OUKA
The University of Osaka
Institutional Knowledge Archive

Title	Exploring new methods in regulating circular RNA production using cellular model
Author(s)	Lu, Ni
Citation	大阪大学, 2024, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/96401
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

Doctoral Dissertation

Exploring new methods in regulating circular RNA production using cellular model

Ni Lu

2024

Preface

The study presented in this thesis was conducted under the direction of Prof. Kazuhiko Nakatani at the School Science Department of Chemistry in Osaka University from April 2021 to March 2024. This thesis is concerned with exploring potential new strategies that could be used to promote circRNA production using a model cellular environment. All the results shown in this thesis have been obtained by the author himself.

Table of Contents

General Introduction

Circular RNAs: What are they? And what is their purpose?

Generally, most genes in higher eukaryotes are arranged in a split manner, where transcribed RNA (pre-mRNA) is spliced to join the coding regions (exons) to one another in a consecutive manner by the removal of intervening non-coding regions (introns) in order to generate a functioning messenger RNA (mRNA). A process that has been long appreciated to be commonly exploited by cells via splicing together exons in various alternative manners to generate a diverse set of mRNAs from a given gene. Producing new mRNA sequences potentially coding for proteins with new unique functions and localizations.[1-3] In human genes, more than 95% is thought to undergo such alternative-splicing,[4] with the splicing decisions heavily regulated by a combination of various cis-regulatory and trans-regulatory elements.[5,6] Though it is commonly thought that alternative-splicing generates exclusively linear mRNA products, a noncanonical form has been discovered that splices exons in a tail-to-head manner, surprisingly producing "circular RNAs" (circRNAs).[7-10]

Broadly speaking, circRNAs are produced when such splicing reaction goes in the reverse direction, or "back-splices", to join a downstream splice donor site (5′ splice site or 5′ss) to an upstream splice acceptor site (3′ splice site or 3′ss) (discussed in more detail below), resulting in a single-stranded, closed-loop RNA molecule without 5′ end caps or 3′ polyadenylation.[11,12] As a result, circRNAs are generally resistant, though not completely, to exonuclease digestion,[13] endowing it with a much longer half-life in cells (~48 h) compared to linear mRNAs (~10 h).[11] Furthermore, as a product of alternative pre-mRNA splicing, circRNAs are most commonly comprised purely from a subset of exons from the transcribed host gene (EcircRNA)[14]; though, more recently, several subtypes have been discovered where intervening introns were retained in some circRNAs, exon-intron-circRNA (EIcircRNA),[15] or circRNAs that originate purely from intronic sequences, circular intronic RNA (IcircRNA)[16].

Interestingly, despite the recent surge in our understanding and interest in circRNAs, the existence of such circular forms of RNA molecules has been known in the form of viral RNA genomes and viroids since 1978 and described in mammalian cells as "scrambled exons" since 1991.[17,18] However, they were often dismissed as simple low occurrence splicing errors. Even with the advent and proliferation of high throughput next-generation sequencing (NGS) technologies, circRNAs often eluded detection due to poly(A)-tail purification being commonly practiced as a preferred means to deplete ribosomal RNAs from NGS library, a

feature circRNAs distinctly lacks.[11,19] Despite the setback, recent advances in ribosomal depletion techniques and utilization of random DNA primer-based cDNA preparation from transcriptome allowed not only the identification of a diverse set of new non-polyadenylated linear RNAs, but also abundant unexpected junctions sequences that do not align linearly to the reference transcriptome (e.g. tail sequence of exon 3 joined to head sequence of exon 2, etc.) indicating the presence of circular RNA species and suggesting that circRNAs are neither low-occurrence nor splicing errors, but a common product of alternative splicing.[14] Many genes express multiple circRNAs while others seemingly are expressed specifically for the purpose of circRNA production, showing several-fold higher expression of circRNA compared to the cognate mRNA.[10,20,21] Today, several biochemical methods[11] and computational methods[22] have been developed to identify and quantify their presence in cells, and many databases[23-25] have been built to accommodate the increasing number of newly discovered circRNAs. The results firmly establish the ubiquitous presence of circRNAs, with many showing high conservation throughout evolution,[8,10] great variation of expression between tissues,[26,27] and significant dysregulation in diseases,[28,29] implying the potentially essential regulatory role circRNAs play in cell development and differentiation.[9,30]

Despite their great diversity and stability in cells, however, most circRNAs are produced in low numbers without a clearly defined functional role in cells.[8,9,14,26] Currently, of the several potential functionalities reported for circRNAs expressed in mammalian cells, the most well-established circRNA roles are: microRNA (miRNA) sponges,[31] RNA binding protein (RBP) sponges,[32,33] and acting as translation templates[34]. The details of which are introduced below.

circRNAs regulate miRNA activity levels via sponging

miRNAs are ~21-nt long small RNAs that hybridize with mRNAs and function as a repressor to protein production, achieved by triggering mRNA degradation through the formation of RNA-induced silencing complex (RISC) in conjunction with members of the Argonaute (AGO) family of proteins and cleavage of the target mRNA.[35] Through thorough analysis of circRNA sequences for conserved miRNA binding sites, it was revealed that some circRNAs host an abundant array of such sequences. One of the first and most well-cited examples that exhibit such a function is the circular RNA CDR1as (also known as ciRS-7), which has been shown to contain >70 evolutionarily conserved binding sites for miRNA-7 (miR-7) in addition to 1 conserved binding site for miRNA-671 (miR-671).[36,37] However, due to none of the sites on CDR1as being perfectly complementary to miR-7, CDR1as escapes RISC induced cleavage

despite the high miR-7 binding density, hence allowing CDR1as to essentially "sponge" miR-7. The sponged miR-7 can then be released in the desired subcellular location in the presence of miR-671, for which CDR1as carries a fully complementary site. The ability of CDR1as to sponge and act as a "cargo" for miR-7 has been shown to have significant implications for the development of neuronal tissues in mice, with CDR1as knockout mice exhibiting impaired sensorimotor gating. [37]

Other notable examples include circZNF91, which contains 24 and 23 binding sites for miRNA-23b-3p and miRNA-766-3p, respectively, and is involved in regulating the differentiation of human epidermal stem cells.[38] While circHIPK3 has shown the ability to bind 9 different types of miRNAs with 18 potential binding sites, involved in regulating the proliferation of hepatocellular carcinoma cells.[39]

circRNAs regulate RBP activity levels via sponging

RBPs are a diversely expressed family of proteins that contain various structural motifs, such as RNA recognition motif (RMM), dsRNA binding domains, zinc fingers etc., and are known to control all aspects of RNA metabolism and functions in cells.[51] The interaction between RBPs and circRNAs has been well-described previously, as evidenced by the presence of potential RBP binding sequence within many circrRNAs and the co-localization of some circRNAs with RBPs. [32,33] CircRNAs exploit their resistance to degradation and stably "sponge" various RPBs in cells to modulate their function, similar to miRNAs. circAMOTL1, for example, has shown the ability to interact and sponge c-myc, STAT3, PDK1, ATK1, and AMOTL1, a protein produced from its cognate mRNA, to facilitate their nuclear translocation and regulate the transcription of their respective target genes.[46] While some circRNAs act as an endogenous competitive RNA (ecRNA), such as in the case of circPABPN1, which competes with its cognate mRNA PABPN1 for the binding of HuR protein, with circPABPN1-HuR interaction resulting in the suppression of PABPN1 translation.[52] Moreover, more recent examples have shown circRNAs containing binding sites for multiple different proteins functioning as a hub molecule, similar to the reported functions of some long noncoding RNAs (lncRNAs).[53] Facilitating the assembly of large protein complexes to modulate protein functions and activity. Such as circFOXO3, which was reported to bind to both CDK2 and p21 to form a ternary complex to inhibit the function of CDK2,[54] and circEXOC6B which was reported to enhance the interaction between RBMS1 and HuR to increase the translation of AKAP12.[55]

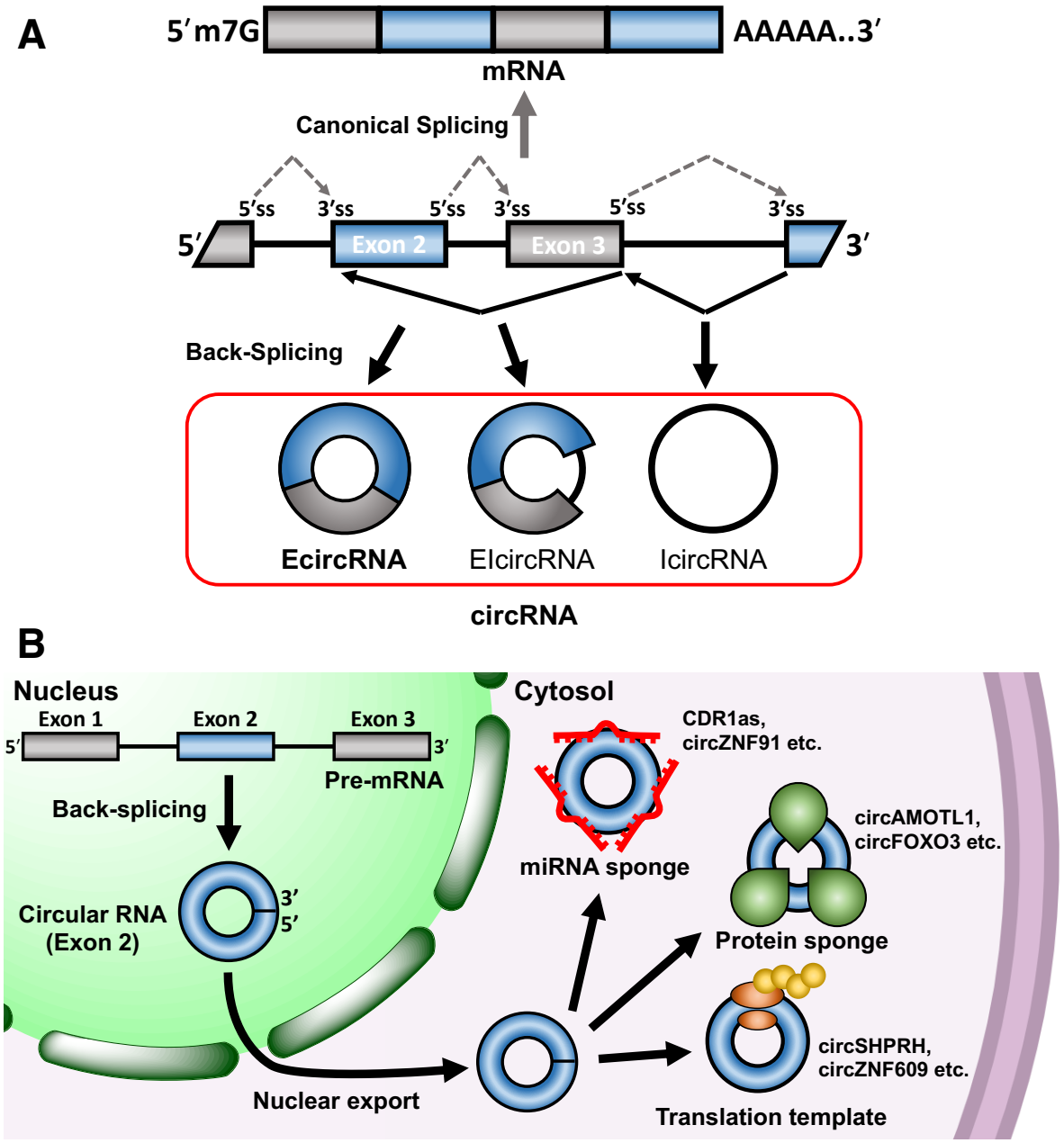


Figure 1. Schematic representation of **A)** how a pre-mRNA can be spliced to generate linear mRNA through the canonical splicing process, where the resulting mRNA was also characteristically capped at the 5′-end and polyadenylated at the 3′-end (top). Alternatively, the pre-mRNA can undergo splicing in the reversed direction, "back-splicing", and generate circRNAs, for which the termini are covalently linked by a 3′-5′ phosphodiester bonding (bottom). The arrows indicate the direction of the splicing process. **B)** Localization and reported functions of circRNA in cells. Representative examples of circRNAs are listed next to their respective functions.

Table 1: Non-exhaustive list of examples of various circRNAs and their functions in diseases. (HCC: Hepatocellular carcinoma, ESCC: Esophageal squamous cell carcinoma, GC: Gastric cancer, AD: Alzheimer's disease)

Disease	CircRNA	Role	Regulation	Ref.
Glioma	circTTBK2	miR-217 sponging	Up	40
	circFBXW7	Expression of FBXW7-185aa short peptide	Down	41
HCC	circZKSCAN1	miR-1178-3p sponging/ZKSCAN1-206aa short peptide	Down	42
ESCC	circITCH	miR-17/ 7/ 214 sponging	Down	43,44
Lung cancer				43
GC	circPVT1	miR-124 sponging	Up	45
Breast cancer	circAMOTL1	c-myc protein sponging	Up	46
	circABCB10	miR-1271 sponging	Up	47
	circFOXO3	ID-1, E2F2, FAK, and HIF1a sponging	Down	48
Cardiac senescence			Up	49
AD	CDR1as	miR-7 sponging	Down	50

circRNAs act as translation templates to produce short peptides

In general, circRNAs are considered non-coding,[56,57] due to their lack of 5′ end cap which allows for ribosome recognition and efficient translation. In spite of such limitations, recent studies have shown that a subset of endogenously circRNAs has the ability to undergo cap-independent translation in the presence of either internal ribosome entry site (IRES) sequences, such as the case of circZNF609[58] and circSHPRH[59], or m6A modification allowing for m6A-dependent translation, such as the case of circARHGAP35[60]. Although the presence of such ribosomal entry motifs within circRNA sequences are currently considered to be the hallmark for the efficient translation of protein-coding circRNA.[61-63] However, the exact necessity of such sequences for the translation of circRNAs is still debated, as *in vitro* studies have demonstrated that efficient translation can be initiated with only the presence of 10-nt long Kozak sequence[64], while in some cases, though less efficient, even the Kozak sequence can be omitted, only requiring the presence of an open reading frame (ORF) for the initiation of protein production[34].

CircRNA which are currently known to encode proteins, produce shorter trimmed variants of the protein encoded by their cognate mRNA. They are distinguished from the full-length protein due to the presence of unique terminus amino acid (aa) sequences stemming from translation through the junction sequence of circRNAs.[59,60] Functionally, however, proteins produced from circRNA partake in a much different, sometimes contrasting, role compared to its cognate mRNA. The prime example of such divergence in function is circARHGAP35,

circARHGAP35 encodes a 1289 amino acid (aa) long protein which was demonstrated to interact with transcription factor TFII-I to promote tumor proliferation, migration, and invasion; while its cognate mRNA ARHGAP35 encodes the same protein with an additional RhoA binding domain, and functions to inhibit Rho GTPase function resulting in the inhibition of tumor migration and invasion.[60]

They are important, so how are they produced?

Currently, the mounting evidence suggests that circRNAs are produced as a result of a special form of alternative-splicing mechanism, as circRNAs derived from exons of pre-mRNAs accounting for ~80% of the annotated circRNA population and the presence of canonical AG-GU sequence at the junction between the circularized exons.[63,65,66] Displaying the capacity to preferentially recognize exons and indicating the involvement of the canonical splicing machinery in the production of circRNAs. Due to the presence of the characteristic inverted "back-spliced" tail-to-head junction sequence in circRNAs (hereon as BSJ), this specific alternative-splicing process is termed "back-splicing".

Mechanistically, the back-splicing process is thought to occur in a similar manner to that of the canonical linear splicing mechanism which generates mRNAs, albeit in a reversed direction. Briefly, in canonical linear splicing, the initial attack of the 2′-OH of the branch point adenosine (BPA) in the intron occurs at the phosphodiester of the upstream 5′splice site (ss), generating a 3′-OH at the terminus of the preceding exon. Which then attacks the phosphodiester of the downstream 3′ss, linearly joining of the two exons and splicing out of the intron as a lariat, completing the splicing reaction.[67] However, in the case of back-splicing, as opposed to the upstream, the initial 2′-OH attack occurs at the downstream 5′ss, resulting in a 3′-OH at the terminus of a proceeding exon, which then attacks the phosphodiester of the upstream 3′ss, joining the downstream tail of exon(s) to upstream head to form circRNA.[12,68] The general consensus on the matter suggests that the reversal of the splicing direction requires the proximal localization of the upstream BPA to the downstream 5′ss, a feature which was recently confirmed by cryo-EM of yeast's early-stage spliceosome complex, where the spliceosomes have demonstrated the capacity to complex between the upstream BPA site with downstream 5′ss site of the circularizing exon in question.[68] Due to the relatively simplistic requirements for the production of circRNAs, theoretically, all internal exons (exons that have splicing signals at their 5′ and 3′ end) within a pre-mRNA can undergo circularization. However, analysis of sequence data reveals, only a subset of exons has been shown to undergo

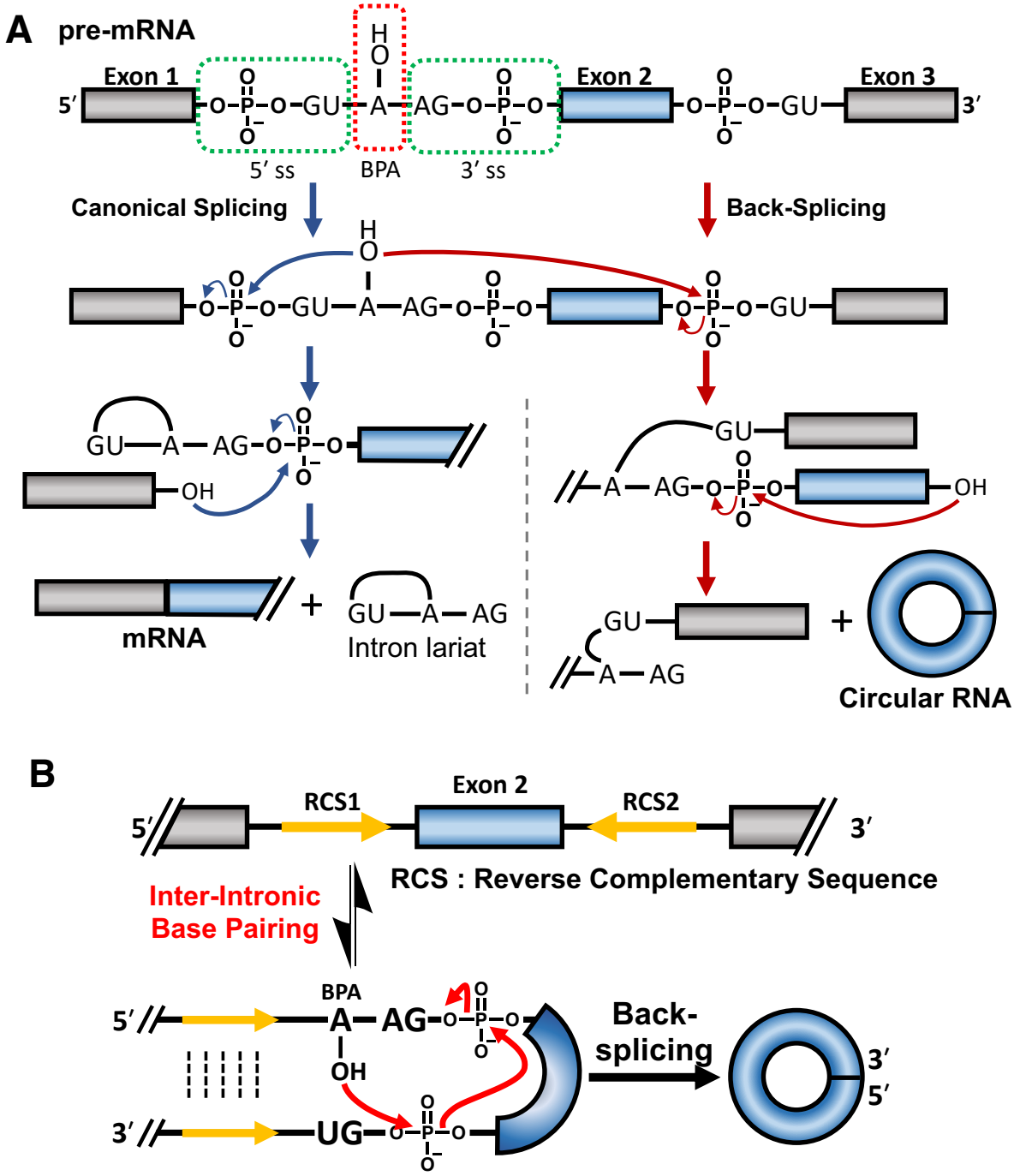


Figure 2. A) Schematic representation and the comparative difference between the canonical splicing process and back-splicing process. **B)** The presence of RSC in introns flanking the circularizing exon facilitates the back-splicing reaction by forming inter-intronic base pairs to bring the upstream BPA in proximity to the downstream 5′ss site.

circularization and, in most cases, the circRNA involves more than one exon, with the most common circRNA having 2~3 exons.[69] How spliceosome manages to loop back across such a long distance to form a splicing complex and the mechanisms involved in their regulation are

still poorly understood. Currently, this is thought to be achieved through the cooperation of various cis- and trans-regulatory elements within the intronic sequence flanking the circularizing exon(s). Undergoing stable inter-intronic interactions to generate a large hairpin-like structure with the circularizing exon(s) looping back to form the loop region, bringing upstream and downstream splice sites in proximity.[12,70] Towards this end, several factors have been proposed to partake in the production of circRNAs, with the most well-studied mechanisms being namely: inter-intronic base-pairing interaction between flanking reverse-complementary sequences (RCS)[69] and cross-intron dimerization and regulation by RBPs[71-74]; details of which will be introduced below. Additional mechanisms, such as the potential involvement of intron lariats in the production of circRNAs have been demonstrated in fission yeast (S. *pombe*),[75] such processes have yet to be confirmed in mammalian cells, they will not be further discussed here.

The presence of reverse-complementary sequences in flanking introns promotes circRNA production

The involvement of flanking RCS in the production was initially described when one of the first mammalian-expressed circRNA was discovered, stemming from the mouse Sry gene, which was expressed as a dominant form of the transcript.[76] Further studies attributed the apparent dominant expression of the circRNA isoform to the presence of ~50 knt of RCS within the introns flanking the circularizing exon of the Sry gene, which forms a highly stable extended hairpin structure necessary for the circularization of exon. While circSry represents an extreme case for flanking RCS length, bioinformatic studies have suggested that shorter RCS pairs that stem from the inverted arrangement of repetitive elements are a major contributor to facilitating the expression of circRNA.[77] Specifically among primates, the ~300 nt long Alu family of short interspersed nuclear elements (SINE: Alu) [77] in addition to other repeat elements such as mammalian interspersed repeats (SINE: MIRs) [78], is cited to be most prominently involved in the production of circRNAs. Though further studies have demonstrated in both expression constructs and endogenous pre-mRNAs, that the length of RCS paring required for the efficient production of circRNAs is likely much shorter, from 36 nt in artificial constructs[79] to as short as 11 nt for endogenous circRNAs (Plt-circR4[80] and circEXOC6B[81] etc.), and the presence of but not the content of RCS pairs promotes circRNA production in cells[69,79].

However, due to splicing occurring in a co-transcriptional manner, the mere presence of such RCSs is not always sufficient for the efficient production of circRNAs. This is due to most introns prior to splicing being subjected to extensive regulation by various RBPs such as ADAR1[21,82,83] and remodeling of interactions by RNA helicases such as DHX9[84]. Furthermore, as most introns contain multiple repetitive elements, the competitive base pairing between the repetitive elements also plays a major role in regulating the production of circRNAs in cells.[69] All of which may destabilize and inhibit the capacity of the RCSs to promote circRNA production.

RBPs promote circRNA production via the formation of cross-intron dimers

Initially, the role of inter-intronic interaction was thought to be singly attributed to the presence of RCSs in the flanking introns, as described above. However, recent advancements revealed the prominent involvement of splice regulatory RBPs in the role of promoting circRNA production. In general, the RBPs are thought to function in a similar manner to RCSs, where the RBPs prompt the formation of the large hairpin structure by binding to target sequences present in the introns flanking the circularizing exon(s) and forming cross-intron dimers.[12,30,70] Examples of RBPs which has shown such functions include the STAR family quaking proteins (QKIs)[71], Mbl[65], NF90/NF110[72], ESRP1[73], and FUS[74].

On a separate note, besides the above-mentioned direct role of promoting circRNA production, RBPs have been also shown to affect the expression of circRNAs in a more indirect manner by affecting the retention time of the intron. Proteins such as splicing factor proline/glutamine-rich (SFPQ) bind to long intronic sequences and have been shown to promote the production of a subset of circRNA by increasing the retention time of the intron within the pre-mRNA, thereby allowing sufficient time for the base pairing of RCSs between the flanking introns.[85]

Overall, the production of circRNAs occurs in a highly combinatorial manner, where many factors in the likes of RBPs and RCS competition tightly regulate the efficiency and production of circRNAs. In general, however, the formation and stability of the intron-exon-intron hairpin structure plays an essential role in the production of circRNAs in cells.[69,79,82]

The current state of circRNAs as a therapeutic target

Due to their ubiquitous presence and characteristic functionalities in cells, circRNAs have recently garnered much interest in the field of biology and medical sciences.[9,20,21,26,86] This has advanced our understanding of their involvement in the development and progression of

various diseases significantly over the past years, with many reports indicating the dysregulation of circRNAs being a major factor.[28-30] As such, circRNAs have become a new potential ncRNA target for therapeutics, and the capacity to regulate their expression has become a particular interest.[87-89] Towards this end, several methods have been developed to achieve the desired regulation of circRNAs in the cellular environment.

Briefly, there are currently two major methodologies to achieve the upregulation of circRNA in cells. The first method relies on in vitro preparation of the circRNA followed by transfection into cells.[90,91] To achieve this, the current state-of-the-art method utilizes a specifically designed pre-mRNA which is comprised of the desired circRNA sequence flanked by two halves of a self-splicing intron. Following the in vitro transcription, the introns undergo self-removal from the pre-mRNA in the presence of a co-factor (usually a guanosine monophosphate), and the circularization of the sequence in between the self-splicing intron occurs as a result. The circRNA is then purified and transfected into cells to achieve upregulation. The second method relies on the cellular splicing machinery to directly overexpress the circRNA in cells.[92,93] Where the overexpression vector is encoded to transcribe a pre-mRNA containing a desired "exon" sequence containing the canonical 5′ and 3′ splice sites flanked by a pair of RCSs. The circRNAs are produced in cells following the cellular splicing mechanism described previously.

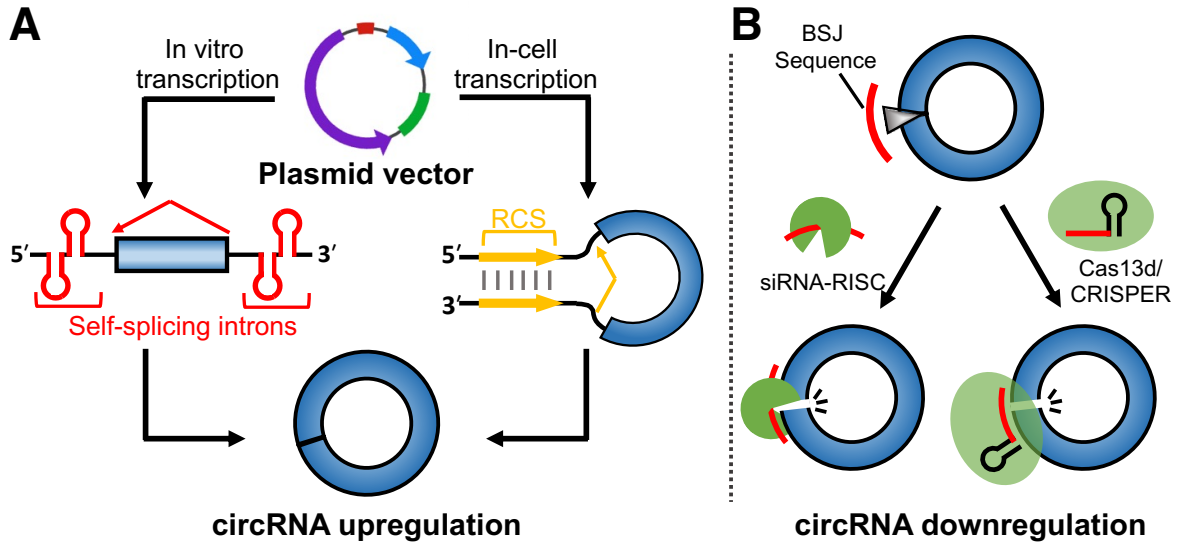


Figure 3. The various current state-of-the-art methodologies used to **A)** upregulate and **B)** downregulate circRNA in cells.

The downregulation of circRNAs on the other hand, has been demonstrated to be achievable by traditional targeted RNA knockdown strategies, such as the utilization of siRNAs or shRNAs, and more recently, Cas13d/CRISPER systems.[94-97] However, as circRNAs share the

same sequence with their cognate mRNAs, the designs of such knockdown systems are limited to targeting a stretch of sequences that span the BSJ junction to achieve circRNA selective downregulation.

The general aim of this thesis: Can we directly regulate circRNAs?

The characteristic potentials of circRNAs put forth a new challenge to selectively target and modulate specific circRNA expression, and it is poised to provide a new vector for regulating cellular function and progression of diseases. However, despite the recent advancements in circRNA modulation as described above, current approaches still lack the capacity to directly modulate circRNAs expression at the pre-mRNA level. Towards this end, based on the above-mentioned characteristic mechanism involved in the production of circRNA, the back-splicing reaction, this thesis aims to explore the feasibility of utilizing externally introduced trans-acting factors to facilitate and stabilize the inter-intronic interaction between flanking introns. Ultimately upregulating circRNA production from a pre-mRNA. The model examples demonstrated in this thesis hope to potentially provide a stepping-stone for new strategies to regulate the production of biologically relevant endogenous circRNAs.

Survey of this thesis

The content of this thesis is divided into two chapters, with each chapter dedicated to exploring the applicability of small molecules and oligonucleotides (ONs), respectively, in regulating circRNA production in cells. To demonstrate the concept, a previously reported model pre-mRNA producing a well-studied circRNA, circZKSCAN1, was used.[79] Overall, it covers the two currently major modalities in targeted modulation of RNA functions, demonstrating the broad applicability of the proposed concept.

Chapter 1 explores the feasibility of utilizing an RNA binding small molecule to promote circRNA production in a cellular environment. Here, naphthyridine carbamate dimer (**NCD**) was prepared and used as the small molecule, and the RCS pair in the introns flanking the circularizing exons in the model pre-mRNA was modified to contain the target sequence of **NCD**. Firstly, the affinity of **NCD** towards the target sequence introduced within the RCS pair of the modified pre-mRNA was evaluated using surface plasmon resonance. **NCD**'s capacity to upregulate circRNA production was then evaluated using HeLa cells expressing the

modified model pre-mRNA, and the upregulation of circZKSCAN1 was quantified using RT-qPCR.

Chapter 2 explores the feasibility of utilizing a modified ON to promote circRNA production in a cellular environment. Firstly, a 25-nt long ON, fully 2′-OMe, phosphorothioate backbone-modified, was designed to target and bridge the two distal intronic sequences flanking the pair of RCSs, a CircuLarIzation Promoting OligoNucleotide (**CLIP-ON**). The capacity for the designed **CLIP-ON** to upregulate circZKSCAN1 in HeLa cells expressing the model pre-mRNA was then evaluated by RT-qPCR. The formation of the tertiary structure, the bridged structure, was confirmed *in vitro* using gel-electrophoresis and hybridizing affinity by isothermal calorimetry and surface plasmon resonance. The broad applicability of **CLIP-ON** design for promoting circRNA production was further demonstrated using model pre-mRNA with destabilized RCS, and exons of other circRNAs. Finally, the involvement of various design elements in **CLIP-ON**'s biological effectiveness was elucidated.

References:

1 U. Braunschweig, S. Gueroussov, A. M. Plocik, B. R. Graveley and B. J. Blencowe, *Cell*, 2013, **152**, 1252–1269.

2 T. W. Nilsen and B. R. Graveley, *Nature*, 2010, **463**, 457–463.

3 M. C. Wahl, C. L. Will and R. Lührmann, *Cell*, 2009, **136**, 701–718.

4 E. T. Wang, R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S. F. Kingsmore, G. P. Schroth and C. B. Burge, *Nature*, 2008, **456**, 470–476.

5 J. F. Cáceres and A. R. Kornblihtt, *Trends Genet.*, 2002, **18**, 186–193.

6 C. W. J. Smith and J. Valcárcel, *Trends Biochem. Sci.*, 2000, **25**, 381–388.

7 S. Braun, H. Domdey and K. Wiebauer, *Nucleic Acids Res.*, 1996, **24**, 4152–4157.

8 W. R. Jeck, J. A. Sorrentino, K. Wang, M. K. Slevin, C. E. Burd, J. Liu, W. F. Marzluff and N. E. Sharpless, *RNA*, 2013, **19**, 141–157.

9 S. Memczak, M. Jens, A. Elefsinioti, F. Torti, J. Krueger, A. Rybak, L. Maier, S. D. Mackowiak, L. H. Gregersen, M. Munschauer, A. Loewer, U. Ziebold, M. Landthaler, C. Kocks, F. Le Noble and N. Rajewsky, *Nature*, 2013, **495**, 333–338.

10 J. Salzman, C. Gawad, P. L. Wang, N. Lacayo and P. O. Brown, *PLoS One*, 2012, **7**, e30733.

11 W. R. Jeck and N. E. Sharpless, *Nat. Biotechnol.*, 2014, **32**, 453–461.

12 L. S. Kristensen, M. S. Andersen, L. V. W. Stagsted, K. K. Ebbesen, T. B. Hansen and J. Kjems, *Nat. Rev. Genet.*, 2019, **20**, 675–691.

13 X. Ma, S. Zhai and L. Yang, *Trends Genet.*, 2023, **39**, 897–907.

14 J. U. Guo, V. Agarwal, H. Guo and D. P. Bartel, *Genome Biol.*, 2014, **15**, 409.

15 Z. Li, C. Huang, C. Bao, L. Chen, M. Lin, X. Wang, G. Zhong, B. Yu, W. Hu, L. Dai, P. Zhu, Z. Chang, Q. Wu, Y. Zhao, Y. Jia, P. Xu, H. Liu and G. Shan, *Nat. Struct. Mol. Biol.*, 2015, **22**, 256–264.

16 K. Abdelmohsen, A. C. Panda, S. De, I. Grammatikakis, J. Kim, J. Ding, J. H. Noh, K. M. Kim, J. A. Mattison, R. de Cabo and M. Gorospe, *Aging (Albany. NY).*, 2015, **7**, 903–910.

17 H. J. Gross, H. Domdey, C. Lossow, P. Jank, M. Raba, H. Alberty and H. L. Sänger, *Nature*, 1978, **273**, 203–208.

18 J. M. Nigro, K. R. Cho, E. R. Fearon, S. E. Kern, J. M. Ruppert, J. D. Oliner, K. W. Kinzler and B. Vogelstein, *Cell*, 1991, **64**, 607–613.

19 L. Szabo, R. Morey, N. J. Palpant, P. L. Wang, N. Afari, C. Jiang, M. M. Parast, C. E. Murry, L. C. Laurent and J. Salzman, *Genome Biol.*, 2015, **16**, 126.

20 P. G. Maass, P. Glažar, S. Memczak, G. Dittmar, I. Hollfinger, L. Schreyer, A. V. Sauer, O. Toka, A. Aiuti, F. C. Luft and N. Rajewsky, *J. Mol. Med.*, 2017, **95**, 1179–1189.

21 A. Rybak-Wolf, C. Stottmeister, P. Glažar, M. Jens, N. Pino, S. Giusti, M. Hanan, M. Behm, O. Bartok, R. Ashwal-Fluss, M. Herzog, L. Schreyer, P. Papavasileiou, A. Ivanov, M. Öhman, D. Refojo, S. Kadener and N. Rajewsky, *Mol. Cell*, 2014, **58**, 870–885.

22 L. Szabo and J. Salzman, *Nat. Rev. Genet.*, 2016, **17**, 679–692.

23 P. Glažar, P. Papavasileiou and N. Rajewsky, *Rna*, 2014, **20**, 1666–1670.

24 X. O. Zhang, R. Dong, Y. Zhang, J. L. Zhang, Z. Luo, J. Zhang, L. L. Chen and L. Yang, *Genome Res.*, 2016, **26**, 1277–1287.

25 D. B. Dudekula, A. C. Panda, I. Grammatikakis, S. De, K. Abdelmohsen and M. Gorospe, *RNA Biol.*, 2016, **13**, 34–42.

26 J. Salzman, R. E. Chen, M. N. Olsen, P. L. Wang and P. O. Brown, *PLoS Genet.*, 2013, **9**, e1003777.

27 S. Xia, J. Feng, L. Lei, J. Hu, L. Xia, J. Wang, Y. Xiang, L. Liu, S. Zhong, L. Han and C. He, *Brief. Bioinform.*, 2017, **18**, 984–992.

28 B. Han, J. Chao and H. Yao, *Pharmacol. Ther.*, 2018, **187**, 31–44.

29 J. N. Vo, M. Cieslik, Y. Y. Zhang, S. Shukla, L. Xiao, Y. Y. Zhang, Y. M. Wu, S. M. Dhanasekaran, C. G. Engelke, X. Cao, D. R. Robinson, A. I. Nesvizhskii and A. M. Chinnaiyan, *Cell*, 2019, **176**, 869-881.e13.

30 L. S. Kristensen, T. Jakobsen, H. Hager and J. Kjems, *Nat. Rev. Clin. Oncol.*, 2022, **19**, 188–206.

31 T. B. Hansen, T. I. Jensen, B. H. Clausen, J. B. Bramsen, B. Finsen, C. K. Damgaard and J. Kjems, *Nature*, 2013, **495**, 384–388.

32 J. Zang, D. Lu and A. Xu, *J. Neurosci. Res.*, 2020, **98**, 87–97.

33 A. Huang, H. Zheng, Z. Wu, M. Chen and Y. Huang, *Theranostics*, 2020, **10**, 3503–3517.

34 N. R. Pamudurti, O. Bartok, M. Jens, R. Ashwal-Fluss, C. Stottmeister, L. Ruhe, M. Hanan, E. Wyler, D. Perez-Hernandez, E. Ramberger, S. Shenzis, M. Samson, G. Dittmar, M. Landthaler, M. Chekulaeva, N. Rajewsky and S. Kadener, *Mol. Cell*, 2017, **66**, 9-21.e7.

35 R. Shang, S. Lee, G. Senavirathne and E. C. Lai, *Nat. Rev. Genet.*, 2023, **24**, 816–833.

36 B. Kleaveland, C. Y. Shi, J. Stefano and D. P. Bartel, *Cell*, 2018, **174**, 350-362.e17.

37 M. Piwecka, P. Glažar, L. R. Hernandez-Miranda, S. Memczak, S. A. Wolf, A. Rybak-Wolf, A. Filipchyk, F. Klironomos, C. A. Cerda Jara, P. Fenske, T. Trimbuch, V. Zywitza, M. Plass, L. Schreyer, S. Ayoub, C. Kocks, R. Kühn, C. Rosenmund, C. Birchmeier and N. Rajewsky, *Science* **357** (6357), eaam8526.

38 L. S. Kristensen, T. L. H. Okholm, M. T. Venø and J. Kjems, *RNA Biol.*, 2018, **15**, 280–291.

39 Q. Zheng, C. Bao, W. Guo, S. Li, J. Chen, B. Chen, Y. Luo, D. Lyu, Y. Li, G. Shi, L. Liang, J. Gu, X. He and S. Huang, *Nat. Commun.*, 2016, **7**, 11215.

40 J. Zheng, X. Liu, Y. Xue, W. Gong, J. Ma, Z. Xi, Z. Que and Y. Liu, *J. Hematol. Oncol.*, 2017, **10**, 52.

41 Y. Yang, X. Gao, M. Zhang, S. Yan, C. Sun, F. Xiao, N. Huang, X. Yang, K. Zhao, H. Zhou, S. Huang, B. Xie and N. Zhang, *JNCI J. Natl. Cancer Inst.*, 2018, **110**, 304–315.

42 Z. Yao, J. Luo, K. Hu, J. Lin, H. Huang, Q. Wang, P. Zhang, Z. Xiong, C. He, Z. Huang, B. Liu and Y. Yang, *Mol. Oncol.*, 2017, **11**, 422–437.

43 S. Ghafouri-Fard, T. Khoshbakht, M. Taheri and E. Jamali, *Front. Oncol.*, 2021, **11**, 1–10.

44 F. Li, L. Zhang, W. Li, J. Deng, J. Zheng, M. An, J. Lu and Y. Zhou, *Oncotarget*, 2015, **6**, 6001–6013.

45 J. Chen, Y. Li, Q. Zheng, C. Bao, J. He, B. Chen, D. Lyu, B. Zheng, Y. Xu, Z. Long, Y. Zhou, H. Zhu, Y. Wang, X. He, Y. Shi and S. Huang, *Cancer Lett.*, 2017, **388**, 208–219.

46 Q. Yang, W. W. Du, N. Wu, W. Yang, F. M. Awan, L. Fang, J. Ma, X. Li, Y. Zeng, Z. Yang, J. Dong, A. Khorshidi and B. B. Yang, *Cell Death Differ.*, 2017, **24**, 1609–1620.

47 H.-F. Liang, X.-Z. Zhang, B.-G. Liu, G.-T. Jia and W.-L. Li, *Am. J. Cancer Res.*, 2017, **7**, 1566–1576.

48 W. Yang, W. W. Du, X. Li, A. J. Yee and B. B. Yang, *Oncogene*, 2016, **35**, 3919–3931.

49 W. W. Du, W. Yang, Y. Chen, Z.-K. Wu, F. S. Foster, Z. Yang, X. Li and B. B. Yang, *Eur. Heart J.*, 2016, **38**, ehw001.

50 W. J. Lukiw, *Front. Genet.*, 2013, **4**, 1–2.

51 M. W. Hentze, A. Castello, T. Schwarzl and T. Preiss, *Nat. Rev. Mol. Cell Biol.*, 2018, **19**, 327–341.

52 K. Abdelmohsen, A. C. Panda, R. Munk, I. Grammatikakis, D. B. Dudekula, S. De, J. Kim, J. H. Noh, K. M. Kim, J. L. Martindale and M. Gorospe, *RNA Biol.*, 2017, **14**, 361–369.

53 T. Yamazaki, S. Souquere, T. Chujo, S. Kobelke, Y. S. Chong, A. H. Fox, C. S. Bond, S. Nakagawa, G. Pierron and T. Hirose, *Mol. Cell*, 2018, **70**, 1038-1053.e7.

54 W. W. Du, W. Yang, E. Liu, Z. Yang, P. Dhaliwal and B. B. Yang, *Nucleic Acids Res.*, 2016, **44**, 2846–2858.

55 C. Zhang, S. Wang, F. Chao, G. Jia, X. Ye, D. Han, Z. Wei, J. Liu, G. Xu and G. Chen, *Mol. Ther.*, 2023, **31**, 1705–1721.

56 T. B. Hansen, *Methods*, 2021, **196**, 68–73.

57 L. V. W. Stagsted, K. M. Nielsen, I. Daugaard and T. B. Hansen, *Life Sci. Alliance*, 2019, **2**, e201900398.

58 I. Legnini, G. Di Timoteo, F. Rossi, M. Morlando, F. Briganti, O. Sthandier, A. Fatica, T. Santini, A. Andronache, M. Wade, P. Laneve, N. Rajewsky and I. Bozzoni, *Mol. Cell*, 2017, **66**, 22-37.e9.

59 M. Zhang, N. Huang, X. Yang, J. Luo, S. Yan, F. Xiao, W. Chen, X. Gao, K. Zhao, H. Zhou, Z. Li, L. Ming, B. Xie and N. Zhang, *Oncogene*, 2018, **37**, 1805–1814.

60 Y. Li, B. Chen, J. Zhao, Q. Li, S. Chen, T. Guo, Y. Li, H. Lai, Z. Chen, Z. Meng, W. Guo, X. He and S. Huang, *Adv. Sci.*, 2021, **8**, 13, 2001701

61 R. Chen, S. K. Wang, J. A. Belk, L. Amaya, Z. Li, A. Cardenas, B. T. Abe, C. K. Chen, P. A. Wender and H. Y. Chang, *Nat. Biotechnol.*, 2023, **41**, 262–272.

62 C. K. Chen, R. Cheng, J. Demeter, J. Chen, S. Weingarten-Gabbay, L. Jiang, M. P. Snyder, J. S. Weissman, E. Segal, P. K. Jackson and H. Y. Chang, *Mol. Cell*, 2021, **81**, 4300-4318.e13.

63 Y. Wang and Z. Wang, *RNA*, 2015, **21**, 172–179.

64 N. Abe, K. Matsumoto, M. Nishihara, Y. Nakano, A. Shibata, H. Maruyama, S. Shuto, A. Matsuda, M. Yoshida, Y. Ito and H. Abe, *Sci. Rep.*, 2015, **5**, 1–9.

65 R. Ashwal-Fluss, M. Meyer, N. R. Pamudurti, A. Ivanov, O. Bartok, M. Hanan, N. Evantal, S. Memczak, N. Rajewsky and S. Kadener, *Mol. Cell*, 2014, **56**, 55–66.

66 S. Starke, I. Jost, O. Rossbach, T. Schneider, S. Schreiner, L. H. Hung and A. Bindereif, *Cell Rep.*, 2015, **10**, 103–111.

67 M. M. Konarska, P. J. Grabowski, R. A. Padgett and P. A. Sharp, *Nature*, 1985, **313**, 552–557.

68 X. Li, S. Liu, L. Zhang, A. Issaian, R. C. Hill, S. Espinosa, S. Shi, Y. Cui, K. Kappel, R. Das, K. C. Hansen, Z. H. Zhou and R. Zhao, *Nature*, 2019, **573**, 375–380.

69 X.-O. Zhang, H.-B. Wang, Y. Zhang, X. Lu, L.-L. Chen and L. Yang, *Cell*, 2014, **159**, 134–147.

70 L.-L. Chen, *Nat. Rev. Mol. Cell Biol.*, 2016, **17**, 205–211.

71 S. J. Conn, K. A. Pillman, J. Toubia, V. M. Conn, M. Salmanidis, C. A. Phillips, S. Roslan, A. W. Schreiber, P. A. Gregory and G. J. Goodall, *Cell*, 2015, **160**, 1125–1134.

72 X. Li, C. X. Liu, W. Xue, Y. Zhang, S. Jiang, Q. F. Yin, J. Wei, R. W. Yao, L. Yang and L. L. Chen, *Mol. Cell*, 2017, **67**, 214-227.e7.

73 D. Liu, B. K. Dredge, A. G. Bert, K. A. Pillman, J. Toubia, W. Guo, B. J. A. Dyakov, M. M. Migault, V. M. Conn, S. J. Conn, P. A. Gregory, A. Gingras, D. Patel, B. Wu and G. J. Goodall, *Nucleic Acids Res.*, 2023, 1–17.

74 L. Errichelli, S. Dini Modigliani, P. Laneve, A. Colantoni, I. Legnini, D. Capauto, A. Rosa, R. De Santis, R. Scarfò, G. Peruzzi, L. Lu, E. Caffarelli, N. A. Shneider, M. Morlando and I. Bozzoni, *Nat. Commun.*, 2017, **8**, 14741.

75 S. P. Barrett, P. L. Wang and J. Salzman, *Elife*, 2015, **4**, 1–18.

76 B. Capel, A. Swain, S. Nicolis, A. Hacker, M. Walter, P. Koopman, P. Goodfellow and R. Lovell-Badge, *Cell*, 1993, **73**, 1019–1030.

77 W. R. Jeck, J. A. Sorrentino, K. Wang, M. K. Slevin, C. E. Burd, J. Liu, W. F. Marzluff and N. E. Sharpless, *RNA*, 2013, **19**, 141–157.

78 R. Yoshimoto, K. Rahimi, T. B. Hansen, J. Kjems and A. Mayeda, *iScience*, 2020, **23**, 101345.

79 D. Liang and J. E. Wilusz, *Genes Dev.*, 2014, **28**, 2233–2247.

80 C. Preußer, L. H. Hung, T. Schneider, S. Schreiner, M. Hardt, A. Moebus, S. Santoso and A. Bindereif, *J. Extracell. Vesicles*, **7**, 1424473

81 C. Zhang, S. Wang, F. Chao, G. Jia, X. Ye, D. Han, Z. Wei, J. Liu, G. Xu and G. Chen, *Mol. Ther.*, 2023, **31**, 1705–1721.

82 A. Ivanov, S. Memczak, E. Wyler, F. Torti, H. T. Porath, M. R. Orejuela, M. Piechotta, E. Y. Levanon, M. Landthaler, C. Dieterich and N. Rajewsky, *Cell Rep.*, 2015, **10**, 170–177.

83 L. Shi, P. Yan, Y. Liang, Y. Sun, J. Shen, S. Zhou, H. Lin, X. Liang and X. Cai, *Cell Death Dis.*, 2017, **8**, e3171–e3171.

84 T. Aktaş, I. A. Ilik, D. Maticzka, V. Bhardwaj, C. Pessoa Rodrigues, G. Mittler, T. Manke, R. Backofen and A. Akhtar, *Nature*, 2017, **544**, 115–119.

85 L. V. W. Stagsted, E. T. O'leary, K. K. Ebbesen and T. B. Hansen, *Elife*, 2021, **10**, 1–26.

86 E. W. Ottesen, D. Luo, J. Seo, N. N. Singh and R. N. Singh, *Nucleic Acids Res.*, 2019, **47**, 2884–2905.

87 L. S. Kristensen, T. B. Hansen, M. T. Venø and J. Kjems, *Oncogene*, 2018, **37**, 555–565.

88 L. M. Holdt, A. Kohlmaier and D. Teupser, *Front. Physiol.*, **9**, 1262

89 L. Santer, C. Bär and T. Thum, *Mol. Ther.*, 2019, **27**, 1350–1363.

90 R. A. Wesselhoeft, P. S. Kowalski and D. G. Anderson, *Nat. Commun.*, 2018, **9**, 2629.

91 R. Chen, S. K. Wang, J. A. Belk, L. Amaya, Z. Li, A. Cardenas, B. T. Abe, C. K. Chen, P. A. Wender and H. Y. Chang, *Nat. Biotechnol.*, 2023, **41**, 262–272.

92 D. Liu, V. Conn, G. J. Goodall and S. J. Conn, *Methods Mol. Biol.*, 2018, **1724**, 97–105.

93 B. W. Stringer, M. Gabryelska, S. Marri, L. Clark, H. Lin, R. Liu, J. E. Wilusz, V. M. Conn and S. J. Conn, *bioRxiv*, 2023, DOI: 10.1101/2023.11.21.568171

94 N. R. Pamudurti, I. L. Patop, A. Krishnamoorthy, R. Ashwal-Fluss, O. Bartok and S. Kadener, *Cell Discov.*, 2020, **6**, 52.

95 Q. Yang, F. Li, A. T. He and B. B. Yang, *Mol. Ther.*, 2021, **29**, 1683–1702.

96 S. Li, X. Li, W. Xue, L. Zhang, L.-Z. Yang, S.-M. Cao, Y.-N. Lei, C.-X. Liu, S.-K. Guo, L. Shan, M. Wu, X. Tao, J.-L. Zhang, X. Gao, J. Zhang, J. Wei, J. Li, L. Yang and L.-L. Chen, *Nat. Methods*, 2021, **18**, 51–59.

97 Y. Zhang, T. M. Nguyen, X.-O. Zhang, L. Wang, T. Phan, J. G. Clohessy and P. P. Pandolfi, *Genome Biol.*, 2021, **22**, 41.

Chapter 1

Mismatch binding ligand upregulated back-splicing reaction producing circular RNA in a cellular model

Abstract

One of the most peculiar features of circRNAs is the process in which it's produced in cells, involving a specialized splicing process which was termed: back-splicing. The major defining feature of this splicing reaction involves the pre-requisite requirement of inter-intronic interaction between the introns flanking the circularizing exons to facilitate the process. In this chapter, we demonstrate that RNA binding small molecules such as naphthyridine carbamate dimer (**NCD**), can be used to stabilize and facilitate the necessary inter-intronic interaction and upregulate the production of a circRNA using a model pre-mRNA in a cellular environment. Overall, it demonstrates the feasibility of small-molecule mediated circRNA production.

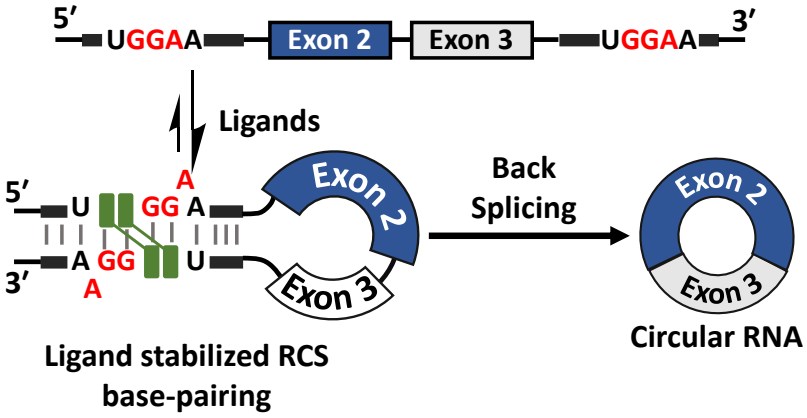


Figure 1. Schematic representation of the core concept explored in this chapter. Here, RNA binding small molecule promoted circRNA production from a pre-mRNA via stabilization of interactions between the introns flanking the circularizing exons.

Introduction

As described in the previous section, "General Introduction", circRNAs as a class of ncRNA have recently been demonstrated to have diverse functions in the cellular environment, and their dysregulation was shown to play major roles in the development and progression of various illnesses. Hence with such characteristic potentials for circRNAs, regulation of the circRNA content in cells has become an attractive target for therapeutic development, to eventually provide a new vector in regulating cellular function.

Recently, there has been significant development in the designs and application of small molecular designs that could selectively target and bind to specific RNA sequences, modulating the functionality of the target RNAs.[1-3] Of which, we have recently reported a series of mismatch binding ligands (MBL) which have demonstrated the capacity to regulate the functions of RNAs by facilitating the formation of ligand-stabilized RNA structures.[3] The potency of the molecules and the applicability of such a concept were demonstrated previously, where small molecule stabilized tertiary contacts between RNA sequences induced the activation of ribozymes[4], and activation of programmed pseudoknot transcriptional frameshifting[5]. Considering the unique biosynthetic mechanism, back-splicing, involved in the production of circRNAs is promoted in the presence of stable inter-intronic interactions between introns flanking the circularizing exons.[6,7] I hypothesized that MBLs could potentially be utilized as an externally introduced trans-factor to induce and stabilize such intronic interactions in the presence of suitable mismatch sites and promote the production of circRNA in a cellular environment.

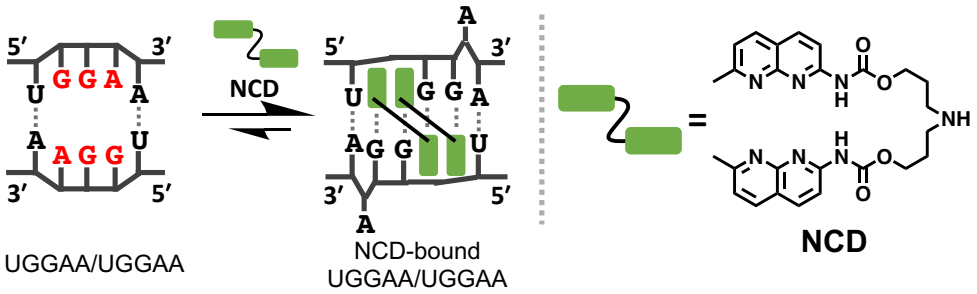


Figure 2. Schematic diagram of NCD binding to RNA 5′-UGGAA-3′/5′-UGGAA-3′ duplex sequence

We recently reported RNA-binding MBL, **NCD**, which was reported to form a 2:1 complex with 5′-UGGAA-3′/5′-UGGAA-3′ motif (hereafter as UGGAA/UGGAA) in duplex RNAs.[8] For which **NCD** was shown to stabilize the putative 5′-GGA-3′/5′-GGA-3′ trinucleotide mismatch present in the motif. The complex structure was determined by NMR, which revealed the binding of two **NCD** molecules to four guanines of the internal loop present in the UGGAA/UGGAA with a concomitant flipping out of the adenine bases (Fig. 2). In this chapter, we report that **NCD** binding to the engineered RCSs containing the UGGAA/UGGAA motif increased the production of circRNA from exons flanked by those RCSs in HeLa cells. We also show that the upregulation of circRNA production is sensitively dependent on the presence of the UGGAA/UGGAA motif and **NCD** concentration, demonstrating that the increased formation of circRNA is due to the proximal location of BPA to the 5′ss in back-splicing.

Results:

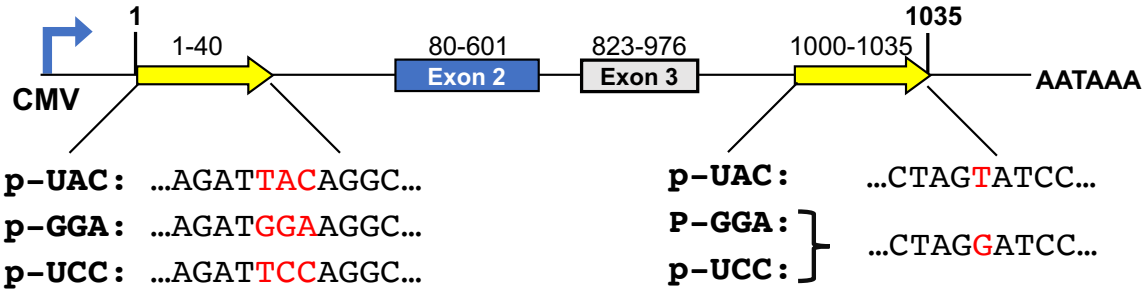
Model pre-mRNA design, and in vitro binding assessment of NCD towards target intron sequence

Model pre-mRNA and experimental design

To demonstrate the feasibility of the concept, a model pre-mRNA previously reported by Wilusz et.al. was used.[9] Briefly, the pre-mRNA consists of exons 2 and 3 of gene ZKSCAN1, flanked by a pair of 36-nt RCSs (Scheme 1, Sup. Fig. 1A, the expression plasmid hereon as ***p*-UAC** and transcribed pre-mRNA as **pre-mUAC**), which was determined to be the minimum length necessary for promoting the production of a well-studied circRNA: circZKSCAN1 via an RCS-facilitated back-splicing pathway.[9,10] Furthermore, it was reported that the introduction of additional destabilizing factors, e.g. introduction of mismatch sites, resulted in impaired circRNA production. The plasmids that expressed the target model pre-mRNA containing **NCD** responsive motifs in its RCSs were then promptly designed using the ***p*-UAC** as the template. Briefly, primer-directed mutagenesis was used to introduce a point mutation, T to G, and a trinucleotide mutation, TAC to GGA, in the complementary region of the introns down and upstream of the circularizing exons, respectively. This results in a plasmid that expresses a model pre-mRNA that contains the potential motif for **NCD**, UGGAA/UGGAA, within the reverse complementary sequences (the expression plasmid hereon as ***p*-GGA** and transcribed pre-mRNA as **pre-mGGA**). Furthermore, a plasmid expressing the control pre-mRNA of **pre-mGGA** containing the full match 5′-UUCCA-3′/5′-UGGAA-3′ motif in the same position, was also prepared in a similar manner. (the expression plasmid hereon as *p*-UCC and transcribed pre-mRNA as **pre-mUCC**). The schematic representation of the plasmid designs and modification are shown in **Scheme 1**, the details of the preparation are described in the experimental section.

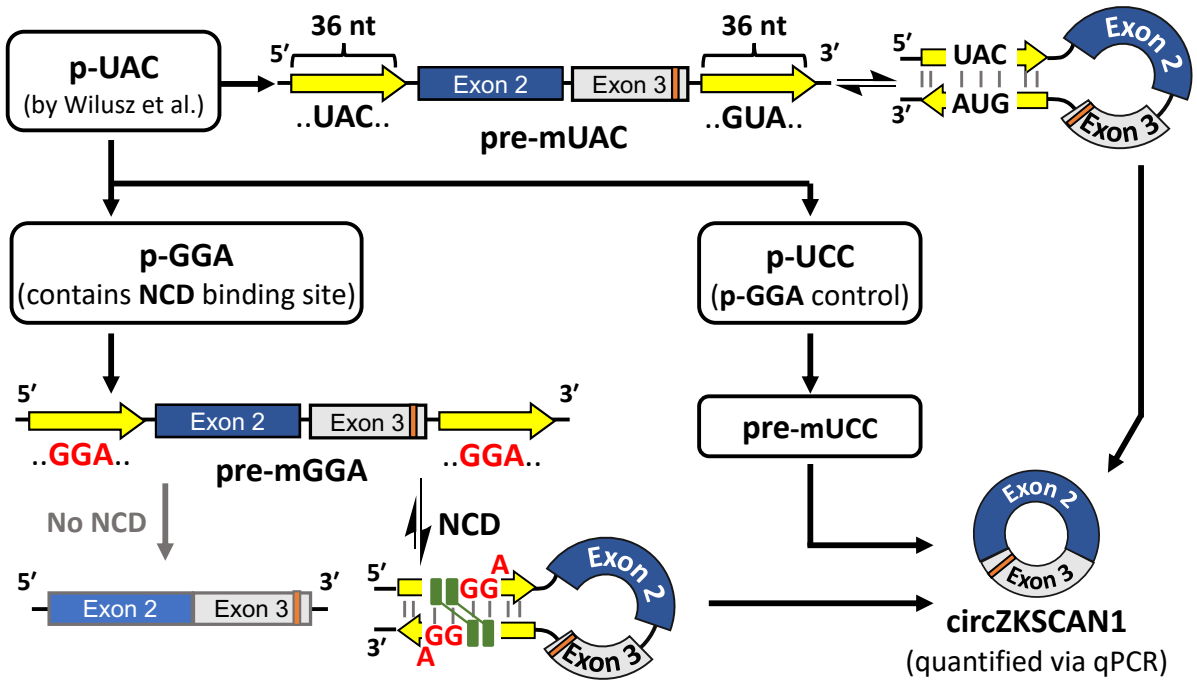
Here, the introduced mismatches in **pre-mGGA** are expected to reduce the thermal stability of the base pairing between the RCSs in the flanking introns and impair the formation of the necessary hairpin for circRNA production as reported previously.[9] However, the mismatch is expected to be recognized by **NCD**, and in the presence of the compound, the binding of **NCD** is expected to stabilize the stable hairpin structure and rescue the impaired production. On the other hand, the full-match counterpart **pre-mUCC** is not expected to show any changes

regardless of the presence of **NCD**. The schematic representation of the overall experimental flow is shown in **Scheme 2**.



Plasmid	Resulting RNA transcript	Resulting circRNA
p-UAC	pre-mUAC	circZKSCAN1 containing a single nucleotide difference at 901
p-GGA	pre-mGGA	
p-UCC	pre-mUCC	

Scheme 1: Specific mutation introduced to generate the **pre-mGGA** and **pre-mUCC** expressing construct ***p*-GGA** and ***p*-UCC** respectively, from the original design, ***p*-UAC**, previously reported by Wilusz et al., the yellow arrows denote the RCSs in intronic sequences, the mutated nucleotides are highlighted in red, and the length of the sequences is denoted underneath individual elements.



Scheme 2: Schematic representation of plasmid designs and general flow of the experiment.

In vitro binding assessment of NCD towards target RCSs via surface plasmon resonance assay (SPR)

Prior to starting in-cell biological experiments, SPR experiments were conducted to confirm the **NCD** binding to the UGGAA/UGGAA motif which was introduced into the complementary region of **pre-mGGA**'s RCS pair. On streptavidin-coated sensors (SA chip, GE healthcare), hairpin RNAs containing the partial sequence of the RCS pair from the intron of **pre-mGGA** or **pre-mUCC** were immobilized, respectively (Fig. 3A). **NCD** was prepared accordingly as described in previous reports.[8,10-12] In addition, an analog of **NCD**, quinoline carbamate dimer **QCD**, was also prepared as a compound control, as it was reported to lack the capacity to bind to RNA.[8]

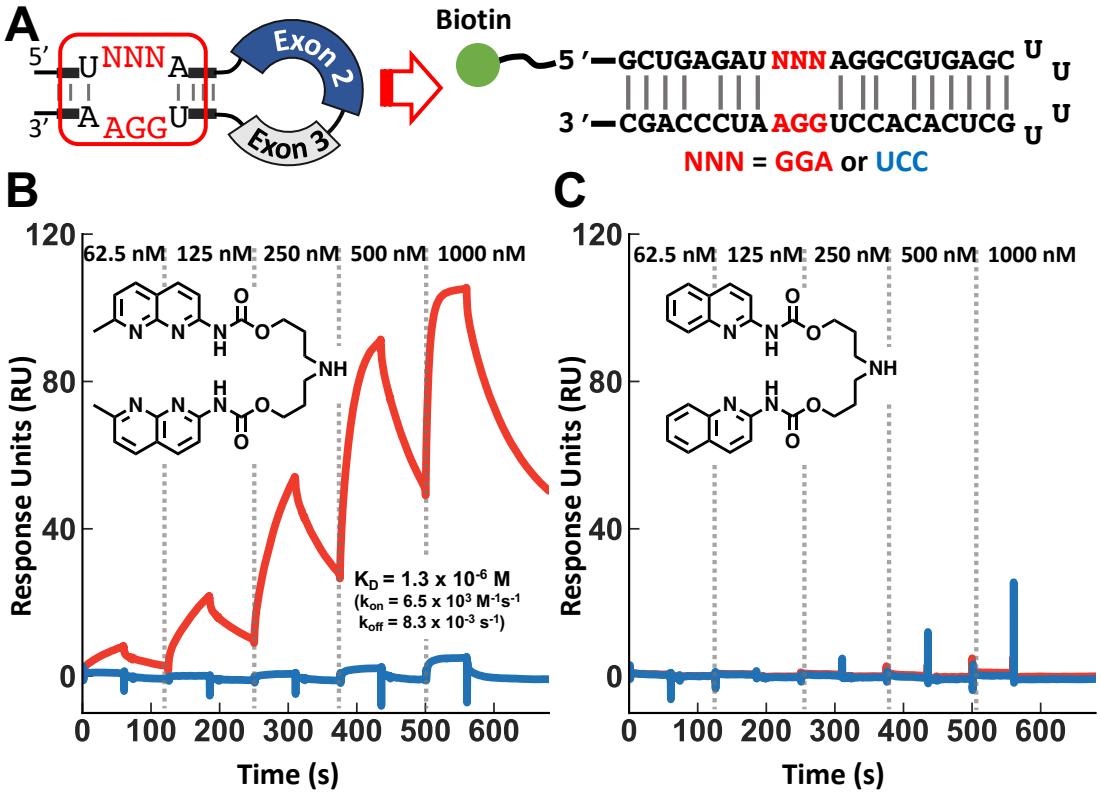


Figure 3. A) The full sequence of the immobilized hairpin RNA used, the mutated area is highlighted in red. (where NNN = GGA or UCC), B) SPR response of GGA mismatch (red) and full match (blue) RNA in the presence of **NCD** (left). C) SPR response of GGA mismatch (red) and full match (blue) RNA in the presence of **QCD** (right)

The results expectedly showed **NCD** invoking a strong SPR response for the immobilized RNA containing the partial sequence of **pre-mGGA**. Indicating **NCD**'s high binding affinity towards the mismatch motif within **pre-mGGA** (Fig. 3B). The equilibrium dissociation constant which was determined using the full range of the SPR response data resulted in the micromolar range (apparent K_D = 1.29 x 10^{-6} M). However, upon further analysis, it was revealed that the

response curves obtained in the two lowest concentration ranges (62.5 nM and 125 nM) showed a much lower affinity (apparent K_D = 1.69 x 10^{-5} M) compared to the higher concentration ranges (250 nM to 1000 nM) which showed a much higher affinity akin to previously reported values (apparent K_D = 1.66 x 10^{-7} M). The observed 2-order deviation in the K_D was attributed to the 2:1 binding nature of NCD towards the target site, as in the lower concentration ranges such binding arrangements may not be favorable, resulting in the observed lower binding affinity. On the other hand, full-match RNA counterpart, the partial **pre-mUCC** sequence, showed a weak SPR response towards the presence of **NCD**. While **QCD** was not able to induce any response from either partial RNA sequence used. (Fig. 3C)

Overall, these results confirm the selectivity and the high affinity of **NCD** towards RNA containing the target UGGAA/UGGAA mismatch motif and lack the capacity to bind to the control full match sequence. Additionally, confirming the lack of **QCD**'s capacity to bind to the partial RNA sequence of **pre-mGGA** or **pre-mUCC**. Demonstrating the suitability of utilizing **pre-mUCC** and **QCD** as a control for **pre-mGGA** and **NCD**, respectively.

Cellular experiments

Primer design for reverse transcription quantitative PCR (RT-qPCR) experiment

In order to investigate the effect of **NCD** on the upregulation of circZKSCAN1 production from either **pre-mGGA** or **pre-mUCC** in the cellular environment, HeLa cells were used as the model cell line and RT-qPCR was used to quantify the changes in circZSKCAN1 expression. However, since circZKSCAN1 shares its sequence identity fully with its cognate mRNA counterpart, which is also expressed in HeLa cells.[13] To specifically amplify circZKSCAN1, a "divergent" primer was designed.[14,15] Briefly, unlike commonly used primer pairs in qPCR, which are designed to amplify sequence segments towards each other, or in a "convergent" manner; divergent primer pairs amplify away from each other, resulting in selective amplification of segments that contain the head-to-tail junction, or "back-splice" junction (shown as a triangle in Fig 4A), specific to circular RNA species. Furthermore, the presence of endogenous circZKSCAN1 expression in HeLa cells further complicates the matter. Toward this end, a single nucleotide difference which was present in the ZKSCAN1 sequence of ***p*-GGA** and ***p*-UCC** (Shown in orange in Fig 4A), where thymine in exon 3 of endogenous

ZKSCAN1 gene was cytosine (The C is highlighted in red in Fig. S1A) was taken advantage of, and the 3' terminus of the forward primer (circZKSCAN1_Fw in Table S1) was set to end at the 1-nt difference to selectively amplify circZKSCAN1 stemming from these plasmids (Fig. 4A, S1).[16] The selectivity of the divergent primer design in detecting circZKSCAN1 stemming from either ***p*-GGA** or ***p*-UCC** in the presence of endogenous circZKSCAN1 was first confirmed by comparing the RT-qPCR amplification curves of the non-transfected HeLa cells against HeLa cells transfected with ***p*-GGA** treated with 0 or 5 μM **NCD** (Fig 4B, Table 1). The results show that while the difference in cycle threshold (C_t) values for the internal standard, β-actin: a housekeeping gene, showed similar values for both non-transfected (black line), ***p*-GGA** transfected cells (light gray) (19.5, 20.0 respectively), confirming that the same amount of total RNAs was used for analysis. The non-transfected HeLa cells (green line) showed a 7.2 cycle difference in circZKSCAN1 Ct value compared to those transfected with ***p*-GGA** (red line). This indicates that overall, the PCR amplification of circZKSCAN1 produced from **pre-mGGA** is approximately 147-fold more efficient (2-7.2 = 1/147) than endogenous circZKSCAN1 using the above-mentioned primer design. Hereon all descriptions regarding circZKSCAN1 refer to circZKSCAN1 stemming from ***p*-GGA** or ***p*-UCC**.

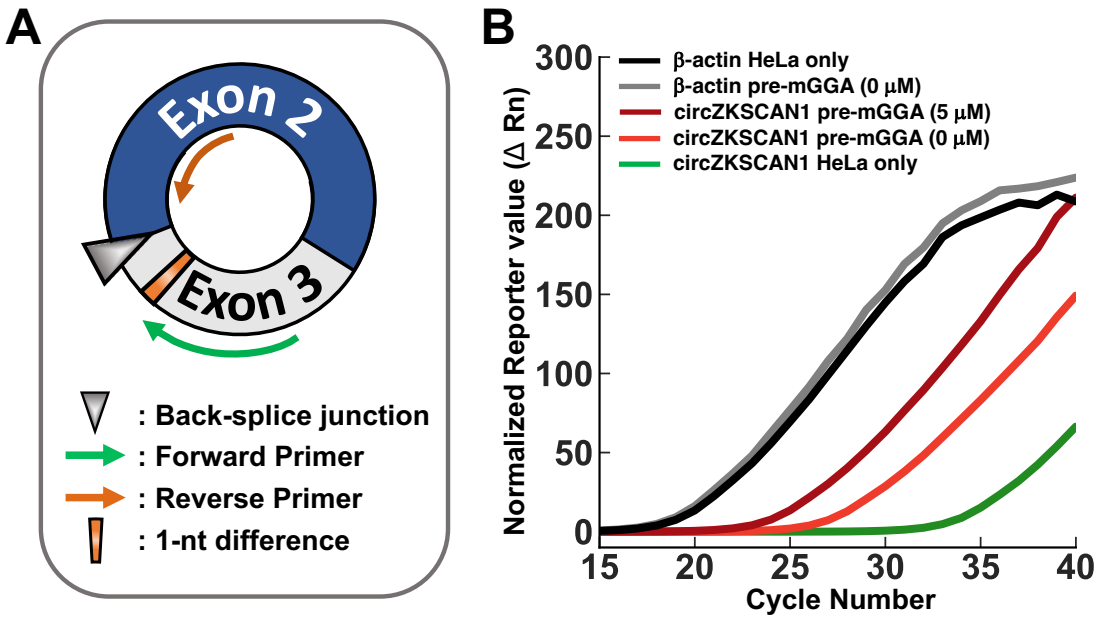


Figure 4. A) Schematic representation of circZKSCAN1. The site of the back splice-junction is shown with a triangle, the green and red arrows represent the primers used for the qPCR, and the orange highlight represents the location of 1-nt difference between circZKSCAN1 and its endogenous counterpart, and B) the RT-qPCR amplification curves of beta-actin and circZKSCAN1 using the designed primers. Each amplification curve is represented as follows: beta-actin in non-transfected HeLa (Black), beta-actin in ***p*-GGA** transfected cell (light gray), beta-actin in *p*-GGA transfected cell 5 μM **NCD** (dark gray), circZKSCAN1 in non-transfected HeLa (green line), circZKSCAN1 in ***p*-GGA** transfected HeLa (red line), circZKSCAN1 in ***p*-GGA** transfected HeLa in the presence of 5 μM **NCD** (blue line).

Table 1: Comparison between the calculated C_t values of HeLa only and ***p*-GGA** transfected cells (0 and 5 μM **NCD**). (Std. Dev. = standard deviation, n=3)

Samples	Average Ct	Std. Dev.
***p*-GGA** transfected HeLa beta-actin	19.5	0.16
HeLa only beta-actin	20.0	0.14
***p*-GGA** transfected HeLa circZKSCAN1 (5 μM **NCD**)	24.6	0.38
***p*-GGA** transfected HeLa circZKSCAN1 (0 μM **NCD**)	27.2	0.30
Non-transfected HeLa circZKSCAN1	34.4	1.1

The specificity of the divergent primer pair used was further confirmed by conducting a polyacrylamide gel electrophoresis (PAGE) analysis of the qPCR amplicon obtained from ***p*-GGA** transfected cells treated with 0 and 5 μM **NCD**, and 5 μM **QCD** (Fig. 5AB). The results show only a single DNA band of the expected size (approx. 140 bp) on the gel under each condition, while no band was observed for the non-reverse transcribed sample (RT(-)). Suggesting that there was minimal contamination signal from non-reverse transcribed residues. Furthermore, sequencing of the amplicon isolated from the gel showed the back-splicing junction sequence specific to circZKSCAN1 (Fig. 4CDE), corroborating the suitability of our divergent primer for detecting the desired target circRNA.[9,10]

Elucidation of the effects of NCD on the expression of circZKSCAN1 from model pre-mRNA in cellular environment

The relative change in expression of the target circZKSCAN1 produced from the model pre-mRNA was quantified using the comparative C_t ($\Delta\Delta C_t$) method.[17] Where the relative quantity (RQ) of circZKSCAN1 expressed in cells under different concentrations of ligand, relative to the untreated sample, was determined. Briefly, the raw C_t value of circZKSCAN1 obtained from qPCR experiments was first normalized against the C_t value of internal control, mRNA of β-actin, to give a ΔC_t value. Then, ΔC_t values of treated samples were then further normalized against those in the absence of ligands (**NCD** = 0 μM) to provide $\Delta\Delta C_t$. The RQ value is then calculated from the obtained ΔΔCt value ($RQ = 2^{-\Delta\Delta Ct}$).

The results show cells transfected with ***p*-GGA**, expressing **pre-mGGA**, showed a concentration-dependent increase in the RQ of circZKSCAN1 in the presence of **NCD**. Where cells treated with 3 μM **NCD** showed an apparent 3.1-fold increase in RQ of circZKSCAN1 ($p < 1 \times 10^{-6}$), while at the highest concentration of 5 μM **NCD**, a 5.6-fold increase was observed ($p < 1 \times 10^{-6}$, Fig. 6A). In contrast, no statistically significant change in RQ was observed for cells treated with 5 μM **QCD**. On the other hand, the cells transfected with ***p*-UCC**, expressing

pre-mUCC, the fully matched counterpart to **pre-mGGA**, were also treated similarly with the ligands (Fig. 6B). However, no statistically significant change in RQ of circZKSCAN1 was observed either upon **NCD**- or **QCD**- treatment of the cells expressing ***p*-UCC**.

Furthermore, to compare the difference in expression of circZKSCAN1 between **pre-mGGA** and **pre-mUCC**, RQ of circZKCAN1 for cells transfected with ***p*-GGA** was calculated relative to the cells transfected with the control ***p*-UCC** (**NCD** 0 μM, Fig. 6C). The results show that in

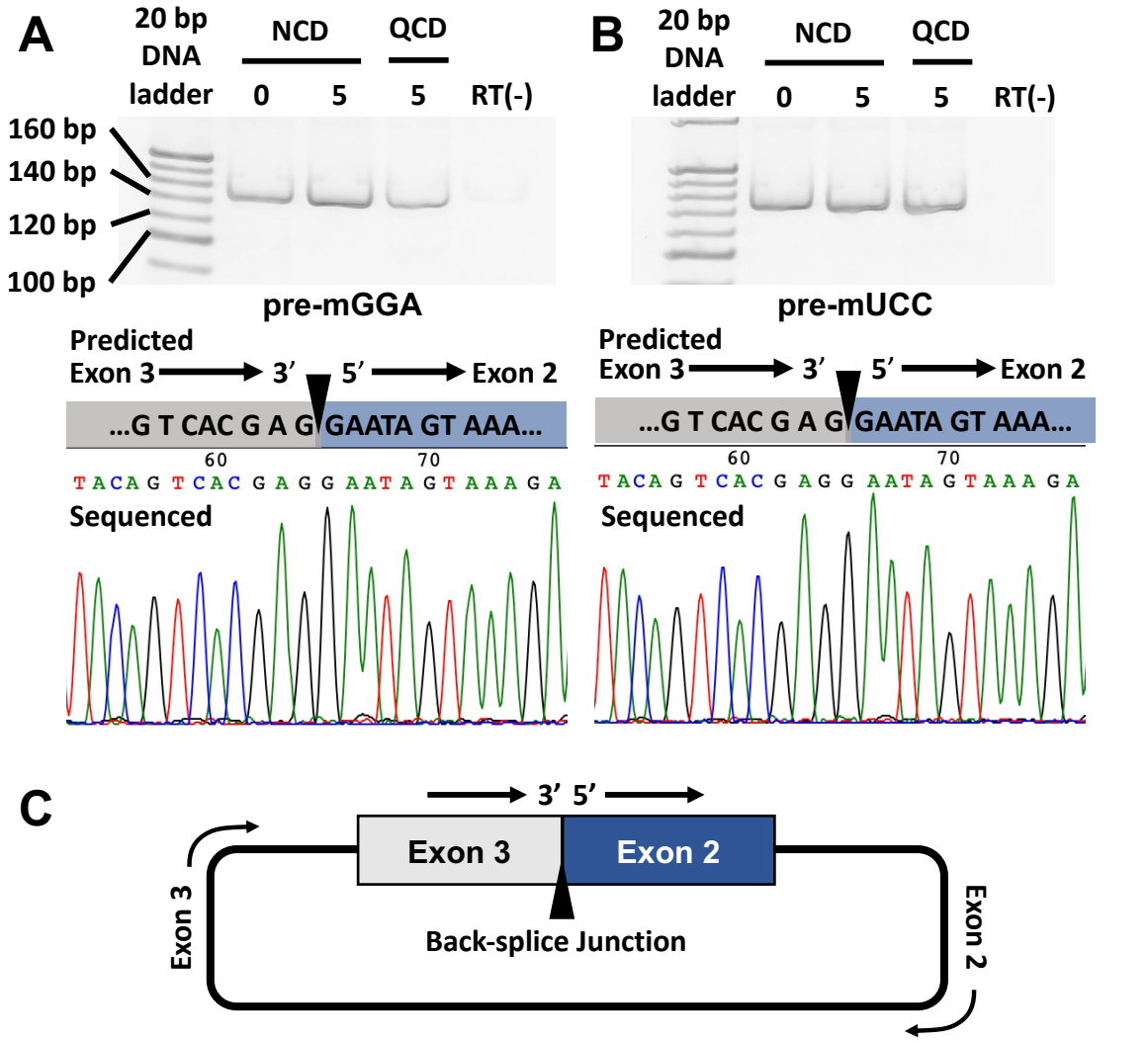


Figure 5. A) polyacrylamide gel electrophoresis of qPCR product from cells expressing **pre-mGGA**. (0 and 5 μM **NCD** sample, and 5 μM **QCD** sample and Non-RT sample), and the representative sequencing data of the qPCR product isolated from expressing **pre-mGGA**. (5 μM **NCD**). B) Polyacrylamide gel electrophoresis of qPCR product from cells expressing **pre-mUCC**, (0 and 5 μM **NCD** sample, and 5 μM **QCD** sample and Non-RT sample) and the representative sequencing data of the qPCR product isolated from expressing **pre-mUCC**. (0 μM **NCD**) C) Schematic description of the amplified sequence segment and its relative location within circZKSCAN1. For Sanger sequencing data, the top schematic denotes the predicted sequence and the bottom actual Sanger sequencing result, black triangle indicates the junction site.

the absence of **NCD**, cells transfected with ***p*-GGA** expectedly showed a significantly reduced expression of the target circZKSCAN1 compared to cells transfected with ***p*-UCC**, with an RQ of 0.074 ($p < 3 \times 10^{-6}$), due to the destabilization of RCS pairing resulting from the introduction of GGA/GGA mismatches. However, in the presence of 5 μM **NCD**, the cells transfected with ***p*-GGA** showed an RQ of 0.41 ($p < 0.005$), suggesting that under the highest concentration of 5 μM, **NCD** was capable of inducing 34% (0.41 – 0.074 = 0.34) recovery in the expression of circZKSCAN1 for **pre-mGGA**, with regards to **pre-mUCC**, and rescue the expression of circZKSCAN1 up to 41% of the original full-match **pre-mUCC** expression level.

Ribonuclease R (RNase R) treatment to determine the circularity of the detected RNA species

In order to determine whether the increased detection stems from an actual increase in circZKSCAN1 expression, the total RNA obtained from ***p*-GGA** transfected cells which were treated with 5 μM **NCD** was then incubated with or without RNase R to verify the circularity of the expressed product.[18,19] Since RNase R is a 3′-5′ RNA exonuclease, the circRNA product is expected to be resistant to its digestion. The qPCR results following RNase R treatment showed the detected circZKSCAN1 to be resistant to RNase R digestion as compared to the control linear beta-actin. Confirming that the increased expression does indeed arise from the desired circular product (Fig. 6D).

Verification of the correct splicing of the expressed circZKSCAN1 species using RT-PCR

The divergent primer used to this point has been able to demonstrate the capacity to selectively detect and quantify the presence of the exon 3 -exon 2 junction sequence characteristic of the target circZKSCAN1 as reported previously. However, due to the limits of optimal amplicon size for qPCR detection,[20] does not show whether the internal intron present between exon 2 and 3 is correctly spliced out or not, a process that may be affected due to non-specific binding of **NCD**. Toward this end, a new set of divergent primers that amplify a longer sequence segment encompassing the entirety of the smaller exon 3 was designed. The RT-PCR of the previously **NCD** and **QCD**-treated samples were then conducted, and the results were analyzed using PAGE. (Fig. 7A) The PAGE results show the amplification of two clearly unique bands, one of shorter amplicon lengths around 100 and the longer around 300 bp. Here, the longer amplicon corresponds to the expected amplicon fragment size when circZKSCAN1 is

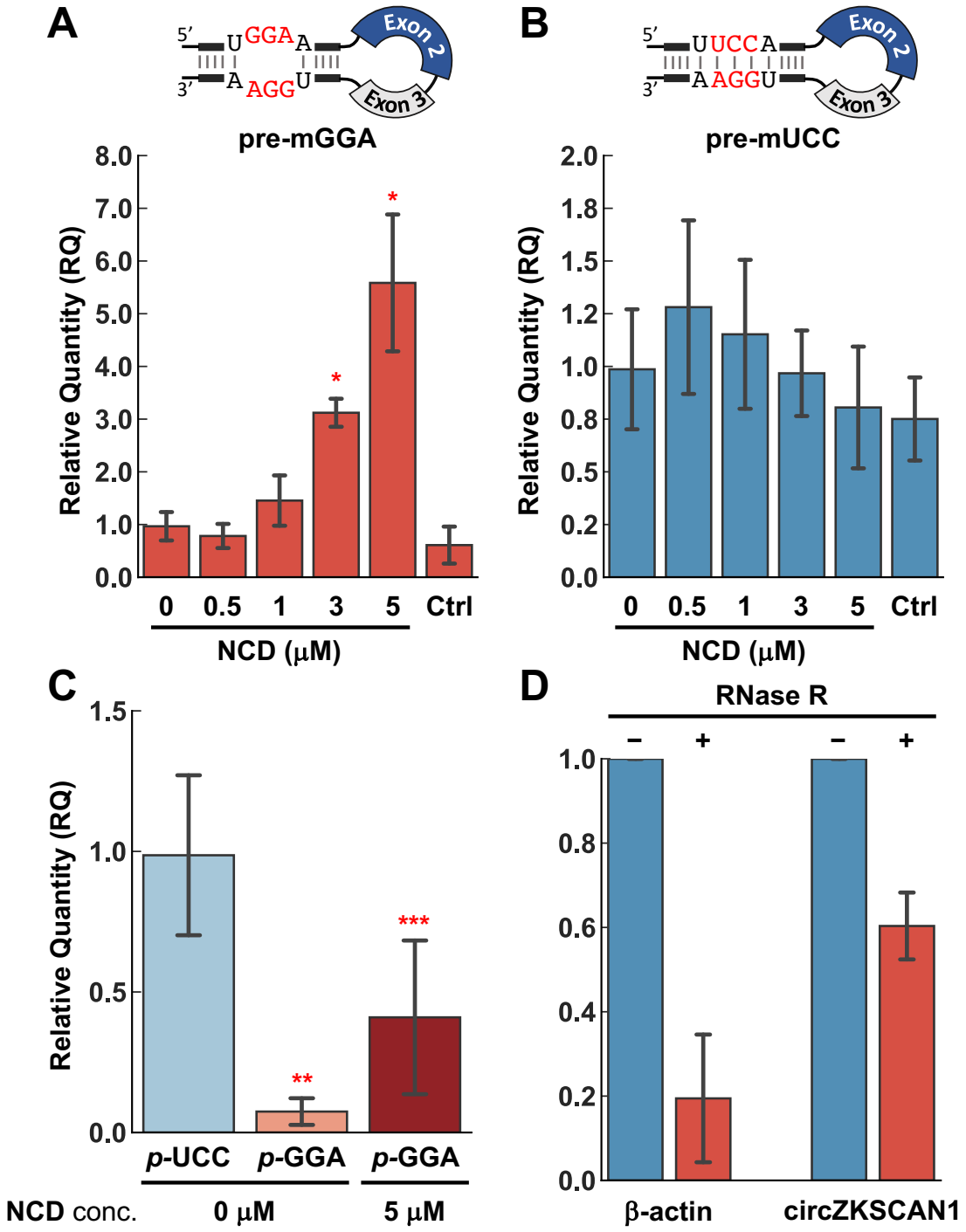
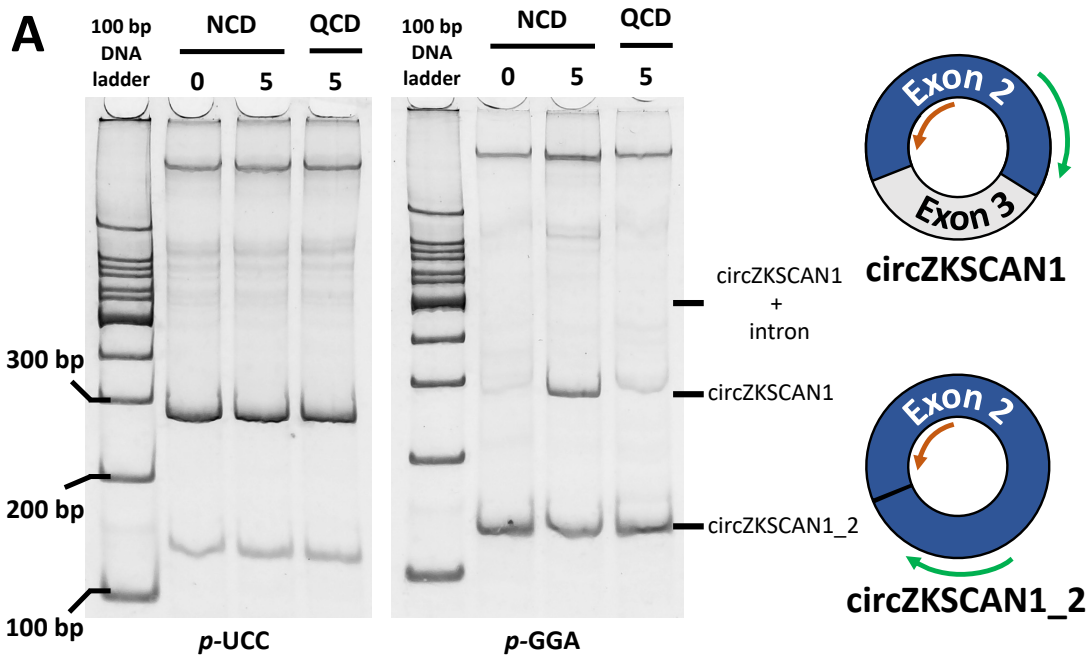
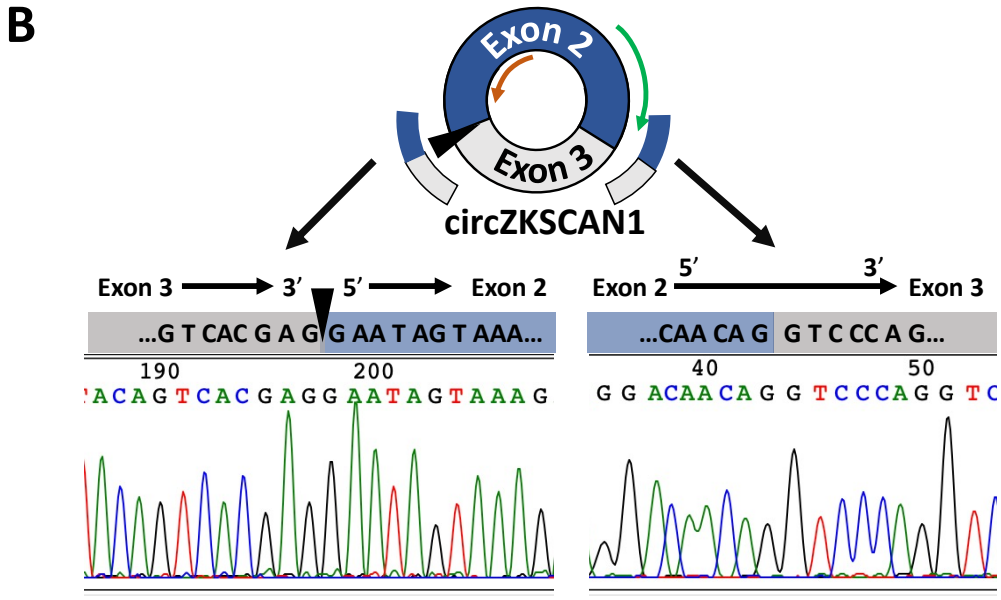


Figure 6. RT-qPCR results after **A) pre-mGGA** and **B) pre-mUCC** expressing Hela cells were incubated under various **NCD** concentrations, the respective schematic depiction of the expressed pre-mRNA is shown on top. (*: $p < 0.05$, fold change > 2, he significance was obtained using two tailed student's t test, n=7) **C)** Relative expression of circZKSCAN1 in cells transfected with ***p*-GGA** when compared to ***p*-UCC**. RQ value was obtained by normalizing the ΔCt value against the *p*-UCC transfected cell (**NCD** 0 μM) sample. **D)** RNase R treatment of total RNA obtained from cells transfected with *p*-GGA (5 μM **NCD**) followed by RT-qPCR detection of beta-actin and circZKSCAN1. **: $p < 3 \times 10^{-6}$, ***: $p < 0.005$, ****: $p < 0.02$, the significance was obtained using two tailed student's t test, n=3)


A
100 bp DNA ladder
NCD
QCD
0
5
5
100 bp DNA ladder
NCD
QCD
0
5
5
300 bp
200 bp
100 bp
p-UCC
p-GGA
circZKSCAN1 + intron
circZKSCAN1
circZKSCAN1_2
Exon 2
Exon 3
circZKSCAN1
Exon 2
circZKSCAN1_2


B
Exon 2
Exon 3
circZKSCAN1
Exon 3 → 3′ 5′ → Exon 2
...G T CAC G A G G AA T AG T AAA...
190
200
Exon 2 5′ → 3′ Exon 3
...CAA CA G G T C CC A G...
40
50

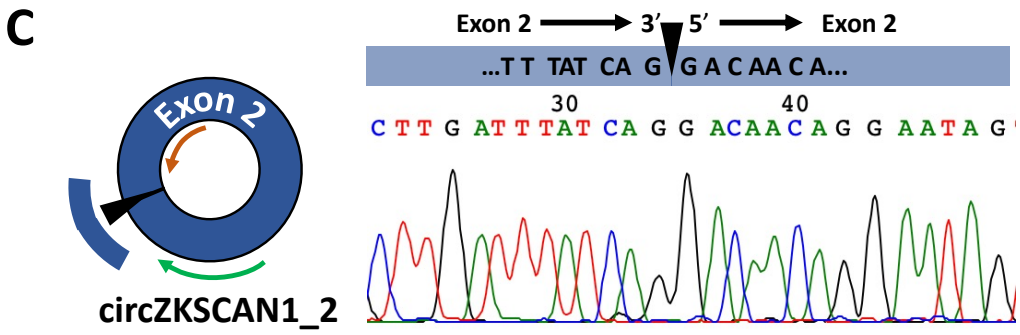

C
Exon 2
circZKSCAN1_2
Exon 2 → 3′ 5′ → Exon 2
...T T TAT CA G G A C AA C A...
30
40

Figure 7. **A)** Polyacrylamide gel electrophoresis of qPCR product using primers amplifying across the entirety of exon 3, using samples from cells transfected with ***p*-UCC** and ***p*-GGA** with or without compound treatment (concentration in μM). The samples used are denoted above their respective lanes. The schematic representation of primer positions and the suspected identities of the two amplified species are shown on the right. **B)** The representative sequencing data of qPCR fragments isolated from the PAGE corresponding to the band of the slower mobility B) The representative sequencing data of qPCR fragments isolated from the PAGE corresponding to the band of the faster mobility. The predicted splice site sequence is shown above the raw Sanger sequencing data. The black triangle denotes the back-spliced junction site.

comprised solely of exons 2 and 3 and the internal intron was spliced out as expected; while the identity of the observed shorter amplicon is unknown, the length corresponds to an alternative species of circZKSCAN1, hereon as circZKSCAN1_2, which is comprised only of exon 2 (Fig. 7A right bottom).[21] In all cases, however, no band of larger amplicon size (~ 500 bp), which correspond to the retention of the internal intron between exons 2 and 3, was detected. Indicating that the application of **NCD** or **QCD** did not affect the efficacy of splicing other than that of the targeted back-splicing.

The identities of the observed bands were then further clarified by Sanger sequencing.

The sequencing results showed that, of the two unique amplicon species detected, the longer amplicon contained, in addition to the entirety of exon 3 sequence, both the back-splicing and canonical splicing junction (Fig. 7B), further confirming that the intron between exon 2 and 3 is indeed removed and circZKSCAN1 is a fully exonic circRNA. While the sequencing result of the shorter amplicon showed the presence of exon 2 back-spliced junction, indicating that the amplified target is indeed the alternatively spliced circZKSCAN1_2, containing only a singular exon 2 of ZKSCAN1 pre-mRNA (Fig. 7C). Furthermore, as the band intensity stemming from circZKSCAN1_2 did not seem to change regardless of the expression of circZKSCAN1, it is likely a naturally occurring species since the divergent primer used does not discriminate between circRNA originating from the transfected plasmid and endogenously pre-existing circZKSCAN1.

Elucidation of potential off-target effects of NCD on other endogenous circRNAs

The changes in endogenous circRNAs following **NCD** treatment were also studied to elucidate any potential off-target effects of **NCD**. Here, the expression changes of four other well-known circRNAs (circHIKP3[22], circEPHB4[23], circFOXO3[24], and circPVT1[25]) were investigated in the

presence of 3 or 5 µM **NCD** concentration and 5 µM **QCD** (Fig. 8). The results showed that in general no significant changes were observed in the expression of the above-mentioned endogenous circRNA species for both ***p*-GGA** and ***p*-UCC** transfected cells. However, a significant 2-fold down-regulation of circEPHB4 was observed under 5 µM **NCD** in both cases. While we searched for potential **NCD** binding sites within 1000 bp up/downstream intron of the circEPHB4 transcript within its host pre-mRNA, we could not identify any relevant motifs that may cause the fluctuations mentioned above in expression level in the presence of **NCD**.

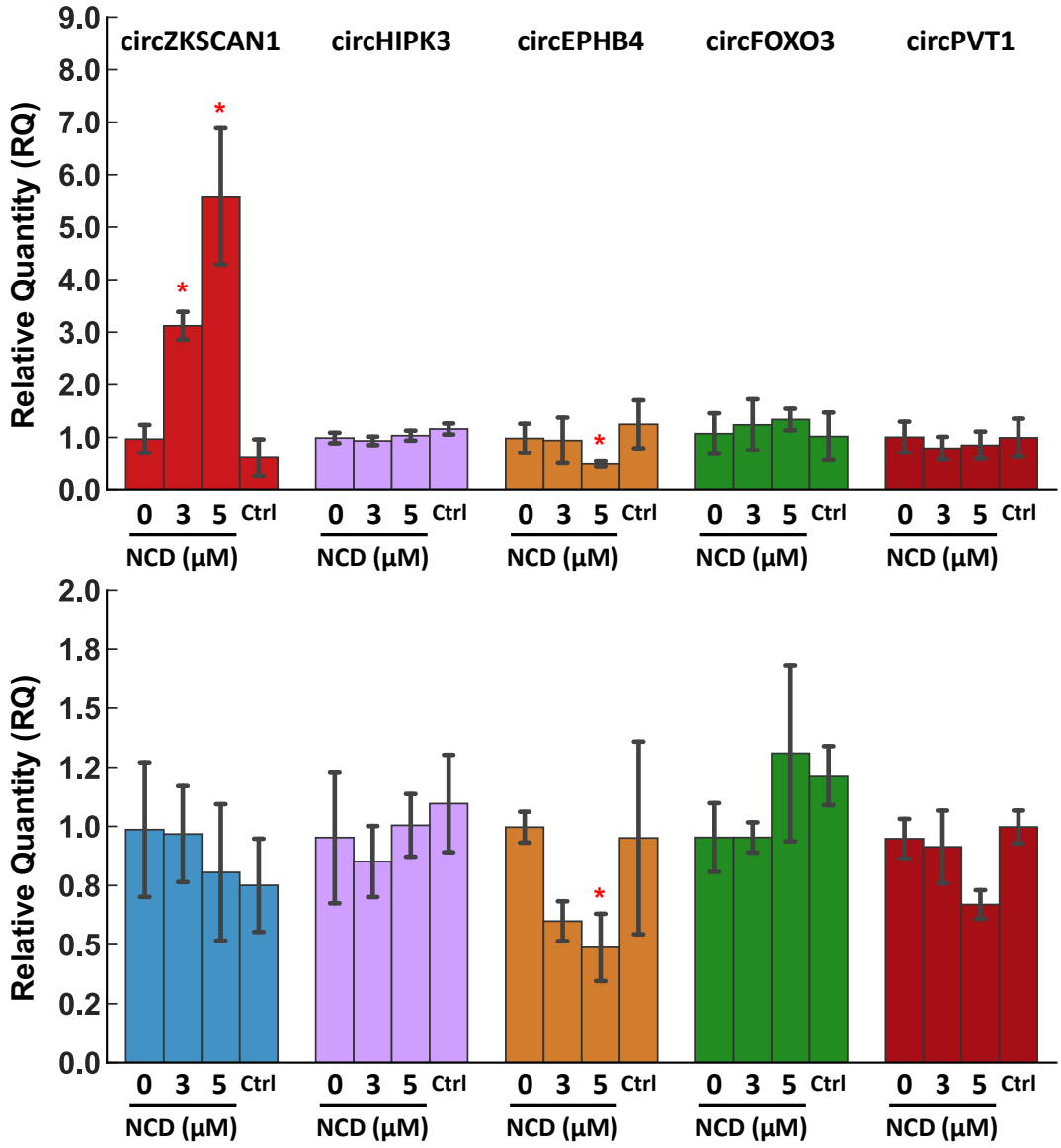


Figure 8. The relative change in expression of other commonly studied endogenously expressed circRNAs under the two highest treated **NCD** concentrations (3 and 5 µM) and **QCD** (Ctrl, 5 µM) in ***p*-GGA** transfected cells (top) and ***p*-UCC** transfected cells (bottom). The relative expression of circZKSCAN1 as discussed previously, was included for comparison. (*: $p < 0.05$, fold change ≥ 2, the significance was obtained against their respective **NCD** 0 µM treated sample, using a two-tailed t-test, n=4)

Conclusion

Here, we demonstrated the first instance of utilizing a small molecular probe to modulate the biogenesis of circRNA, using a model based on previously reported minimal pre-mRNA sequence for circZKSCAN1 biogenesis, we were able to modulate the biogenesis of circZKSCAN1 in Hela cells with **NCD** in a concentration-dependent manner. As recently it has become evident that circRNAs play an increasingly important regulatory role in various cellular functions[26-29] and disease progression[22,30,31]. This result not only serves as a basis for a new overexpression vector design for which **NCD** acts as an on/off switch to the biogenesis of circRNAs and a new route to study their functions but also as a strong evidence that an RNA binding molecule such as **NCD** could facilitate formation of hairpin structure necessary for successful back-splicing reaction by bringing the two flanking introns together. Since many RCSs and repeat sequences found in introns adjacent to exons forming circRNAs have many unpaired regions, such as stem-loops and bulge loops, when they form duplex structures. These unpaired regions in dsRNA provide potential small molecule binding sites and opportunities for new molecular designs. Thus, the demonstrated model provides a steppingstone in regulating the expression of circRNA in response to the externally added small molecules, presenting one possible strategy to regulate circRNA biogenesis. Recent progress in molecular designs that target particular RNA motifs should accelerate the discovery of RNA-binding small molecules that can regulate the biogenesis of biologically important circRNAs.

Experimental Section

General information:

Surface plasmon resonance (SPR) measurements were conducted using BIAcore T200 instrument (GE Healthcare). All oligonucleotides were purchased from Invitrogen. The original plasmid construct used in these experiments: pcDNA3.1(+) ZKSCAN1 nt 400-1782 delta440-500 delta1449-1735, was a gift from Jeremy Wilusz (Addgene plasmid # 60633 ; http://n2t.net/addgene:60633 ; RRID:Addgene_60633).[9] Escherichia coli (E. coli) strain DH5α was used to amplify plasmid DNA and the amplified plasmid DNA was extracted and purified using NucleoBond® Xtra Midi EF (Macherey-Nagel). HeLa cells (RIKEN BRC, RCB0007) were used for transfection and were maintained at 37 °C under 5% CO2 in Dulbecco's modified eagle's medium (Sigma, D6429) supplemented with 10% (v/v) fetal bovine serum (MP Biomedicals) and penicillin-streptomycin (Gibco). The chemical compounds described in the paper, **NCD** and **QCD**, were synthesized as previously reported.[8,10-12] FuGENE HD™ transfection reagent (Promega) was used for plasmid transfection. For RT-qPCR experiments, cell lysis and reverse transcription was conducted using Superprep® II cell lysis & RT kit (TAKARA), and real-time PCR (qPCR) experiments were conducted using StepOne real-time PCR system (Applied Biosystem) using GoTaq® qPCR master mix (Promega). For Ribonuclease R (RNase R) treatment, cell lysis and total RNA extraction were conducted using ISOGEN (Nippon gene Co.), RNase R was purchased from Epicenter Technologies, and the reverse transcription was conducted using ReverTra Ace® qPCR RT Master Mix with gDNA remover (TOYOBO).

SPR assay:

1) Sensor preparation: For the SPR assay, RNA containing the partial intronic sequence containing 5′-GGA-3′/5′-GGA-3′ and 5′-UCC-3′/5′-GGA-3′ was immobilized onto the sensor chip SA (Cytiva) where the surface was coated with streptavidin. The surface of the sensor chip was first washed three times with 50 mM NaOH and 1 M NaCl for 60 sec with a flow rate of 30 μL min^{-1}. Then the biotin-labeled RNAs, biotin-TEG-5′-GCUGAGAU GGA AGGCGUGAGCUUUUGCUCACACCU GGA AUCCCAGC-3′ and biotin-TEG-5′-GCUGAGAU UCC AGGCGUGAGCUUUUGCUCACACCU GGA AUCCCAGC-3′ (The 5′-GGA-3′/5′-GGA-3′ and 5′-UCC-3′/5′-GGA-3′ regions are underlined), were immobilized onto the surface under the following conditions: 1 μM RNA in 10 mM HEPES, 500 mM NaCl, pH

7.4, the flow rate of 5 μL min^{-1} for the flow of 60 sec. The amount of RNA immobilized on the sensor chip SA was 1217 RU and 1075 RU respectively.

2) SPR analysis protocol: SPR analysis for the binding of **NCD** and **QCD** to the RNA-immobilized surface was conducted on BIAcore T-200 and carried out by subsequently flowing 0.0625, 0.125, 0.25, 0.5, and 1 μM of each compound for 60 sec of contacting time in HBS-EP+ buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.05 % v/v Surfactant P20, pH 7.4) (Cytiva) with the flow rate of 30 μL min^{-1} at 25 °C followed by dissociation of bound compounds by flowing running buffer for 120 sec. Surface regeneration was carried out after the assay of each compound using 1.2 mM NaOH, 0.2 M NaCl, and 0.1 mM EDTA solution, with a contact time of 180 sec and a flow rate of 30 μL min^{-1}. Kinetic parameters of the binding of **NCD** and **QCD** to the RNA-immobilized surface were obtained by using the single-cycle kinetics method.

Plasmid construct preparation:

The desired mutation was introduced via PCR mutagenesis into the region coding for the circZKSCAN1 expressing pre-mRNA of the original construct (pcDNA3.1(+) ZKSCAN1 nt 400-1782 delta440-500 delta1449-1735, ***p*-UAC**). The specific location of the introduced mutation is shown in Scheme 1. First, a guanine point mutation was introduced into one of the complementary intronic sequences (Scheme 1, yellow arrow 1000-1035 nt region), using the primers (5′-GTA GCT CAC ACC TGG AAT CCC AGC AGC GG-3′) and (5′- CCG CTG CTG GGA TTC CAG GTG TGA GCT AC-3′) and PfuUltra High-fidelity DNA polymerase (Agilent) to produce the intermediate construct. The PCR product was then treated with DpnI before it was transformed into E. coli DH5α cells for amplification and isolated, as described above in General information. Additional nucleotide mutation was then further introduced into the intermediate construct in the other complementary intronic sequence (Scheme 1, yellow arrow 1-40 nt region). For ***p*-GGA** plasmid, the primer pair (5′-GAA TTC AAA GTG CTG AGA TGG AAG GCG TGA GCC ACC ACC-3′) and (5′- GGT GGT GGC TCA CGC CTT CCA TCT CAG CAC TTT GAA TTC-3′) was used, while for ***p*-UCC** plasmid, the primer pair (5′- GAA TTC AAA GTG CTG AGA TTC CAG GCG TGA GCC ACC ACC-3′) and (5′-GGT GGT GGC TCA CGC CTG GAA TCT CAG CAC TTT GAA TTC-3′) was used. The respective constructs were then amplified via E. coli DH5α cells and collected as described previously, and the final sequence of the region coding for **pre-mGGA** and **pre-mUCC** are shown in Fig. S1C and D.

Reverse-transcription quantitative PCR (RT-qPCR experiments):

The HeLa cells were plated in a 96-well plate at a density of 1 x10^4 cells/well and were incubated for 4 hrs to allow the cells to adhere to the well bottom. It was then transfected with ***p*-GGA** and ***p*-UCC** expressing plasmid obtained previously using Fugene HD^{TM} with the ratio (FuGENE : plasmid = 5:2, v/w) according to the manufacturer's instructions. After 24 hrs incubation under 37 oC with 5% CO2, the medium was exchanged with fresh medium containing **NCD** or **QCD** (**NCD**: 0, 0.5, 1, 3, and 5 μM; **QCD**: 5 μM), and the cells were further incubated for 24 hrs. The transfected cells were lysed and the total RNA was reverse transcribed into cDNA using the Superprep® II cell lysis & rt kit according to the manufacturer's instructions. The resulting cDNA was diluted 5-fold before performing qPCR analysis.

qPCR experiments were conducted with StepOne real-time PCR system (Applied Biosystem) and GoTaq® qPCR master mix (Promega) using the cDNA as the template. Each 10 μl reaction contained 1 μl of the cDNA solution, 0.2 μl each of forward- and reverse- primers (10 μM), 0.1 μl of 100x CXR dye, and 5 μl of GoTaq® qPCR Master Mix. A negative control without cDNA template or samples without reverse transcriptase was included in each assay. The primers used for the detection of β-actin and the various target circRNAs are listed below in Table S1.

Relative fold change of circRNA was calculated using the comparative ΔΔCt method and beta-actin was used as endogenous control. Of the 9 data points obtained for each sample; the outliers at the minimum and maximum values were omitted. The qPCR products were analyzed by 8% (29:1, acylamide/bis-acrylamide) polyacrylamide gel electrophoresis to check the amplicon size. The specificity of primer pairs was also verified with the presence of a single peak in the melting curve after PCR amplification and sequencing of the amplicon.

Ribonuclease R (RNase R) treatment:

HeLa cells plated in a 24-well plate (5 x10^4 cells/well) were first transfected with the ***p*-GGA** expression plasmid and treated in the same manner as described above. Total RNA was extracted using the ISOGEN RNA extraction kit, according to manufacturer's instructions. The obtained total RNA was then treated with DNase I, followed by ethanol precipitation. The RNA pellet was dissolved in water, and 360 ng of the total RNA was then incubated for 15 min at 37 °C with or without 1U RNase R. The resulting solution was directly subjected to reverse transcription using ReverTra Ace® qPCR RT Master Mix with gDNA remover, according to

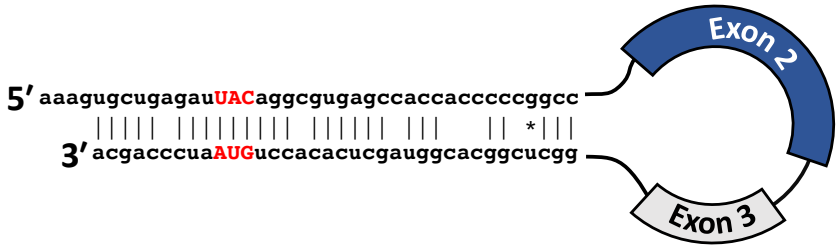
manufacturer’s instructions. The expression was then analyzed by qPCR as mentioned above, the experiment was conducted 3 independent times.

Supplementary Figures

A

Original construct from Addgene reported by Wilusz et. al. (***p*-UAC**):

aaagtgctgagatTACaggcgtgagccaccacccccggccCACTTTTTGTAAAGGTACGTAC
TAATGACTTTTTTTTTTATACTTCAGGAATAGTAAAGAAACACATCATAAAACCTCCCAGGAC
<-3′-TGGAGGGTCCTG
TATTTCCACTC-5′: circZKSCAN1_Rev
ATAAAGGTGAGCACAGACCCTGTTTGGATCAAGTCAGTTCCTGGAGCCTGAATGATGACTGC
TGAATCACGGGAAGCCACGGGTCTGTCCCCACAGGCTGCACAGGAGAAGGATGGTATCGTAA
TAGTGAAGGTGGAAGAGGAAGATGAGGAAGACCACATGTGGGGGCAGGATTCCACCCTACAG
GACACGCCTCCTCCAGACCCAGAGATATTCCGCCAACGCTTCAGGCGCTTCTGTTACCAGAA
CACTTTTGGGCCCCGAGAGGCTCTCAGTCGGCTGAAGGAACTTTGTCATCAGTGGCTGCGGC
CAGAAATAAACACCAAGGAACAGATCCTGGAGCTTCTGGTGCTAGAGCAGTTTCTTTCCATC
CTGCCCAAGGAGCTCCAGGTCTGGCTGCAGGAATACCGCCCCGATAGTGGAGAGGAGGCCGT
GACCCTTCTAGAAGACTTGGAGCTTGATTTATCAGGACAACAGGTAAAAAGAGGTGAAACCT
ATTATGTGTGAGCAGGGCACAGACGTTGAAACTGGAGCCAGGAGAAGTATTGGCAGGCTTTA
GGTTATTAGGTGGTTACTCTGTCTTAAAAATGTTCTGGCTTTCTTCCTGCATCCACTGGCAT
ACTCATGGTCTGTTTTTAAATATTTTAATTCCCATTTACAAAGTGATTTACCCACAAGCCCA
ACCTGTCTGTCTTCAGGTCCCAGGTCAAGTTCATGGACCTGAGATGCTCGC
circZKSCAN1_Fw: 5′-CAGTTCAGGAGTCCTCGAGCC-3′->
AAGGGGGATGGTGCCTCTGGATCCAGTTCAGGAGTCCTCGAGCCTTGACCTTCATCACGAGG
CCACCCAGTCCCACTTCAAACATTCGTCTCGGAAACCCCGCCTCTTACAGTCACGAGGTAAG
AAGCAAGGAAAAGAATTAggctcggcacggtagctcacacctGTAatccca



B

aaagtgctgagatGGAaggcgtgagccaccacccccggccCACTTTTTGTAAAGGTACGTAC
TAATGACTTTTTTTTTTATACTTCAGGAATAGTAAAGAAACACATCATAAAACCTCCCAGGAC
ATAAAGGTGAGCACAGACCCTGTTTGGATCAAGTCAGTTCCTGGAGCCTGAATGATGACTGC
TGAATCACGGGAAGCCACGGGTCTGTCCCCACAGGCTGCACAGGAGAAGGATGGTATCGTAA
TAGTGAAGGTGGAAGAGGAAGATGAGGAAGACCACATGTGGGGGCAGGATTCCACCCTACAG
GACACGCCTCCTCCAGACCCAGAGATATTCCGCCAACGCTTCAGGCGCTTCTGTTACCAGAA
CACTTTTGGGCCCCGAGAGGCTCTCAGTCGGCTGAAGGAACTTTGTCATCAGTGGCTGCGGC
CAGAAATAAACACCAAGGAACAGATCCTGGAGCTTCTGGTGCTAGAGCAGTTTCTTTCCATC
CTGCCCAAGGAGCTCCAGGTCTGGCTGCAGGAATACCGCCCCGATAGTGGAGAGGAGGCCGT
GACCCTTCTAGAAGACTTGGAGCTTGATTTATCAGGACAACAGGTAAAAAGAGGTGAAACCT
ATTATGTGTGAGCAGGGCACAGACGTTGAAACTGGAGCCAGGAGAAGTATTGGCAGGCTTTA
GGTTATTAGGTGGTTACTCTGTCTTAAAAATGTTCTGGCTTTCTTCCTGCATCCACTGGCAT
ACTCATGGTCTGTTTTTAAATATTTTAATTCCCATTTACAAAGTGATTTACCCACAAGCCCA
ACCTGTCTGTCTTCAGGTCCCAGGTCAAGTTCATGGACCTGAGATGCTCGCAAGGGGGATGG
TGCCTCTGGATCCAGTTCAGGAGTCCTCGAGCCTTGACCTTCATCACGAGGCCACCCAGTCC
CACTTCAAACATTCGTCTCGGAAACCCCGCCTCTTACAGTCACGAGGTAAGAAGCAAGGAAA
AGAATTAggctcggcacggtagctcacacctGGAatcccagca

C

aaagtgctgagatTCCaggcgtgagccaccacccccggccCACTTTTTGTAAAGGTACGTAC
TAATGACTTTTTTTTTTATACTTCAG**GAATAGTAAAGAAACACATCATAAAACCTCCCAGGAC**
ATAAAGGTGAGCACAGACCCTGTTTGGATCAAGTCAGTTCCTGGAGCCTGAATGATGACTGC
TGAATCACGGGAAGCCACGGGTCTGTCCCCACAGGCTGCACAGGAGAAGGATGGTATCGTAA
TAGTGAAGGTGGAAGAGGAAGATGAGGAAGACCACATGTGGGGGCAGGATTCCACCCTACAG
GACACGCCTCCTCCAGACCCAGAGATATTCCGCCAACGCTTCAGGCGCTTCTGTTACCAGAA
CACTTTTGGGCCCCGAGAGGCTCTCAGTCGGCTGAAGGAACTTTGTCATCAGTGGCTGCGGC
CAGAAATAAACACCAAGGAACAGATCCTGGAGCTTCTGGTGCTAGAGCAGTTTCTTTCCATC
CTGCCCAAGGAGCTCCAGGTCTGGCTGCAGGAATACCGCCCCGATAGTGGAGAGGAGGCCGT
GACCCTTCTAGAAGACTTGGAGCTTGATTTATCAGGACAACAGGTAAAAAGAGGTGAAACCT
ATTATGTGTGAGCAGGGCACAGACGTTGAAACTGGAGCCAGGAGAAGTATTGGCAGGCTTTA
GGTTATTAGGTGGTTACTCTGTCTTAAAAATGTTCTGGCTTTCTTCCTGCATCCACTGGCAT
ACTCATGGTCTGTTTTTAAATATTTTAATTCCCATTTACAAAGTGATTTACCCACAAGCCCA
ACCTGTCTGTCTTCAG**GTCCCAGGTCAAGTTCATGGACCTGAGATGCTCGCAAGGGGGATGG**
TGCCTCTGGATCCAGTTCAGGAGTCCTCGAGCCTTGACCTTCATCACGAGGCCACCCAGTCC
CACTTCAAACATTCGTCTCGGAAACCCCGCCTCTTACAGTCACGAGGTAAGAAGCAAGGAAA
AGAATTA**ggctcggcacggtagctcacacctGGAatcccagca**

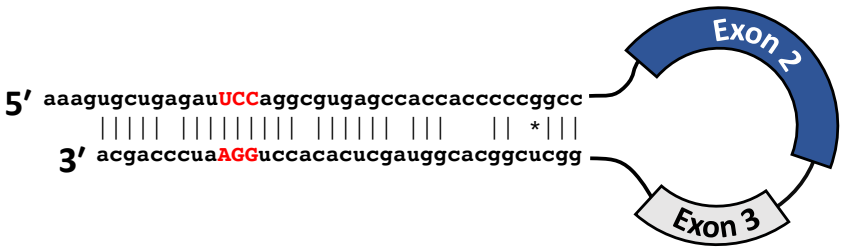


Figure S1. A) Whole sequence of the region coding for the original circZKSCAN1 in ***p*-UAC** (top), and the predicted resulting **pre-mUAC** hairpin structure (bottom). The divergent primer alignment is also shown, the arrow indicate the direction of PCR amplification. B) The whole sequence of the region coding for the ***p*-GGA** (top) and the predicted resulting **pre-mGGA** hairpin structure (bottom). C) The whole sequence of the region coding for the ***p*-UCC** (top) and the predicted resulting **pre-mUCC** hairpin structure (bottom). Characters in blue denote the exon 2 region, and the ones highlighted in gray denote the exon 3 region. The yellow highlighted lower-case letters in bold denote the RCSs which is predicted to form the stem of the hairpin. The single nucleotide difference in exon 3 is highlighted in red, in reference genome, it is expected to be T instead of C. The introduced mutations are highlighted in bold uppercase red characters.

Table S1: Primers used for RT-qPCR

Probe name	Sequence
Beta_actin_mRNA_Fw	5′-CTCTTCCAGCCTTCCTTCCT-3′
Beta_actin_mRNA_Rev	5′-AGCACTGTGTTGGCGTACAG-3′
circZKSCAN1_Fw	5′-CAGTTCAGGAGTCCTCGAGCC-3′
circZKSCAN1_Rev[a]	5′-CTCACCTTTATGTCCTGGGAGGT-3′
circZKSCAN1_splice_chk_Fw	5′-AATACCGCCCCGATAGTGGAG-3′
circZKSCAN1_splice_chk_Rev[a]	5′-CTCACCTTTATGTCCTGGGAGGT-3′
circHIPK3_Fw	5′-TATGTTGGTGGATCCTGTTCGGCA-3′
circHIPK3_Rev	5′-TGGTGGGTAGACCAAGACTTGTGA-3′

circEPHB4_Fw	5′-ATTCCCTCACAGCCTCATTAGGGT-3′
circEPHB4_Rev	5′-TGCCCGTCATGATTCTCACAGAGT-3′
circFOXO3_Fw	5′-GTGGGGAACTTCACTGGTGCTAAG-3′
circFOXO3_Rev	5′-GGGTTGATGATCCACCAAGAGCTCTT-3′
circPVT1_Fw	5′-GGTTCCACCAGCGTTATTC-3′
circPVT1_Rev	5′-CAACTTCCTTTGGGTCTCC-3′

[a] The same forward primer was used for circZKSCAN1_Rev and circZKSCAN1_splice_chk_Rev, to detect the canonical splice site between exon 2 and 3.

References

1 A. Granzhan, N. Kotera and M. P. Teulade-Fichou, *Chem. Soc. Rev.*, 2014, **43**, 3630–3665.

2 J. L. Childs-Disney, X. Yang, Q. M. R. Gibaut, Y. Tong, R. T. Batey and M. D. Disney, *Nat. Rev. Drug Discov.*, 2022, **21**, 736–762.

3 K. NAKATANI, *Proc. Japan Acad. Ser. B*, 2022, **98**, PJA9801B-03.

4 C. Dohno, M. Kimura and K. Nakatani, *Angew. Chemie - Int. Ed.*, 2018, **57**, 506–510.

5 S. Matsumoto, N. Caliskan, M. V. Rodnina, A. Murata and K. Nakatani, *Nucleic Acids Res.*, 2018, **46**, 8079–8089.

6 L. S. Kristensen, M. S. Andersen, L. V. W. Stagsted, K. K. Ebbesen, T. B. Hansen and J. Kjems, *Nat. Rev. Genet.*, 2019, **20**, 675–691.

7 L.-L. Chen, *Nat. Rev. Mol. Cell Biol.*, 2016, **17**, 205–211.

8 T. Shibata, K. Nagano, M. Ueyama, K. Ninomiya, T. Hirose, Y. Nagai, K. Ishikawa, G. Kawai and K. Nakatani, *Nat. Commun.*, 2021, **12**, 236.

9 D. Liang and J. E. Wilusz, *Genes Dev.*, 2014, **28**, 2233–2247.

10 Z. Yao, J. Luo, K. Hu, J. Lin, H. Huang, Q. Wang, P. Zhang, Z. Xiong, C. He, Z. Huang, B. Liu and Y. Yang, *Mol. Oncol.*, 2017, **11**, 422–437.

11 T. Peng, T. Murase, Y. Goto, A. Kobori and K. Nakatani, *Bioorganic Med. Chem. Lett.*, 2005, **15**, 259–262.

12 L. Ni, T. Yamada and K. Nakatani, *Chem. Commun.*, 2020, **56**, 5227–5230.

13 L. Szabo and J. Salzman, *Nat. Rev. Genet.*, 2016, **17**, 679–692.

14 W. R. Jeck and N. E. Sharpless, Nat. Biotechnol., 2014, 32, 453–461.

15 A. C. Panda, K. Abdelmohsen and M. Gorospe, in *Oncogene-Induced Senescence*, ed. M. A. Nikiforov, Springer New York, New York, NY, 2017, vol. 1534, pp. 79–87.

16 S. Lefever, F. Pattyn, J. Hellemans and J. Vandesompele, Clin. Chem., 2013, 59, 1470–1480.

17 K. J. Livak and T. D. Schmittgen, Methods, 2001, 25, 402–408.

18 W. R. Jeck, J. A. Sorrentino, K. Wang, M. K. Slevin, C. E. Burd, J. Liu, W. F. Marzluff and N. E. Sharpless, *RNA*, 2013, **19**, 141–157.

19 H. Suzuki, *Nucleic Acids Res.*, 2006, **34**, e63–e63.

20 W. Van Holm, J. Ghesquière, N. Boon, T. Verspecht, K. Bernaerts, N. Zayed, I. Chatzigiannidou and W. Teughels, *Appl. Environ. Microbiol.*, 2021, **87**, 1–11.

21 circZKSCAN1_2 "hsa_circ_0081379":http://www.circbase.org/cgibin/singlerecord.cgi?id=hsa_circ_0081379

22 Q. Zheng, C. Bao, W. Guo, S. Li, J. Chen, B. Chen, Y. Luo, D. Lyu, Y. Li, G. Shi, L. Liang, J. Gu, X. He and S. Huang, *Nat. Commun.*, 2016, **7**, 11215.

23 C. Jin, J. Zhao, Z. P. Zhang, M. Wu, J. Li, B. Liu, X. Bin Liao, Y. X. Liao and J. P. Liu, *Mol. Oncol.*, 2021, **15**, 596–622.

24 W. W. Du, W. Yang, E. Liu, Z. Yang, P. Dhaliwal and B. B. Yang, *Nucleic Acids Res.*, 2016, **44**, 2846–2858.

25 S. Qin, Y. Zhao, G. Lim, H. Lin, X. Zhang and X. Zhang, *Biomed. Pharmacother.*, 2019, **111**, 244–250.

26 M. Piwecka, P. Glažar, L. R. Hernandez-Miranda, S. Memczak, S. A. Wolf, A. Rybak-Wolf, A. Filipchyk, F. Klironomos, C. A. Cerda Jara, P. Fenske, T. Trimbuch, V. Zywitza, M. Plass, L. Schreyer, S. Ayoub, C. Kocks, R. Kühn, C. Rosenmund, C. Birchmeier and N. Rajewsky, *Science* **357**, eaam8526.

27 I. Legnini, G. Di Timoteo, F. Rossi, M. Morlando, F. Briganti, O. Sthandier, A. Fatica, T. Santini, A. Andronache, M. Wade, P. Laneve, N. Rajewsky and I. Bozzoni, *Mol. Cell*, 2017, **66**, 22-37.e9.

28 C. Arcinas, W. Tan, W. Fang, T. P. Desai, D. C. S. Teh, U. Degirmenci, D. Xu, R. Foo and L. Sun, *Nat. Metab.*, 2019, **1**, 688–703.

29 X. You, I. Vlatkovic, A. Babic, T. Will, I. Epstein, G. Tushev, G. Akbalik, M. Wang, C. Glock, C. Quedenau, X. Wang, J. Hou, H. Liu, W. Sun, S. Sambandan, T. Chen, E. M. Schuman and W. Chen, *Nat. Neurosci.*, 2015, **18**, 603–610.

30 J. N. Vo, M. Cieslik, Y. Y. Zhang, S. Shukla, L. Xiao, Y. Y. Zhang, Y. M. Wu, S. M. Dhanasekaran, C. G. Engelke, X. Cao, D. R. Robinson, A. I. Nesvizhskii and A. M. Chinnaiyan, *Cell*, 2019, **176**, 869-881.e13.

31 U. Dube, J. L. Del-Aguila, Z. Li, J. P. Budde, S. Jiang, et al., *Nat. Neurosci.*, 2019, **22**, 1903–1912.

Acknowledgement

A part of Chapter 1 is reproduced from ref. 12 with permission from the Royal Society of Chemistry.

Chapter 2

Utility of oligonucleotide in upregulating circular RNA production in a cellular model

Abstract

The demonstration of the capacity for mismatch binding small molecules, such as **NCD**, to upregulate circRNA in cellular environment. This has shown the feasibility of the concept of utilizing RNA binding ligands as trans-acting factors to modulate the production of circRNAs in cells by targeting and stabilizing the inter-intronic interactions. In this chapter, the concept is expanded further to explore the feasibility of other modalities that have previously been used to target and modulate RNA functions. Here, based on the mechanism involved in back-splicing reaction, we explored the feasibility utilizing a newly designed oligonucleotide, CircuLarIzation Promoting OligoNucleotide (**CLIP-ON**), to promote the production of circRNA in cells. Based on the sequence of model pre-mRNA used in the previous chapter, **CLIP-ON** was designed to hybridize with and physically bridge two distal sequences in the flanking introns of the circularizing exons. The capacity of **CLIP-ON**'s ability to upregulate circRNA was then confirmed in HeLa cells that expressed the model pre-mRNA. Demonstrating the applicability of ONs as a trans-acting modulator of a new modality to upregulate the production of circRNAs in a cellular environment.

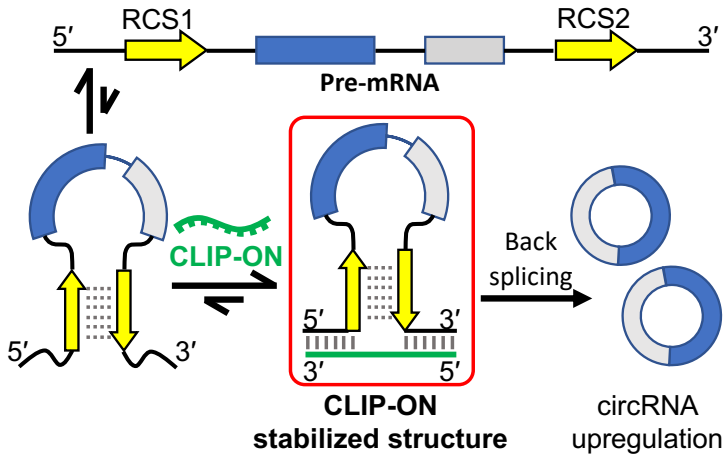


Figure 9. Schematic representation of the core concept explored in this chapter. Here, the newly designed **CLIP-ON** promoted circRNA production from a pre-mRNA via physically bridging the introns flanking the circularizing exons and stabilizing the of interactions between the RCSs.

Introduction

In the previous chapter, I demonstrated selective upregulation of circRNA production in cells utilizing a small molecule, **NCD**.[1] Demonstrating the potency of small molecules as an externally introduced trans-regulatory factor to upregulate the production of circRNAs in cells. In which the capacity for **NCD** to stabilize the necessary RCS interactions and facilitate the formation of a prerequisite hairpin structure through the selective binding of a designated target sequence in the flanking introns played a major role in upregulating the production of circRNA from a model pre-mRNA. However, despite the observed potency and selectivity shown by such small molecules, finding and screening of suitable compounds for other sequences, whilst avoiding potential off-target effects, could be bottlenecks to their application in regulating circRNA production.

Toward this end, based on the key feature in the back-splicing reaction, the use of oligonucleotides (ONs) was explored as a new trans-acting modality for targeting circRNA upregulation. In this chapter, I demonstrate that ONs could potentially be designed to stabilize and facilitate inter-intronic interactions and promote circRNA production in a cellular environment in a similar manner to the previously reported **NCD**. ONs were chosen for the nature of their modality, as they can be designed swiftly in a sequence-dependent manner to target a broad range of RNA sequences. Furthermore, the large library of available nucleotide/sugar/ phosphate backbone modifications allows for a fine-tuning of ON binding affinity and functionality.[2,3] Design-wise, **CLIP-ON** targets two separate distal sequences in the introns flanking the circularizing exon(s), and its hybridization is expected to physically bridge and facilitate the interaction between the flanking introns, facilitating the formation of the prerequisite hairpin structure for exon-circularization (Fig. 9). This design is a departure from the traditional ONs previously used for exon inclusion/skipping and RNase H antisense ONs, which typically only targeted continuous adjoining sequences.[3,4]

The efficacy of **CLIP-ON** as an externally introduced trans-factor in promoting circRNA production was examined in HeLa cells utilizing a pre-mRNA model, **pre-mUAC**,[5] used in the previous chapter.

Results:

Designs of ONs:

First, to ensure the selective and stable binding of the two distal target RNAs, **CLIP-ON** was designed according to the upper end of the typical ON design length, 25-nt.[6] Briefly, the design involves two 12-nt segments targeting the sequences flanking the above-mentioned RCSs, which were then bridged by a 1-nt uracil linker as a spacer between the two independent hybridizing sections. The **CLIP-ON** sequence was further optimized to minimize extended self-complementarity and fulfill the standard requirements of ON designs, where the predicted energy was > – 4 kcal mol^{-1} for ON and > – 15 kcal mol^{-1} for ON-ON complexes.[7] **CLIP-ON** was then further fully modified with a phosphorothioate backbone and 2′-O-methyl, to improve its RNA nuclease stability[2,3], as it will be used for cellular experiments. In addition, as control ONs, a partially hybridizing control (**part-CLIP-ON**), where one of the hybridizing 12-nt segments was fully scrambled, and a fully scrambled control (**Scram-ON**) was utilized. All ON sequences that were used in this chapter are listed in **Table 2**.

In vitro evaluation of CLIP-ON design in its capacity to bind two distal RNA sequences:

Native PAGE analysis:

Due to the significant departure of **CLIP-ON**'s design philosophy compared to traditional ONs up until this point, the hybridization and formation of the desired bridged structure between **CLIP-ON** and the two distal intronic target sites were first evaluated via native polyacrylamide gel-electrophoresis (native PAGE). Here, the wells were loaded with an equimolar concentration of partial RNA sequence from **pre-mUAC** (RNA1 and RNA2) and the designed ONs, **CLIP-ON**, **part-CLIP-ON**, and **Scram-ON**, respectively. The results show, that in the absence of **CLIP-ON** (Fig. 10A lane 4), two distinct bands were observed corresponding to the respective RNAs (Fig. 10A lanes 1 and 2). The observed slower mobility of the band corresponding to RNA2, as compared to RNA1 despite similar fragment length, was attributed to the presence of a high GC content palindromic sequence which may potentially form stable self-dimers (Sup. Fig. 2). By itself, **CLIP-ON** showed a smeared band (Fig. 10A lane 3), which was attributed to the racemic nature of the phosphorothioate backbone modification.

Table 2. Oligonucleotide sequences used in experiments, the linker uracil is highlighted in bold red. All ONs were fully 2′-O-methyl phosphorothioate modified.

Oligonucleotide Name	Sequence
CLIP-ON	5′-GACUCGAGCGGC **U** CUUUGAAUUCCA-3′
CLIP-ON-10	5′-CUCGAGCGGC **U** CUUUGAAUUC-3′
CLIP-ON-8	5′-CGAGCGGC **U** CUUUGAAU-3′
CLIP-ON-NoLink	5′-GACUCGAGCGGC CUUUGAAUUCCA-3′
CLIP-ON-3Link	5′-GACUCGAGCGGC **UUU** CUUUGAAUUCCA-3′
Scram-ON	5′-AUGAUUCCUUCGAGAGCUCCGCGAU-3′
Part-CLIP-ON	5′-GACUCGAGCGGCAUCUCUCUGUAUA-3′

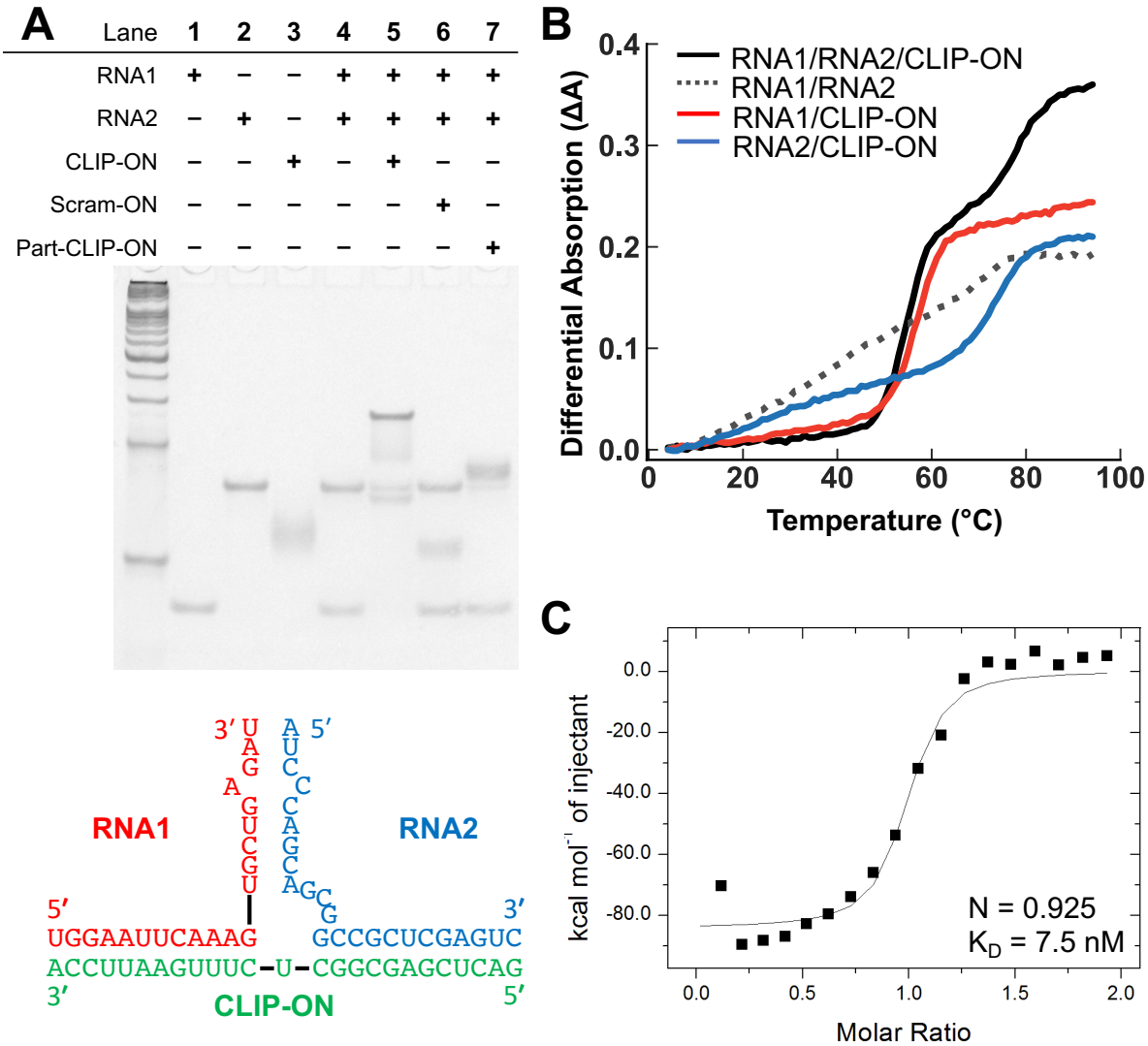


Figure 10. A) Native PAGE results obtained when various RNA and/or ONs mixed under equimolar conditions. The table (top) lists the components which were loaded in each lane. The predicted hybridized structure and the sequence of RNA1/RNA2 used for the native PAGE are shown (bottom). For the size marker, 20 bp DNA ladder was used. B) The representative thermal melting (T_m) profile of RNA1/RNA2/**CLIP-ON** mixture (black line) compared to RNA1/RNA2 (dotted line), RNA1/**CLIP-ON** (red line), and RNA2/**CLIP-ON** (blue line). C) ITC analysis of **CLIP-ON** and target RNA sequence hybridization.

Table 3. Calculated Tm and TA values for the thermal melting profile in Fig. 3B. (S.D. : Standard deviation.)

Sample	T_{m1} (°C)	S.D.	T_{m2} (°C)	S.D.	TA_1 ($°C^{-1}$)	TA_2 ($°C^{-1}$)
RNA1/RNA2/**CLIP-ON**	53.7	0.9	75.8	1.8	0.021	0.008
RNA1/RNA2	---	---	---	---	---	---
RNA1/**CLIP-ON**	56.5	---	---	---	0.016	---
RNA2/**CLIP-ON**	---	---	72.8	---	---	0.008

In the absence of **CLIP-ON** RNAs by themselves showed no hybridization under PAGE conditions (Fig. 10A lane 4). However, upon addition of **CLIP-ON**, a new low-mobility band was observed in addition to the depletion of faster mobility bands corresponding to RNA1, RNA2, and **CLIP-ON** (Fig. 10A lane 5), while a small fraction of band corresponding to remaining uncomplexed RNA2 dimer and RNA1/**CLIP-ON** was also observed. By contrast, **Scram-ON** showed distinct bands corresponding to the two RNAs and **Scram-ON** (Fig. 10A lane 6), whereas **part-CLIP-ON** showed a slight shift in the RNA2 band associated with**part-CLIP-ON**/RNA2 hybridization (Fig. 10A lane 7) and no changes were observed for RNA1. The observed shift in the band mobility in the presence of **CLIP-ON** indicated the successful formation of the RNA1/RNA2/**CLIP-ON** bridged structure. Confirming the feasibility of the **CLIP-ON** design concept and its ability to hybridize and facilitate interaction between two sequences that are located spatially distant from each other.

RNA thermal melting assay:

The thermal stability of the resulting RNA/**CLIP-ON** complex was then further evaluated using the thermal melting assay, as the overall thermal stability of the resulting bridged complex likely plays a major role in determining the biological efficacy of **CLIP-ON**.

In the presence of an equimolar concentration of the two RNAs and **CLIP-ON**, a biphasic sigmoidal curve in the Tm profile was observed (Fig. 10B, corresponding T_m values in Table 3), for which T_{m1} = 53.7 °C (S.D. = 0.9 °C) and T_{m2} = 75.8 °C (S.D. = 1.8 °C), respectively. While no apparent Tm was observed when evaluating the melting curve of the equimolar

mixture of the two RNA fragments in the absence of **CLIP-ON**. The two observed T_m values for the RNA1/RNA2/**CLIP-ON** mixture were shown to correspond to the dissociation of the individual strands from the complex, as when T_m of RNA1/**CLIP-ON** and RNA2/**CLIP-ON** was measured independently, the obtained T_m values were 56.5 °C and 72.8 °C respectively. Furthermore, the observed deviation of the T_{m1} and T_{m2} of RNA1/RNA2/**CLIP-ON** complex from the independently measured Tm of RNA1/**CLIP-ON** and RNA2/**CLIP-ON** by –2.8 °C and +3.0 °C, respectively, indicates the presence of cooperativity in the formation of the RNA1/RNA2/**CLIP-ON** complex. Based on this observation, the melting profiles were further evaluated by the largest tangent angle (TA) at the respective Tm values. Briefly, TA corresponds to the slope of the sigmoidal curve at the T_m value, for which larger values are associated with more abrupt dissociation of duplex, which is indicative of higher cooperativity.[8,9] Here, The TA corresponding to the first dissociation of the RNA1/RNA2/**CLIP-ON** complex, T_{m1} (0.021 °C^{-1}), is larger than that of RNA1/**CLIP-ON** (0.016 °C^{-1}), whereas TA corresponding to the second dissociation, T_{m2} (0.008 °C^{-1}), remained unchanged from that of RNA2/**CLIP-ON** (0.008 °C^{-1}).

The observation of initial dissociation temperature corresponding to one of the RNA strands at T_{m1} of 53.7 °C (S.D. = 0.9 °C), suggests the complex has very high thermal stability and that the equilibrium strongly favors the formation of the desired bridged structure at physiological temperature. Additionally, the larger TA associated with the initial stage of complex dissociation, T_{m1} (0.021 °C^{-1}), compared to the corresponding TA of RNA1/**CLIP-ON** (0.016 °C^{-1}), indicating the presence of a stronger cooperative effect present in the formation of the bridged complex when compared to simple duplex formation.

Isothermal titration calorimetry and surface plasmon resonance assay:

Next, the apparent binding affinity of **CLIP-ON** toward its target RNA sequences was further clarified via isothermal titration calorimetry (ITC, Fig. 10C), as the formation of an otherwise unorthodox complex was expected to reduce the overall binding affinity of **CLIP-ON**. An RNA comprised of a pair of partial RCS sequences and the relevant flanking sequences joined by a 4-uracil linker was used to simulate a 1:1 binding of **CLIP-ON** toward the target pre-mRNA site (RNA3, sequence listed in Table S2). The results show **CLIP-ON** binding to the RNA target with an equilibrium dissociation constant (K_D) of 7.5 nM. This value was also further confirmed with a surface plasmon resonance (SPR) assay. On streptavidin-coated sensors (SA chip, GE healthcare), a hairpin RNAs containing the partial sequence of the RCS

pair from the intron of **pre-mUAC** were immobilized. The results showed a K_D of 10 nM (Sup. Fig. 3).

Overall, the in vitro results indicated that the designed **CLIP-ON** could bind to two distal sequences with high affinity and thermal stability, providing strong evidence for the potential of **CLIP-ON** to form a stable ternary complex, bridged structure, to induce the upregulation of circRNAs in the cellular environment.

In cellular evaluation of CLIP-ON activity:

The efficacy of **CLIP-ON** as an externally introduced trans-factor in promoting circRNA production was examined in HeLa cells utilizing a pre-mRNA model expressing a well-studied circRNA: circZKSNCA1, which was previously reported by Wilusz et al. and was introduced in the previous chapter (expression plasmid: ***p*-UAC**, pre-mRNA: **pre-mUAC**).[5,10,11] The changes in circZKSCAN1 production in cells expressing **pre-mUAC**, following treatment with various **CLIP-ON** in cells were assessed using RT-qPCR. The relative quantity (RQ) of circZKSCAN1 in cells treated with **CLIP-ON** compared with untreated cells (**CLIP-ON** 0 nM treated cells) was then obtained by using the comparative C_t ($\Delta\Delta C_t$) method.[12] The "divergent" primer utilized specifically to amplify the circZKSCAN1 stemming from the model was designed as described in previous chapter (Fig. 11A).[1,13,14] Where a single nucleotide difference in exon 3 of plasmid-expressed circZKSCAN1 was utilized to allow the discrimination from endogenous circZKSCAN1 via primer design.[15]

CLIP-ON promotes circRNA production in model cellular environment and is sensitively dependent on its capacity to bridge distal intronic sequences:

Upon treatment with **CLIP-ON**, HeLa cells expressing the model pre-mRNA (**pre-mUAC**) showed an increase in the expression of circZKSCAN1 in a **CLIP-ON** concentration-dependent manner. Compared with untreated samples, cells treated with **CLIP-ON** achieved up to an apparent 5.2-fold ($p < 5 \times 10^{-8}$), 4.0-fold ($p < 5 \times 10^{-7}$), and 2.6-fold ($p < 5 \times 10^{-4}$) increase in circZKSCAN1 expression at the treated **CLIP-ON** concentration of 25, 13, and 6.3 nM, respectively (Fig. 11B). However, no statistically significant changes were observed for cells treated with control ONs, **part-CLIP-ON**, or **Scram-ON**, under the same 25 nM concentration. The qPCR amplified target and primer specificity was further confirmed via

PAGE analysis followed by sequencing, confirming the presence of the back-spliced junction sequence stemming from circZKSCAN1, and the circularity was then further confirmed by its resistance to RNase R treatment relative to the linear mRNA of β-actin (Fig. 11C and D). The observed **CLIP-ON** concentration-dependent upregulation of circZKSCAN1 demonstrates the capacity of the newly designed **CLIP-ON** to promote circRNA production in cells. Furthermore, the inability of both fully and partially scrambled control ONs to upregulate circZSKCAN1, provides evidence that the physical bridging induced by **CLIP-ON** played a crucial role in the observed upregulation.

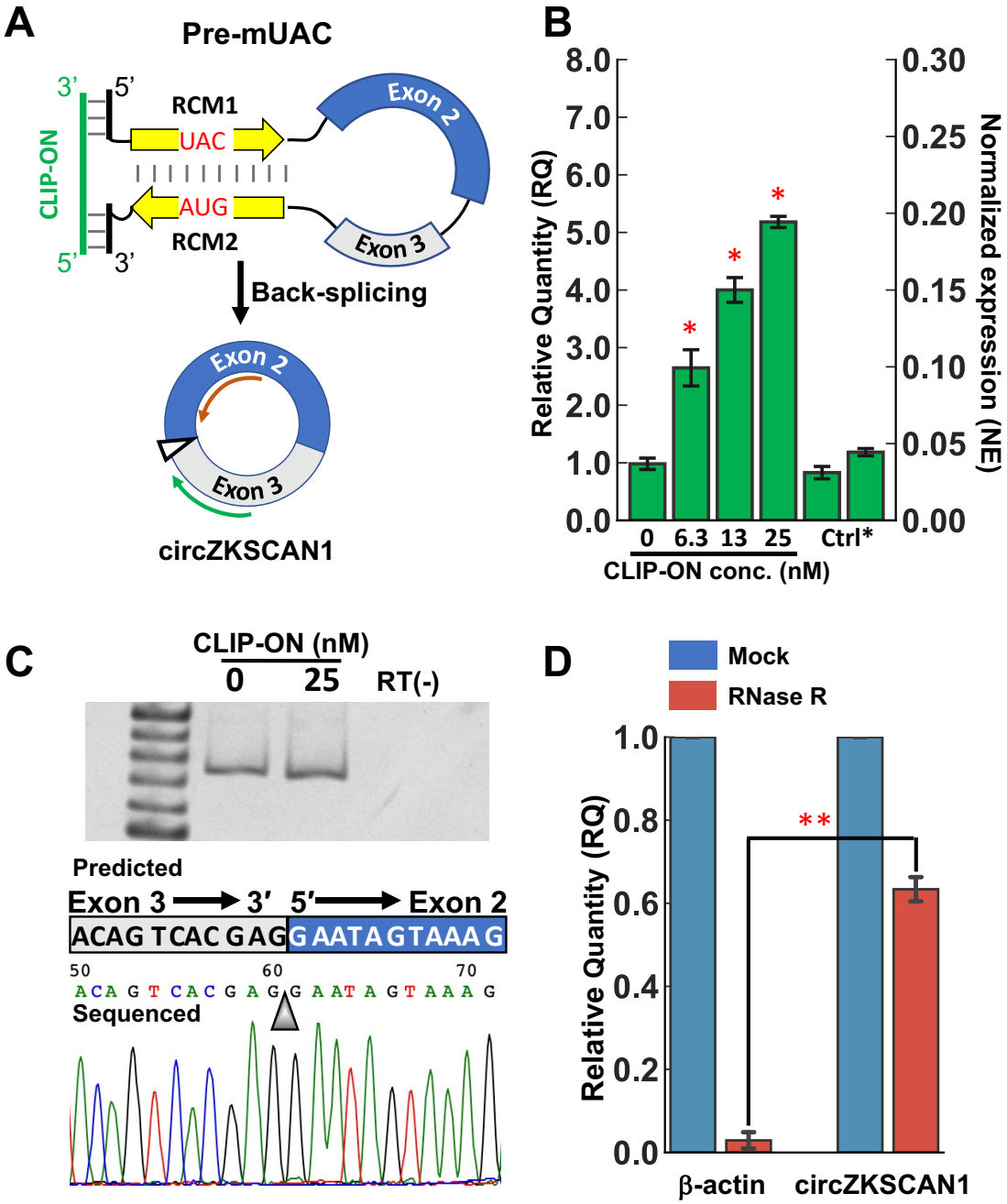


Figure 11. (A) Schematic representation of predicted **CLIP-ON**/**pre-mUAC** complex, followed by the schematic representation of circZKSCAN1. The site of the back-splice junction is shown with a triangle, while the green and orange arrows represent the primers used for the qPCR. The table denotes the nucleotide sequences modified between **pre-mUAC** and **pre-mGGA**. (B) **CLIP-ON** concentration-dependent change of circZKSCAN1 in HeLa cells transfected with ***p*-UAC**, expressing **pre-mUAC**. (C) PAGE analysis of amplified qPCR products (nontreated, 25 nM **CLIP-ON** treated, and non-RT), and the typical sequencing result of the amplified PCR product. (D) qPCR results following RNase R treatment of total RNA obtained from **pre-mUAC** expressing HeLa cells. (Ctrl*: 25 nM **Scram-ON** (left) and **part-CLIP-ON** (right) treated samples; *:$p < 0.005$, RQ > 2; the significance was obtained against the sample treated with 0 μM **CLIP-ON**, using two-tailed t-test, n = 4; **: $p < 10^{-5}$, two-tailed t-test, n=3).

CLIP-ON can promote circRNA production from pre-mRNA with destabilized RCS interaction and different exons:

To clarify whether bridging by **CLIP-ON** requires the distal target sites in the introns to be prearranged by stable RCS pairing, as it may potentially be the case in **pre-mUAC**. A model pre-mRNA where the inter-intronic interaction between the RCSs was disrupted by a 5′-GGA-3′/5′-GGA-3′ trinucleotide mismatch was used (Fig. 12 A, expression plasmid: ***p*-GGA**, pre-mRNA: **pre-mGGA**). The introduced mismatch site was shown in previous chapters to destabilize the hybridization of RCSs and impair the production of circZKSCAN1.[1,5]

Upon treatment with **CLIP-ON**, HeLa cells expressing the **pre-mGGA** showed **CLIP-ON** concentration-dependent increase in the expression of circZKSCAN1 similar to that of **pre-mUAC**. Compared with untreated samples, cells treated with **CLIP-ON** achieved up to an apparent 5.8-fold ($p < 5 \times 10^{-3}$), 4.2-fold ($p < 5 \times 10^{-5}$), and 3.6-fold ($p < 5 \times 10^{-3}$) increase in circZKSCAN1 expression at the treated **CLIP-ON** concentration of 25, 13, and 6.3 nM, respectively, and no statistically significant changes were observed for cells treated with control ONs under the same 25 nM concentration (Fig. 12B). The identity of the amplified target was confirmed via PAGE and Sanger sequencing (Sup. Fig. 5)

Due to **CLIP-ON** inducing a similar fold upregulation of circZKSCAN1 for cells expressing either **pre-mUAC** or **pre-mGGA**, β-actin normalized expression ($2^{-\Delta Ct}$, hereon as NE) was also calculated in addition to RQ to allow cross-examination of the relative expression between cells expressing the two model pre-mRNAs following ON treatment. The results show, that despite a similar fold increase in upregulation in circZKSCAN1, the NE of circZKSCAN1 in HeLa cells expressing **pre-mGGA** was expectedly 13-fold lower (NE = 0.0028, Fig. 12B) than that of **pre-mUAC** (NE = 0.037, Fig 11B), in untreated samples as demonstrated previously. The difference was attributed to the lower stability of RCS hybridization in **pre-mGGA** due to

the presence of a mismatch. While at the highest treated concentration, **CLIP-ON** showed the ability to compensate for the reduced stability and rescue the expression of circZKSCAN1 from **pre-mGGA** up to 43% (NE = 0.016) of HeLa cells expressing the full match counterpart.

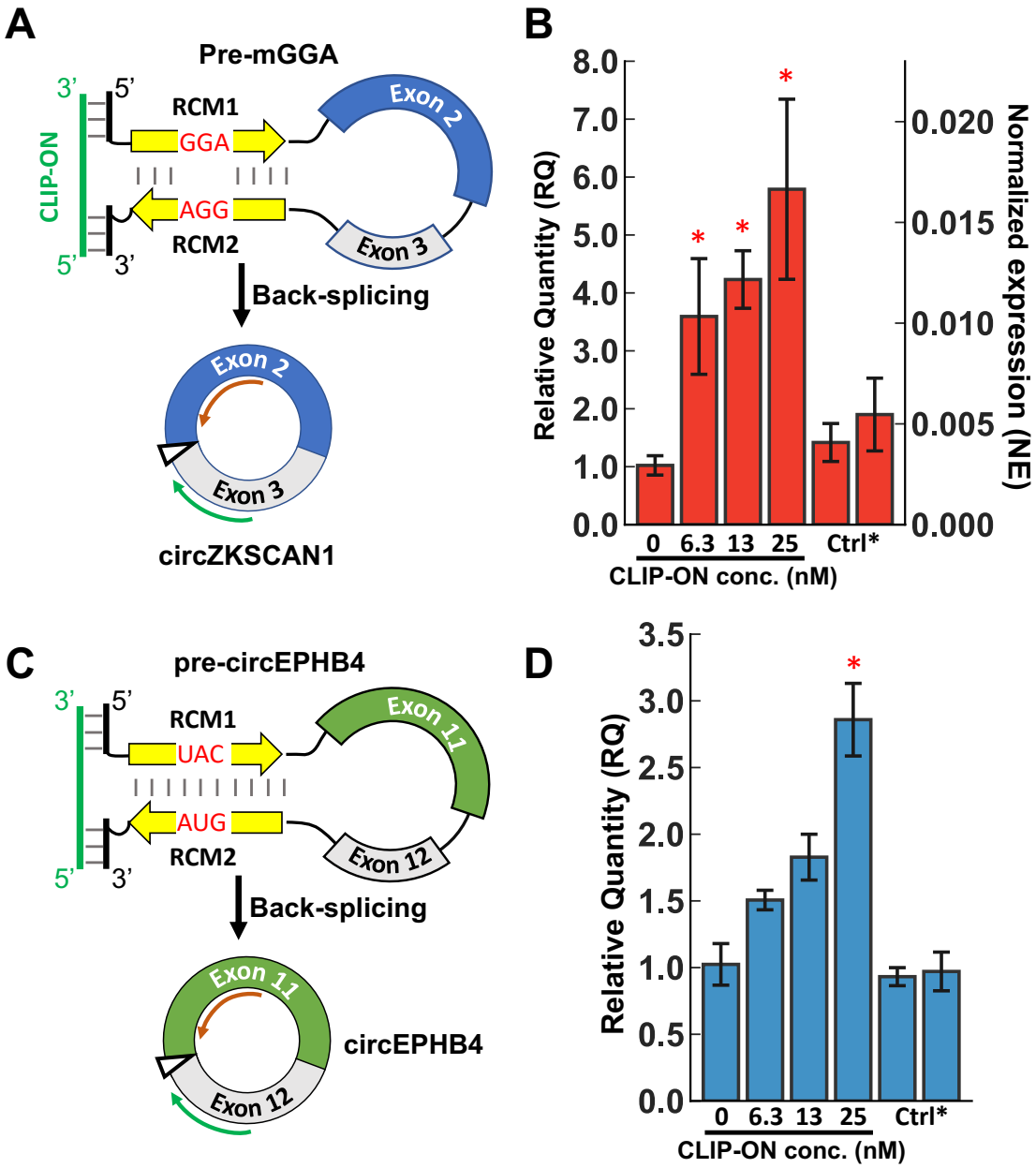


Figure 12. A) Schematic representation of predicted **CLIP-ON/pre-mGGA** complex, followed by the schematic representation of circZKSCAN1. The site of the back-splice junction is shown with a triangle, while the green and orange arrows represent the primers used for the qPCR. The table denotes the nucleotide sequences modified between **pre-mUAC** and **pre-mGGA**. (B) **CLIP-ON** concentration-dependent change of circZKSCAN1 in HeLa cells transfected with ***p*-GGA**, expressing **pre-mGGA**. C) Schematic representation of pre-mRNA modified to include exon 11 and 12 from gene EPHB4, **pre-circEPHB4**, and resulting circRNA, circEPHB4. D) **CLIP-ON** concentration-dependent circEPHB4 change for HeLa cells transfected with ***p*-circEPHB4** (Ctrl*: 25 nM **Scram-ON** (left) and **part-CLIP-ON** (right) treated samples; *:$p < 0.005$, RQ > 2; the significance was obtained against the sample treated with 0 μM **CLIP-ON**, using two-tailed t-test, n = 4).

Although in absolute terms relative to β-actin the upregulation was significantly higher for cells expressing **pre-mUAC** (ΔNE = 0.16) compared to **pre-mGGA** (ΔNE = 0.013).

Furthermore, to determine the effects of exon on the efficacy of **CLIP-ON**, the ZKSCAN1 exons were also replaced by exons that comprise circEPHB4, exons 11 and 12 from gene EPHB4 (Fig. 12C) .[5,16] They were also modified to allow discrimination from endogenously expressed circEPHB4 in a similar manner to circZKSCAN1, as described above. The divergent primers were then designed with an additional mismatch in the third position from the 3′ end of the reverse primer for specific amplification of the desired segment, (Sup. Fig. 4, the plasmid hereon as ***p*-circEPHB4** and transcribed pre-mRNA as **pre-circEPHB4**). Upon treatment, **CLIP-ON** similarly showed the ability to concentration-dependently upregulate the expression of circEPHB4, achieving a 2.9-fold ($p < 5 \times 10^{-4}$) increase in the production of circEPHB4 compared with untreated samples (Fig. 12D). No statistically significant changes were observed for cells treated with control ONs under the same 25 nM concentration. The qPCR amplified target and primer specificity was further confirmed via PAGE analysis followed by sequencing, confirming the presence of the back-spliced junction sequence stemming from circEPHB4 and the circularity was further confirmed by its resistance to RNase R treatment (Fig. 13AB).

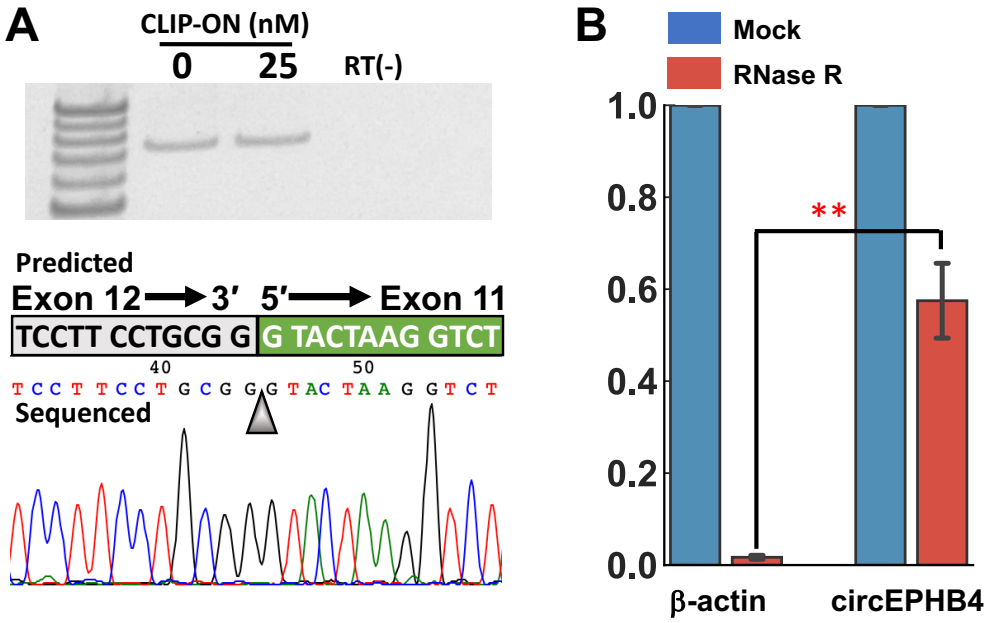


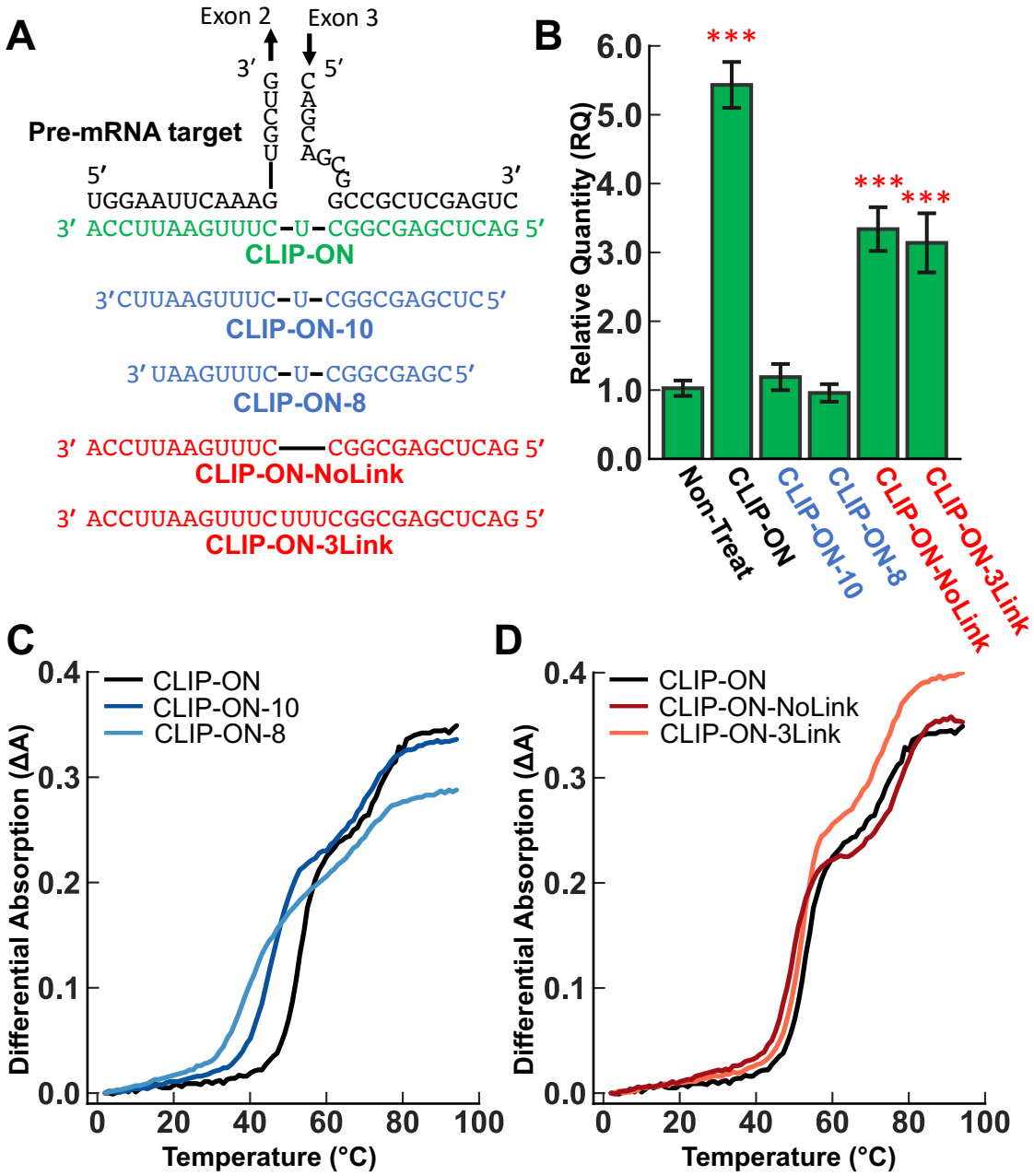
Figure 13. **A)** PAGE analysis of amplified qPCR products of cells transfected with ***p*-circEPHB4** (nontreated, 25 nM **CLIP-ON** treated, and non-RT), and the typical sequencing result of the amplified PCR product. **B)** qPCR results following RNase R treatment of total RNA obtained from **pre-circEPHB4** expressing HeLa cells. (**: $p < 10^{-5}$, two-tailed t-test, n=3).

Overall, **CLIP-ON** was able to promote circZKSCAN1 production from model pre-mRNA with destabilized RCSs, **pre-mGGA**, and achieve a similar upregulation of 5.8-fold at the highest treated **CLIP-ON** concentration, compared to the original **pre-mUAC** (Fig. 11B and 12B). This result demonstrates **CLIP-ON**'s capacity to bind and compensate for the loss in RCS hybridization stability introduced by the mismatches, indicating that the stable hybridization of RCSs is not a necessary requirement for **CLIP-ON**'s biological activity. Further experimentation using pre-mRNAs containing exons of other circRNA, **pre-circEPHB4**, showed the capacity of **CLIP-ON** to circularize exons other than ones comprising circZKSCAN1. Albeit with differing efficiencies, achieving a 2.9-fold upregulation of circEPHB4 (Fig. 12D). Suggests that the ability of **CLIP-ON** to upregulate circRNAs in the cellular environment is independent of the circularizing exon in question.

Elucidation of factors associated with CLIP-ON activity:

A variation of **CLIP-ON** with either 10-nt or 8-nt segments (**CLIP-ON-10** and **CLIP-ON-8**, respectively), or with 0-nt or 3-nt uracil linker (**CLIP-ON-NoLink** and **CLIP-ON-3Link**, respectively) was also prepared to further elucidate the effects of various components on the viability of **CLIP-ON** design (Figure 14A, Table 2). Towards this end, HeLa cells expressing **pre-mUAC** were treated with 25 nM of the above variants of **CLIP-ON**, and the changes in RQ of circZKSCAN1 were then determined using qPCR (Fig. 14B). The results show that for **CLIP-ON** variants with reduced hybridizing segment length, **CLIP-ON-10** and **CLIP-ON-8**, the ability to induce upregulation in circZKSCAN1 was completely abolished in both cases. On the other hand, despite having the same hybridizing segment as the original design **CLIP-ON** variants with different linker lengths, **CLIP-ON-NoLink** and **CLIP-ON-3Link**, showed a reduced biological activity, providing an apparent 3.3-fold ($p < 5 \times 10^{-5}$) and 3.1-fold ($p < 5 \times 10^{-4}$) upregulation of circZKSCAN1 at 25 nM, respectively.

To elucidate the underlying factors behind the observed differences in biological activity between the variants of **CLIP-ON**, a thermal melting analysis was conducted. For **CLIP-ON** variants with reduced hybridizing segments, the thermal melting profile corresponding to the dissociation of the bridged structure (T_{m1}) was reduced to 45.6 °C (S.D. = 0.1 °C) and 40.0 °C (S.D. = 0.4 °C) for **CLIP-ON-10** and **CLIP-ON-8**, respectively (Fig. 14C, Table 4). Interestingly, despite having the same hybridizing segment length, **CLIP-ON** variants with different linker lengths also showed a decrease in T_{m1}, with 49.5 °C (S.D. = 0.1 °C) and 51.8°C (S.D. = 0.3 °C) for **CLIP-ON-NoLink** and **CLIP-ON-3Link**, respectively (Fig. 14D, Table 4)



Oligonucleotide	T_{m1} (°C)	S.D.
CLIP-ON	53.7	0.9
CLIP-ON-10	45.6	0.1
CLIP-ON-8	40.1	0.4
CLIP-ON-NoLink	49.5	0.1
CLIP-ON-3Link	51.8	0.3

Figure 14. A) Comparison of the different variations of **CLIP-ON** used in this experiment. **B) CLIP-ON** variant dependent change in expression of circZKSCAN1 in HeLa cells expressing **pre-mUAC**, in all cases HeLa cells were treated with 25 nM of the respective ONs. **C)** The representative T_m profile of **CLIP-ON** variants with truncated hybridizing segments, **CLIP-ON-10** and **CLIP-ON-8**, with RNA1 and RNA2. **D)** The representative T_m profile of **CLIP-ON** variants with modified uracil linker, **CLIP-ON-NoLink** and **CLIP-ON-3Link**, with RNA1 and RNA2. The calculated respective T_m values for both (B) and (C) are shown in the table below, while the standard deviation (S.D.) is listed within brackets. (***:$p < 5 \times 10^{-4}$, RQ > 2; the significance was obtained against the sample treated with 0 μM **CLIP-ON**, using two-tailed t-test, n = 4).

Table 4. Calculated T_{m1} and TA values for the thermal melting profile in Fig. 12C and D (S.D. : Standard deviation.)

To elucidate the reason behind the observed difference in Tm between **CLIP-ON** variants with differing linker lengths, their binding characteristics were assessed using ITC (Sup. Fig. 6, Calculated values in Table 5). The results show that these **CLIP-ON** variants bind with a K_D of similar order and free energy *(ΔG*) as the **CLIP-ON** design. Where K_D was 8.7 nM and 2.8 nM for **CLIP-ON-NoLink** and **CLIP-ON-3Link** respectively; while calculated *ΔG* was –11.1 kcal mol^{-1}, –11.0 kcal mol^{-1}, and –11.7 kcal mol^{-1} for **CLIP-ON**, **CLIP-ON-NoLink**,and **CLIP-ON-3Link** respectively. However, further examination showed a shift in the thermodynamic parameters of hybridization upon changing the linker length of **CLIP-ON**. Relative to the **CLIP-ON** design, which showed ΔH = –84.1 kcal mol^{-1} with ΔS = –245 cal mol^{-1} $°C^{-1}$, **CLIP-ON-NoLink** binding was more enthalpically favorable with ΔH = –93.1 kcal mol^{-1} but entropically less favorable with ΔS = –275 cal mol^{-1} $°C^{-1}$; on the other hand, **CLIP-ON-3Link** binding was less enthalpically favorable with ΔH = –71.3 kcal mol^{-1} but was compensated by being more entropically favorable with ΔS = –200 cal mol^{-1} $°C^{-1}$.

Table 5. The thermodynamic parameters obtained from ITC measurements for **CLIP-ON** variants with different linker lengths.

Oligonucleotide	K_D (nM)	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	ΔS (cal mol^{-1} $°C^{-1}$)
CLIP-ON	7.5	-11.1	-84.1	-245
CLIP-ON-NoLink	8.7	-11.0	-93.1	-275
CLIP-ON-3Link	2.8	-11.7	-71.3	-200

These results suggest that despite not actively contributing to the hybridization and formation of the bridging structure, the linker length likely plays a crucial role in determining the stability of the resulting complex by modulating the spatial arrangement of hybridizing segments and

relaxation of rotational strain involved in hybridization with the two non-continuous distal sequences. This notion is further supported by the results of molecular dynamic simulation conducted using Maestro 11.1.012 (Sup. Fig. 7). Where, following the minimization of the molecular structure, the typical RNA A-helix structure was distorted near the bridge junction formed by CLIP-ON, in all cases (Sup. Fig 7A and BCD). However, of the three variants, the original CLIP-ON with one uracil linker showed the least amount of helical distortion and almost no crowding near the core of the bridge structure (Sup. Fig. 7B). On the other hand, CLIP-ON-NoLink showed a slight unwinding of the helical structure (Sup. Fig. 7C); while CLIP-3Link showed obvious deformation and elongation of the helical structure stemming from over-crowing near the core of the bridge structure (Sup. Fig. 7D). These results further revealed that the biological activity of **CLIP-ON** was sensitively dependent on the thermal stability of the resulting bridging structure, in which both the hybridizing segment and the linker selection play a major role.

Conclusion:

Overall, based on the obtained results showing the upregulation of circRNA in a cellular environment, the formation of bridged ternary structure in vitro, and the sensitive dependence of biological activity on the thermal stability of the formed bridged structure, we have demonstrated that our new ON design, **CLIP-ON**, can upregulate circRNA production in a cellular model likely in accordance with the hypothesized mechanism. These results further demonstrate the applicability of externally introduced trans-acting modulators in upregulating circRNA production in a cellular environment and present a stepping-stone in molecular designs that regulate circRNA production. It provides interesting prospects for future ON designs, as it demonstrates the utility of modified ONs for targeting distal sequences to promote or invoke biologically relevant secondary structures in a cellular environment, representing a departure from traditional ON designs. Future developments in the elucidation of pre-mRNA secondary structures, the flexibility of ON in target selection and designs could be fully leveraged to offer potential new strategies to regulate biologically relevant circRNAs.

Experimental Section:

General information:

Surface plasmon resonance (SPR) measurements were conducted using BIAcore T200 instrument (GE Healthcare). Tm melting temperature measurements were conducted on UV-2700 UV-Vis spectrophotometer (SHIMADZU). Tm and TA analysis was done using a custom Python code (Sup. Code 1, https://github.com/BenNiLu/Simple-Scripts-for-data-analysis/blob/main/Tm%20analysis/Tm_multiple_curvefit.py). ITC was conducted using MicroCal iTC200 calorimeter (Malvern Panalytical). The original plasmid construct, ***p*-UAC**, used in these experiments: pcDNA3.1(+) ZKSCAN1 nt 400-1782 delta440-500 delta1449-1735, was a gift from Jeremy Wilusz (Addgene plasmid # 60633 ; http://n2t.net/addgene:60633 ; RRID:Addgene_60633)24 and ***p*-GGA** was prepared as described previously.[1] HeLa genomic DNA was purchased from New England Biolabs Inc. and plasmid construct ***p*-circEPHB4** was prepared using in-Fusion® Snap Assembly Master Mix. Escherichia coli (E. coli) strain DH5α was used to amplify plasmid DNA and the amplified plasmid DNA was extracted and purified using NucleoBond® Xtra Midi EF (Macherey-Nagel). HeLa cells (RIKEN BRC, RCB0007) were used for transfection and were maintained at 37 °C under 5% CO2 in Dulbecco's modified eagle's medium (Sigma, D6429) supplemented with 10% (v/v) fetal bovine serum (MP Biomedicals) and penicillin-streptomycin (Gibco). All primers were purchased from Invitrogen. All 2′-omethyl phosphorothioate modified oligonucleotides (2′-OMe PS ON) were purchased from GeneDesign Inc. The feasibility of ON designs was assessed and predicted RNA structures were visualized using ViennaRNA package 2.6.1.39 FuGENE HDTM transfection reagent (Promega) was used for plasmid transfection and Lipofectamine 3000 (Thermo Fisher Scientific) was used for ON transfection. For RT-qPCR experiments, cell lysis and reverse transcription were conducted using Superprep® II cell lysis & RT kit (TAKARA), and real-time PCR (qPCR) experiments were conducted using QuantStudio™ 1 system (Applied Biosystem) using GoTaq® qPCR master mix (Promega). For Ribonuclease R (RNase R) treatment, cell lysis and total RNA extraction were conducted using NucleoSpin RNA (TAKARA), RNase R was purchased from Epicenter Technologies, and the reverse transcription was conducted using ReverTra Ace® qPCR RT Master Mix with gDNA remover (TOYOBO).

SPR assay:

1) Sensor preparation: For the SPR assay, RNA containing the partial intronic sequence containing the **CLIP-ON** target sites was immobilized onto the sensor chip SA (Cytiva), where the surface was coated with streptavidin. The surface of the sensor chip was first washed three times with 50 mM NaOH and 1 M NaCl for 60 sec with a flow rate of 30 μL min^{-1}. Then the biotin-labeled RNAs, biotin-TEG-5′-UGGAAUUCAAAGUGCUGAGAUUACAGGCGU GAGCUUUUGCUCACACCUGUAAUCCCAGCAGCGGCCGCUCGAGUC-3′ (b-TEG-RNA3, The **CLIP-ON** target regions are underlined), were immobilized onto the surface under the following conditions: 0.5 μM RNA in 10 mM HEPES, 500 mM NaCl, pH 7.4, the flow rate of 5 μL min^{-1} for the flow of 60 sec. The amount of RNA immobilized on the sensor chip SA was 990 RU.

2) SPR analysis protocol: SPR analysis for the binding of **CLIP-ON** to the RNA-immobilized surface was conducted on BIAcore T-200 and carried out by subsequently flowing 62.5,125, 250, 500, and 1000 nM of each compound for 60 sec of contacting time in HBS-EP+ buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.05 % v/v Surfactant P20, pH 7.4) (Cytiva) with the flow rate of 30 μL min^{-1} at 25 °C followed by dissociation of bound compounds by flowing running buffer for 120 sec. Surface regeneration was carried out after the assay of each compound using 1.2 mM NaOH, 0.2 M NaCl, and 0.1 mM EDTA solution, with a contact time of 180 sec and a flow rate of 30 μL min^{-1}. Kinetic parameters of the binding of **CLIP-ON** to the RNA-immobilized surface were obtained by using the single-cycle kinetics method.

Plasmid construct preparation:

The insert of EPHB4 (exon 11 and 12) was introduced into the original plasmid (pcDNA3.1(+) ZKSCAN1 nt 400-1782 delta440-500 delta1449-1735, ***p*-UAC**) using the in-Fusion® Snap Assembly Master Mix. Briefly, ***p*-UAC** fragment was amplified using the primers (5′-ACA GGC GGA GGA GCC CAG AGG TAA GAA GCA AGG AAA AGA ATT AGG C-3′) and (5′-TAT AAA AAA AAA GTC ATT AGT ACG TAC CTT TAC AAA AAG TG-3′), and the EPHB4 insert was amplified from HeLa genome using primers (5′-TTT TTT ATA CTT CAG GTA CTA AGG TCT ACA TCG ACC CCT TCA CTT ATG AAG ACC CTA ATG AGG CTG CGA GGG-3′) and (5′-TTC CTT GCT TCT TAC CCG CAG GAA GGA GTC CAG GGC GCC GTT C-3′). The fragments were ligated under the manufacture's recommended conditions and the constructs were then amplified via E. coli DH5α cells and collected as

described previously. The final sequence of the region coding for **pre-circEPHB4** are shown in Fig. S4.

Reverse-transcription quantitative PCR (RT-qPCR experiments):

The HeLa cells were plated in a 96-well plate at a density of 1 x10^4 cells/well and were incubated for 4 hrs to allow the cells to adhere to the well bottom. It was then transfected with either ***p*-UAC**, ***p*-GGA**, or ***p*-circEPHB4** expressing plasmid using Fugene HD™ with the ratio (FuGENE : plasmid = 5:2, v/w) according to the manufacturer's instructions. After 24 hrs incubation under 37 °C with 5% CO2, the medium was exchanged with fresh medium, before being transfected with ONs using Lipofectamine 3000, according to the manufacturer's instructions, and the cells were further incubated for 24 hrs. The transfected cells were lysed and the total RNA was reverse transcribed into cDNA using the Superprep® II cell lysis & rt kit according to the manufacturer's instructions. The resulting cDNA was diluted 2-fold before performing qPCR analysis.

qPCR experiments were conducted with QuantStudio™ 1 system (Applied Biosystem) and GoTaq® qPCR master mix (Promega) using the cDNA as the template. Each 10 µl reaction contained 1 µl of the cDNA solution, 0.2 µl each of forward and reverse primers (10 µM), 0.1 µl of 100x CXR dye, and 5 µl of GoTaq® qPCR Master Mix. A negative control without cDNA template or samples without reverse transcriptase was included in each assay. The primers used for the detection of β-actin and the various target circRNAs are listed below in Table S1.

For comparison of the fold change in expression from cells transfected with the same plasmid under various concentrations of **CLIP-ON**, the relative quantity (RQ) was obtained using the comparative Ct (ΔΔCt) method by subtracting the cycle threshold value (Ct) of target circRNA twice, first from that of β-actin then from target circRNA of untreated sample. The RQ was then obtained accordingly through $RQ = 2^{-\Delta\Delta Ct}$. To compare the relative difference in circZKSCAN1 upregulation between **pre-mUAC** and **pre-mGGA**, the relative expression of circZKSCAN1 to the endogenous control β-actin was used. For which the cycle threshold value (C_t) of circZKSCAN1 was first subtracted from that of β-actin, and normalized expression value (NE) was then obtained through $NE = 2^{-\Delta Ct}$. The qPCR products were analyzed by 8% (29:1, acylamide/bis-acrylamide) polyacrylamide gel electrophoresis to check the amplicon size. The specificity of primer pairs was also verified with the presence of a single peak in the melting curve after PCR amplification and sequencing of the amplicon.

For RNase R treatment, HeLa cells plated in a 24-well plate (5 x 10^4 cells/well) were first transfected with the expression plasmid and treated in the same manner as described above. Total RNA was extracted using NucleoSpin RNA according to the manufacturer's instructions. 500 ng of the total RNA was then incubated for 10 min at 37 °C with or without 1 U/μg RNase R. The resulting solution was directly subjected to reverse transcription using ReverTra Ace® qPCR RT Master Mix with gDNA remover, according to manufacturer's instructions. The expression was then analyzed by qPCR as mentioned above, the experiment was conducted 3 independent times.

Isothermal titration calorimetry (ITC) assay:

A solution of hairpin RNA containing partial RCS and **CLIP-ON** binding sequence (RNA3, 1 μM) was titrated with a solution containing **CLIP-ON** and its variants (10 μM) at 25 °C cell temperature in sodium phosphate buffer (10 mM, pH 7.0) containing NaCl (100 mM). The temperature inside the cell was kept at 25 °C during the experiment. The titration experiment was performed by 19 times injections of the titrant (0.4 μL for the first and 2 μL for all following injections) from the syringe into the sample cell at 25 °C (stirred at 750 rpm). The initial time prior to the first injection was 60 seconds. The duration for each injection was 4 seconds, and the interval between the two nearest injections was 150 seconds. Additionally, the titrant was also injected into the cell containing only buffer solution in the same manner to obtain the data for the heat of dilution of the titrant. Prior to fitting analysis, the data point corresponding to the first injection was removed, and the heat data were corrected by subtracting the heat of dilution. Binding isotherms were fitted to the Origin models by least-squares analysis.

Molecular dynamic simulation of CLIP-ON bridged structure:

The bridged structure of CLIP-ON, CLIP-ON-NoLink, and CLIP-ON-3Link was molded in Maestro 11.1.012 using typical RNA A-helix as the starting structure. The molecular dynamics simulation and energy minimization were then conducted under the following parameters. Force field: AMBER*; Solvent: Water; Maximum iterations: 125000. Default values were used for other parameters.

The figures were then exported to PyMOL 2.5.4 to produce the figures.

Supplementary Figures:

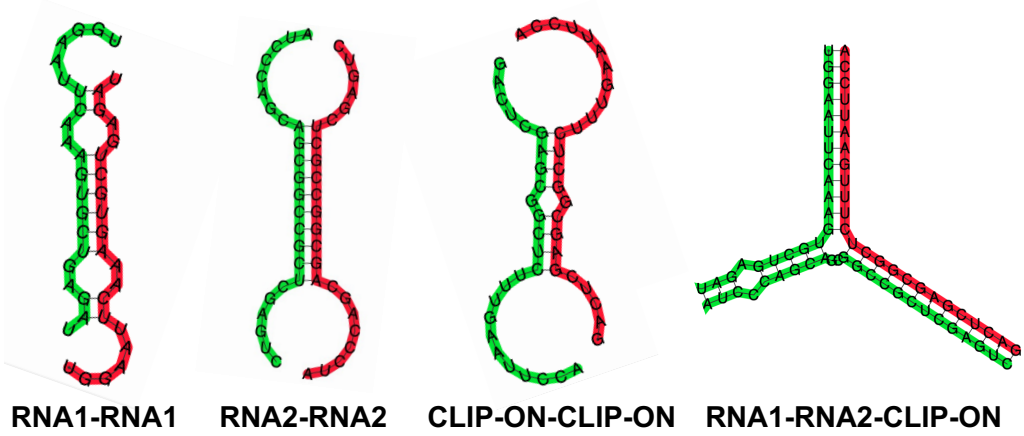
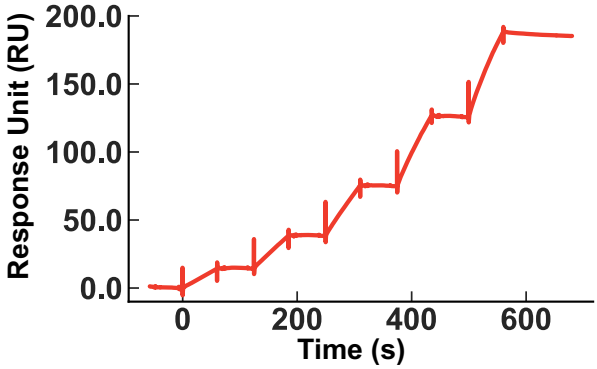


Figure S2. The visualized potential secondary structure of **A)** RNA1 dimer, B) RNA2 dimer, C) **CLIP-ON** dimer, and D) RNA1/RNA2/**CLIP-ON** complex.



k_a	1.53 x 10^4 $M^{-1}s^{-1}$
k_d	1.57 x 10^{-4} s^{-1}
K_D	1.02 x 10^{-8} M

Figure S3. SPR assay to determine the affinity of **CLIP-ON** towards model hairpin RNA, b-TEG-RNA3 (Table S1), was immobilized on the gold surface containing its target sequence. Kinetic parameters of the binding of **CLIP-ON** to the RNA-immobilized surface were obtained using the single-cycle kinetics method and shown in the table on the left.

TGGAATTCaaagtgctgagatTACaggcgtgagccaccacccccggccCACTTTTTGTAAAG
GTACGTACTAATGACTTTTTTTTTATACTTCAGGTACTAAGGTCTACATCGACCCCTTCACT
TATGAAGACCCTAATGA

<-3'-GCACCCTTAAACGTTTTCTCTAG-5': circEPHB4_Rev
GGCTGCGAGGGAATTTGCAAAAGAGATCGATGTCTCCTACGTCAAGATTGAAGAGGTGATTG
GTGCAGGTGAGAGCCGAAGGCTGCCCGGGCACCTGGGAACGAAGCGGGGGTGGGCAGGGCCA
CACTGGAGCGGGAGAGCTGATGACCTCTGCGTCCTTGTTTGAAGGTGAGTTTGGCGAGGTGT
GCCGGGGGCGGCTCAAGGCCCCAGGGAAGAAGGAGAGCTGTGTGGCAATCAAGACCCTGAAG
GGTGGCTACACGGAGCGGCAGCGGCGTGAGTTTCTGAGCGAGGCCTCCATCATGGGCCAGTT
CGAGCACCCCAATATCATC
5'-GCGTGGTCACCAACAGCAT-3'-> :circEPHB4_Fw
CGCCTGGAGGGCGTGGTCACCAACAGCATGCCCGTCATGATTCTCACAGAGTTCATGGAGAA
CGGCGCCCTGGACTCCTTCCTGCGGGTAAGAAGCAAGGAAAAGAATTAggctcggcacggta
gctcacacctGTAatcccagcagcgGCCGCTCGAGTC

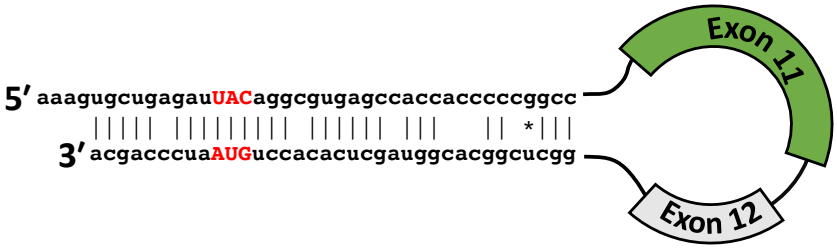


Figure S4. Whole sequence of the region coding for the circEPHB4 expressing ***p*-circEPHB4**, and the predicted resulting pre-circEPHB4 hairpin structure (bottom). The divergent primer alignment is also shown and the arrow indicates the direction of PCR amplification. Characters in green denote the exon 11 region, and the ones highlighted in gray denote the exon 12 region. The yellow highlighted lower-case letters in bold denote the RCSs which are predicted to form the stem of the hairpin. Characters in bold underlined green denote the target sequences of **CLIP-ON**. The single nucleotide difference in exon 11 is highlighted in red, in the reference genome, it is expected to be T instead of C. A single mismatch within the designed circEPHB4_Rev primer is denoted to discriminate against another potential primer seed sequence located in exon 12 (exon 12, underlined sequence).

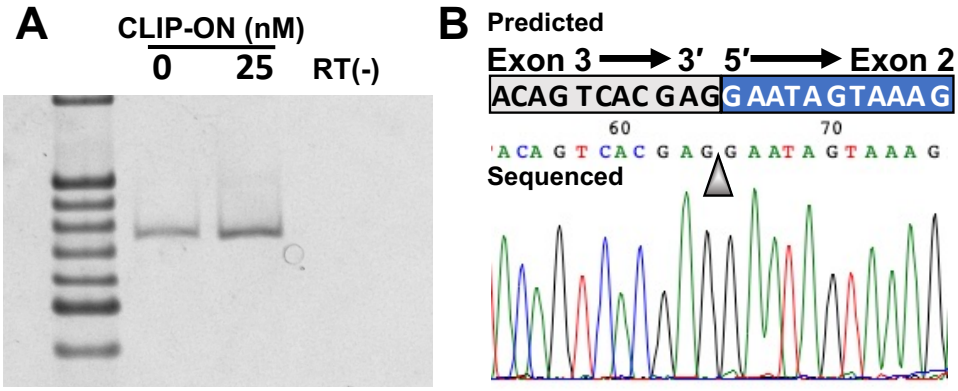


Figure S5. **A)** PAGE analysis of amplified qPCR products of cells transfected with ***p*-GGA** (nontreated, 25 nM **CLIP-ON** treated, and non-RT). **B)** the typical sequencing result of the amplified PCR product.

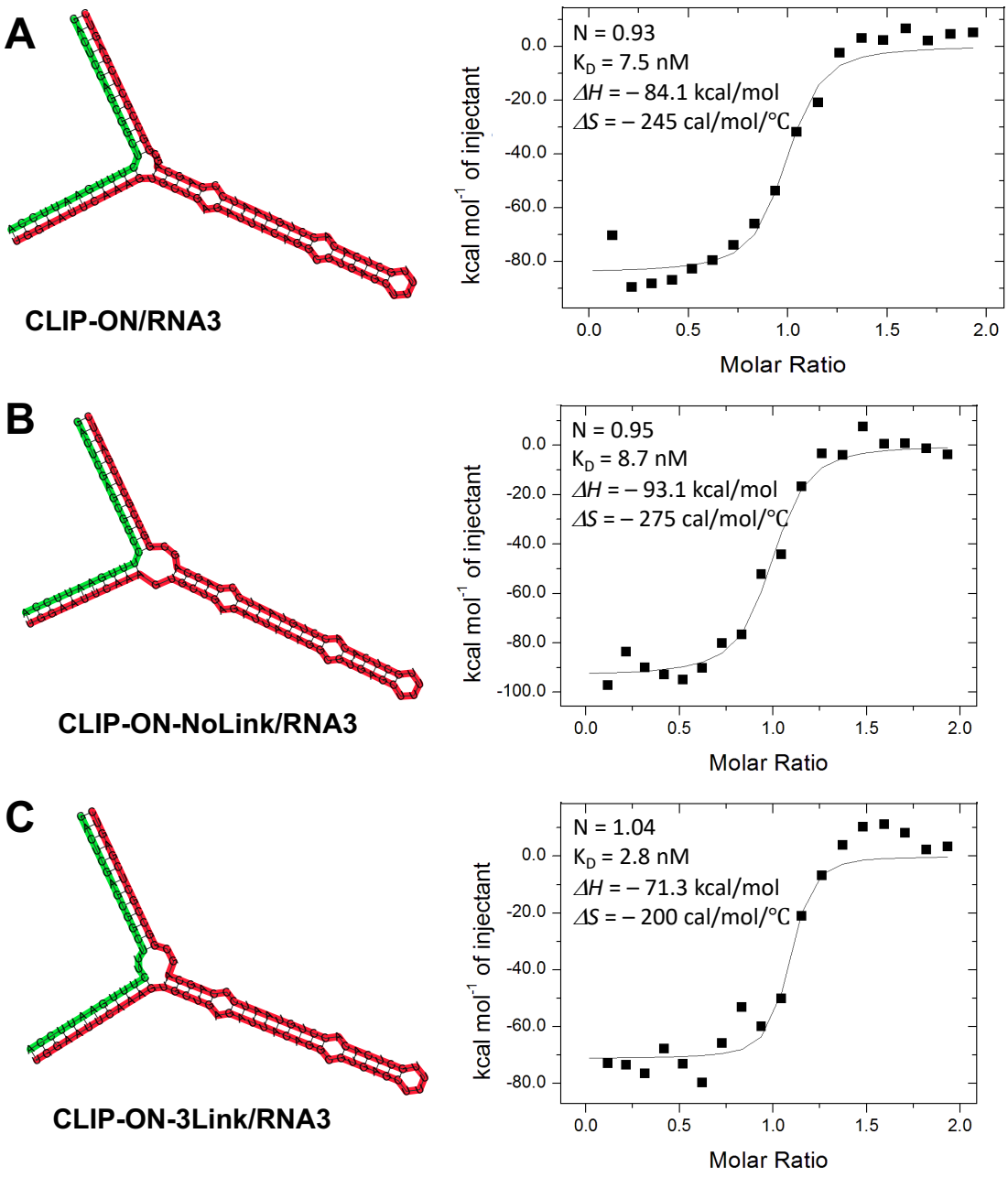


Figure S6: Visualization of the hybridized structure of **A) CLIP-ON**, **B) CLIP-ON-NoLink**, or **C) CLIP-ON-3Link** in green and the target RNA3 sequence in red, their respective ITC profiles are shown on the left.

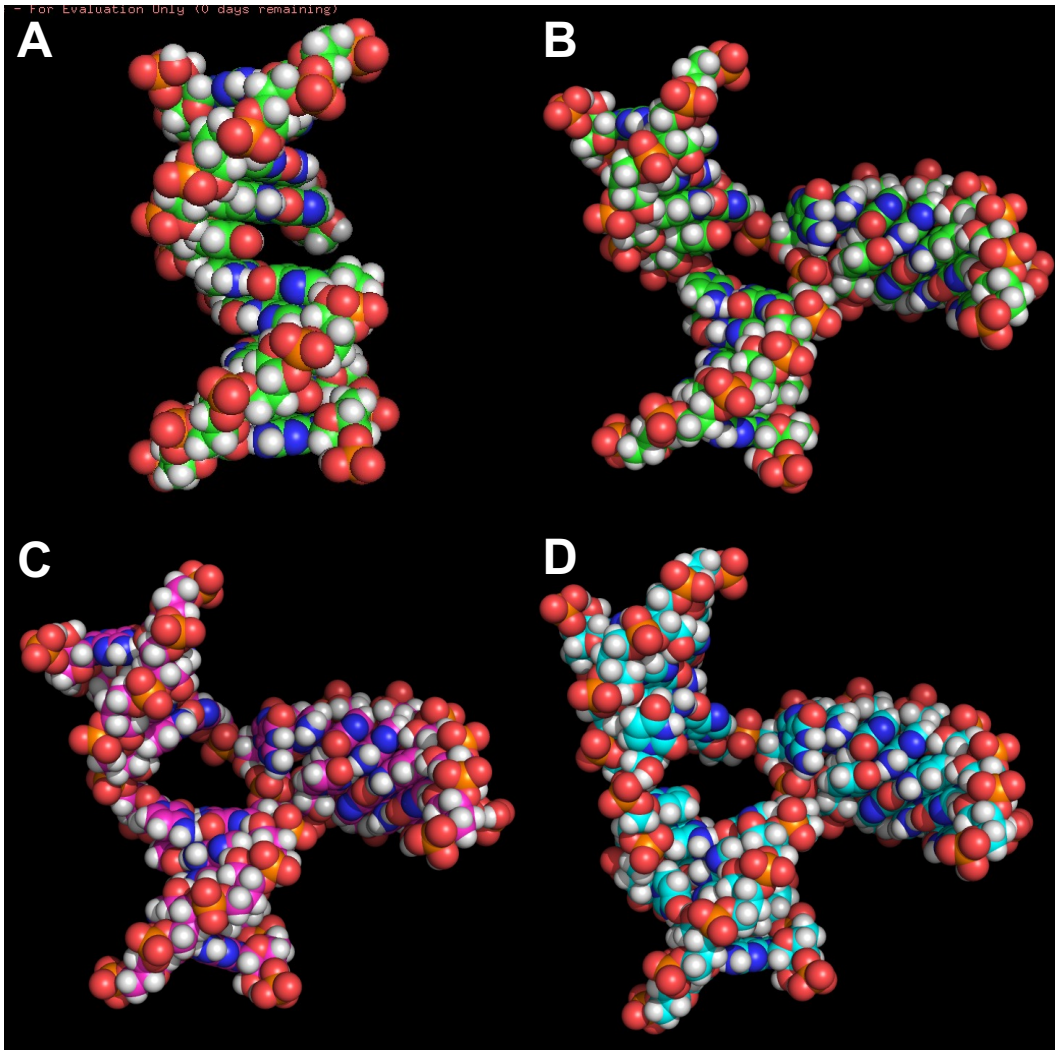


Figure S7: Simulated hybridization structure of **A)** Typical RNA A-Helix, **B) CLIP-ON** bridged structure, or **C) CLIP-ON-NoLink** bridged structure **D) CLIP-ON-3Link** bridged structure.

Table S2. Primers and RNAs used for the experiments, the linker uracil is highlighted in bold red, **CLIP-ON** target sequences are highlighted in bold blue, and parital RCS sequence is highlighted in bold green.

Sequence Name	Sequence
Beta_actin_mRNA_Fw	5′-CTCTTCCAGCCTTCCTTCCT-3′
Beta_actin_mRNA_Rev	5′-AGCACTGTGTTGGCGTACAG-3′
circZKSCAN1_Fw	5′-CAGTTCAGGAGTCCTCGAGCC-3′
circZKSCAN1_Rev	5′-CTCACCTTTATGTCCTGGGAGGT-3′
circEPHB4_Fw	5′-GCGTGGTCACCAACAGCAT-3′
circEPHB4_Rev	5′-GATCTCTTTTGCAAATTCCCACG-3′

RNA1	5′-UGGAAUUCAAAGUGCUGAGAU-3′
RNA2	5′-AUCCCAGCAGCGGCCGCUCGAGUC-3′
b-TEG-RNA3	Biotin-TEG-5′- UGGAAUUCAAAGUGCUGAGAUUACAGGCGUGAGC UUUU GCUCACACCUGUAAUCCCAGCAGCGGCCGCUCGA GUC-3′
RNA3	5′- UGGAAUUCAAAGUGCUGAGAUUACAGGCGUGAGC UUUU GCUCACACCUGUAAUCCCAGCAGCGGCCGCUCGA GUC-3′

Supporting Code 1:

Packages used: Numpy 1.24.4, Scipy 1.10.1, Pandas 2.0.3, Matplotlib 3.7.2, Sympy 1.12, Argument used to generate the data: --sn 2 --l 30

```
def main (input_DF, file_name):
    xdata = np.array(input_DF['Temp'].values.tolist())
    ydata = np.array(input_DF['ABS'].values.tolist())
    ydata = ydata-ydata.min() #subtract baseline absorption

    def sigmoid(x, L=max(ydata), x0=21, k=0.6, b=5): #define sigmoidal curve function to fit
        y = (L / ((1 + np.exp(-k*(x-x0)))))+b
        return (y)

    p0 = [max(ydata), np.median(xdata), 1, 0] #initial guess condition for sigmoid curve

    popt, pcov = curve_fit(sigmoid, xdata, ydata, p0, method='dogbox',  maxfev=100000) #fit
    L=popt[0]
    x0=popt[1]
    k=popt[2]
    b=popt[3]
    e = np.e

    solved_sigmoid_func = (L/(1+e**(-k*(x-x0))))+b #best fit sigmoid function (for use below)
    print(solved_sigmoid_func)

    Sigmoid_deriv = solved_sigmoid_func.diff(x) #first differentiation (for Tangent angle value)
    Sec_Sig_deriv = solved_sigmoid_func.diff(x, 2) #second differentiation (for Tm value)

    local_max = solve([x >= xdata.min(), x <= xdata.max(),Sec_Sig_deriv],x)
#solve the second derivative function -> Tm value
    TM_value = float(str(local_max)[5:19]) #Parse Tm from output
    TA_value=Sigmoid_deriv.subs({x:TM_value}) #input Tm value to solve for the largest Tangent Angle

    Summary = 'Tm = '+str(TM_value)+', TA = '+str(TA_value)
    print(Summary)

    ydd = []
    for i in xdata: #out put data for plotting first derivative line
        ydd.append(Sigmoid_deriv.subs({x:i}))
```

```
    yd = sigmoid(xdata, *popt)

    plt.plot(xdata, ydata, 'o', label='data') #plot the data
    plt.plot(xdata,ydd, '-', label='fit_deriv') #plot the first derivative curve
    plt.plot(xdata,yd, label='fit') #plot the fit line
    plt.legend(loc='best')
    plt.savefig(f'{file_name}_Sigmoid_fit.png')
    df2 = pd.DataFrame({'name':str(file_name),'Tm':TM_value,'TA':TA_value}, index=['i',])
    df2.to_csv("TA.csv", mode='a', header=True, index=False)

def find_Tm_range(Input_DF, reference_frame_lower, reference_frame_upper, lower_limit, higher_limit):
    temp_Df=Input_DF.drop(Input_DF[Input_DF.Temp < lower_limit].index)
    temp_Df.drop(temp_Df[temp_Df.Temp > higher_limit].index, inplace=True)
    temp_Df['diff']=temp_Df['ABS'].diff()
    idxmax=temp_Df['diff'].idxmax()
    Temp=temp_Df['Temp']
    Tm=int(Temp[idxmax])
    if Tm-int(reference_frame_lower) < int(Input_DF['Temp'].iat[0]):
        start_temp = Input_DF['Temp'].iat[0]
    else:
        start_temp = Tm-int(reference_frame_lower)
    if Tm+int(reference_frame_upper) > int(Input_DF['Temp'].iat[-1]):
        end_temp = Input_DF['Temp'].iat[-1]
    else:
        end_temp = Tm+int(reference_frame_upper)

    target_range = [start_temp, end_temp]
    print(target_range)
    return target_range

def Fit_sigmoid(infile=None, no_sig=1, reference_frame_lower=15, reference_frame_upper=15,
lower_limit=30, higher_limit=90):
    df=pd.read_csv(infile, names=('Temp', 'ABS')) #Parse data
    for i in range(int(no_sig)):
        Target_Fit_range = find_Tm_range(df, reference_frame_lower,reference_frame_upper, lower_limit,
higher_limit)
        df2 = df.drop(df[df.Temp < Target_Fit_range[0]].index)
        df2.drop(df2[df2.Temp > Target_Fit_range[1]].index, inplace=True)
        main(df2, infile)
        if Target_Fit_range[0]-int(df['Temp'].iat[0]) <= int(df['Temp'].iat[-1])-Target_Fit_range[1]:
            df.drop(df[df.Temp < Target_Fit_range[1]].index, inplace=True)
        elif Target_Fit_range[0]-int(df['Temp'].iat[0]) > int(df['Temp'].iat[-1])-Target_Fit_range[1] and
Target_Fit_range[0] > lower_limit:
            df.drop(df[df.Temp > Target_Fit_range[0]].index, inplace=True)
        else:
            df.drop(df[df.Temp < Target_Fit_range[1]].index, inplace=True)

if __name__ == '__main__':
    parser = argparse.ArgumentParser(description="Determine Tm and TA of melt curve, adapted for
single/multiphasic curves directly from raw txt file")
    parser.add_argument("--f", "--inFile", help="The name of TXT file to be imported. E.g. something.txt")
    parser.add_argument("--sn","--No_Sigmoid", type=int, default=1, help="Number of Sigmoidal curve
expected in the data, default is 1")
    parser.add_argument("--rfl","--ref_frame_l", type=int, default=15, help="Lower range distance [data point]
from the Tm to be taken account when fitting curve, default is 15")
    parser.add_argument("--rfh","--ref_frame_h", type=int, default=15, help="Higher range distance [data point]
from the Tm to be taken account when fitting curve, default is 15")
    parser.add_argument("--l", "--lower_limit", type=int, default=15, help="The minimum temperature below
which it will not look for a Tm curve, default is 15")
```

```
    parser.add_argument("--h", "--higher_limit", type=int, default=80, help="The maximum temperature above
which it will not look for a Tm curve, default is 90")
    args = parser.parse_args()
    inFile = args.f
    no_sig = args.sn
    reference_frame_lower = args.rfl
    reference_frame_upper = args.rfh
    lower_limit = args.l
    higher_limit = args.h
    Fit_sigmoid(inFile, no_sig, reference_frame_lower, reference_frame_upper, lower_limit, higher_limit)
```

References

1 L. Ni, T. Yamada, A. Murata and K. Nakatani, *Chem. Commun.*, 2022, **58**, 3629–3632.

2 T. C. Roberts, R. Langer and M. J. A. Wood, *Nat. Rev. Drug Discov.*, 2020, **19**, 673–694.

3 J. H. P. Chan, S. Lim and W. F. Wong, *Clin. Exp. Pharmacol. Physiol.*, 2006, **33**, 533–540.

4 A. Aartsma-Rus, L. van Vliet, M. Hirschi, A. A. M. Janson, H. Heemskerk, C. L. de Winter, S. de Kimpe, J. C. T. van Deutekom, P. A. C. 't Hoen and G. J. B. van Ommen, *Mol. Ther.*, 2009, **17**, 548–553.

5 D. Liang and J. E. Wilusz, *Genes Dev.*, 2014, **28**, 2233–2247.

6 P. Jafar-nejad, B. Powers, A. Soriano, H. Zhao, D. A. Norris, J. Matson, B. DeBrosse-Serra, J. Watson, P. Narayanan, S. J. Chun, C. Mazur, H. Kordasiewicz, E. E. Swayze and F. Rigo, *Nucleic Acids Res.*, 2021, **49**, 657–673.

7 W. M. C. van Roon-Mom and A. Aartsma-Rus, in *Methods in Molecular Biology*, 2012, vol. 867, pp. 79–96.

8 L. Ni, T. Yamada and K. Nakatani, *Chem. Commun.*, 2020, **56**, 5227–5230.

9 T. Yamada, S. Miki, L. Ni and K. Nakatani, *Chem. Commun.*, 2018, **54**, 13072–13075.

10 D. Liang and J. E. Wilusz, *Genes Dev.*, 2014, **28**, 2233–2247.

11 Z. Yao, J. Luo, K. Hu, J. Lin, H. Huang, Q. Wang, P. Zhang, Z. Xiong, C. He, Z. Huang, B. Liu and Y. Yang, *Mol. Oncol.*, 2017, **11**, 422–437.

12 K. J. Livak and T. D. Schmittgen, Methods, 2001, 25, 402–408.

13 W. R. Jeck and N. E. Sharpless, Nat. Biotechnol., 2014, 32, 453–461.

14 A. C. Panda, K. Abdelmohsen and M. Gorospe, in *Oncogene-Induced Senescence*, ed. M. A. Nikiforov, Springer New York, New York, NY, 2017, vol. 1534, pp. 79–87.

15 S. Lefever, F. Pattyn, J. Hellemans and J. Vandesompele, Clin. Chem., 2013, 59, 1470–1480.

16 C. Jin, J. Zhao, Z. P. Zhang, M. Wu, J. Li, B. Liu, X. Bin Liao, Y. X. Liao and J. P. Liu, *Mol. Oncol.*, 2021, **15**, 596–622.

Note

A part of Chapter 2 is under submission to Scientific Reports entitled "Utility of oligonucleotide in upregulating circular RNA production in a cellular model." (https://doi.org/10.21203/rs.3.rs-3849465/v1) This work is licensed under a CC BY 4.0 License.

Future Prospective

The concepts described in this thesis provide a conceptual basis for utilizing externally introduced ligands, whether they be small molecules or medium-sized molecules such as oligonucleotides, to upregulate circRNA production in the cellular environment. Given a strong enough capacity to stabilize or facilitate the formation of RNA-RNA interactions, such molecules have the potential to be utilized to modulate long-distance intronic interactions present within pre-mRNAs. Ultimately promoting the formation of the necessary large-hairpin-like structure for circRNA production. At the current rate of development seen in the field of RNA binding molecules and designs, with new molecules becoming ever more available, the future potential of ligand-modulated circRNA production is seemingly limitless.[1,2]

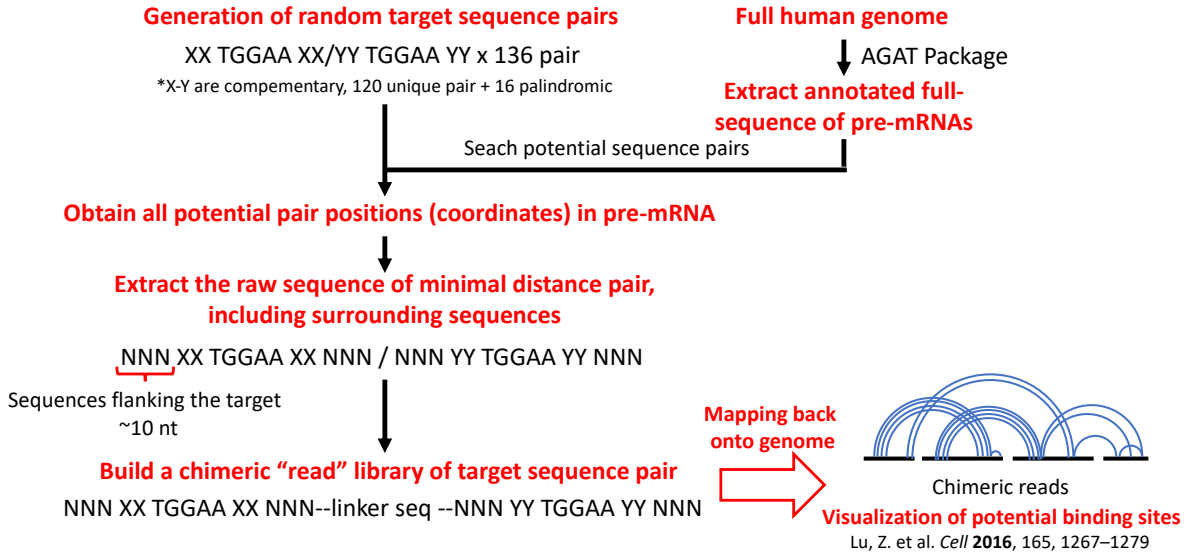
However, despite the simplicity of the concept, the applicability of the concept beyond model studies toward targeting endogenously expressed circRNA targets still faces significant difficulties and obstacles. The most obvious of which, would be the identification of potential target sites within the desired pre-mRNAs. As there are no robust databases that can be used to assist in the identification of any potential target sites.[3,4] Admittedly, certain genomic features, such as the presence of SINE:Alu RCS pairs can provide clues on potential target sites, and have been the prime indicator of circRNA production.[5] However, recent evidence has suggested that most abundant circRNAs are not correlated with such long repeats, with evidence showing more involvement from shorter non-repeating RCSs and proteins.[6-8]

To remedy such limitations, attempts have been made to train machine learning algorithms to identify such interacting sequences within introns, the inherent lack of information regarding such intronic interactions resulted in an overall high false positive rate, greatly limiting its applicability.[9]

Fortunately, in the case of locating potential small molecule targeting sites within pre-mRNAs, it is likely possible to utilize purely the sequence data of pre-mRNA. This is due to the nature of its target selection and the mechanics involved in binding. As small molecules are generally designed to directly recognize and stabilize RNA duplex regions. As such, whether it directly intercalating within the helix or through the presence of mismatches,[10] like that of NCD, their binding generally is not affected by sequence features beyond the few nucleotides surrounding its binding site. Hence allowing us to potentially filter for candidate binding sites purely by searching for the presence of limited sequence features and disregarding any other surrounding sequence features that may be present.

Here, I will attempt to provide a general overview of how it can be potentially achieved. However, it must be understood that underneath this provided proposal belies several core assumptions that may not be appropriate for all situations. The assumptions are: 1. Target pre-mRNA structure is highly dynamic, 2. Pre-mRNA is considered a naked RNA. 3. Distance between the two interacting sequences is minimized.

The first two assumptions are required to provide sufficient freedom for the desired sequences to interact and form the target sites for the small molecules and allow for conformational selection towards the final desired structure. The last assumption limits the potential candidates to a few potential pairs to simplify the process of analysis.



Scheme FP 1. Schematic representation of the processing involved in obtaining candidate binding sequences for small molecules in pre-mRNAs. Here, the NCD target site was used as an example.

The general flow of the analysis method is shown in Scheme FP 1. Briefly, all potential pairs of target sites for the target molecule are generated. The coordinates corresponding to each of the pair sequences are then searched and extracted from the full-length pre-mRNA sequence (extracted from the genomic sequences through the AGAT package[11]). The distance between all potential pairs is computed, and only the minimal distance pairs are retained for further analysis. The raw sequence of the minimal distance pairs is then extracted from the pre-mRNA, in addition to ~10 nt of flanking sequences on each side, this is to allow for unique alignment back to the genomic position later. The extracted sequence pairs are then joined with a simple known linker sequence, and the single sequence is then stored as a chimeric "read" sequence. This is then aligned back to the genomic sequence using the DG process reported previously,[12]

and the data can then be visualized as a genomic feature in genome viewers (e.g. IGV[13] etc.) for further analysis.

The above method, however, may not be applicable for CLIP-ON designs as oligonucleotides target non-hybridized sequences, and purely screening for potentially interacting short-sequence pairs is no longer sufficient, and requires detailed interactome analysis of surrounding sequences. Although not impossible computationally, it may be simpler and more efficient to develop methodologies to experimentally elucidate pre-mRNA structures and interactions, and design CLIP-ON accordingly. Although methods such as RIC-seq[14] and PARIS[12] exist, elucidating intronic interactions seems to be of great difficulty, likely due to their low abundance in the cellular environment.

Overall, I believe that the model cases demonstrated in this thesis are just the beginning and its applicability may expand extensively in the future in conjunction with new and more precise computational and experimental methods. The concepts provided here hopefully built a potential foundation and new perspective for both chemists and biologists when studying circRNAs in the future.

References:

1 K. NAKATANI, *Proc. Japan Acad. Ser. B*, 2022, **98**, PJA9801B-03.

2 S. M. Meyer, C. C. Williams, Y. Akahori, T. Tanaka, H. Aikawa, Y. Tong, J. L. Childs-Disney and M. D. Disney, *Chem. Soc. Rev.*, 2020, **49**, 7167–7199.

3 P. C. Bevilacqua, L. E. Ritchey, Z. Su and S. M. Assmann, *Annu. Rev. Genet.*, 2016, **50**, 235–266.

4 Y. Jin, Y. Yang and P. Zhang, *RNA Biol.*, 2011, **8**, 450–457.

5 W. R. Jeck, J. A. Sorrentino, K. Wang, M. K. Slevin, C. E. Burd, J. Liu, W. F. Marzluff and N. E. Sharpless, *RNA*, 2013, **19**, 141–157.

6 L. V. W. Stagsted, K. M. Nielsen, I. Daugaard and T. B. Hansen, *Life Sci. Alliance*, 2019, **2**, e201900398.

7 C. Zhang, S. Wang, F. Chao, G. Jia, X. Ye, D. Han, Z. Wei, J. Liu, G. Xu and G. Chen, *Mol. Ther.*, 2023, **31**, 1705–1721.

8 S. J. Conn, K. A. Pillman, J. Toubia, V. M. Conn, M. Salmanidis, C. A. Phillips, S. Roslan, A. W. Schreiber, P. A. Gregory and G. J. Goodall, *Cell*, 2015, **160**, 1125–1134.

9 S. Kalmykova, M. Kalinina, S. Denisov, A. Mironov, D. Skvortsov, R. Guigó and D. Pervouchine, *Nat. Commun.*, 2021, **12**, 1–17.

10 J. R. Thomas and P. J. Hergenrother, *Chem. Rev.*, 2008, **108**, 1171–1224.

11 Dainat J. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format. (Version v0.8.0). Zenodo. https://www.doi.org/10.5281/zenodo.355271

12 Z. Lu, Q. C. Zhang, B. Lee, R. A. Flynn, M. A. Smith, J. T. Robinson, C. Davidovich, A. R. Gooding, K. J. Goodrich, J. S. Mattick, J. P. Mesirov, T. R. Cech and H. Y. Chang, *Cell*, 2016, **165**, 1267–1279.

13 J. T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz and J. P. Mesirov, *Nat. Biotechnol.*, 2011, **29**, 24–26.

14 Z. Cai, C. Cao, L. Ji, R. Ye, D. Wang, C. Xia, S. Wang, Z. Du, N. Hu, X. Yu, J. Chen, L. Wang, X. Yang, S. He and Y. Xue, *Nature*, 2020, **582**, 432–437.

List of Publications

Chapter 1:

Mismatch binding ligand upregulated back-splicing reaction producing circular RNA in a cellular model

Ni, L., Yamada, T., Murata, A., and Nakatani, K.

Chem. Commun. **2022**, *58*, 3629–3632.

Chapter 2:

Utility of oligonucleotide in upregulating circular RNA production in a cellular model

Ni, L., Yamada, T., and Nakatani, K.

Submitted to *Sci. Rep.*, and is currently available as pre-print (https://doi.org/10.21203/rs.3.rs-3849465/v1)

Patents

「一本鎖核酸及びその用途」

基礎出願番号：特願 2023-014938　国際出願番号：PCT/JP2024/003377

発明者：山田剛史、倪露、中谷和彦

List of Presentations

1. Regulation of circular RNA biogenesis via nucleic acid binding small molecule
   Ni, L., Yamada, T., Murata, A., and Nakatani, K.
   101st CSJ Annual Meeting
   March 2021, JAPAN (Online oral presentation)
2. Regulation of circular RNA biogenesis via nucleic acid binding small molecule in cells
   Ni, L., Yamada, T., Murata, A., and Nakatani, K.
   48th International Symposium on Nucleic Acid Chemistry (ISNAC2021)
   November 2021, JAPAN (Online poster presentation)
3. Regulation of circular RNA biogenesis using nucleic acid binding small molecule in cellular environment

Ni, L., Yamada, T., Murata, A., and Nakatani, K.

International Chemical Congress of Pacific Basin Societies (Pacifichem)

December 2021, Honolulu, Hawaii, USA (Online, Poster Presentation)

4. Regulating circular RNA production in cells using 3-way junction forming antisense oligonucleotide

   Ni, L., Yamada, T., Murata, A., and Nakatani, K.

   XXIV International Round Table on Nucleosides, Nucleotides, and Nucleic Acids (IRT2022)

   August 2022, Karolinska Institute, Stockholm, SWEDEN (Poster Presentation)

5. Regulating circular RNA production in cellular model using bridging antisense oligonucleotide

   Ni, L., Yamada, T., and Nakatani, K.

   49th International Symposium on Nucleic Acid Chemistry (ISNAC2022)

   November 2022, Tokyo University of Science, Tokyo, JAPAN (Poster Presentation)

6. Exploring methods in regulating circular RNA production in cellular model

   Ni, L., Yamada, T., and Nakatani, K.

   9th Special Lectures from Outstanding Young Scientists

   March 2023, FIBER, Konan University, Kobe, JAPAN

   (Oral Presentation, **Invited Lecture**)

7. Exploring new methods in regulating circular RNA production using cellular model

   Ni, L., Yamada, T., Murata, A., and Nakatani, K.

   103rd CSJ Annual Meeting

   March 2023, Tokyo University of Science, JAPAN (Oral presentation)

Awards and Fellowship

1. Research Fellow of the Japan Society for the Promotion of Science (JSPS DC2, April 2022 – March 2024)
2. Outstanding Poster Award (49th International Symposium on Nucleic Acid Chemistry)
3. Student Presentation Award (103rd CSJ Annual Meeting)

Acknowledgment

I am forever indebted to Prof. Kazuhiko Nakatani for supporting me to freely pursue the research of my interest for the past 6 years of my time in the laboratory. Throughout my time in the laboratory, Prof. Nakatani has offered me invaluable encouragement and fruitful discussions, even at the most trying times when nothing seems to go according to plan, without which I do not believe I would have been able to complete this thesis in a timely manner. Furthermore, I am grateful to Prof. Nakatani for holding my ideas to the highest standards of scrutiny, teaching me to look at things from a different perspective, and reminding me to constantly think of the social impact of my research. It has been a great honor and privilege to have had the chance to be under the guidance of Prof. Nakatani.

I am very grateful for Assoc. Prof. Chikara Dohno for providing fruitful discussions and ideas when I initially started my project and during the seminars. Furthermore, I must give my special thanks to Assoc. Prof. Asako Murata for teaching me various techniques used in molecular biology and providing extensive support in setting up the necessary environment for me to pursue my endeavor in molecular biology. I offer my utmost gratitude to Asst. Prof. (now Lecturer) Takeshi Yamada for guiding me through the roughest initial years of my research in the laboratory, collaborating very closely with me on the projects, and putting aside time during holidays to help me correct my various writings. Many things, including obtaining the DC2 fellowship, may not have been possible without Lecturer Yamada's tireless assistance. I specially thank Assis. Prof. Tomonori Shibata for building the foundation for my research in this thesis, and engaging with me in fruitful discussions. I am grateful to specially appointed Assis. Prof. Bimolendu Das for always engaging in fruitful and elaborate discussions on research and career paths.

I would like to give special thanks to Dr. Eitaro Murakami for his kind assistance and fruitful discussions throughout the entirety of my 6 years of stay in the laboratory; I have learned many things in chemistry thanks to you. I would also like to give special thanks Dr. Yusuke Takashima for engaging with me in fruitful discussions on RNA biology and bioinformatics, and giving me important insights into navigating the doctorate course. I thank Dr. Yusuke Fujiwara for his cheerful demeaner and sharing his experiences in cellular experiments with me.

I am very thankful for my peers Dr. Anisa Ul'Husna, Dr. Qingwen Chen and Dr. Syuhei Sakurabayashi for offering their helpful opinions regarding my research and providing information on nucleic acid biology and bio/chemo-informatic which I may have overlooked. I am also grateful to all other past and present members of Prof. Nakatani's Laboratory for their

helpful discussions, friendly support, and for creating a receptive atmosphere for conducting research.

I would like to thank the Japan Society for the Promotion of Science (JSPS) for financial support (Research Fellowship for Young Scientist).

Finally, I wish to express my utmost gratitude to my mother Dr. Youmei Lu for bringing me around the globe to different countries and into the world of laboratory science; my father Dr. Henmei Ni for initially bringing the family out of China to Japan, providing me with the foothold to grow further internationally, and giving me the first spark to ignite my interest in science. Both of them have fully supported my wish to pursue a scientific degree and provided invaluable encouragement and support throughout my studies; without them, I would not be the person I am today. I also wish to express my utmost gratitude to my now-wife Mrs. Amihan Leana Losanta Anajao, who has provided me with tireless encouragement and support to push further during my most trying times, engaging with me in delightful discussions regarding my research, and always being safe harbor for my soul.