The Distance Between Exons and *Alu* Elements Influences RNA Circularization Efficiency

By

Jiacheng "Coco" Liu

Senior Honors Thesis

Marzluff Lab, Department of Biology

University of North Carolina at Chapel Hill

April 16th, 2019

Approved by:

Dr. William Marzluff, Thesis Advisor

Dr. Terry Furey, Reader

Dr. Scott Hammond, Reader

Abstract

Circular RNA (circRNA) is a category of RNA that is created when the spliceosome back-splices an exon, thereby forming an RNA covalent circle. A few circRNAs have been shown to have regulatory functions, but the functions of most circRNAs are not known. Previous studies have demonstrated that repetitive elements flanking the exon(s), such as *Alu* elements, facilitate circularization, and have identified the minimal size of repetitive elements needed to drive circularization. We studied how the distance between exon splice donors/acceptors and *Alu* elements affects the efficiency of RNA circularization. To create the distance gradient, we inserted and/or deleted sequences between the splice donors/acceptor and *Alu* elements. We engineered the circular RNA so that it would express GFP after circularization. To measure the circularization efficiency, we conducted Western blots and Northern blots on the proteins and RNA harvested from cells. We showed that in HEK293 cells the distance of the repetitive element upstream of the exon has a large effect on circularization, while the distance downstream has little effect. Combining these observations, we created a minimal construct that can be circularized efficiently and expressed much more protein than our original construct. Overall, our study further contributed to the understanding of the *cis* elements that affect circular RNA formation *in vivo,* and design of vectors to efficiently express proteins from very stable RNAs.

Introduction

Ribonucleic acid (RNA) is a single stranded polymer of ribonucleotides. As described in the central dogma, messenger RNA (mRNA) is transcribed from DNA templates, then subsequently processed and translated into proteins. In eukaryotes, after pre-mRNA is transcribed, sequences called introns are spliced out by spliceosomes made up of snRNA-protein complexes (snRNPs). The spliceosomes are assembled on the exons, facilitated by a polypyrimidine tract, a pyrimidine rich sequence located upstream of the exon (Coolidge, Seely, and Patton 1997; Reed 1989), and a splice donor sequence downstream of the exons. First, a branch site in the intron attacks the 5' splice site (splice donor sequence) to form a lariat intermediate; in the second step, the splice donor attacks the 3' splice site (splice acceptor) to ligate two exons together (Black 2003). Most of the time, the exons are spliced head-to-tail to form a linear mRNA, along with a lariat intron removal.

In 2012, Will Jeck, a M.D.-Ph.D. student at UNC, published the finding that there are a large number of circular RNAs present in mammalian cells (Jeck et al. 2012). These arise because certain canonical splicing sites of pre-mRNAs are back-spliced into circRNAs; a splice donor at the 3' end of the exon, is joined to the splice acceptor at the 5' end of the same exon, or a more 5' exon (Jeck et al. 2012; Vicens and Westhof 2014). This can occur when the reverse complementary matches (RCM) present in the intronic regions on the two ends of an exon base pair with each other and form a stem loop (Jeck et al. 2012; Ivanov et al. 2014). The stem loop can therefore bring splice donors and acceptors physically closer to allow splicing.

Although some circRNAs may arise as a result of errors in pre-mRNAs splicing, and are present in very low amounts relative to their linear counterpart (Chen 2016), the amount of circular RNA

in the cell is significant, as 15% of actively transcribed genes have a circRNA product in fibroblasts (Jeck et al. 2012; Liang and Wilusz 2014). The ratio of circRNA to linear RNA varies enormously between genes, with some genes producing more circular RNA than linear RNA (Jeck et al. 2012). While the function, if any, of most circRNAs are not known, there are a few studies that have shown that circRNAs function as regulatory molecules. CircRNA can serve as miRNA sponges, presenting miRNA target sites to sequester miRNAs, such as for miR-7 (Hansen et al. 2013). CircRNAs can also serve as protein sponges and hence can act as regulators of pre-mRNA linear splicing (Ashwal-Fluss et al. 2014). Apart from the endogenous functions discovered in cells, circRNA has also been studied as a potential vector for gene therapy delivered by adeno-associated virus (AAV). CircRNAs are much less susceptible to exonucleases compared to their linear counterparts, due to the absence of free 3' ends, and therefore generally have very long half-lives, surviving at least twice as long (Enuka et al. 2016; Meganck et al. 2018).

Because of their extraordinary stability and potential important functions in gene expression regulation, the study of the factors affecting circularization is significant. Previous studies have shown that the *Alu* family elements of human short interspersed nuclear elements (SINE) can facilitate RNA circularization (Jeck and Sharpless 2014). The *Alu* elements present in the intronic region of a pre-mRNA can base-pair and form a stem-loop structure. Liang (Liang and Wilusz 2014) further identified that only a portion of the *Alu* element is required for the stem-loop and circRNA formation. In our study, we further asked how the distances between the splice sites and *Alu* elements influences RNA circularization. Our first gene model was the human homeodomain-interacting protein kinase 3 gene (HIPK3), whose circRNA has been shown to be highly abundant (Jeck et al. 2012), and for which a minimal intron pair has been described (Zheng et al. 2016;

Liang and Wilusz 2014). A circRNA reporter construct with an internal ribosomal entry site (IRES) was used that allowed us to visualize circularization with GFP expression (Meganck et al. 2018; Wang and Wang 2015). The distance between exon and *Alu* elements was modified by inserting randomly generated sequences of different lengths or deleting the original sequences that were not key factors for circularization. The circRNA levels were measured directly by northern blot and by GFP expression, in comparison to the original construct. Our study contributes to knowledge of what intronic elements are required to drive circularization, and what properties regulate this process.

Results

Circular RNA Reporter Design

To distinguish circular RNA from its linear counterpart, a reporter construct with a split GFP was modified from Meganck et al. (Meganck et al. 2018). Figure 1 demonstrates the schematics of the construct. Cytomegalovirus (CMV) was the promoter for the sequence, followed by the upstream flanking intron that contains the *Alu* sequence of 35 nucleotides. The green fluorescent protein (GFP) sequence was split into two parts. The second half (marked "FP") was inserted after the upstream intron, whereas the first half (marked "G") was inserted after the poliovirus internal ribosome entry site (IRES). An IRES site served as a ribosome attachment and translation initiation site for the GFP protein when a 5' cap was not present. Downstream flanking intron sequences were inserted after the "G" half of GFP and before the terminator sequence. An *Alu* family sequence of 91 nucleotides was also inserted in the downstream intron. The SV40 Poly-A sequence allowed transcription termination and polyadenylation formation of the premature RNA. If the

mRNA was circularized, the split GFP gene can be connected and expressed; if mRNA remained in the linear form, then no green fluorescence was expressed.

Deleting the Alu sequences on either side of the intron drastically decreases RNA circularization and circRNA expression

In order to confirm the findings by Liang et. al (Liang and Wilusz 2014) in terms of the necessity of *Alu* elements, we deleted the *Alu* sequence on the upstream (UPnoAlu) and downstream (DWnoAlu) exon, respectively. Each construct, along with the control, were transfected into HEK293 cells, and were taken for an assessment four days after. Minimal GFP expression could be assessed with fluorescent images for both deletions, and the Western and Northern blot quantifications showed substantially decreased circularization (figure 2).

Increasing upstream distance of HIPK3 decreases circularization efficiency drastically in HEK293 cells

We inserted 100 nt (UP100), 500 nt (UP500), 1000 nt (UP1000), and 1500 nt (UP1500) of random sequences at a position 100 nucleotides away upstream of the HIPK3 engineered exon, transfected each construct into the HEK293 cells and took GFP microscopic images of each construct (figure 3B). Surprisingly, it showed a drastic decrease in GFP expression; inserting only 100 nt already greatly decreased GFP expression, and the three larger inserts had minimal GFP expression. Western blot GFP quantification of all inserts showed very significant decrease compared to the control constructs (figure 3C-D). Similarly, Northern blot probed against IRES-GFP showed significant decrease for the UP100 insert and very significant decrease for the latter three. In order to consider our inserts' possible interference with the polypyrimidine tract, which is crucial for

spliceosome assembly, we made a new group of inserts much farther away from the exon. We inserted 100 nt (UP100F), 500 nt (UP500F), 1000 nt (UP1000F), and 1500 nt (UP1500F) into a site 325 nucleotides away from the exon. Both Western and Northern blot quantifications complemented the GFP microscopic images to show significant decrease when distances increased. The similarity of results between the two groups of constructs excluded the possible interference of the polypyrimidine tract, further illustrating that the decrease of circularization efficiency with increasing distance on the upstream side was not affected by the location of insertion, and that circRNA formation and expression efficiency were very sensitive to upstream distance increase.

Increasing downstream distance of HIPK3 did not significantly influence RNA circularization efficiency in HEK293 cells

We proceeded to study the influence of increased exon-*Alu* on the downstream side of the HIPK3 exon. Similarly, we inserted a random sequence of 100 nt (DW100), 500 nt (DW500), 1000 nt (DW1000), 1500 nt (DW1500) into a site 515 base pairs away from the end of the exon. Qualitatively, significant decrease of GFP expression was not observed from the microscopic GFP images (figure 5B). The Western and Northern blot quantification confirmed that there was no significant change of GFP expression or circRNA across the distance gradient. Comparing the downstream insertions to upstream insertion, we saw very different influence by distance gradient.

Shortening the Exon-Alu distance upstream and downstream individually had different outcomes

Given the insights from above results, we next asked if decreasing the distance between exon and *Alu* elements would give different results. We started with deleting the entire sequence between

the polypyrimidine tract and the *Alu* sequence upstream and between exon and the *Alu* sequence downstream of HIPK3 exon, respectively. For the upstream intron, we deleted 220 nucleotides (UPdel), and for downstream 652 nucleotides (DWdel) (figure 6A-B). We transfected the constructs into HEK293 cells, and took fluorescent images four days after transfection. While the fluorescent images of UPdel did not show significant change, those of DWdel showed a drastic decrease. We then proceeded to create a new construct that deleted 500 nucleotides downstream from the end of the exon (DWdpartial), whose fluorescent images showed in general, a similar amount of GFP expression. We then compared the Western and Northern blot quantifications for our control, UPdel, DWdel, and DWdpartial (figure 6E-H). The Western blot for GFP expression and Northern blot probing for circRNA showed an increase of circularization for upstream deletion (UPdel), but very significant decrease for downstream full deletion (DWdel), as expected. There was with no significant change of GFP expression or circular RNA amount for the downstream partial deletion construct (DWdpartial).

In light of the different circularization and expression levels assessed from the upstream deletions and downstream partial deletions, we proceeded to ask whether the combination of upstream and downstream deletions can lead to significantly higher circle expression. We constructed the double deletion construct, deleting 220 nucleotides from upstream of the exon and 500 nucleotides from the downstream of the exon (same deletions as UPdel and DWdpartial), named UP+DWdel. We transfected the control, UPdel, DWdpartial, and UP+DWdel in HEK293 cells, and took fluorescent images four days after. The GFP images of UP+DWdel saw much higher expression compared to the control group. The Western blotting against GFP and Northern blotting against circRNA

showed an increase of GFP expression and circular RNA amount compared to the control group (figure 7C-F), and the increase was also significant compared to downstream partial deletions only.

Engineered Laccase 2 gene showed similar but less significant results

Following, we next asked if the trend we found with HIPK3 gene can be generalized to other genes, so we expanded our query reporter to the Laccase 2 gene from *Drosophila*. Instead of *Alu* elements that belong to human SINE element, the Laccase 2 intron has flanking sequences that acts similarly to *Alu* elements. We tested the two extreme inserts in the laccase 2 gene. Since we found no influence on the maximum (1500 nt) insert downstream of the exon in the HIPK3 gene, we inserted 1500 nt on the downstream of the laccase 2 gene (LAC2DW). However, because we observed a sharp decrease of GFP expression and circularization with only 100 nt inserts on the upstream of the HIPK3 exon, we inserted 100 nt on the upstream of the laccase 2 gene (LAC2UP). We asked if these insertions have the same effects as those in the HIPK3 gene. Similar to HIPK3 constructs, we transfected control, LAC2UP, and LAC2DW into HEK293 cells, and harvested cells four days later. We found a decrease in GFP expression and circular RNA amount in upstream 100 nt insert (LAC2UP), although the effect was not as large as that of HIPK3. We did not find a significant difference in GFP expression or circularization efficiency with 1500 nt downstream insertion. (figure 8D-E).

Discussion

The study of circular RNA has gained more attention due to multiple newly discovered potential functions and future applications. In our study using engineered HIPK3 as circular RNA reporters and HEK293 as tissue culture, we observed that modifying the distance between exonic region and

Alu elements has different consequences. The upstream *Alu* was exceptionally sensitive to an increased distance to the exon, whereas the downstream *Alu* was resistant to increasing distance. Shortening the *Alu*-exon distance on either side can increase RNA circularization efficiency, and we were able to engineer a double deletion construct which has the highest circularization and circRNA expression efficiency.

Our study further confirmed the requirement of flanking repeats (in the case of humans, *Alu* family sequences) for facilitating RNA circularization (Liang and Wilusz 2014), and our results fit the model proposed by Barrett et al. (Barrett, Wang, and Salzman 2015), that circRNAs are generated by exons containing lariat, and Zhang et al. (Zhang et al. 2014) that base-pairing between flanking introns on each side of exon brings splice donors and acceptors physically closer to each other. Furthermore, our study provided a basis for using circRNA as a therapeutic platform proposed by Meganck et al. (Meganck et al. 2018). We showed the possibility to increase packaging efficiency by shortening intronic sequences on both sides to be packaged in the AAV genome, while at the same time reached higher expression of circRNA.

The generalizability of the trends we found is still in need of further study. Although Laccase 2 engineered constructs showed generally similar results, the effect size of the results was much smaller. Due to the differences observed between HIPK3 and Laccase 2, we are bringing another gene that was identified to have relatively high circularization rate, the human ZKSCAN1 (Zinc Finger With KRAB And SCAN Domains 1) gene to our study. We hope to transfer the same modulation from the first two genes to ZKSCAN1 in order to study the generalizability of the

pattern we discovered. We are also going to use the U87 cell line as another model for this study, in order to test whether the circularization with exon-*Alu* distance is cell type dependent.

In light of the results we observed from modulating intronic distance between exon and *Alu* sequences, in the future, we hope to further look at other factors that can be modulated to influence circularization efficiency. We will be looking at the exonic sequences can influence circularization and expression efficiency. To do this, we will modulate sizes of exons by inserting sequences in to making bigger circRNA, and test whether increasing the sizes will have an effect on circularization efficiency.

Acknowledgement

I want to thank the Marzluff and Asokan Lab for their support on this project. I thank Dr. Marzluff for his guidance in improving my research methods and training me to become an entry-level scientist. I especially thank Rita Meganck from the Marzluff lab for her close mentorship, guidance on experimental techniques, and scientific thought process. I thank Dr. Marzluff and Rita Meganck for their instructions and help on editing this thesis. I thank Noah Legall for providing the random sequences for the inserts. This project is supported by UNC Summer Undergraduate Research Fellowship (2018) and NIH grant R01NS099371.

Methods

Cloning the construct

Two random sequences of 1500 nucleotides with no secondary structure were generated by Noah Legall for upstream and downstream inserts, respectively (Table 1). Primer pairs were designed to

generate 100 nucleotides (nt), 500 nt, 1000 nt, and 1500 nt sequence inserts from the random sequence, and restriction enzyme sites overhang were added（Table 2). Two genes which produce circular RNAs, Homeodomain-interacting protein kinase 3 (HIPK3) from human and Laccase 2 (LAC2) from *Drosophila*, were used as the backbones of the experiments. Both the genes and inserts were digested with restriction enzymes (Table 2) and ligated with T4 ligase (M0202L, New England Biolabs, Ipswich, MA) to generate the constructs with a distance gradient. Site-directed mutagenesis was used in order to delete sequences between *Alu* elements and exons (Table 2). The constructed HIPK3 gene was then cloned into a pcDNA3.1 plasmid where the plasmids were transformed into DH10B competent cells for mass production. The plasmid DNAs were sequenced after extracted them from the cells using a miniprep kit (27106, QIAprep Spin Miniprep Kit, Hilden, Germany).

In vivo experiment: tissue culture

HEK293 was the cell line chosen for this study. Cells were cultured in DMEM (1X, Gibco, Fisher Scientific, Hampton, NH) supplemented with 10% Fetal Bovine Serum (HyClone, Fisher Scientific, Hampton, NH) and 1% Pen-Strep. The sequenced plasmid containing different constructs were transfected into HEK293 cells in a 6-well plate (2.5 micrograms of total DNA in 12.5 microliters, 12.5 microliters of 1% PEI-MAX 40K, diluted to 125 microliters with serum free DMEM). Equimolar amounts of the test DNAs ere used, adjusted the amount of DNA with empty vector where necessary. Green fluorescent images of the cells were taken (EVOS FL Imaging System, AMF4300, Thermo Fisher Scientific, Waltham, MA) four days after transfection. The cells were harvested with 1 ml PBS, divided in half and recovered by centrifugation, extracted with 100 microliters of Trizol reagent for RNA (15596026, Thermo Fisher Scientific, Waltham,

MA) or and proteins were harvested with Passive Lysis Buffer (5X, E194A. Promega, Madison, WI).

Quantification of circularized RNA

Northern blotting

RNA extracted from cells (5000 nanograms) was resuspended in denaturing buffer (67% deionized formamide, 6.7% formaldehyde, 1×3-morpholinopropane-1-sulfonic acid [MOPS] running buffer (Meganck et al. 2018)) and Northern dye (10X). RNA electrophoresis was run on a 1.15% denaturing agarose gel in 1X MOPS buffer (35V, 22 hours), and was then transferred to a positively charged nylon transfer membrane (RPN303B, GE Healthcare, Chicago, IL) in 10X SSC buffer (overnight). After prehybridization, radiolabeled ^{32}P probe complementary to the reporter (GFP sequence) was generated with Prime-It II Random Primer Labeling Kit (Agilent Technologies, Santa Clara, CA) and hybridized with the RNA in Rapid-hyb buffer (RPN1636, GE Healthcare, Chicago, IL). The blots were washed with wash buffers (2X SSC, 0.1% SDS; 0.1X SSC, 0.1% SDS) and was exposed to film. The radioactivity was quantified with Phospho-imager.

Western blotting

Cells were harvested in Passive Lysis Buffer to collect the proteins translated in the cell. Equal volumes of protein samples were prepared with 2X SDS loading dye (100 mM Tris-Cl (pH 6.8), 4% (w/v) SDS, 0.2% (w/v) bromophenol blue, 20% (v/v) glycerol, and 200 mM DTT (dithiothreitol) ("SDS Gel-Loading Buffer (2X)" 2006)), and were separated on a 10% Tris-glycine SDS-Page gel in 1X SDS buffer (120V, 50min). The proteins were transferred to a nitrocellulose membrane (1620115, Bio-Rad Laboratories, Hercules, CA) (100V, 80min) and

blocked with 5% milk (M0841, LabScientific, Highlands, NJ). GFP was visualized using an anti-GFP antibody (1:1000) (sc-9996, Santa Cruz Biotechnology, Dallas, TX) and the loading control actin was visualized using an anti-actin antibody (1:10000) (GT5512, GeneTex, Irvine, CA). Sheep anti-mouse HRP secondary antibody (1:3000) (NA931V, GE Healthcare, Chicago, IL) was used with the Femto substrate (PI34094, Thermo Scientific, Waltham, MA) to visualize the proteins, and exposed with an ImageQuant camera (General Electric, Boston, MA).

Statistical analysis

GFP expression levels from Western blots were quantified by ImageJ, and circularization levels from Northern blots were quantified by Typhoon software. Actual expression/circularization levels were normalized to the control group to produce relative GFP expression/circularization efficiency. Unpaired two-tailed Student's *t* tests were conducted between each experimental group and control group to test for significant difference.

Reference

Ashwal-Fluss, Reut, Markus Meyer, Nagarjuna Reddy Pamudurti, Andranik Ivanov, Osnat Bartok, Mor Hanan, Naveh Evantal, Sebastian Memczak, Nikolaus Rajewsky, and Sebastian Kadener. 2014. "CircRNA Biogenesis Competes with Pre-MRNA Splicing." *Molecular Cell* 56 (1): 55–66. https://doi.org/10.1016/j.molcel.2014.08.019.

Barrett, Steven P, Peter L Wang, and Julia Salzman. 2015. "Circular RNA Biogenesis Can Proceed through an Exon-Containing Lariat Precursor." *ELife* 4: 1–18. https://doi.org/10.7554/elife.07540.

Black, Douglas L. 2003. "Mechanisms of Alternative Pre-Messenger RNA Splicing." *Annual*

Review of Biochemistry 72 (1): 291–336. https://doi.org/10.1146/annurev.biochem.72.121801.161720.

Chen, Ling-ling. 2016. "Ling-Ling Chen -The Biogenesis and Emerging Roles of Circular RNAs" 17 (April): 15–17.

Coolidge, Candace J., Raymond J. Seely, and James G. Patton. 1997. "Functional Analysis of the Polypyrimidine Tract in Pre-MRNA Splicing." *Nucleic Acids Research* 25 (4): 888–95. https://doi.org/10.1093/nar/25.4.888.

Enuka, Yehoshua, Mattia Lauriola, Morris E. Feldman, Aldema Sas-Chen, Igor Ulitsky, and Yosef Yarden. 2016. "Circular RNAs Are Long-Lived and Display Only Minimal Early Alterations in Response to a Growth Factor." *Nucleic Acids Research* 44 (3): 1370–83. https://doi.org/10.1093/nar/gkv1367.

Hansen, Thomas B., Bettina H. Clausen, Bente Finsen, Jørgen Kjems, Jesper B. Bramsen, Trine I. Jensen, and Christian K. Damgaard. 2013. "Natural RNA Circles Function as Efficient MicroRNA Sponges." *Nature* 495 (7441): 384–88. https://doi.org/10.1038/nature11993.

Ivanov, Andranik, Sebastian Memczak, Emanuel Wyler, Francesca Torti, Hagit Porath, and Marta Orejuela. 2014. "Analysis of Intron Sequences Reveals Hallmarks of Circular RNA Biogenesis in Animals." *Cell Reports* 10 (2): 170–77. https://doi.org/10.1016/j.celrep.2014.12.019.

Jeck, William R., and Norman E. Sharpless. 2014. "Detecting and Characterizing Circular RNAs." *Nature Biotechnology* 32 (5): 453–61. https://doi.org/10.1038/nbt.2890.

Jeck, William R., Jessica A. Sorrentino, Kai Wang, Michael K. Slevin, Christin E. Burd, Jinze Liu, William F. Marzluff, and Norman E. Sharpless. 2012. "Circular RNAs Are Abundant, Conserved, and Associated with ALU Repeats." *Rna* 19 (2): 141–57.

https://doi.org/10.1261/rna.035667.112.

Liang, Dongming, and Jeremy E. Wilusz. 2014. “Short Intronic Repeat Sequences Facilitate Circular RNA Production.” *Genes and Development* 28 (20): 2233–47. https://doi.org/10.1101/gad.251926.114.

Meganck, Rita M., Jeremy E. Wilusz, Aravind Asokan, Ruth M. Castellanos Rivera, William F. Marzluff, Miranda L. Scalabrino, and Erin K. Borchardt. 2018. “Tissue-Dependent Expression and Translation of Circular RNAs with Recombinant AAV Vectors In Vivo.” *Molecular Therapy - Nucleic Acids* 13 (December): 89–98. https://doi.org/10.1016/j.omtn.2018.08.008.

Reed, Robin. 1989. “The Organization of 3’ Splice-Site Sequences in Mammalian Introns.” *Genes and Development* 3 (12 B): 2113–23. https://doi.org/10.1101/gad.3.12b.2113.

“SDS Gel-Loading Buffer (2X).” 2006. *Cold Spring Harbor Protocols* 2006 (1): pdb.rec407. https://doi.org/10.1101/pdb.rec407 .

Vicens, Quentin, and Eric Westhof. 2014. “Biogenesis of Circular RNAs.” *Cell* 159 (1): 13–14. https://doi.org/10.1016/j.cell.2014.09.005.

Wang, Yang, and Zefeng Wang. 2015. “Efficient Backsplicing Produces Translatable Circular MRNAs.” *Rna* 21 (2): 172–79. https://doi.org/10.1261/rna.048272.114.

Zhang, Xiao Ou, Hai Bin Wang, Yang Zhang, Xuhua Lu, Ling Ling Chen, and Li Yang. 2014. “Complementary Sequence-Mediated Exon Circularization.” *Cell* 159 (1): 134–47. https://doi.org/10.1016/j.cell.2014.09.001.

Zheng, Qiupeng, Chunyang Bao, Weijie Guo, Shuyi Li, Jie Chen, Bing Chen, Yanting Luo, et al. 2016. “Circular RNA Profiling Reveals an Abundant CircHIPK3 That Regulates Cell Growth by Sponging Multiple MiRNAs.” *Nature Communications* 7: 1–13.

https://doi.org/10.1038/ncomms11215.

Figures

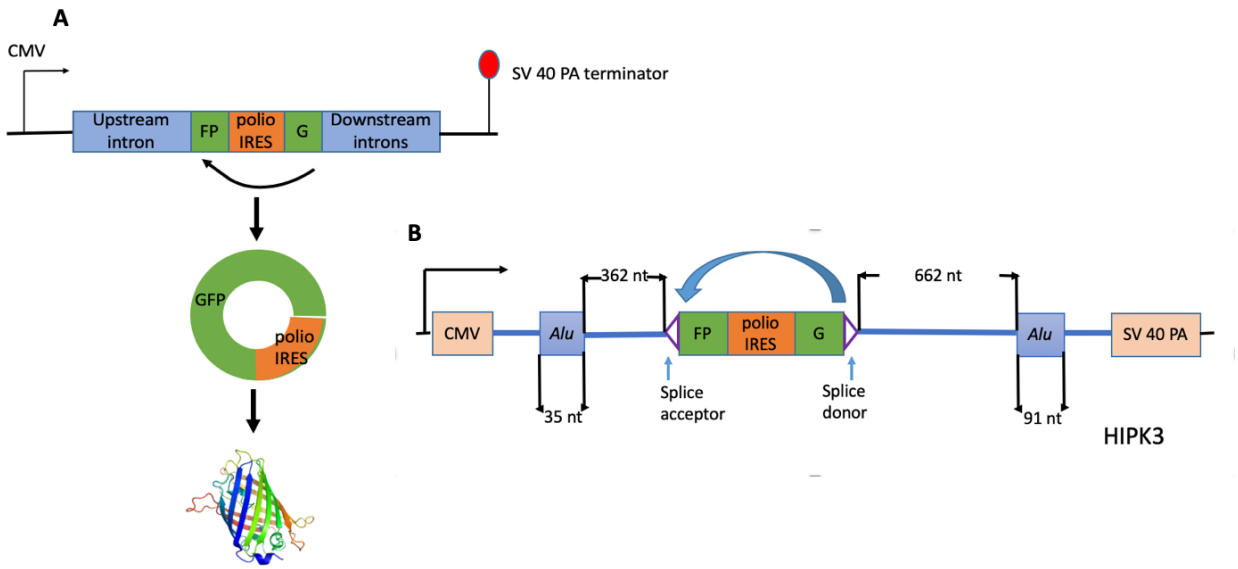

Figure 1. **CircRNA reporter schematics for circRNA.** A). A split GFP sequence was inserted on both ends of the exon. The circularized RNA will have a complete GFP open reading frame, and will translate into green fluorescent protein with IRES. B) A schematics of the HIPK3 control group. A sequence of *Alu* identity (35 nt) is located 362 nt upstream of the exon, and another sequence of *Alu* identify (91 nt) is located 662 nt downstream of the exon.

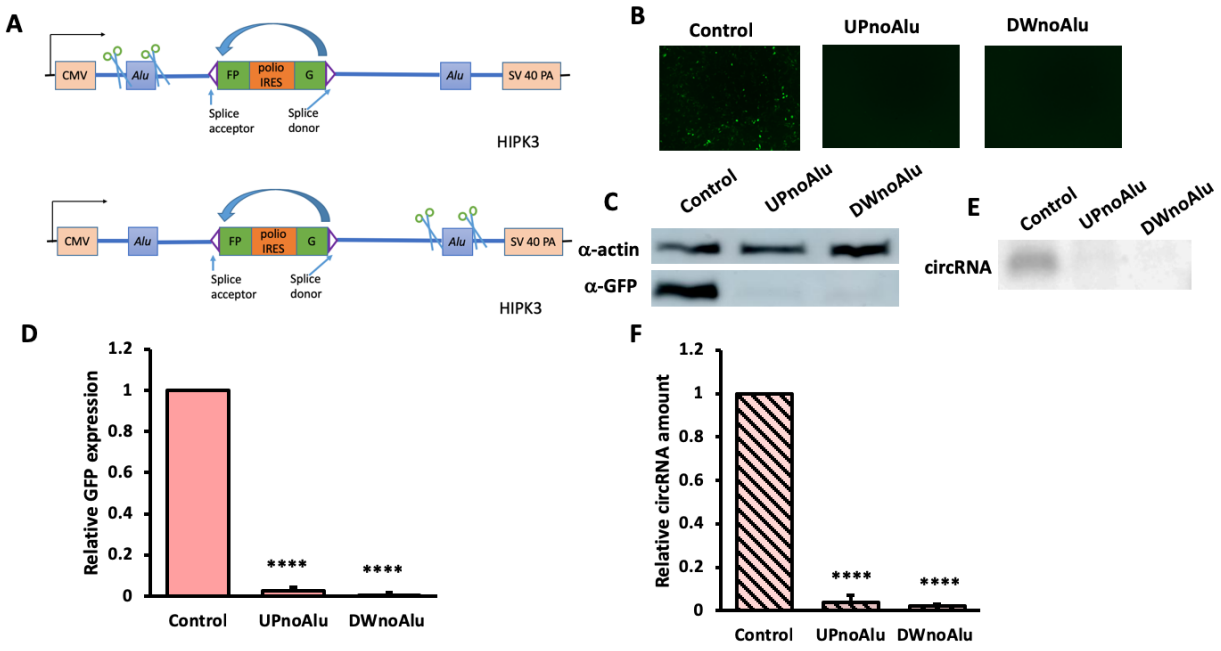

Figure 2. **Deleting *Alu* sequences on each side of exon drastically decreased circularization and GFP expression.** Constructs were transfected into HEK293 cells, and were harvested four days after. A) A schematics for *Alu* sequence deletion; B)fluorescent images of cells; C) Western blot and D) quantifications of GFP of each construct; E) Northern blot and F) quantifications of circularized RNA . (N=3; Error bars indicated standard deviation; significance level using two-tailed Student's t test was represented by [*]: [*] $p<0.05$, [**] $p<0.005$,[***]$p<0.0005$,[****]$p<0.00005$.)

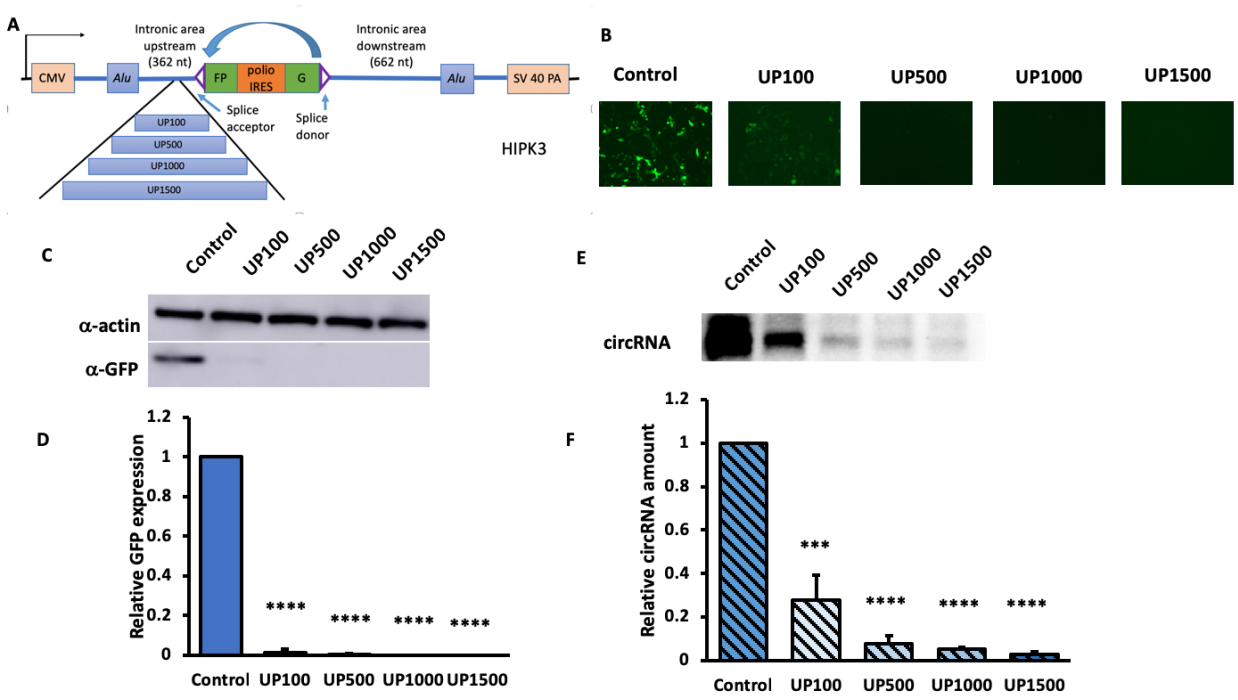

Figure 3. **Increasing upstream *Alu*-exon distance drastically decreased circularization and expression efficiency.** Constructs were transfected into HEK293 cells, and were harvested four days after. A) A schematics for upstream inserts; B)fluorescent images of cells with upstream insertion; C) Western blot and D) quantifications of GFP of each insert; E) Northern blot and F) quantifications of circularized RNA . (N=3; Error bars indicated standard deviation; significance level using two-tailed Student's t test was represented by [*]: [*] $p<0.05$, [**] $p<0.005$,[***]$p<0.0005$,[****]$p<0.00005$.)

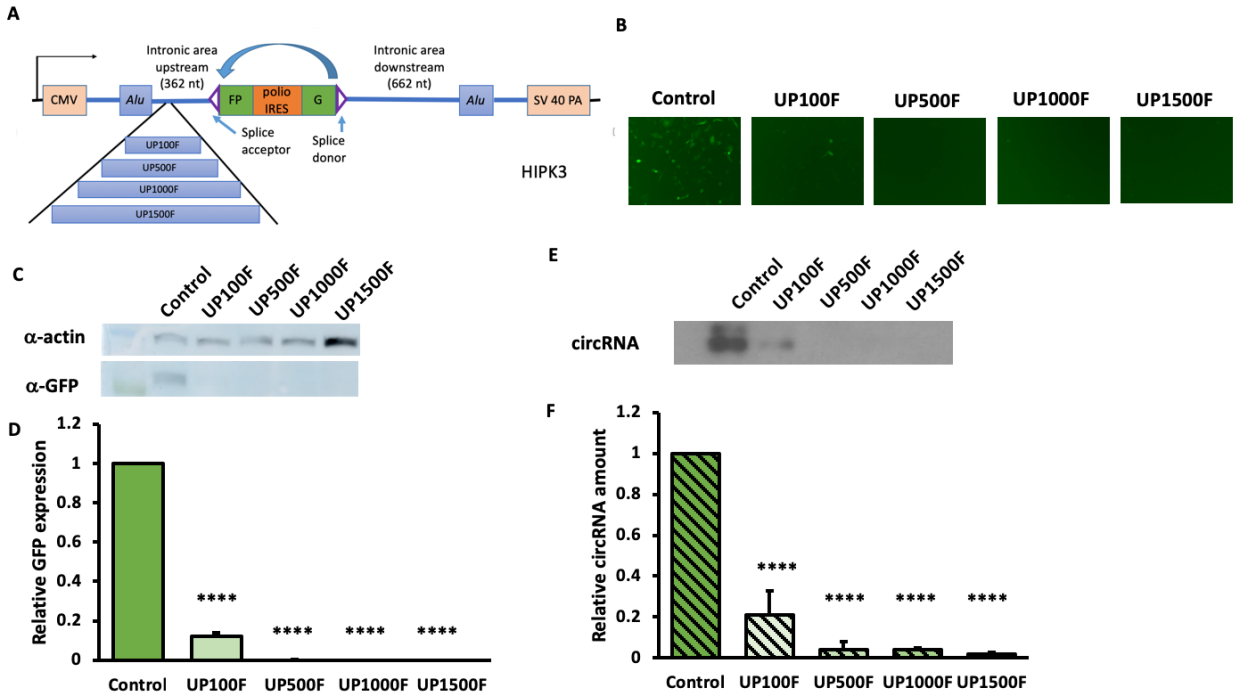

Figure 4. **Increasing upstream *Alu*-exon distance at a farther site showed a similar trend as the more downstream insertion in figure 3.** Constructs were transfected into HEK293 cells, and were harvested four days after. A) A schematics for upstream inserts at a farther site; B)fluorescent images of cells with upstream insertion; C) Western blot and D) quantifications of GFP of each insert; E) Northern blot and F) quantifications of circularized RNA . (N=3; Error bars indicated standard deviation; significance level using two-tailed Student's t test was represented by [*]: [*] p<0.05, [**] p<0.005,[***]p<0.0005,[****]p<0.00005.)

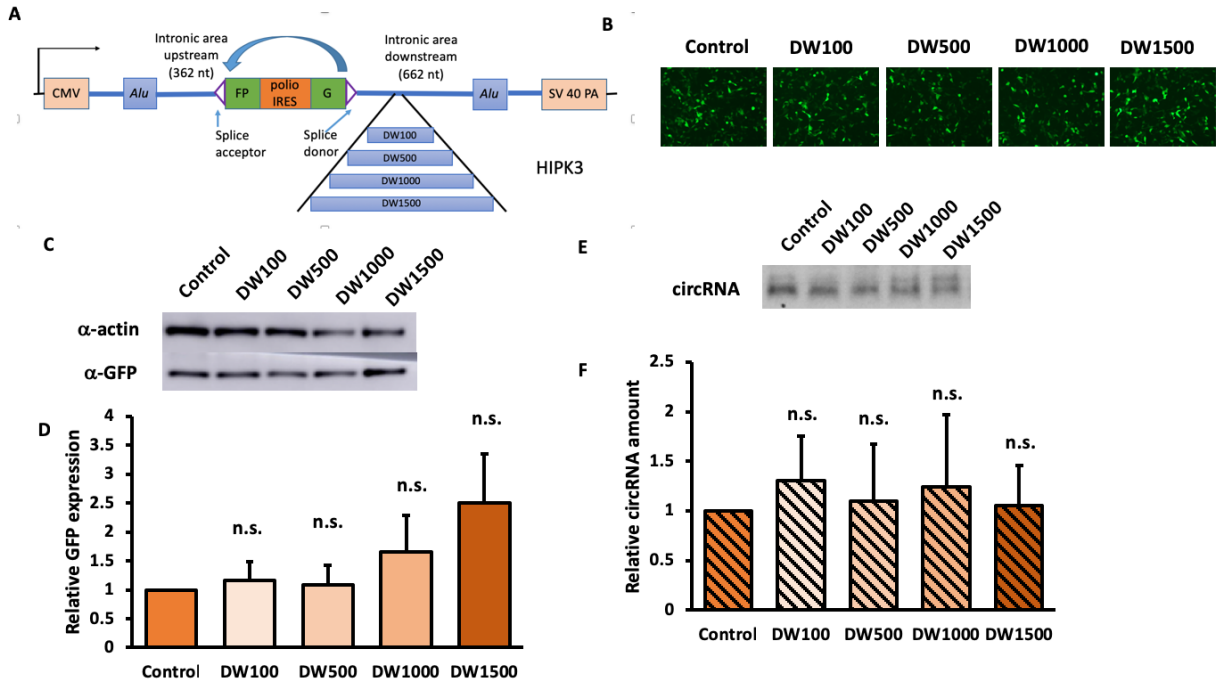

Figure 5. **Increasing downstream *Alu*-exon distance did not significantly affect circularization and expression efficiency.** Constructs were transfected into HEK293 cells, and were harvested four days after. A) A schematics for downstream inserts; B)fluorescent images of cells with downstream insertion; C) Western blot and D) quantifications of GFP of each insert; E) Northern blot and F) quantifications of circularized RNA . (N=3; Error bars indicated standard deviation; significance level using two-tailed Student's t test was represented by [*]:[*] $p<0.05$, [**] $p<0.005$,[***]$p<0.0005$,[****]$p<0.00005$.)

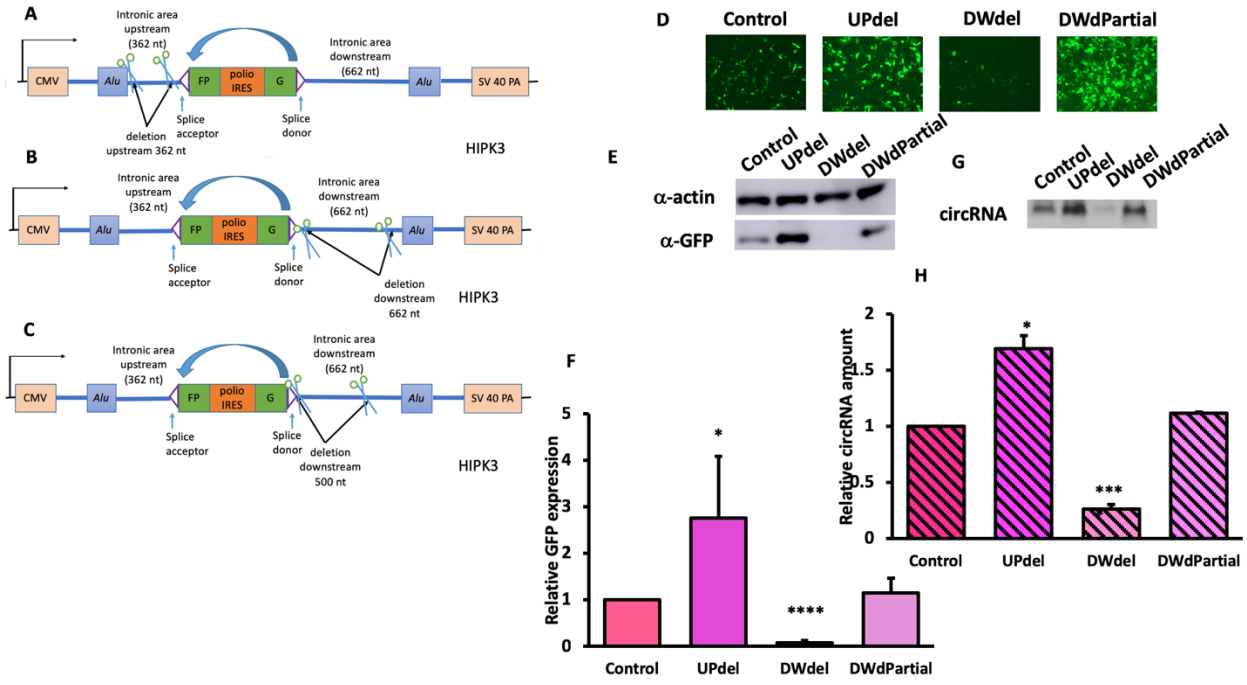

Figure 6. **Decreasing intronic distance on two sides of exon had different effects.** Constructs were transfected into HEK293 cells, and were harvested four days after. A)Schematics of upstream deletion construct (UPdel); B) Schematics of downstream full deletion construct (DWdel); C) Schematics of upstream partial deletion construct (UPdPartial); D)fluorescent images of cells; E) Western blot and F) quantifications of GFP of each deletion; G)Northern blot and H) quantifications of circularized RNA. (N=3; Error bars indicated standard deviation; significance level using two-tailed Student's t test was represented by [*]: [*] $p<0.05$, [**] $p<0.005$,[***]$p<0.0005$,[****]$p<0.00005$.)

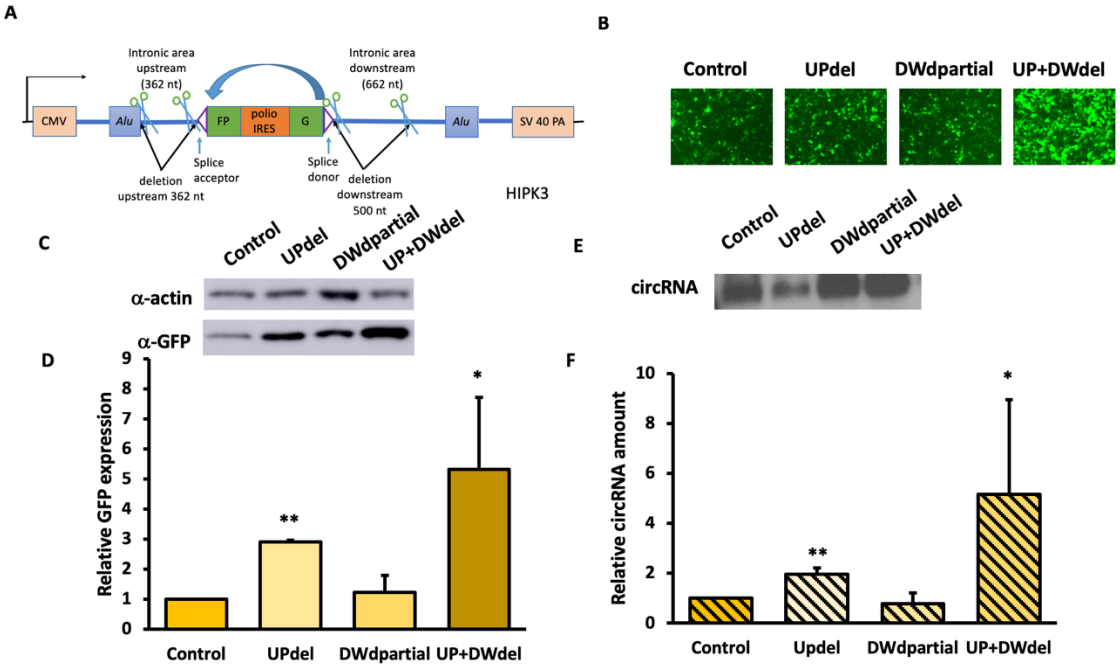

Figure 7. **Combining upstream and downstream deletion significantly increased GFP expression.** Constructs were transfected into HEK293 cells, and were harvested four days after. A) A schematics for combining upstream full deletion and downstream partial deletion; B)fluorescent images of cells; C) Western blot and D) quantifications of GFP of each construct; E) Northern blot and F) quantifications of circularized RNA . (N=3; Error bars indicated standard deviation; significance level using two-tailed Student's t test was represented by [*]: [*] $p<0.05$, [**] $p<0.005$,[***]$p<0.0005$,[****]$p<0.00005$.)

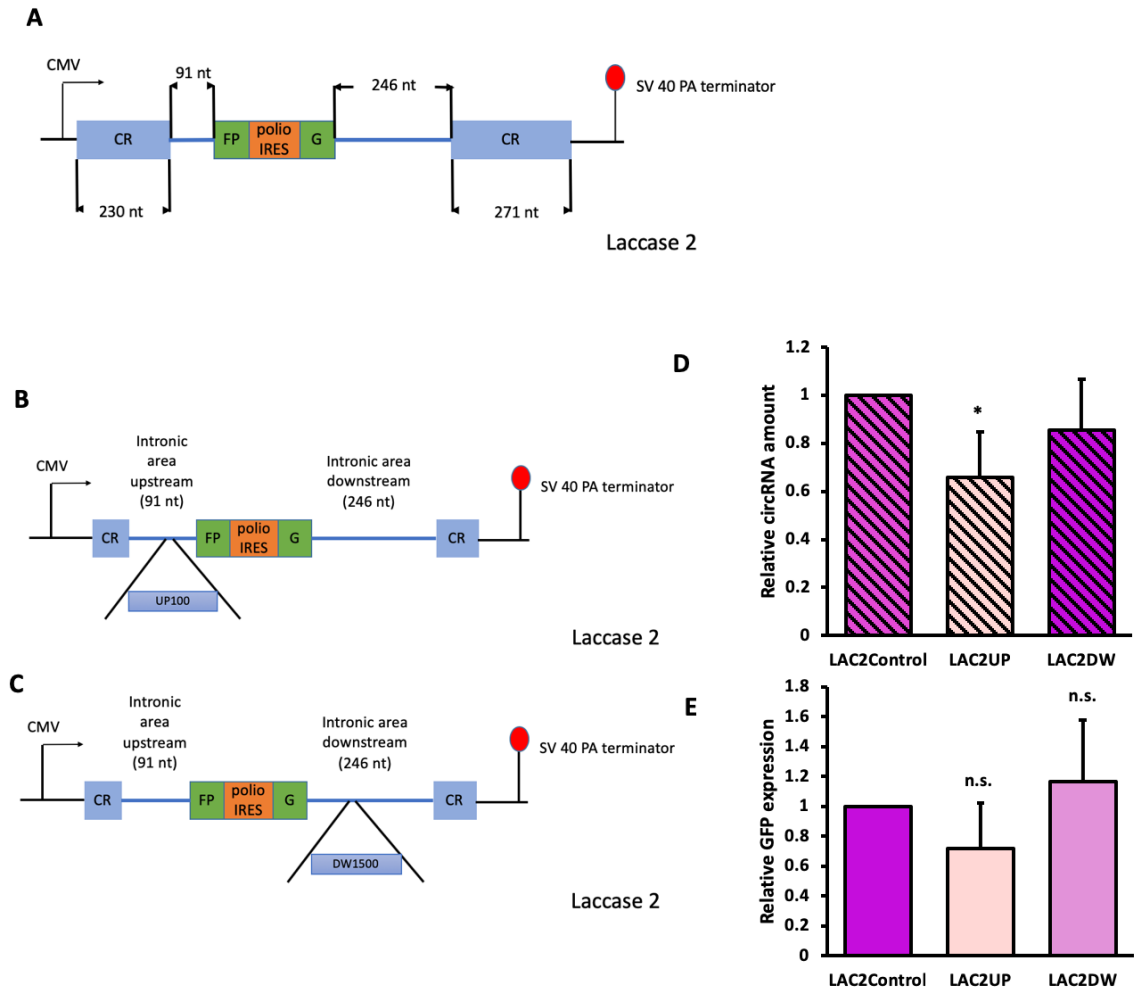

Figure 8. **Engineered laccase 2 constructs had similar but less significant effect.** Constructs were transfected into HEK293 cells, and were harvested four days after. A) A schematics for laccase 2 control group, CR stands for complementary region; B) A schematics for laccase 2 upstream insert; C) A schematics for laccase 2 downstream inserts; D) Northern blot quantification and E) Western blot quantification for laccase inserts; (N=3; Error bars indicated standard deviation; significance level using two-tailed Student's t test was represented by [*]: [*] $p<0.05$, [**] $p<0.005$,[***]$p<0.0005$,[****]$p<0.00005$.)

Tables

Table 1. The random sequences generated for insertions upstream and downstream

Location	Random sequence
Upstream to exons	TTCTTTAACCTTTTAATTGCTCTCTTAGAAGAAAGACAGATAACTT CTTACCAATATATCATCATAGGTAATACGATCGCATGTCCCATGTA AATGATTGATAAACTCAGTGTTTTATGAGCGACAAAAACTTAAAT AAGAAATACGCGCTCATAACTTGGTACGAATACAGATTGTAGCAA TGTTCGTCTGACTATGATTTATATATTACAGGCGGTATGTTTATTTT AATCAATTTTTAATAGCTTATAAAATAATACAGCTACTGGTGATCA CTCAATAATCTCGACTTCCCATTACAACTATGAGGATTTTTAAAGA GCCAATTATATTCGCTAATGTGAGGATAATGTAGTATTAGCAAAC AATAAGTTTCGAACTAGTTGTAACCTAACAAAAAGTGAATTTCAT AATACGTGTTATCCCATGCATATGGTAAATTTGAATAAAATTAAAT AGAGTTTGATCAATCTTTACATCGATCTAAAATCGAATGCAATGAT TTTTCAGGTGCAAATCAAAAATATTAGGTAACTAGAAGATTTGTA ATATTCTAAGTGTTGATCCATATGAATCATCATCTAGGATTATGTC GCTCTAAAAAAAAGATATTAGTAACTCTCTTCTTTAGCAGTCTGGT CTATGGAAACTACAAAACTAACCTTCCTAGTAACCGAGGACTAAG AATCTATGATATGAGTCAAGATATTTACTCAGTAATTTATGCTTTA GGTATTTAATTATTCTCATTGTCTTAAAGAGACCTATATTTTCTGCT TGTCAATCTATAAAATTCATATTAATGCGCAGATTTAATTCGAAAT AAAATGTCCAGAACCAAAAACAACCAGCATTTCGCATCTTGCCTA ATCCTCCTACATATTGTTATAAATAATCAGTAGAAATTTAATGTTA GATGATGGAATGATCTTAAATTAAATAGAAAATTAAAGGGAATGT ATATTCAATGTAATGAAGTTGGAGGATTAACATGGGAATCGTGCT TCTGTTTAAACAAGAATGGATATAAAGTAATAACCATTCCCCTAA CGTATAGGGTGCATTTTGTAATAATTTGAGAATCCAAAAACTTGCT ATTTTTGAAATTTTTCTTTAAGCACAAGTATTGAACTAAGCTTATA TCTAAAATCGTAGCAAGCAGATTTAAATAAAATATATTTTACCCGC GTTACAAATAAAATTAGTTAAAAGTTATGGAATATATTAATATGT AGATGGCCACTGGTGAGTTGTTACACCTCTACGGCAATGTTGAAA TTCTTAAATTATTCTGGTTAAATTTAAGCTGTAACACCCGTTTTACT TCATAACCATTTGTAATTCATAGCTTGATCTAGATTGGATTGTCAT TTTCTCAAAGTATTATGCAGACTGACGTACGCATCCCATATAAACT TATCATAATTTATCTGAAATTACTTAAAAATGTAGCTAGATTTTTA CCCACGCACCTAA
Downstream to exons	ATTGTGTTTTTATTATAATATCAGAATCTTTAAGTCGAGTCAATTA AACTCGGATTACAGTATTTACCGCATCTTGTGATTACTCACAAATT ATAATTCATCACAAGTCAAGCCATTACCTCTTTGAAATGCCGTATG AATTAATATGTAAACTTTGTGCGAATTTACTATGATTTGTTTAGTTT CGTTTAAAGGTACAATCAAATTTCTATTTATATGTTCAGCTAACTT TTACCCATCCCCCAAAATTTAGTAGGTTGTGAGATGTTATAGAAGT

	TCTTATTCATCTCGTAGGACATCAAGCTTTACTTTAATAAAGCATT TCATTCGAGTATAACAGAAAAAACGCTTACTGAATTGTGCAATTCT TCTACCTTAACTAAGATAGCTATCAATATTTAGTTTTTTAGCCTTGC AATAAATTTCTTATTTAGATTGCCACATATTGAGCTAGTGAATCAG TAATAAGCATGACACGCTTTCAAACGTCACGAATATGTGAATTAA GGCTCTGAACAGGACTATATACTTGAATTTGATTTCGCCCTGACAA CTGCAAACCTCAACATTTATAGATTATAAGATTAGCCGAAATTGTA CGTGATAACGTCCGTTAACTGCTCCCCGAGTGTGGCTCTTTGATTT GATAATATGCAACCTCTATCATAATTGATTATTTCTACGAACCATG TTATTTTCATAGTTTGGGCATATTTCTGTTGTAGGAGTGAAATCAC TTAACTTTGTGCCGTAGTCTTAATGAAAAATCTATGGACTTTGTTT TAGGTAACATTAAAAATCTAAACCTTATAAATGTGAAGGTCGCAT GCATAGATTTTTATCTTCGATTCAAGTTAAGTATGAGGCTACATGC TATATTATTACATCTACACTACTCAAAGTAAATATAGGAAGTGCAC GGCCTGGCCTGAGGTGTTTCGCATCATCATGTATTCGTTAATTGTT AATTGATGACACATAAACAATATTGTAGTCTCTCAAATTCAGCTCT ATTATCTTGAGCGTTATGTGTTAAATAGCGTAGAACAGAATTGACT GTTTAATACTAATTAGTGTTCGGTTTGGTAATGAAGAATCTATAGG ACTATGTCACTAATACTTTCGAAATACCTTATATCGATACTGAACA AATTGATGCAAACTTCCATCTTTGAATAGAGATAAATATACAAGT CGATAGAAAATGGGTAGGGGCTTCTAATTCATTCAACACTCTACG TCTTCTTCAAGAATTAGTAGAGTATCCTGCAGTTGAAAAAGAAATT ATTTCGTAAAGTAAGCTCATACTGTTTTTCTTGCGAAAGACTTAAC ATGATAAGAAATTAGAATAGTTTCAAATGATAGTTATTAATCCTA ATAACGGAACGTTATTTAAAGAATAAGTGTGACAAAGTATAACTC GATGAGTTATCCATTAATTGAACTAAGCGAGAGATTCCAGTGCTA ATGCACTCAATCC

Table 2. The sequence source, primers and restriction enzymes for construct preparation

Gene	Construct name	Source	Forward primer	Reverse primer	Restriction enzymes
HIPK3	Control				
HIPK3	UP100	upstream random	CTCGCTTAAG TTCTTTAACC TTTTAATTGC TCTC	CTAGCTTAAGCA ATCATTTACATG GGACATGCGATC G	AflII
HIPK3	UP500	upstream random	CTCGCTTAAG TTCTTTAACC TTTTAATTGC TCTC	CTAGCTTAAGAT TGCATTCGATTT TAGATCGATGTA AAGATTGAT	AflII
HIPK3	UP1000	upstream random	CTCGCTTAAG TTCTTTAACC	CTAGCTTAAGGA TTCCCATGTTAA	AflII

			TTTTAATTGCTCTC	TCCTCCAACTTCATTAC	
HIPK3	UP1500	upstream random			AflII
HIPK3	DW100	downstream random	GCTAACTAGTATTGTGTTTTTATTATAATATCAGAATCTTTAAGTCG	CGATACTAGTTGAATTATAATTTGTGAGTAATCACAAGATGCGG	SpeI
HIPK3	DW500	downstream random	GCTAACTAGTATTGTGTTTTTATTATAATATCAGAATCTTTAAGTCG	CGATACTAGTACATATTCGTGACGTTTGAAAGCGTG	SpeI
HIPK3	DW1000	downstream random	GCTAACTAGTATTGTGTTTTTATTATAATATCAGAATCTTTAAGTCG	CGATACTAGTGAGAGACTACAATATTGTTTATGTGTCATCAATTAACAA	SpeI
HIPK3	DW1500	downstream random			SpeI
HIPK3	UP100F	upstream random	GCCAACATTTTCCCTCAAATTCTTTAACCTTTTAATTGCTCTCTTAGAAG	CTAGACCGGTCAATCATTTACATGGGACATGCGATCG	AgeI
HIPK3	UP500F	upstream random	GCCAACATTTTCCCTCAAATTCTTTAACCTTTTAATTGCTCTCTTAGAAG	CTAGACCGGTATTGCATTCGATTTTAGATCGATGTAAAGATTGAT	AgeI
HIPK3	UP1000F	upstream random	GCCAACATTTTCCCTCAAATTCTTTAACCTTTTAATTGCTCTCTTAGAAG	CTAGACCGGTGATTCCCATGTTAATCCTCCAACTTCATTAC	AgeI
HIPK3	UP1500F	upstream random	GCCAACATTTTCCCTCAAATTCTTTAACCTTTTAATTGCTCTCTTAGAAG	CTAGACCGGTTTAGGTGCGTGGGTAAAAATCTAGC	AgeI
HIPK3	UPdel		TATACAATA	GTTACCTACCTG	

			ATCCCAGCAACTTGGGAGG	GACGTAGCCTT	
HIPK3	DWdel				
HIPK3	DWdPartial		GCCAACTAGTTAAGCACCCCC	GTTACCTACCTGGACGTAGCCTT	SpeI
HIPK3	UP+DWdel				
HIPK3	UPnoAlu				
HIPK3	DWnoAlu				
Laccase2	Control				
Laccase2	UP100	upstream random	CTAAGCTAAATCGAGACTAAGTTTTATTGT	CATTATACCCATAAATCGTACAGTAAAAAGGTATA	BlpI
Laccase2	DW1500	downstream random	GGAACTATAAAAGCTAGAAGGATGAGTTTTA	GGATGTCTCGGCGGTAATAAAATAAT	ACCIII