHAPE-like reactivity profiles and supplied to RNAstructure as pseudo-SHAPE constraints. The resulting predicted secondary structures for the truncated ribozymes are shown in Extend Data Figs. 4a-d. Consistent with the fitness profiles, each ribozyme contains only a short segment that is critical for self-cleavage. Strikingly, the essential core of all four ribozymes adopts a simple hairpin-loop architecture (Figs. 3e-h). In three of them (GS46, GS59, and GS76), the terminal loop is dispensable for both structure and activity, whereas GS69 folds into a compact, 21-nucleotide hairpin that constitutes its entire functional core.

Biochemical characterization of the identified ribozymes. To biochemically characterize the ribozymes and validate the predicted minimal functional cores, we designed bimolecular constructs based on the fitness profiles and computationally predicted secondary structures. For three of the ribozymes (GS46, GS59, and GS76), the minimal core was rationally divided into a separate substrate strand and a separate enzyme strand (Fig. 4a). The short, self-contained GS69 ribozyme, however, could not be feasibly partitioned and was therefore assayed as a unimolecular, self-cleaving RNA. Fig. 4b demonstrates that the bimolecular construct of four novel ribozymes (GS46, GS59, GS69, GS76) could self-cleave in the reaction buffer environment. Initial experiments in water or in a complete reaction buffer (1× TKMS buffer: 50 mM Tris, 25 mM KCl, 10 mM $MgCl_2$, 2 mM spermidine) showed that cleavage of the 5′-FAM/3′-TAMRA dual-labeled substrates, visualized by the appearance of a faster-migrating fluorescent product band, occurred only in the TKMS buffer after a 6-hour incubation, confirming an obligate dependence on both the ribozyme and specific buffer components. Systematic dissection of the TKMS buffer revealed that the self-cleavage activity of GS46, GS59, and GS76 was absolutely dependent on the presence of Mg^{2+}, as reactions containing only Tris (T) or Tris+KCl (TK) showed no product formation (Figs. 4c and 4d). However, the GS69 ribozyme could utilize K^+ to initiate catalysis, and its cleavage efficiency was further enhanced by the addition of spermidine (Fig. 4d).

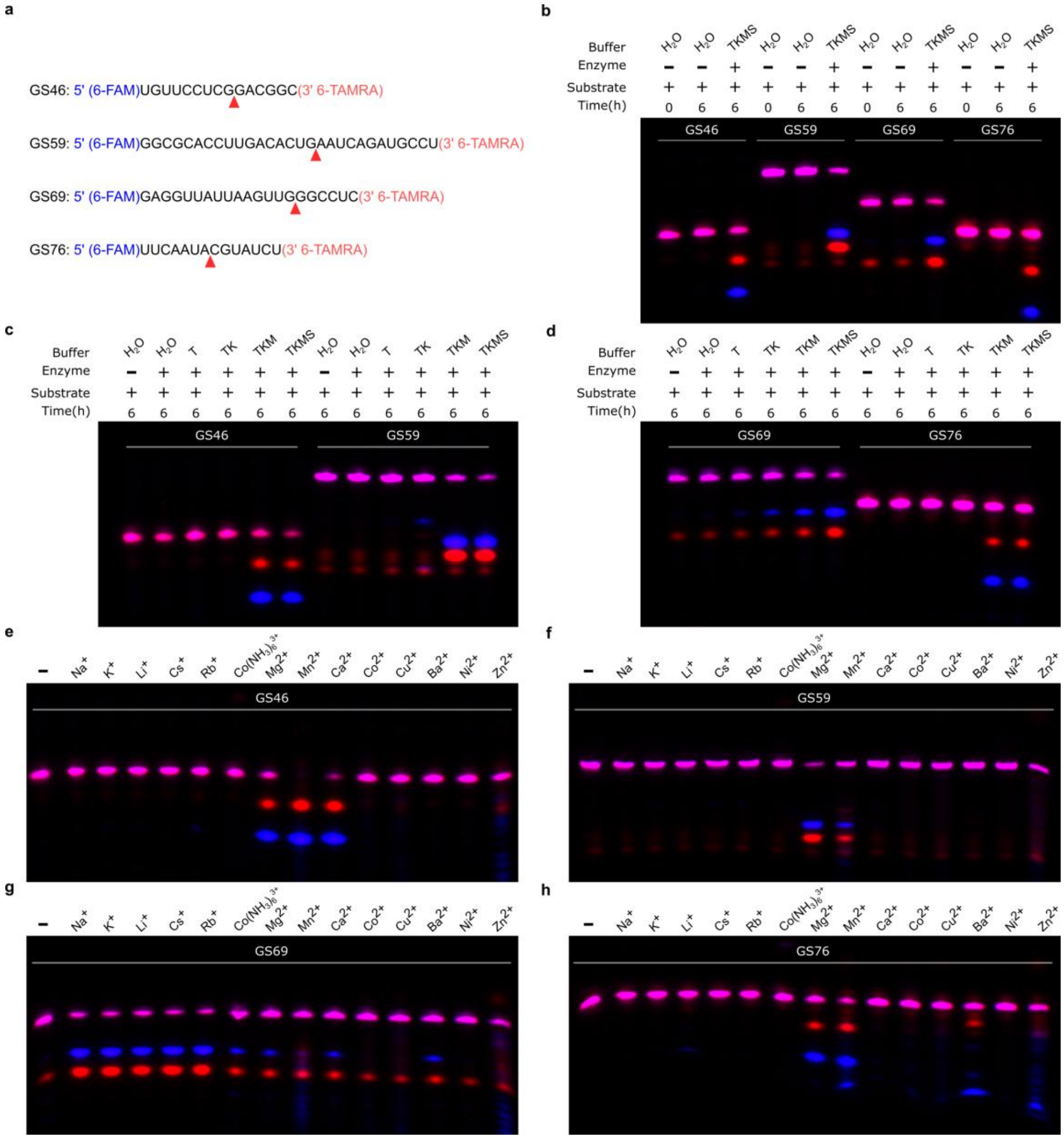


Fig. 4: Analysis of ribozyme cleavage activity under various conditions.

a, Schematic of the four fluorescently labeled RNA substrates (GS46, GS59, GS69, GS76) used in the cleavage assays. The cleavage site is indicated by a red arrowhead. Each substrate RNA is 5′-labeled with 6-FAM and 3′-labeled with TAMRA. **b-h**, Time-course cleavage assays monitored by denaturing gel electrophoresis. For each panel, the tested ribozyme is indicated. All reactions contained 3 μM substrate (S) and 4.5 μM enzyme (E) strand unless otherwise noted, except for the GS69 ribozyme, which was present as a single RNA strand at 3 μM. **b**, Control reactions with substrate alone or ribozyme alone in the reaction buffer. The 0 h and 6 h time points show the progression of the cleavage reaction. **c-d**, Systematic dissection of individual TKMS buffer components on ribozyme cleavage efficiency. The single-letter codes denote the presence of each component added to the base buffer: T, Tris; K, KCl; M, $MgCl_2$; S, spermidine. **e-h**, Cleavage activity in the presence of different monovalent (1M), 10 mM $Co(NH_3)_6^{3+}$, or divalent metal ions (10 mM). Reactions were incubated at 37 °C for 6 hours.

Furthermore, comprehensive screening of monovalent and divalent metal ions (Li^+, Na^+, K^+, Rb^+, Cs^+, $Co(NH_3)_6^{3+}$, Mg^{2+}, Mn^{2+}, Ca^{2+}, Mn^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}, Ba^{2+}) established a strict metal-ion specificity. The four ribozymes exhibited distinct metal-ion dependencies, delineating two functional classes. GS46, GS59, and GS76 displayed a strict catalytic requirement for divalent metal ions (Figs. 4e, 4f, and 4h). While Mg^{2+} supported robust cleavage in all three, and Mn^{2+} supported a similar level of activity, only GS46 showed substantial cleavage with Ca^{2+}, and only GS76 exhibited detectable, albeit weak activity with Ba^{2+} and Zn^{2+}. Critically, none of these three ribozymes was active in the presence of $Co(NH_3)_6^{3+}$, a redox-inert structural analog of $Mg(H_2O)_6^{2+}$ whose amine ligands do not readily dissociate. The complete loss of activity with $Co(NH_3)_6^{3+}$ indicates that for GS46, GS59, and GS76, divalent cations are required not only for structural stabilization but also for direct chemical participation in the catalytic mechanism. In stark contrast, GS69 displayed a uniquely broad metal-ion tolerance (Fig. 4g). It was active in the presence of multiple divalent cations and, unlike the other ribozymes, also catalyzed significant cleavage with monovalent ions alone, defining a fundamentally different mechanistic strategy.

To further investigate the biochemical properties of the four newly identified ribozymes, we measured the reaction kinetics of their respective constructs under varying conditions. This analysis was performed using the same bimolecular constructs for GS46, GS59, and GS76, and the unimolecular construct for GS69. Fluorescence-based time-course assays revealed distinct kinetic profiles and biochemical dependencies for the four novel ribozymes (Figs. 5a-d). In a minimal Tris buffer (T), all ribozymes exhibited negligible self-cleavage. The addition of 10 mM Mg^{2+} (TM condition) activated all four ribozymes. Strikingly, inclusion of 2 mM spermidine (TMS condition) dramatically enhanced the reaction rate for all ribozymes (Fig. 5e). GS59 showed the most pronounced activation, with its observed rate constant (k_{obs}) increasing more than 10-fold upon addition of spermidine. The strong stimulation by spermidine likely reflects its dual role in RNA catalysis: as a polycation, it neutralizes the repulsive negative charge of the RNA backbone, promoting the compact folding of the active structure; it may also directly stabilize specific transition states or tertiary interactions within the catalytic core.

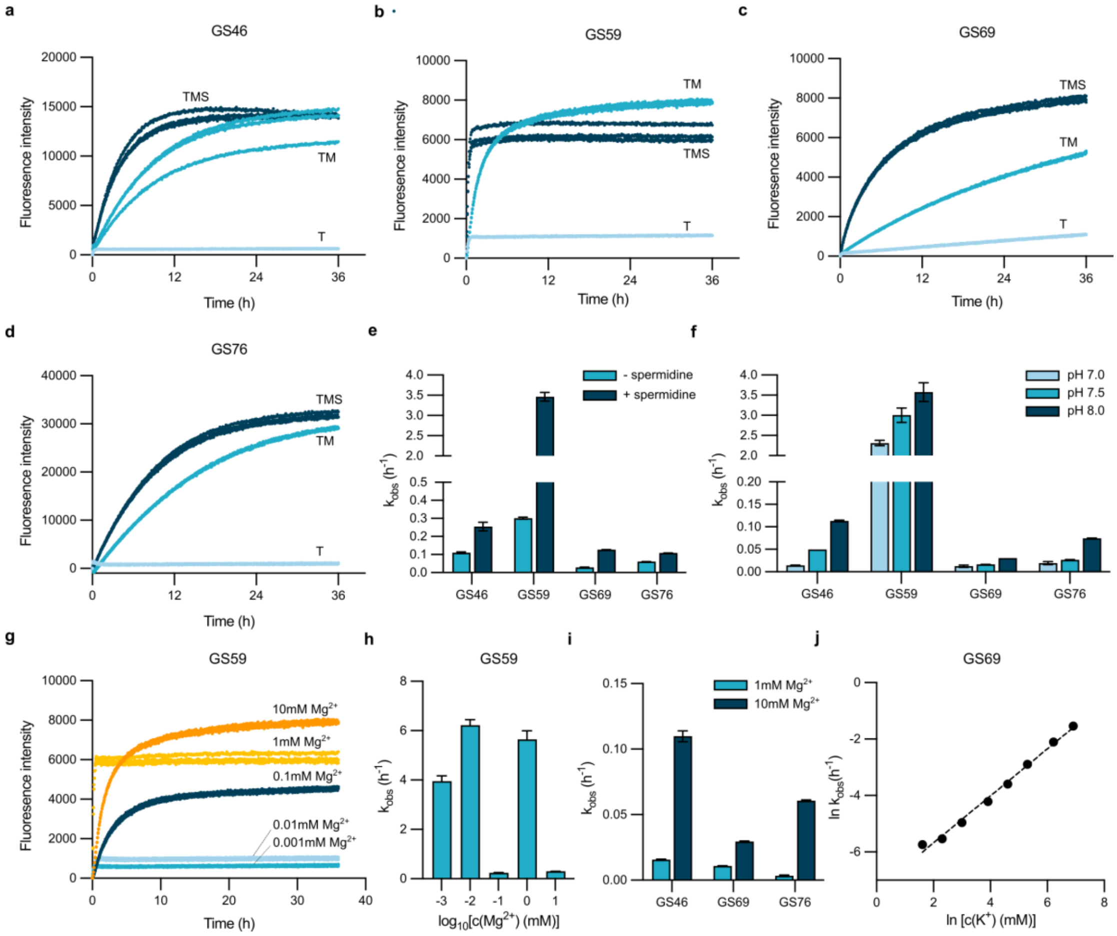


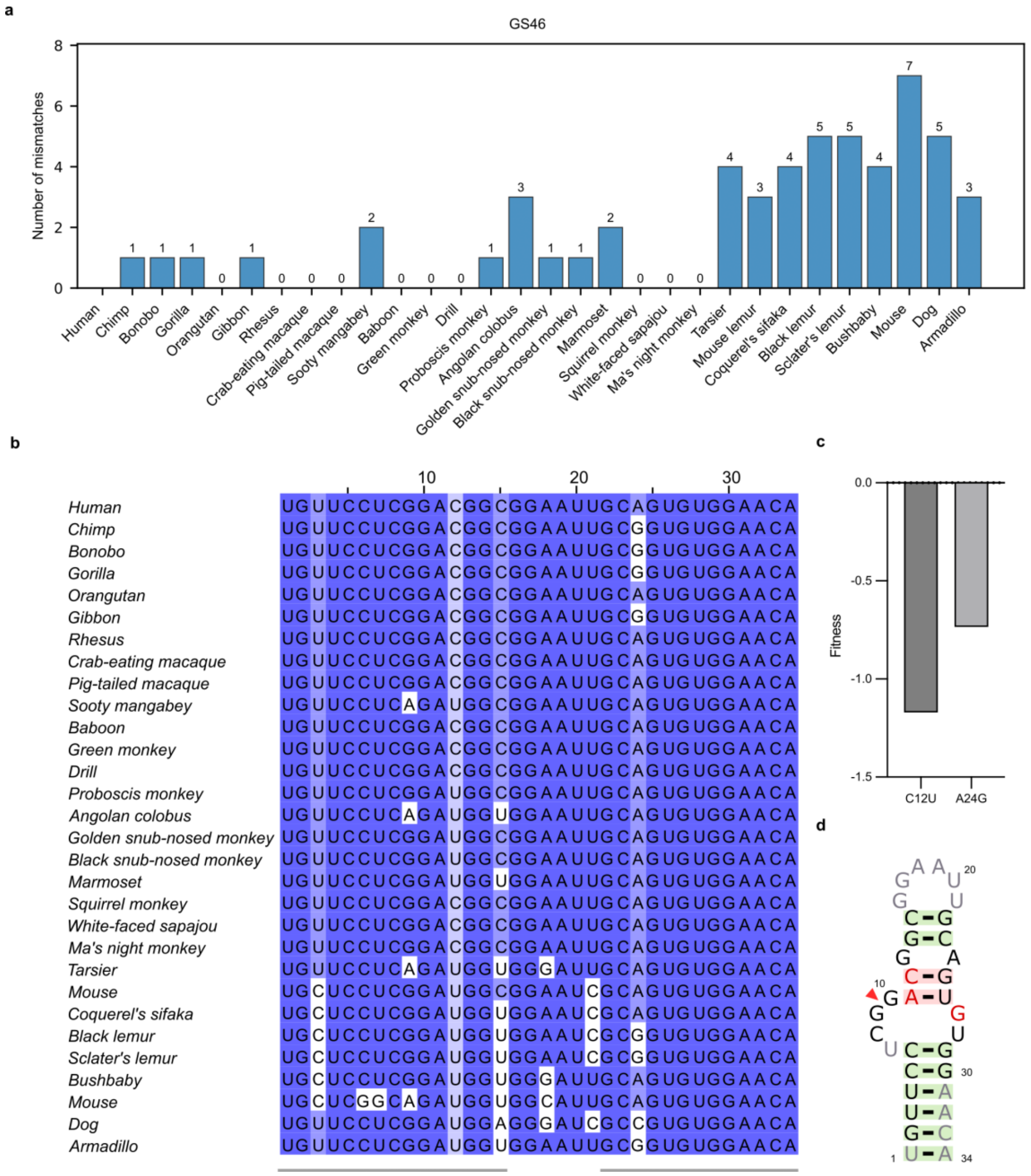
Fig. 5: Kinetic characterization of ribozyme cleavage under varying reaction conditions.

a-d, Time-course fluorescence curves showing the cleavage progress of ribozymes GS46, GS59, GS69, and GS76 in three different buffer conditions: T (20 mM Tris, pH 8.0), TM (T + 10 mM $MgCl_2$), and TMS (TM + 2 mM spermidine). Each condition was represented by 3 biological replicates. **e-f**, Bar graphs summarizing the calculated k_{obs} values under variations of reaction components. Data are presented as mean ± SD (n = 3 biological replicates). **g-i**, Effects of Mg^{2+} concentration on cleavage kinetics. **j**, Effects of K^+ concentration on cleavage kinetics of GS69.

The cleavage activity was also pH-dependent, with optimal cleavage at pH 8.0 for GS46, GS69, and GS76; GS59 displayed a broader, less pronounced pH optimum (Fig. 5f). A detailed Mg^{2+} titration for GS46 revealed a biphasic concentration dependence characterized by a hyperbolic activation profile that transitions to inhibition at supra-optimal concentrations (Figs. 5g, 5h), where the observed rate constant (k_{obs}) peaked near 1 mM Mg^{2+} and diminished sharply at both lower (0.1 mM) and higher (10 mM) levels; this non-monotonic kinetic profile suggests Mg^{2+} exerts multiple, opposing effects, likely by stabilizing the active ribozyme fold at optimal concentrations while inducing non-productive conformations or inhibitory complexes in excess. Concomitantly, the reaction yield, reflected in the final fluorescence amplitude, increased progressively with Mg^{2+} concentration, indicating that Mg^{2+} also thermodynamically stabilizes the reaction equilibrium or product complex, thereby driving the overall conversion forward. The other three ribozymes (GS46, GS69, and GS76) also exhibited a clear, concentration-dependent increase in activity with Mg^{2+}, as seen in their respective kinetic traces and derived k_{obs} values (Fig. 5i). Strikingly, GS69 defined a unique mechanistic class. Building on the finding that multiple monovalent cations support GS69 catalysis (Fig. 4g), we selected K^+ as a representative ion for detailed kinetic analysis. The activity of GS69 increased logarithmically with K^+ concentration across the tested range (1–1000 mM) and showed no sign of saturation (Fig. 5j). Hill analysis of this dependence yielded a slope of approximately 0.26, a value significantly less than 1. This indicates strong negative cooperativity or a high degree of heterogeneity in K^+ interactions[15], suggesting that multiple, weakly coordinating or functionally distinct sites contribute to the observed activation.

Evolutionary conservation of four ribozymes. To investigate whether the identified ribozymes are conserved across species and to understand their evolutionary constraints, we performed a comparative sequence analysis. Figs. 6a-c present a comprehensive genetic and functional analysis of the novel ribozyme GS46, which is notably the only ribozyme located within a protein-coding exon. The data reveal its strong evolutionary conservation and define its core functional structure.

Comparative genomics across 30 species (including 27 primate species) demonstrates that the GS46 sequence is highly conserved, particularly in humans and closely related primates, as evidenced by the low number of mismatches (Fig. 6a) and the high degree of sequence identity in the multiple alignment (Fig. 6b). This conservation is notably stringent, consistent with its location within a protein-coding exon, suggesting it is under strong purifying selection. The functional importance of specific nucleotides is highlighted in Fig. 6c, where mutational scanning result shows that single-nucleotide substitutions at two less conserved positions (C12U and A24G) result in significant fitness defects, directly linking sequence conservation to catalytic necessity. The GS46 ribozyme is predicted to be functionally active in all species shown with zero mismatch in Fig. 6a, which include human, orangutan, rhesus, crab-eating macaque, pig-tailed macaque, baboon, green monkey, drill, squirrel monkey, white-faced sapajou and ma's night monkey. The perfect sequence conservation suggests that the structural and catalytic integrity of the ribozyme is fully maintained in these primates. For species with one or more mismatches, the likelihood of retaining identical activity decreases, as even single nucleotide substitutions could potentially disrupt the essential folding or chemical mechanism of the catalytic core. Finally, the consensus secondary structure model derived from fitness data of artificial mutants (Fig. 6d) reveals that the essential core of GS46 forms a simple hairpin, with the critical nucleotides implicated in the fitness map forming part of its base-paired stem. In the predicted secondary structure (Fig. 6d), two specific base pairs (G:C and A:U) are highlighted with red shading, which indicates that no significant covariation was detected for these pairs. This indicates that either the predicted pairing for these positions is incorrect, or the bases are so highly conserved that they are never substituted, thus leaving no evolutionary trace of covariation.



Human	UGUUCCUCGGACGGCGGAAUUGCAGUGUGGAACA
Chimp	UGUUCCUCGGACGGCGGAAUUGCGGUGUGGAACA
Bonobo	UGUUCCUCGGACGGCGGAAUUGCGGUGUGGAACA
Gorilla	UGUUCCUCGGACGGCGGAAUUGCGGUGUGGAACA
Orangutan	UGUUCCUCGGACGGCGGAAUUGCAGUGUGGAACA
Gibbon	UGUUCCUCGGACGGCGGAAUUGCGGUGUGGAACA
Rhesus	UGUUCCUCGGACGGCGGAAUUGCAGUGUGGAACA
Crab-eating macaque	UGUUCCUCGGACGGCGGAAUUGCAGUGUGGAACA
Pig-tailed macaque	UGUUCCUCGGACGGCGGAAUUGCAGUGUGGAACA
Sooty mangabey	UGUUCCUCAGAUGGCGGAAUUGCAGUGUGGAACA
Baboon	UGUUCCUCGGACGGCGGAAUUGCAGUGUGGAACA
Green monkey	UGUUCCUCGGACGGCGGAAUUGCAGUGUGGAACA
Drill	UGUUCCUCGGACGGCGGAAUUGCAGUGUGGAACA
Proboscis monkey	UGUUCCUCGGAUGGCGGAAUUGCAGUGUGGAACA
Angolan colobus	UGUUCCUCAGAUGGUGGAAUUGCAGUGUGGAACA
Golden snub-nosed monkey	UGUUCCUCGGAUGGCGGAAUUGCAGUGUGGAACA
Black snub-nosed monkey	UGUUCCUCGGAUGGCGGAAUUGCAGUGUGGAACA
Marmoset	UGUUCCUCGGAUGGUGGAAUUGCAGUGUGGAACA
Squirrel monkey	UGUUCCUCGGACGGCGGAAUUGCAGUGUGGAACA
White-faced sapajou	UGUUCCUCGGACGGCGGAAUUGCAGUGUGGAACA
Ma's night monkey	UGUUCCUCGGACGGCGGAAUUGCAGUGUGGAACA
Tarsier	UGUUCCUCAGAUGGUGGGAUUGCAGUGUGGAACA
Mouse	UGCUCCUCGGAUGGCGGAAUCGCAGUGUGGAACA
Coquerel's sifaka	UGCUCCUCGGAUGGUGGAAUCGCAGUGUGGAACA
Black lemur	UGCUCCUCGGAUGGUGGAAUCGCGGUGUGGAACA
Sclater's lemur	UGCUCCUCGGAUGGUGGAAUCGCGGUGUGGAACA
Bushbaby	UGCUCCUCGGAUGGUGGGAUUGCAGUGUGGAACA
Mouse	UGCUCGGCAGAUGGUGGCAUUGCAGUGUGGAACA
Dog	UGUUCCUCGGAUGGAGGGAUCGCCGUGUGGAACA
Armadillo	UGUUCCUCGGAUGGUGGAAUUGCGGUGUGGAACA

Fig. 6: Comparative genetic analysis of the GS46 ribozyme.

a, Assessment of GS46 ribozyme conservation across 30 species. The bar chart shows the number of mismatches relative to the human sequence, computed from the UCSC 30 Primates Multiz Alignment track. Lower values indicate higher sequence conservation. **b**, Nucleotide sequence alignment of the GS46 minimal region across selected species. Nucleotide positions are numbered above. Conserved bases are highlighted in blue. **c**, Fitness values of the two most common single-nucleotide substitutions (C12U and A24G) identified in primate homologs. Fitness values were derived from prior high-throughput mutational profiling experiments. **d**, Consensus secondary structure model of the GS46 catalytic core. The model was generated in R2R using artificial homologous sequences that retained functional fitness (fitness > −0.5) in prior mutational screens. In the diagram, nucleotides are colored according to their evolutionary conservation: gray (95%), black (98%), and red (100%). Predicted base pairs are highlighted with green shading if supported by covariation analysis, or with red shading if no significant covariation was detected. A red arrowhead denotes the cleavage site.

Cross-species conservation analysis of the other three novel ribozymes reveals a pattern of less stringent evolutionary constraint within primates and increased sequence divergence in more distant mammals. The number of nucleotide mismatches relative to the human reference sequence was quantified for ribozymes GS59, GS69, and GS76 across a series of species spanning primates and other vertebrates (Extend Data Fig. 5). All three ribozymes exhibit very few mismatches in closely related primates such as chimpanzee, gorilla, and orangutan, indicating strong purifying selection within the primate lineage. This conservation breaks down in more evolutionarily distant species. Several species, marked by red "X" symbols, lack detectable homologs, indicating that the ribozyme sequence is not present or is too divergent to align in those genomes. Collectively, these mismatch profiles demonstrate that the identified ribozymes are highly conserved primate sequences, with their evolutionary constraint weakening predictably with increasing phylogenetic distance from humans.

Discussion

Self-cleaving ribozymes are ubiquitous in lower organisms, yet their representation in complex eukaryotic genomes remains enigmatic. Although prior studies have identified a handful of examples in humans—including the HDV-like CPEB3 ribozyme[6], two hammerhead ribozymes[5], and the hovlinc ribozyme[8]—these discoveries relied heavily on known sequence motifs, specific biochemical properties or genome-scale search according to the discovery of hovlinc ribozyme[8]. This raises a fundamental question: is the scarcity of known human ribozymes biological reality, or a reflection of methodological constraints?

To address this, we developed a high-throughput, unbiased discovery pipeline centered on a defining catalytic product: RNA fragments with 2′,3′-cyclic phosphate and 5′-hydroxyl termini. Our key innovation is the dual-selection of both termini using RtcB ligase, which concurrently requires both cleavage signatures. This strategy dramatically enhances specificity over single-end capture methods and provides a generalizable framework for genome-wide ribozyme mining. Our method represents a significant advance in discovery efficiency. From 96 high-confidence candidates, we validated six active sequences—including two known and four novel ribozymes—substantially outperforming previous motif-, similarity-based and biochemical enrichment-based searches. However, the presence of false positives underscores the necessity for rigorous experimental validation. These signals may originate from trans-cleavage events, non-specific enzymatic reactions, or other RNA processing activities; their resolution will be crucial for future pipeline optimization. A technical limitation of our RtcB-based capture is variable detection sensitivity across ribozyme classes, as evidenced by the weak signal for hovlinc and the absence of signal for CPEB3 in 5OH-seq data. Stable 3′-terminal secondary structures or post-ligation cleavage may impair adapter ligation efficiency, suggesting areas for methodological refinement.

Nevertheless, we discovered four novel self-cleaving ribozymes from the human genome. Notably, their genomic contexts suggest diverse functional and evolutionary origins. Ribozyme GS46 resides in a protein-coding exon of WDFY1, implying a potential role in post-transcriptional regulation, possibly through cleavage-mediated nonsense-mediated decay of its host mRNA. In contrast, GS59 and GS69 are embedded within repetitive elements—an intronic DNA repeat in PLD5 and an LTR retrotransposon, respectively—hinting at mobilization via genomic dynamics. The antisense localization of GS76 within an SYNJ2BP intron suggests possible cis-regulation of the sense transcript through mechanisms like RNA interference. Truncation analysis defined minimal, functional catalytic cores ranging from ~70 to 100 nucleotides, with GS69 being the most compact at only 21 nucleotides. Mutational profiling revealed that these cores adopt simple structural folds, predominantly single hairpin loops.

In contrast to many previously reported ribozymes, these structurally simple motifs exhibited relatively low in vitro catalytic activity. Under all tested conditions, the observed rate constants (k_{obs}) for three of the ribozymes (GS46, GS69, and GS76) did not exceed 0.3 h^{-1}. The most active one, GS59, achieved a maximum k_{obs} of 3.46 h^{-1} with the presence of spermidine, which is notably higher than the ~0.69 h^{-1} reported for the CPEB3 ribozyme[6]. Despite their modest in vitro kinetics, a series of biochemical assays (Figs. 4, 5) conclusively demonstrated their intrinsic self-cleavage capability. Importantly, cellular activity assays showed that three of these novel ribozymes possess intracellular activity comparable to that of the CPEB3 ribozyme (Fig. 2j). It is particularly noteworthy that GS46 is the only identified ribozyme located within a protein-coding exon and is highly conserved in primates closely related to humans. This suggests that self-cleaving ribozymes may fulfill novel biological functions in higher eukaryotes.

In summary, this work significantly expands the catalog of human self-cleaving ribozymes and provides a powerful, adaptable strategy for their discovery. The unique genomic associations and putative novel structures of the identified ribozymes open new avenues for exploring RNA catalysis in eukaryotic regulation. Future efforts to determine their tertiary structures, precise mechanisms, and biological functions will deepen our understanding of the functional complexity of the human RNA world.

Methods

Experimental process for ribozyme identification. HeLa genomic DNA libraries were constructed using the tagmentation method, which simultaneously fragments genomic DNA (NEB, no. N4006S) and incorporates adapter sequences via the Tn5 transposase. The process was performed using the Nextera XT DNA Library Preparation Kit (Illumina, no. FC-131-1024), following the manufacturer's guidelines. PCR amplification of the tagmentation product was carried out using primers T7p-s7 and T7t-s5. These primers were designed to anneal to the Nextera s7 and s5 adapter regions, respectively, and incorporated the T7 promoter sequence at the 5′ ends of the amplicons. Sequences of all oligonucleotides used in this study are listed in Supplementary Table 1. Separation of the DNA library fragments (500-1000 bp) was achieved by agarose gel electrophoresis and subsequent gel purification.

Purified DNA library was then used as the template for in vitro transcription by incubation with 100 U of T7 RNA Polymerase (NEB, no. M0251L) in 1×T7 RNA polymerase buffer (40 mM Tris-HCl pH 7.9, 6 mM $MgCl_2$, 1 mM DTT, 2 mM spermidine) with 2 mM NTP mix (NEB, no. N0466S) and 20 U Murine RNase Inhibitor (NEB, no. M0314S) in a 30 μl reaction volume for 5 h at 37 °C. DNA templates were removed by incubation with 2 U of DNase I (NEB, no. M0303S) and 1× DNase I reaction buffer in a 100 μl volume at 37 °C for 15 min, followed by purification with RNA Clean & Concentrator-5 kit (Zymo Research, no. R1016).

2.5 µg purified RNA were incubated with 100 pmol linker (rM13F_3desBio for 3P-seq, rM13R_5desBio_3P for 5OH-seq) in 1× RtcB reaction buffer with 22.5 pmol RtcB ligase (NEB, no. M0458S), 0.1 mM GTP and 1 mM $MnCl_2$ in 30 µl at 37 °C for 1 h. Ligation products were purified with Sera-Mag Streptavidin-Coated Magnetic Beads (Cytiva, no. 30152103010150) and eluted in 20 µl TNB buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 4 mM biotin). Before reverse transcription, ligation products were exchanged into H_2O by column purification.

Ligation products were reversed transcribed using primer (RT1_m13f_adp1 for 3P-seq, RT1_Tn5ME_adp1 for 5OH-seq) and the ProtoScript II Reverse Transcriptase (NEB, no. M0368S). Ligation product was first denatured at 65 °C for 5 min in a 19.5 µl volume containing 1 µl of 100 µM primer and 1.5 µl of 10 mM dNTP mix, followed by rapid snap-cooling on ice for 2 min. Then, 6 µl of 5 × ProtoScript II Reverse Transcriptase buffer, 3 µl of 0.1 M DTT, 0.5 µl of Murine RNase Inhibitor (NEB, no. M0314S) and 1 µl of ProtoScript II Reverse Transcriptase were added and incubated at 42 °C for 1 h and then 65 °C for 20 minutes to inactivate the enzyme, followed by enrichment with Sera-Mag Streptavidin-Coated Magnetic Beads (Cytiva, no. 30152103010150) and clean-up by using RNA Clean & Concentrator-5 kit (Zymo Research, no. R1016).

Purified cDNA was amplified by touch-up PCR using primers (P7R2_Tn5ME/P5R1_adp1 for 3P-seq, P7R2_m13r/P5R1_adp1 for 5OH-seq) with Q5 Hot Start High-Fidelity DNA Polymerase (NEB, no. M0493S) to construct the high-throughput sequencing library. The PCR products were purified with a homemade magnetic bead mix and sequenced on an Illumina HiSeq X sequencer with 20% PhiX control by Novogene Technology Co., Ltd.

High-throughput sequencing data analysis. A standard bioinformatic pipeline was implemented to process the sequencing data. Briefly, adapter sequences were trimmed from the raw reads using Cutadapt[16]. To ensure the stringent removal of the short adapter sequences flanking our library constructs, the Cutadapt-processed reads were further filtered with a custom script. This step involved removing the last 5 nucleotides from all reads longer than 113 nt. The resulting high-quality paired-end reads were then aligned to the UCSC human reference genome (hg38) using Bowtie2[17] in the --very-sensitive mode. The subsequent SAM file was processed to retain only uniquely mapped, properly paired reads with a MAPQ score of 30 or higher. Finally, the alignment data for each read was converted into cleavage site information, and all potential sites were statistically tallied and ranked.

Cleavage sites that appeared only once were filtered out, retaining only those sites with a read count of at least two. Noise reduction was then performed using a custom script. This method functions by analyzing the signal intensity ratios within sequencing reads of a defined window size. All potential cleavage sites within a defined genomic window (50nt) are removed if more than one site is present and the signal intensity ratio between them is less than 10. Each potential cleavage site was subsequently assigned a score calculated as $(N\_5OH + 1) \times (N\_3P + 1)$, where N_5OH and N_3P represent the number of occurrences of the site in the 3P-seq sequencing library and the 5OH-seq sequencing library, respectively. For experimental validation, we selected a final candidate pool comprising the 96 highest-scoring sequences that were identified in both independent screening libraries.

Experimental validation of candidate sequences. The top 96 candidate sequences, each flanked with the T7 promoter and unique index sequences, were synthesized as an oligo pool by IDT (Integrated DNA Technologies). The oligo pool was subsequently PCR-amplified with specific forward and reverse primers to obtain the target genes. PCR products were cloned into the pUC57 vector through homologous ends encoding the T7 promoter and a NotI site by using the HiFi DNA Assembly Master Mix (NEB, no. E2621L). After chemical transformation, single colonies containing the candidate sequences were picked for downstream validation. The NotI linearized candidate DNA templates were transcribed in vitro for 5 hours using T7 RNA Polymerase (NEB, no. M0251L). The transcription products were digested with DNase I (NEB, no. M0303S) and denatured at 70°C for 5 min after adding an equal volume of 2 × RNA Loading Dye (NEB, no. B0363S) for subsequent denaturing PAGE analysis.

Cleavage activity assay of truncated ribozyme variants. All truncated variants were generated via in vitro recombination using a Seamless Cloning kit (Beyotime, no. D7010M). Specifically, each variant was amplified with a pair of primers containing homologous arms flanking the T7 promoter and a NotI restriction site, and then cloned into the pUC57 vector through homologous recombination. The assembled product was subsequently transformed into E. coli DH5α cells. All constructed variants were verified by Sanger sequencing. Plasmid DNA from sequence-verified clones was linearized by NotI (NEB, no. R3189S) digestion and served as the template for in vitro transcription. Transcription reactions were performed using the T7 RNA Polymerase (Beyotime, no. R7012L), following the manufacturer′s protocol. After transcription, the DNA template was removed by DNase I (NEB, no. M0303S) treatment. The RNA products were then incubated in 1× TKMS buffer (50 mM Tris-HCl pH 8.0, 25mM KCl, 10 mM $MgCl_2$, 2 mM spermidine) at 37°C for 2 hours. Finally, the RNA was purified and analyzed by 10% denaturing polyacrylamide gel electrophoresis (PAGE), followed by staining with SYBR Gold nucleic acid gel stain (Thermo Fisher, no. S11494).

Cellular activity assay. All constructs for in vivo activity assay were generated via in vitro recombination using a Seamless Cloning kit (Beyotime, no. D7010M). Each ribozyme construct was amplified using primers containing homology arms to the pmCherry-C1 plasmid (MiaoLing Plasmid Platform, no. P0151) and subsequently cloned into its multiple cloning site via homologous recombination. All key components inside the plasmid were verified by Sanger sequencing. For transfection, plasmids were amplified in E. coli DH5α and purified using Endotoxin Removal Plasmid Mini-prep Plus Reagent Kit (GENSTONE BIOTECH, no. TD435-PLUS). HEK293 cells (Procell, no. CL-0001) were seeded in 24-well plates at a density of 1×10^5 cells per well one day before transfection. Twenty-four hours after seeding, the cells were transfected with the plasmid using PEI MAX (Polysciences, no. 24765) according to the manufacturer's instructions. Cellular RNA was extracted with RNAeasy Animal Long RNA Isolation Kit (Beyotime, no. R0027) 24 hours post-transfection.

Cellular RNA (1 μg) was reverse transcribed into cDNA using the HiScript III 1st Strand cDNA Synthesis Kit (Vazyme, no. R312) with gene-specific primers targeting the mCherry and kanamycin resistance genes, according to the manufacturer′s instructions. The resulting cDNA was diluted 10-fold with nuclease-free water, and 5 μl of the diluted product was used as template for quantitative PCR (qPCR). qPCR was performed using Taq Pro HS U+ Probe Master Mix (Vazyme, no. QN112) on a CFX Opus 96 Real-Time PCR System (Bio-Rad, no. 12011319). Gene-specific primers and corresponding dual-labeled probes (FAM reporter, TAMRA quencher) were used for the detection of mCherry and kanamycin, respectively. Each experimental condition was assessed with five biological replicates, and each biological replicate was analyzed in four technical replicates by qPCR. Gene expression was quantified using the ΔΔCt method, with mCherry as the target gene and kanamycin as the endogenous control. Statistical analysis was performed using ordinary one-way ANOVA followed by Dunnett's multiple comparisons test to compare treatment groups to the control.

Mutational profiling of four identified ribozymes. Shorter versions of the identified ribozymes were defined by sequential truncation of terminal regions that did not significantly reduce cleavage activity. These truncated sequences were used for subsequent mutational-scanning experiments. A random mutagenesis library was generated using a commercial Random Mutagenesis Kit (Biosharp, no. BL1668A). Barcode sequences were appended to the library in three PCR cycles. Approximately 10^6 barcoded molecules were used as template to amplify the final mutant DNA library. The final mutant DNA library served as template for constructing the DNA-seq library. The barcoded DNA library was transcribed with T7 RNA Polymerase (Beyotime, cat. no. R7012L) for 4 h at 37 °C. After transcription, the DNA template was removed by treatment with DNase I (New England Biolabs, no. M0303S). The purified RNA was incubated in 1× TKMS buffer (50 mM Tris-HCl pH 8.0, 25 mM KCl, 10 mM $MgCl_2$, 2 mM spermidine) at 37 °C for 2 h to allow ribozyme cleavage, followed by purification with the RNA Clean & Concentrator-5 kit (Zymo Research, cat. no. R1016). For reverse transcription, ~50 pmol of RNA was mixed with 50 pmol of RT primer and 200 pmol of template-switch oligo (TSO), and reverse-transcribed using M-MLV (H-) Reverse Transcriptase (Vazyme, no. R021) at 42 °C for 1 h. The purified cDNA was used as the template for NGS library amplification. PCR was performed with Q5 High-Fidelity DNA Polymerase (New England Biolabs, no. M0491L) for 12 cycles. The PCR products were purified with a homemade magnetic bead mix and sequenced on a Salus EVO sequencer by Shenzhen Salus BioMed Co., Ltd as the RNA-seq library.

Sequencing data were processed similarly to previously described methods[14] with minor modifications. Briefly, variant counts from the DNA-seq (input) and RNA-seq (selected) libraries were obtained by mapping the associated barcodes. The relative activity (RA) for each variant was then calculated using the following equation: $RA(var) = N_{cleaved}(var)(N_{cleaved}(wt)+N_{uncleaved}(wt))/(N_{cleaved}(var) + N_{uncleaved}(var))N_{cleaved}(wt)$, where $N_{cleaved}$ and $N_{uncleaved}$ represent read counts of the cleaved and uncleaved fractions, respectively, and "wt" denotes the wild-type sequence. Finally, a fitness score for each variant was defined as the natural logarithm of its RA value.

Bioinformatic analysis of four ribozymes. Conservation scores (PhastCons) and multiple sequence alignments (Multiz) for the ribozyme loci were obtained from the UCSC Genome Browser using the 30 Primates track. These data were processed and analyzed with custom Python scripts. To integrate functional data from mutational scanning, positions with a mean single-mutant fitness < -0.5 were masked, as they likely correspond to nucleotides essential for catalysis. For the remaining positions (mean fitness ≥ -0.5), the fitness value was transformed into a pseudo-SHAPE reactivity constraint by adding 1.0, generating a reactivity profile that could be used to guide RNA folding. Secondary-structure predictions were performed with RNAstructure[13] using the pseudo-SHAPE data as constraints, with SHAPE intercept and slope parameters set to –0.6 and 1.3, respectively. From the folding ensemble, the maximum-expected-accuracy (MEA) structure was selected for further analysis. For ribozyme GS46, the consensus secondary-structure diagram (Fig. 6d) was generated with R2R[18] using a multiple sequence alignment of functional mutants (fitness ≥ -0.5) derived from the mutational-scanning data.

PAGE-based biochemical assay. All ribozymes except GS69 were designed as bimolecular constructs, comprising a separate substrate strand and a separate enzyme strand; GS69 functions as a single, self-cleaving RNA strand. Cleavage reactions were performed in 20 µL volumes containing 60 pmol of 5′-FAM-labeled substrate RNA in reaction buffer (final concentration: 20 mM Tris-HCl pH 7.5, unless otherwise noted) supplemented with $MgCl_2$ or other metal ions as indicated. Reactions were initiated by adding 90 pmol of the corresponding enzyme strand. After incubation at 37 °C for the indicated time, reactions were stopped by adding an equal volume of stop solution (90 % formamide, 0.02 % SDS, 0.02 % bromophenol blue, 0.01 % xylene cyanol, 50 mM EDTA). Product analysis was carried out by denaturing PAGE (15 % gel, 8 M urea). Gels were imaged by using the ChemiDoc MP imaging system (Bio-Rad, no. 12003154).

Fluorescence-based kinetic analysis. Fluorescence-based kinetic analysis of the ribozymes was performed as previously described for the CPEB3 ribozyme[14]. Substrate RNAs were labeled with

5′-FAM (fluorophore) and 3′-TAMRA (quencher). Ribozyme-catalyzed cleavage separates the fluorophore from the quencher, generating a time-dependent increase in fluorescence. Standard 100 µL reactions contained 20 pmol of labeled substrate and 100 pmol of enzyme strand in a buffer consisting of 20 mM Tris-HCl (pH 8.0). Reactions were initiated by adding $MgCl_2$ or KCl to the indicated final concentrations. Fluorescence (Ex 488 nm/Em 520 nm) was monitored at 37 °C for 36 h using a BioTek Synergy H1 plate reader. Data were normalized to the fluorescence of a water-only blank control measured at t = 0. The observed cleavage rate constant (k_{obs}) was obtained by fitting the normalized fluorescence (F) versus time (t) to a single-exponential equation: $F = A - Be^{-k_{obs}t}$, where A is the endpoint fluorescence and B is the signal amplitude. Each k_{obs} value reported represents the mean of three independent reactions.

Data availability

All original sequencing data have been deposited in the Gene Expression Omnibus with accession number GSE327123. All scripts and processed data are available for download at https://doi.org/10.5281/zenodo.19430452.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 92370202 and 22350710182) and the National Key R&D Program of China (Grant No. 2021YFF1200400). We also acknowledge the High Performance Computing Cluster at Shenzhen Bay Laboratory (SZBL) and the high-performance computing resources of the Shenzhen Medical Academy of Research and Translation (SMART) for computational support. We also thank Shenzhen Salus BioMed Co., Ltd. for their support with next-generation sequencing.

Conflict of Interest.

All authors declare no financial interest. Zhan is the founder and CEO and Zhou is the scientific founder for Ribopeutic, respectively.

Author contributions

JZ and ZZ conducted the experimental design. ZZ performed wet-lab experiments and data analysis. JZ and YZ initiated and supervised the project. YZ provided the funding support. YZ, ZZ drafted the initial manuscript. All authors are involved in subsequent manuscript improvement.

Reference

1. Cech, T. R. Evolution of biological catalysis: ribozyme to RNP enzyme. *Cold Spring Harb Symp Quant Biol* **74**, 11–16 (2009).
2. Jimenez, R. M., Polanco, J. A. & Lupták, A. Chemistry and biology of self-cleaving ribozymes. *Trends Biochem Sci* **40**, 648–661 (2015).
3. Roth, A. *et al.* A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nat Chem Biol* **10**, 56–60 (2014).
4. Weinberg, Z. *et al.* New classes of self-cleaving ribozymes revealed by comparative genomics analysis. *Nat. Chem. Biol.* **11**, 606–610 (2015).
5. de la Peña, M. & García-Robles, I. Intronic hammerhead ribozymes are ultraconserved in the human genome. *EMBO Rep.* **11**, 711–716 (2010).
6. Salehi-Ashtiani, K., Lupták, 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).
7. Zhang, Z. *et al.* Minimal twister sister-like self-cleaving ribozymes in the human genome revealed by deep mutational scanning. *Elife* **12**, RP90254 (2024).
8. Chen, Y. *et al.* Hovlinc is a recently evolved class of ribozyme found in human lncRNA. *Nat Chem Biol* https://doi.org/10.1038/s41589-021-00763-0 (2021) doi:10.1038/s41589-021-00763-0.
9. Tanaka, N., Chakravarty, A. K., Maughan, B. & Shuman, S. Novel mechanism of RNA repair

by RtcB via sequential 2’,3’-cyclic phosphodiesterase and 3’-Phosphate/5’-hydroxyl ligation reactions. *J. Biol. Chem.* **286**, 43134–43143 (2011).

10. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165 (2016).
11. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**, 1034–1050 (2005).
12. Vogler, C. *et al.* CPEB3 is Associated with Human Episodic Memory. *Front Behav Neurosci* **3**, (2009).
13. Xu, Z. Z. & Mathews, D. H. Experiment-Assisted Secondary Structure Prediction with RNAstructure. *Methods Mol Biol* **1490**, 163–176 (2016).
14. Zhang, Z. *et al.* Accurate inference of the full base-pairing structure of RNA by deep mutational scanning and covariation-induced deviation of activity. *Nucleic Acids Res.* **48**, 1451–1465 (2020).
15. Abeliovich, H. An empirical extremum principle for the hill coefficient in ligand-protein interactions showing negative cooperativity. *Biophys J* **89**, 76–79 (2005).
16. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
17. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).
18. Weinberg, Z. & Breaker, R. R. R2R--software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics* **12**, 3 (2011).